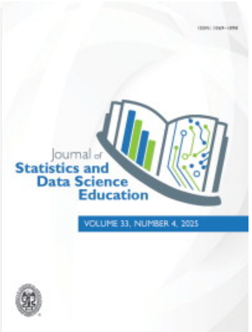


Taylor & Francis
Taylor & Francis Group

Journal of Statistics and Data Science Education

ISSN: 2693-9169 (Online) Journal homepage: www.tandfonline.com/journals/ujse21

# Students' Statistical Thinking when Using Generative AI: A Descriptive Case Study

V.N. Vimal Rao, Amos Jeng, Julianna Drew, Bhuvan Kala & Sanjana Gongati

**To cite this article:** V.N. Vimal Rao, Amos Jeng, Julianna Drew, Bhuvan Kala & Sanjana Gongati (01 Dec 2025): Students' Statistical Thinking when Using Generative AI: A Descriptive Case Study, Journal of Statistics and Data Science Education, DOI: 10.1080/26939169.2025.2596684

**To link to this article:** https://doi.org/10.1080/26939169.2025.2596684



Accepted author version posted online: 01 Dec 2025.

Check for updates

# Students’ Statistical Thinking when Using Generative AI: A Descriptive Case Study

V.N. Vimal Rao*, Amos Jeng, Julianna Drew, Bhuvan Kala, Sanjana Gongati

University of Illinois at Urbana Champaign, Urbana IL 61801, USA

*Corresponding author email: RaoVNV@Illinois.Edu

Abstract

Generative artificial intelligence (AI) technologies are transforming the world of education, but their impact on students' thinking and learning remains unclear. In this study, we investigated AI's potential role in supporting statistical thinking by interviewing undergraduate students ($N$ = 5) as they completed a graphing task using Rtutor.AI, an AI-powered tool that integrates ChatGPT with R. The analysis yielded five key themes that describe students' statistical thinking while using Rtutor.AI. The first theme demonstrated how the iterative and intuitive nature of prompting within Rtutor.AI shaped participants' approaches to problem-solving. The next three themes—"Building statistical understanding through a step-by-step process," "Identifying key elements of a problem to create specific prompts," and "Lowering barriers to completion of statistical tasks"—illustrated how Rtutor.AI facilitated statistical thinking in various ways. The fifth theme showed that students' prior statistical knowledge influenced their ability to interpret and contextualize Rtutor.AI's output, and that Rtutor.AI did not fully absolve students of the need to think statistically. Overall, these results highlight both the potential benefits and risks of incorporating AI into statistics classrooms and can serve as an empirical basis for future scholarship aimed at creating scaffolding around AI technologies to support statistics instruction.

Keywords: statistics education; rtutor.ai; thematic analysis; AI education

# 1. Introduction

The rapid development of AI technologies has galvanized the world to explore its implementation and potential transformative effects, including in education. Recent studies have found that AI tools can support the development of students' thinking and self-regulated learning (Sardi et al. 2025; Zhou et al. 2024), as well as their programming self-efficacy and motivation (Yilmaz and Yilmaz 2023). However, some studies have noted that AI tools may also inhibit students' thinking because it "does everything for you" (Habib et al. 2024, p.4).

Within the realm of statistics and data science education, a few papers have begun to address how AI technologies can be integrated into the classroom. Ellis and Slade (2023) consider how ChatGPT can be used by instructors to generate course content, and how it can be used by students to answer content-related questions such as 'What is a *p*-value?'. Bien and Mukherjee (2025) as well as Bray (2024) describe integrating AI tools in their classrooms to support students' statistical programming, with a focus on instructional considerations and reflections from their teaching experience[1].

The manner in which AI tools may be beneficial to students will certainly vary based on each course's specific learning objectives. In particular, courses with explicit statistical programming learning objectives may be less appropriate places for AI tools than statistics service courses often taken by students to fulfill quantitative general education requirements. These programming-oriented courses often require students to develop programming skills that ought not to be bypassed by the use of AI tools.

In this paper, we focus on the case of programming-free statistics courses and consider the potential role that AI tools can serve in developing students' statistical thinking skills within

[1] Prior to 2023, there is no trace of AI tools to support teaching statistics (Zavez and Harel 2025).

these general education and service-oriented courses. In particular, we consider the potential of Rtutor.AI, a Shiny App that integrates ChatGPT with R (see rtutor.ai for more). To use Rtutor.AI, students first upload a csv file and then type prompts such as "compute the mean of the 'grade' variable" in lieu of R code. When clicking submit, Rtutor.AI then passes the prompt and metadata about the data set to ChatGPT, which then returns R code to Rtutor.AI. Rtutor.AI then runs the code, and displays both the code it has run as well as the output. Students can also download a history of all of the R code that Rtutor.AI ran based on their prompts as an R Markdown file.

For general education and service courses and their students, on one hand, AI tools such as Rtutor.AI may reduce the computational burden to conducting statistical analyses, and AI's ability to make sense of vague prompts may partially ameliorate the hindrance a student's lack of computational skills can have on their statistical problem solving (Woodard and Lee 2021). On the other hand, it is entirely possible, perhaps even likely, that this heightened usability is a double-edged sword that obscures and ultimately undermines the development of students' conceptual understanding of statistics and their statistical thinking.

To establish a foundation for our investigation into these topics, we first define statistical thinking, then discuss the role and potential benefits of incorporating AI into a service course, before outlining the study's research questions and objectives.

## 1.1. Orienting Framework for Statistical Thinking

In this paper, we investigate Rtutor.AI's role in supporting students' statistical thinking, rather than focus on the way it might support the development of students' computational thinking. To distinguish the ways in which context, content, and software interact as part of statistical thinking, and to clearly delineate statistical thinking and the use of Rtutor.AI to achieve statistical tasks from computational thinking about and with Rtutor.AI, we integrate the statistical investigative cycle (Wild and Pfannkuch 1999) into Biehler et al.'s (2015) framework for randomization testing, and generalize the framework for all statistical acts. We visualize the way context, content, and software interact with each other and the investigative cycle as a step-pyramid as seen in Figure 1. Before one can utilize software to conduct statistical tasks, they must step through the real-world context, as well as statistical world content, which together dictate the way in which the software ought to be utilized. This new framework for statistical thinking serves as our orienting framework.

The statistical investigative cycle (a.k.a. the PPDAC cycle) describes "the way one acts and what one thinks about during the course of a statistical investigation" (Wild and Pfannkuch 1999, p. 225). Originally used to describe expert statisticians' thinking, the investigative cycle has been widely adopted into statistical curricula across the world as a framework to guide statistics education (e.g., Arnold 2007; Franklin and Garfield 2006). The cycle as defined by Wild and Pfannkuch (1999) includes five component aspects: defining a problem, creating a statistical plan (i.e., sample and study design, measurement), data collection and management, data analysis, and interpretation and conclusions.

Biehler et al.'s (2015) framework for randomization testing builds off of an initially proposed cycle for "computer-supported statistical problem solving" (p. 138). Biehler et al. (2015) distinguished between three worlds involved in statistical thinking: the 'contextual world', the 'statistical world', and the 'software world.' As "statistics ceases to have meaning if it is not related to any practical problem" (Rao 1975, p. 152), the statistical world rests upon the contextual world, and statistical problems are dependent upon real-world problems. Statistical problems and their associated plans are then achieved through the use of software. The use of software often requires unique considerations for data management and analysis (e.g., how to structure the data, which specific statistical procedure to use, and how to use it), thus distinguishing the software world from the statistical world. Using appropriate procedures generates software output, and interpretation converts software output into statistical results (e.g., the creation of a density plot to visualize the distribution of some quantitative attribute is a software result residing within the software world, while thinking about its shape and diagnosing it as bimodal is a statistical result residing within the statistical world). Finally, these statistical results are re-contextualized to help answer the generating real-world practical problem (e.g., the bimodality indicates group differences, and thinking about what those distinguishing group characteristics might be relies once more on the contextual world).

Applying this framework to the present study, the utilization of Rtutor.AI resides entirely within the software world. Thus, questions about Rtutor.AI's usability and the extent to which students are able to utilize the software to achieve their statistical plans and solve their statistical problems similarly reside entirely within the software world. The conversion of a statistical plan into specific Rtutor.AI procedures, as well as the interpretation of results from Rtutor.AI, necessarily involves the action of stepping between the statistical and software worlds.

It is on these aspects of statistical thinking—stepping from the statistical world to the software world, steps within the software world, and stepping from the software world to the statistical world—that we focus our investigations.

1.2. A Potential Role for AI Tools in Statistics Service Courses

One of the goals from the Guidelines for Assessment and Instruction in Statistics Education College Report (GAISE 2016) is that "students should be able to produce graphical displays and numerical summaries and interpret what these do and do not reveal" (p. 9).

Especially in the case of R, producing graphical displays with ggplot teaches students how to think about statistical graphs in a very structured way through a layered grammar of graphics (Wickham 2010). This structure can help students to “gain insight into the deep structure that underlies statistical graphics” (Wickham 2010, p. 3). In other words, ggplot helps students step through the context world, statistical world, and software world, in a way that not only helps them produce graphical displays, but will likely also result in increased fluency in critically consuming graphical displays.

However, for service courses, introducing statistical programming languages such as R or Python may overwhelm students. Many students are initially reluctant to learn programming (e.g., Tucker et al. 2023), and introducing programming learning objectives without simultaneously removing statistical concept learning objectives may inordinately tax students (e.g., Adams et al. 2021).

For courses devoid of learning objectives specific to the software world, the choice of software is perhaps moot and might be made primarily on the basis of the software’s efficacy in promoting students’ thinking within the statistical world. However, *some* software tool is necessary, as a complete statistical investigation in which students produce statistical analyses does require software[2].

Many such courses might choose to use pedagogical tools such as CODAP or TinkerPlots, but we see these software tools primarily as pedagogical, and not designed or intended for practicing statistics outside of the classroom. Instead, many courses might use drag-and-drop software tools such as StatCrunch or SPSS, but these softwares may not be as suitable for integrating computing concepts into the statistics curriculum as recommended by Nolan and Temple Lang (2010). Specifically, as literacy is both reading and writing, in order to develop

[2] We ignore here the option to produce statistical analyses completely by hand.

students' abilities to both produce and consume graphical displays and determine what they do and do not reveal, we believe that the level of specificity required in programming a statistical software tool to produce graphical displays will best support the development of students' ability to consume graphical displays. This is especially true with R's ggplot, which was intentionally designed to support statistician's thinking about the components in a graphical display, and can thus support students' skills in consuming and producing graphical displays and questioning what they do and do not reveal.

How can we capitalize on the benefits of teaching students programming concepts that will support and complement the development of their statistical understanding while obviating the difficulties inherent in integrating statistical programming into an introductory course, especially for a general student population, many of whom are averse to statistics and programming to begin with?

AI tools present a potential solution to this conundrum. AI tools may serve as a useful scaffold to ease students into statistical programming in R or Python, alongside technologies such as Scratch (Liao 2023) or other pedagogical strategies (e.g., Tucker et al. 2023). However, we consider in this paper the potential of AI tools to replicate the educational benefits of using a programming language without having to use a programming language itself, especially for introductory statistics service courses. That is, AI tools may be able to better balance the need for procedural fluency in statistical software tools against novice usability by increasing students' ease of use (e.g., Sardareh et al. 2020).

### 1.3. Problem Statement and Research Questions

As stated by the 2024 EDUCAUSE Horizon Report, "more and more uses for AI in the classroom are emerging, and these technologies have the potential to change the landscape of teaching and learning and the student experience in a variety of ways, for better or worse" (Pelletier et al., 2024, p.19).

As we continue to explore new ways to integrate AI tools into the statistics classroom, we must aim for the 'better' rather than the 'worse.' To do so, thoughtful curricular design requires an evidence basis with which to evaluate the trade-off between novice usability with AI tools and the benefits to the development of students' conceptual understanding of statistics supported by the utilization of AI-based statistical software tools.

If AI tools have a role in statistics courses (without specific programming of software-world learning objectives), they must be evaluated in terms of how they support students' statistical thinking within the statistical world, by making it easier to step through the software world and facilitating the connection between the software and statistical worlds.

Thus, we present here an empirical trace of students' thinking when utilizing Rtutor.AI to do statistics. We focus on the production and interpretation of graphical displays and numerical summaries. Specifically, we investigate the following research questions:

- RQ1: How do students use Rtutor.AI to complete statistical tasks (i.e., actions in the software world), and how usable is Rtutor.AI for such tasks?
- RQ2: How does students' use of Rtutor.AI help or hinder their thinking about statistical content (i.e., stepping between the statistical world and the software world)?

## 2. Methods

To capture and describe students' thinking when using Rtutor.AI, we employed a multiple descriptive case study method (Merriam 1988). Five undergraduate students participated in semi-structured task-based interviews in which they utilized Rtutor.AI to complete a statistical graphing task[3]. All study materials and data are openly available at the following OSF repository: https://osf.io/za9ds.

2.1. Frameworks for Analysis

As a framework for analyzing students' thinking when using Rtutor.AI, we draw on Moshman and Tarricone's (2016) definition of thinking as metacognitive self-regulation. We operationalize acts of statistical thinking as generally falling into one of three categories of metacognitive self-regulation: planning, monitoring, and evaluating (Schraw et al. 2006; see Table 1). Planning steps typically involve goal setting and strategy selection (e.g., 'I want to change the y-axis limits'). Monitoring steps typically involve self-checking one's progress towards successful implementation of the planned goal and strategy (e.g., 'Will this prompt correctly change the y-axis limits to what I want them to be?'). Evaluating steps typically involve revisions to plans when necessary, either as a function of monitoring steps or as a function of the examination of the results of an execution of an initial plan (e.g., 'This prompt didn't change the y-axis limits the way I wanted, therefore I need to make changes to my prompt').

While the focus of the task is on students' use of Rtutor.AI, students' thinking within the software world necessarily is dependent on their ability to think about the real world context and statistical content world. Thus, we draw on our Framework for Statistical Thinking (see Figure 1 above) as a framework for analyzing how students' statistical thinking is related to their thinking when using Rtutor.AI. We delineate students' actions and thoughts between the real world, statistical world, and software world as follows (see Table 2): Acts of 'contextual' thinking are those related solely to the real world (e.g., 'ACT scores are typically between 0 and 36'); Acts of

[3] All study procedures and materials were reviewed and approved by the Office for the Protection of Research Subjects at the University of Illinois, Project Number IRB24-0602.

‘statistical’ thinking are those related to employing statistical knowledge to achieve one’s contextual or real world objectives (e.g., ‘I should set my y-axis limits to 0 and 36’); Acts of ‘software’ thinking are those related to employing software-specific knowledge to achieve one’s statistical objectives (e.g., in the case of R, ‘ylim = c(0,36)’).

## 2.2. Case Selection and Participants

There were two participant inclusion criteria we utilized to create our sample frame. First, participants must have completed one and only one introductory level statistics course that included instruction or exposure to multivariable scatter plots. Second, participants must not have previous programming experience in any language (e.g., R, Python, Java, etc.). Based on these criteria, we identified a large introductory level class at our institution as a suitable pool for eligible participants.

In order to obtain a diverse sample in terms of participants’ background level of statistics knowledge, we stratified students from the identified class into three tiers: students who earned an A, students who earned a B, and students who earned a C or D. Within each tier, we attempted to recruit 2 participants, verifying that they indeed met all inclusion criteria. However, due to the smaller size of the C/D tier, as well as time constraints related to funding and the overall project timeline, we were only able to recruit one participant from the C/D tier.

Throughout this paper, we will refer to each participant with the following pseudonyms. From the A tier, the two students were Abi and Aino. From the B tier, the two students were Bennu and Bo. From the C tier, the student was Cam.

2.3. Materials

To elicit students' thinking when using Rtutor.AI, we created a three-part task that guides students through an investigation requiring them to generate a multivariable scatter plot (see the OSF repository, Appendix A - Task Instructions). We utilized the context of a couple who recently moved to Chicago and are in need of selecting a school for their children. Participants were told they would be receiving data from 18 different students from three different schools, and would have information related to the students' ACT standardized testing scores as well as the number of hours each student studied for the test.

The first part of the task required participants to plan out what type of graph they wanted to create to help make a recommendation to the couple moving to Chicago. Participants drew a rough sketch of this graph on paper. This was intended to elicit a baseline of participants' statistical thinking, especially as it related to how they converted the real world problem into a statistical problem, and what procedures they planned to execute using software in later parts of the task.

The second part of the task required participants to utilize Rtutor.AI to create a multivariable scatter plot. Participants were given specific instructions to consider specific aesthetic mappings, although they were not instructed to create the mapping in any specific fashion. These included thinking about which attributes should be mapped to which axes, whether color and shape can be used to distinguish between groups, the inclusion of statistical layers such as smoothed trend lines, and other plot layers such as titles. This was intended to elicit an empirical trace of students' thinking when using Rtutor.AI and their procedural fluency to do so, especially when considered in conjunction with their plan from the first part of the task, and also against each of the instructions we provided in this second part.

The third part of the task required participants to compare the final graph they created using Rtutor.AI to their initial by-hand drawing from Part 1. Additionally, participants were asked which school they would recommend that the couple send their children to, and to explain their reasoning. This was intended to elicit a trace of students' thinking as they connect the software output to statistical results and ultimately real world conclusions.

2.4. Procedure

This study was conducted in-person with participants meeting with a member of the research team for approximately 60 minutes. After completing all required consent forms, the research team member explained the think-aloud nature of the interview to participants (see the OSF repository, Appendix B - Task Interview Guide). Participants then were provided scratch paper and a laptop to complete the tasks. The research team member intermittently probed participants' thinking by asking questions such as "what was your thought process when you …", "how did you decide to …", or "why did you …".

Data artifacts included an audio recording of the entire interview, all written materials generated by the participant, the screen recording of the laptop while in use by the participant, and observer notes taken by members of the research team. Transcripts of interviews are included in the OSF repository (see Appendix D - Interview Transcript).

2.5. Analysis Plan

The analysis of interview transcripts followed an Inductive Thematic Analysis (Clarke & Braun, 2017) and proceeded across three iterative stages: preliminary coding, initial coding, and categorization. Preliminary coding involved noting key words, phrases, or sentences that captured initial patterns in the data. These notes served as early impressions to reference later.

For example, when a participant stated, “I need to add a title to my graph,” a researcher jotted down, “Catches something else to add: ‘I need to add a title to my graph.’”

Initial coding involved refining these preliminary codes to develop initial codes, or short words or phrases that captured the essence of participants’ words and behaviors related to the research questions and/or students’ use of Rtutor.AI. This process produced a range of qualitative codes, including in vivo codes (direct participant quotes, e.g., “I think that’s good”), process codes (actions, e.g., “Making prompt more specific”), and descriptive codes (summaries, e.g., “Unexpected result”). Table 3 provides an example of how preliminary and initial codes were generated during analysis.

Categorization involved grouping similar codes into broader themes or “families,” with subcategories added as needed to capture more specific ideas. To ensure the reliability of the coding process, two members of the research team independently coded and categorized each interview transcript. Each pair of coders met to reconcile any differences in the final categories that emerged from analysis. For each participant, the researchers collaboratively created a report summarizing all categories and subcategories, including descriptions of each category and illustrative examples from the transcript. The full participant reports can be found in the OSF repository, Appendix C - Interview Coding Summary.

Finally, participant reports were compared to identify overarching themes that characterize students’ statistical thinking within Rtutor.AI across participants. This process involved synthesizing insights from individual reports by identifying both commonalities and differences in categories across participants. These themes are presented in the Results section.

## 3. Results

A total of five themes emerged to describe students' statistical thinking while utilizing Rtutor.AI to complete statistical tasks. The first theme—"Engaging in planning, monitoring, evaluation, and revision"—provided an overview of students' problem-solving in Rtutor.AI across three aspects of thinking (as defined by Schraw et al., 2006), where each stage of the problem-solving process was shaped by the iterative and intuitive nature of prompting. The next three themes—"Building statistical understanding through a step-by-step process," "Identifying key elements of a problem to create specific prompts," and "Lowering barriers to completion of statistical tasks"—illustrated how Rtutor.AI facilitated statistical thinking by helping students break down complex problems, extract and organize key information, and bypass technical and procedural barriers to problem-solving. The fourth theme—"Lowering barriers to completion of statistical tasks"—also revealed a potential negative repercussion of using the tool; specifically, by automating aspects of the graphing process, Rtutor.AI may have limited students' active engagement in statistical decision-making. Finally, the fifth theme—"Prior knowledge and preferences shaping engagement with AI"—showed that students' prior statistical knowledge and graphing preferences influenced how they interacted with Rtutor.AI, and students' use of Rtutor.AI similarly supported students' statistical thinking, suggesting a bidirectional relationship between how students thought about the statistical content of the graphs and their use of Rtutor.AI.

## 3.1. Engaging in Planning, Monitoring, Evaluation, and Revision

### *3.1.1. Planning*

For multiple participants, the intuitive and iterative nature of prompting in Rtutor.AI encouraged them to adopt a more exploratory approach to planning, in which they refined their strategy over time rather than establishing a specific, statistically-laden plan for how they would achieve their goals. For example, Cam described how they initially developed a strategy for exploring Rtutor's capabilities and limitations by testing its response to general prompts (e.g., "create graph"):

> Yeah, just like any writing or prompts or anything like that, making projects, they say the hardest thing is looking at the blank page. So I just type in a very general prompt and see what AI generates, and then, I'll take one of the very vague explanations and be like, "Oh, I like that," and then build off it and specify and make my project around that.

Similarly, Bennu's first prompt instructed Rtutor.AI to "create a multivariate scatter plot with x-axis as number of hours of studying and y-axis as ACT score." During Bennu's interview, they explained their rationale, saying, "I figured after that, it would just kind of give me a base, and then after that, I can make it more specific and then go off of what else the other instructions and … [what] I wanted to do." Thus, these participants used an iterative prompting process to explore the tool's functionality through relatively broad, unstructured prompts in a search to determine how to use Rtutor.AI to achieve their statistical objectives.

However, not all participants planned their approach in this way. For example, while Cam and Bennu took a more exploratory approach to using Rtutor.AI, Abi demonstrated a more detailed and statistically-laden plan when writing their prompts. Abi's initial prompt was relatively lengthy and contained numerous specific instructions about the components they wanted to see in the output graph (e.g., "Create a multivariate scatter plot for the data", "Put number of hours studying on the x-axis and make the range on the x-axis from 50 to 75", "Make data points from 'Walter Payton College Prep' with gold circles"). As Abi wrote this prompt, they verbalized their thought process as follows:

> So now I'm going to go to Rtutor.AI and basically say, "create a multivariate scatter plot for the data." And then I want to put the number of hours studying on the x-axis. Looks like they're all between … 50 and 75, I think. So I'm going to tell Rtutor.AI to … make the x-axis only for those values … And then I'm going to tell it to put ACT scores on the y-axis. And I'll just make the range what you can get on the ACT score, which is up to 36 … I'm telling it to mark data points from the first school … with gold circles, and I put the name of the school and in quotations, just so it would know exactly what I'm talking about … I'm just doing this for all three schools with different shapes and colors.

Thus, in contrast to Cam and Bennu, whose planning focused on experimenting with Rtutor.AI's capabilities, Abi's planning reflected careful thought about many of the graph's components, including axis setup, ranges, and the desired appearance of data points. Their approach demonstrated a clear expectation of both the graph's structure and how they wanted to use Rtutor.AI to achieve their goals. This difference highlights the variation in how participants approached planning the task, which might also be a function of students' prior statistical knowledge.

*3.1.2. Monitoring*

In addition to shaping how participants approached planning, Rtutor.AI appeared to enable their monitoring of their analyses. For example, Aino initially expressed uncertainty about how to phrase prompts in Rtutor.AI (e.g., "I need to think about how to best word this", "I don't write AI prompts"). When asked about their strategy for navigating this uncertainty, they explained that they actively reflected on the clarity and coherence of their prompts as they wrote them:

> What I've been trying to do is, I've been trying to read it back and see if it makes sense … I've just been trying to read it and, like, "does this make sense?" … If you're given instructions, "does this make sense?" … Yeah, I was trying to understand if the instructions, I guess, were clear and concise. And if they made sense. And then I was also thinking about how that would translate as instructions into the code that it writes.

Likewise, at multiple points, Bennu exhibited monitoring skills by correcting their prompts for greater specificity as they typed and thought aloud. As Bennu verbalized their thought process while writing an initial prompt in Rtutor.AI, they actively assessed and adjusted their wording in the moment, saying, "I'm going to ask it to create a 'multivariate graph' … or 'scatter plot,'" and "I just did 'number of hours'—I guess I'd say '...of studying,' to make it more clear." In this instance, Bennu demonstrated an ability to monitor their progress, recognize when their prompts were overly general (e.g., "multivariate graph," "number of hours"), and quickly correct these prompts to better align with the instructions and dataset before submitting (e.g., by specifying "multivariate scatter plot," "number of hours of studying").

Thus, by continually reflecting on their prompts, Aino and Bennu demonstrated a potential benefit of Rtutor.AI in supporting the development of students' statistical thinking—to ensure that a prompt "makes sense," they needed to reflect on their purpose and plan.

*3.1.3. Evaluation and Revision*

The intuitive nature of prompting Rtutor.AI in English also appeared to facilitate participants' evaluative thinking, as the formation of prompts was accompanied by concrete expectations against which Rtutor.AI's output was compared. In many cases, participants expressed approval of Rtutor's generated results relative to their expectations with comments such as, “I think this compared with what I had in my head” (Aino), “I like that they did that … that’s how my thinking was” (Bennu), and “Yeah, that’s what I expected” (Bo).

However, participants also often recognized areas where Rtutor.AI’s output could be improved. In response, they adjusted and refined their prompts, often multiple times, to achieve their desired results. Thus, the ease of iterative prompting in Rtutor.AI enabled them to repeatedly engage in revision; as Bo put it, “As things needed to be tweaked, I could kind of go in myself and just change little things.”

For example, after prompting Rtutor.AI to create distinct shapes for data points from different schools (“make Whitney Young a star shape dot, Northside a square dot, and Walter Payton a triangle dot”), Bennu noted that the dots were less prominent than they had anticipated:

> It’s a little different than what I was expecting. But I think that’s cool …I was expecting more to see, like, I thought the dots would be a little bit more prominent … I think it would [be helpful to have larger dots], just to see the separate students.

In response, they modified their prompt with ease by adding, “Make the dots more prominent.” This adjustment successfully produced the intended outcome, and facilitated Bennu’s statistical and contextual interpretation of the software output. When evaluating the revised graph, Bennu remarked:

> I like … how they do it in this one … I think it’s interesting to see the individual students because that may play a factor in the parents’ heads, like, “Oh, this one could be an outlier. We shouldn’t take that one student into account but more look at the trend of it.”

Similarly, Abi noted room for improvement in their graph after initially specifying the y-axis limits to be 1 to 36, which caused the data points to cluster toward the top. They stated:

> It’s pretty close [to what I hoped to see]. I didn’t really look at the data values that closely for ACT score, so I was expecting them to maybe be a little more spread out. So if I made it again, I would probably change the y-axis range.

As a result of this evaluation, Abi developed a plan to refine their prompt. In doing so, they also exhibited further monitoring skills by identifying missing graph components (i.e., trend lines) in the moment. They described their revision process as follows:

> I’m going to change the y-axis like I said, just so that the graph will be a little more zoomed in, kind of. So I’m seeing that the lowest value is around 28, and the highest value is 35. So I’ll probably still leave it at 36 but maybe change it to, like, 27 instead. And then I’m also going to add trend lines because I forgot to do that. So I’m going to say, “add a trend line for each school.”

The ease of utilizing Rtutor.AI to adjust graph features appeared to positively impact participants’ statistical thinking. Although Bennu had never been taught how to change the size of the points, and Abi was not explicitly instructed to account for data point clustering, both students intuitively and effortlessly modified these aesthetic elements to enhance their interpretation. As a result, they quickly made sense of the software’s output, both in terms of its statistical results and the real-world conclusions that could be drawn. In this way, Rtutor.AI appeared to effectively support students’ statistical thinking.

### 3.2. Building Statistical Understanding Through Step-by-Step Processes

The iterative and intuitive nature of prompting in Rtutor.AI appears to have helped students enhance their graphical understanding by breaking down complex problems into smaller, more manageable steps. This process allowed them to assess their progress, make necessary adjustments, and better understand how each component contributed to the final graph.

As Bennu explained in their interview, “Seeing each step helps me think about what I should do for the next step or what I should fix in that step.”

Bo’s interview, in particular, illustrated the educational benefits of an iterative approach within Rtutor.AI. At the start of the task, Bo considered copying and pasting all instructions into a single prompt to complete the task quickly, stating, “[Copy-and-pasting is] the easiest way to put it all together at once.” However, they quickly reconsidered and opted to start with a simple prompt that would allow them to add elements incrementally: “Well, I could do just ‘create a multivariate scatter plot’ first and then do that later … maybe I’ll do that.”

Later, Bo reflected on this choice, explaining that a step-by-step approach supported their understanding of the process of graph construction: “I think doing it one at a time can show you, as the student, kind of how it works—like the steps in creating the graph—whereas, all at once just gets you the end result.” They emphasized that the process of iteratively incorporating graph components (e.g., trend lines, differently shaped data points) reinforced their statistical understanding:

> As a student, I like doing it one thing at a time. I wanted to do that, just so I could kind of remind myself how the graph works because I haven’t been in stats for a while. But doing it one at a time kind of allows me as a learner to see how it becomes the final product, and I think that that could be helpful.

Reflecting further, Bo noted that Rtutor.AI would have been valuable in their previous statistics course by helping them grasp how different graph elements come together:

> I wish I knew about this when I was in stats. I think it would have made me understand and learn the material better and kind of get a grasp of putting things together. It helps me when I have to draw it—next time, I’ll know what should go where.

While this iterative nature is not unique to Rtutor.AI, the benefits students perceived from using it demonstrates the tool's curricular potential. Furthermore, existing pedagogical strategies used to teach programming, debugging, and other computational skills can likely be adapted to Rtutor.AI to further enhance its potential as a procedural tool for doing statistics.

### 3.3. Identifying Key Elements of a Problem to Create Specific Prompts

When working within Rtutor.AI, students actively identified, extracted, and organized the most important parts of a problem to accomplish their task. The intuitive nature of prompting in Rtutor.AI supported this process by enabling them to easily incorporate essential components (e.g., relevant variables, graph elements) into their prompts. For example, Abi demonstrated this approach by strategically using quotation marks in their prompts to specify key information they wanted to extract from the dataset (e.g., "Mark 'Northside College Prep' with green squares"), explaining that this strategy helped ensure that Rtutor.AI interpreted their prompt correctly: "I'm afraid that if I don't have quotes … [Rtutor.AI] just won't know what to do with it."

Furthermore, Aino noted that this strategy of identifying key elements was valuable even before using Rtutor.AI, as they first created a rough sketch of the graph. They explained that breaking the problem down into essential components—such as ACT scores, hours spent preparing, and the three schools—made it easier to process and organize information effectively:

> I guess before creating the graph out of the … instructions… I tried to grab the important pieces of the text, like seeing, "Okay, this is ACT scores, and then you have the number of hours." So then you have this data that you need to include, and then these are the three schools. So just pulling that information out of the text before I made the graph … I feel like it's easier to, I guess, process and pop out the stuff if you break down the problem instead of just like, "Oh no, there's a lot of information." But if you

just write down [what you need] … I have that to look at, and I'm like, "Okay, this is what I need to look at," … so you're not distracted by all the other text.

This same approach carried over when Aino worked within Rtutor.AI. They described actively extracting key elements from the instructions, filtering out extraneous details, and organizing the relevant information within their prompts:

> I was still looking at the instructions [when working in Rtutor.AI] … I was like, "Okay, I need to make sure … I have these parts." I think that … extracting info, I feel like that's usually my thought process with all instructions. I typically write off to the side or something, just notes for myself that I can easily understand.

Aino's first prompt reflected this approach, as they incorporated specific variable names and required ranges directly from the data set and instructions (e.g., "Represent act_scores on the Y-axis from 25 to 36. Represent hours_spent_preparing on the X-axis from 55 to 80…"). This demonstrates how they effectively organized information to focus on the most relevant aspects of the task.

This suggests that students developed procedural fluency with Rtutor.AI in a way that both drew on and reinforced their statistical knowledge. This step-by-step approach, similar to pseudocode, can still be elicited and refined through Rtutor.AI. In other words, the AI-driven nature of Rtutor.AI does not completely obviate students' procedural thinking.

### 3.4. Lowering Barriers to Completion of Statistical Tasks

Overall, Rtutor.AI's capabilities appeared to reduce technical and procedural barriers to problem-solving, which may have allowed students to focus more easily on completing statistical tasks. Although some participants expressed uncertainty over how to word prompts appropriately (e.g., Aino remarked, "With Rtutor.AI, I was thinking, 'How do I word this,' which I guess I wasn't really thinking about with [drawing] the graph [by hand]") or noted challenges in how Rtutor.AI interprets prompts (e.g., Cam stated, "Sometimes [AI] won't understand you'll want it to do something very specific, but it'll think you want it to do something else"), it appears the tool generally streamlined the graphing process. In most cases, participants found that the intuitive nature of prompting enabled them to bypass procedural details (e.g., syntax and formatting requirements in R) that might otherwise distract from conceptual understanding when creating graphs (Woodard and Lee, 2021).

For instance, Bennu commented on the ease with which Rtutor.AI allowed them to create visualizations, saying, "I honestly thought [Rtutor.AI] was helpful because … [you] just write in what you need, and … it does give you what you need." Similarly, Abi compared the experience of creating a graph in Rtutor.AI to drawing one by hand, emphasizing how Rtutor.AI simplified the process of translating their statistical plan to a visualization:

> I feel like [Rtutor] made it a lot easier compared to if I was just trying to draw it on paper because … you have to be a little bit careful about how you say stuff, but for the most part, you can just tell it what you want it to do, and it'll do it.

However, by automating key aspects of the graphing process, Rtutor.AI may have also reduced opportunities for deeper engagement in statistical thinking. For instance, Cam noted a lack of effort involved in working within Rtutor.AI, describing the tool as "much easier [than drawing by hand], like I don't have to do anything." Similarly, some participants observed that Rtutor.AI often provided graph elements they had not requested explicitly. In one instance, Bo instructed Rtutor.AI to simply "create a multivariate scatter plot." As a result, Rtutor.AI not only generated a plot but also selected appropriate variables for the axes and included a legend distinguishing data points by school. Reflecting on this, Bo noted:

> It kind of gives me the elements, like hours spent preparing and ACT scores, without me asking, based on the dataset that we uploaded. And then it kind of gives me a key to what schools are recommended based on hours spent studying and their ACT score. And then … yeah, it kind of did the first step for me.

Thus, although Rtutor.AI reduced technical barriers to statistical problem-solving, it may have also limited students' active engagement in statistical decision-making by preemptively completing aspects of the statistical task for them, rather than forcing them to explicate all decisions related to the production of the graph.

3.5. Prior Knowledge and Preferences Shaping Engagement with AI

Although Rtutor.AI influenced students' understanding of statistical concepts, their prior knowledge and graphing preferences also appeared to shape how they engaged with the tool. Thus, our findings suggest a bidirectional relationship between students' use of Rtutor.AI and their statistical thinking. Specifically, during their interviews, participants described how their background knowledge shaped both their approach to working in Rtutor.AI and their interpretation of its output.

For example, Bennu structured their first prompt around three core graph elements: the graph type ("create a multivariate scatter plot…"), x-axis variable (...with x-axis as number of hours of studying…"), and y-axis variable ("...and y-axis as ACT score"). When asked about this choice, Bennu explained:

> I feel like the first thing that you want to know is what data you're getting, and that is the X and the Y. So, for me, that's always been the first obvious step to look at and put on my piece of paper, my computer, or something like that … In my mind, that's what I've been conditioned, either by myself or by school, to do.

Thus, Bennu entered the task with a pre-existing understanding of the appropriate steps involved in graphing, and that understanding, in turn, shaped their structured approach to using Rtutor.AI. Similarly, when asked which school they would recommend for a hypothetical student, Bennu demonstrated an ability to analyze the relationship between the two continuous variables based on Rtutor.AI's output and apply their statistical thinking to the problem's context:

> I would choose Whitney M. Young … they have more average study hours, and I would want my kids to be more driven to study harder and study longer. And they are getting the benefits of it, where they have the higher scores of all the high schools. So I guess I would choose Whitney M. Young, and I would recommend that to the parents.

Bo's interview also demonstrated how prior knowledge influenced engagement with the tool, though in a different way. Unlike Bennu, Bo did not initially frame their prompt around specific graph elements (they simply asked Rtutor.AI to "create a multivariate scatter plot"). When asked what they expected to see, they responded, "I have no idea. Maybe just a scatter plot?" This suggests that Bo entered the task with a less developed understanding of graphing conventions that led them to adopt a more open-ended, exploratory approach.

Furthermore, when asked to interpret the scatter plot created in Rtutor.AI, Bo remarked, "I don't know. I don't really understand what ACT score has to do with the school, but that might just be me. Like, I don't know how to really explain that." This response further suggests that Bo struggled to contextualize the multivariate scatter plot, a limitation that Rtutor.AI alone could not resolve, despite its ability to simplify the graphing process.

Likewise, when asked which school they would recommend to parents, Bo analyzed the two continuous variables independently rather than leveraging the affordances of a scatter plot to examine their relationship: "Probably Whitney Young, just because they're kind of in the middle of hours spent preparing, not too low, not too high. And they kind of have the higher-ish scores … whereas other schools have a more spread-out range." This suggests that while Rtutor.AI facilitated the graphing process, it did not necessarily guide Bo toward a contextualized and integrated evaluation of the tool's output absent further explicit instruction.

Therefore, it appears that a strong conceptual understanding (i.e., knowledge within the statistical world) is a prerequisite for effectively using Rtutor.AI, and that Rtutor.AI cannot on its own replace students' statistical thinking necessary to both generate as well as interpret results.

## 4. Discussion

In this study we investigated the potential of Rtutor.AI to support introductory level statistics students' completion of statistical tasks, specifically whether it is easily usable and whether that usability hinders students' thinking and learning. We intentionally recruited 5 students who completed an introductory level general education statistics course with varying degrees of success, and had them complete a graphing task in Rtutor.AI and interpret the results in order to investigate their thinking.

Overall, we found that Rtutor.AI is a highly effective procedural tool for students with introductory level knowledge of statistics, and no prior programming knowledge. All students found the tool easy to use to achieve their statistical goals. Connecting these findings to our orienting framework (see Figure 1), when utilizing Rtutor.AI to complete statistical tasks, students were easily able to complete steps within the software world. However, the ease of stepping from the statistical world to the software world, or vice versa, appeared to fully depend on students' statistical knowledge.

Furthermore, we found that using Rtutor.AI did not do everything for students, still requiring them to think carefully about the statistical information at hand. That is, Rtutor.AI alone, absent specific instructional scaffolding or other curricular design, was not able to support students lacking in statistical knowledge. These students were unable to clearly formulate a statistical plan, were thus unable to effectively use Rtutor.AI (although were still able to use it easily and efficiently), and were also unable to effectively interpret results from Rtutor.AI. Even though they were able to easily use Rtutor.AI to achieve what they wanted it to, the problem arose because their lack of statistical knowledge led to them not knowing what they wanted Rtutor.AI to do for them. In other words, students lacking strong statistical understanding struggled to take steps out of the statistical world into the software world as well as take steps back into the statistical world from the software world.

However, for students with moderate to strong statistical understanding, there appeared to be a bi-directional relationship between their statistical thinking and their use of Rtutor.AI, with Rtutor.AI's ease supporting iterative investigative behavior that supported their statistical inquiry. For these students, the relative ease with which they were able to step through the software world seemed to facilitate stronger connections between the statistical world and software world, slowly facilitating their ease in stepping out of or back into the statistical world.

## 4.1. Summary of Results

Qualitative analyses of interview transcripts generated five themes related to our primary research questions. We enumerate each of those themes below.

(1) Theme 1: Students' prior statistical knowledge shaped the way they approached using Rtutor.AI. Additionally, Rtutor.AI may have supported the development of students' statistical thinking by facilitating continuous self-reflection on prompts to ensure they

"make sense" and align with goals and expectations – the fact that it was prompts and not programming syntax may have made the connection between statistical concepts and vocabulary an easier connection to make (i.e., near transfer rather than far transfer).

(2) Theme 2: Students benefited from step-by-step problem-solving, a scaffolding known to be widely beneficial for computing education (e.g., Lewis et al. 2018; Priemer et al. 2020).

(3) Theme 3: Students developed procedural skills related to identifying, extracting, and organizing the most important aspects of a problem within Rtutor.AI that both drew on and reinforced their statistical knowledge. In other words, the AI-driven nature of Rtutor.AI does not enable students to "not think."

(4) Theme 4: By lowering barriers to statistical problem-solving, Rtutor.AI may have allowed students to spend more time in the statistical world through its ease of use, thus positively supporting students' statistical thinking. At the same time, however, Rtutor.AI may have limited students' active engagement in statistical decision-making by preemptively completing aspects of the statistical task for them, rather than forcing them to explicate all decisions related to the production of the graph. This was especially true for students who lacked a strong statistical understanding

(5) Theme 5: Students lacking a strong statistical understanding struggled to use Rtutor.AI to generate and interpret graphs effectively, demonstrating that Rtutor.AI could not completely compensate for this lack of understanding.

### 4.2. Implications

While there are many different ways AI tools may help students learn, especially in facilitating information searching as suggested by Ellis and Slade (2023), our empirical results of students' thinking complement the investigations of Bien and Mukherjee (2025) as well as Bray (2024) in using AI tools to support students in conducting statistical analyses.

Like Bien and Mukherjee (2025) discuss in their reflections, we found that students' ability to write specific prompts facilitates ease of use of AI tools. Similarly, we too found that AI tools facilitated a more natural iterative interaction between students and data, as the AI tool enabled them to easily execute new operations for which they had not been explicitly taught prompts or syntax (and without the need to look up such prompts or syntax elsewhere, diverting time away from the statistical task at hand).

Pushing beyond Bien and Mukherjee (2025), by investigating students' thinking, we uncovered new insights into how students' statistical content knowledge interacts with their AI procedural knowledge in the completion of statistical tasks. Perhaps one of the great fears of the implementation of AI tools in the classroom is that it may enable students to thoughtlessly complete tasks, undermining the learning process. Instead, we found the opposite – AI tools can support the development of conceptual understanding for students with foundational knowledge, but did not inherently enable students without such knowledge to be able to make statistical interpretations and conclusions.

Translating these findings for practice, there are three specific implications for future classroom implementation of AI tools that our research suggests:

(1) AI tools will most benefit students with a strong conceptual understanding of statistics. Therefore, a thorough treatment of statistical investigations, including the formulation of statistical plans as well as how to statistically interpret software output, should precede students' use of AI tools to conduct statistical analyses on their own. A strong conceptual understanding should be seen as a prerequisite for the use of AI tools, just as many skill

based certification exams (such as scuba diving or woodworking) require written exams before practical exams.

(2) AI tools are very easy for students to learn, but they may still benefit from direct instruction on how to use AI tools for statistics. Our results indicated that students with moderate levels of statistical understanding benefited by iteratively building a prompt step-by-step—this process helped solidify their conceptual understanding of graphs and the graph creation process. Therefore, direct step-by-step procedural instruction may be especially beneficial for students without a strong conceptual understanding, as it may provide the opportunity for them to make explicit connections between statistical concepts and the prompt procedures. This is likely to be especially true as the complexity of the task increases.

(3) AI tools allow students to spend more time in the 'statistical world' by promoting continuous self-reflection and lowering barriers to statistical problem-solving. As stated by Bien and Mukherjee (2025), AI tools allow students' investigations to be driven by their curiosity, relatively free of technological constraints. Doing so facilitates students' statistical investigative thinking, as they spend more time cycling through all components of the statistical investigative cycle, as well as between the real world and statistical world, in a manner more authentic to statistical practice and expert thinking. Therefore, in courses where the focus is on statistical learning objectives, but that include components of students analyzing their own data, AI tools are likely to be especially beneficial in keeping students' focus on the statistical and real worlds, rather than software.

4.3. Limitations

We believe that the results we found can serve as a helpful empirical basis for research and teaching as we continue to explore the way that AI tools may help or hinder students' learning of statistics. However, our findings must be interpreted with an understanding of the specific nature of the task, which asked students to produce a multivariable scatter plot and to then compare trends between groups. Students were given instruction on scatter plots and trend lines in the course they completed, and were introduced to the idea of interactions. However, instruction did not place a focus on multivariable scatter plots specifically nor in comparing trends between groups.

Therefore, the efficacy of Rtutor.AI as a procedural tool, as well as its limitations as a teaching tool, are likely to vary with the conceptual complexity of the statistical task at hand and based on students' prior knowledge. For example, topics such as hypothesis testing, notoriously confusing for students to understand conceptually (Nickerson, 2000), may give rise to different interactions between students' understanding and Rtutor.AI. Similarly, we expect tasks focused on data wrangling and manipulation to potentially lead to different results. However, future work can explore how students' use of AI tools varies by task, and investigate the instructional scaffolds necessary to support the optimization of AI tools to support students' learning across tasks.

Additionally, we recruited students who had already completed an introductory level course, rather than students currently in an introductory statistics course. While having a strong conceptual understanding is an important foundation for deriving benefits from the use of AI, we believe that with intentional design and alignment between instruction and AI-based activities such as having forced-choice homeworks due before AI-based labs, students will learn the conceptual foundation they need in homeworks before applying it in labs, which will serve to reinforce their knowledge. We see this as akin to the pedagogies employed in teaching students how to read and write. With thoughtful alignment, both skills grow simultaneously. Future research will investigate the pedagogies and scaffolds that might be employed within a course to best support the development of students' understanding and statistical thinking.

We must also comment on the ecological validity of the method by which we collected this data. We found that students with strong conceptual understanding were able to benefit from self-reflection on their Rtutor.AI prompts, as well as benefit from the rapid iterative analysis of the data that Rtutor.AI's usability enabled. However, if this instead were an assignment that students completed at home, some students likely would not have engaged with Rtutor.AI at a similar level in order to reap those benefits. Nevertheless, earnest interactions with Rtutor.AI seem to support students' learning, and future work can explore how best to integrate Rtutor.AI into classroom activities in order to foster such interactions.

### 4.3. Conclusion

AI tools for statistics, such as Rtutor.AI, can be effective as a procedural tool. Furthermore, it is clear that there are ways that these tools can support students' learning of statistics, especially by reinforcing their knowledge through iterative step-by-step investigation of data and its patterns, and through the evaluation and monitoring of AI prompts designed to execute statistical plans that students must necessarily clearly articulate. For students who are unable to clearly articulate a statistical plan, these AI tools are severely limited in their ability to compensate for this lack of conceptual understanding. We believe this is a good thing – it is not easy to be 'mindless' when using AI tools for statistics, and these AI tools cannot think on behalf of students. There is potential in AI tools to support the development of students' statistical thinking, and we continue on the hunt for those pedagogies and scaffolds to maximize AI tools' efficacy in supporting statistics and data science education.

Acknowledgements

The authors would like to acknowledge Dr. Xijin Ge, Dr. Kaylee M. Lukacena, and Dr. Chelsey Legacy for their support and helpful feedback on this project.

Funding

This project is supported by a grant from the University of Illinois Center for Social and Behavioral Sciences.

Disclosure statement

The authors report there are no competing interests to declare.

Data availability statement

Appendix A - Task Instructions, Appendix B - Task Interview Guide, Appendix C - Interview Coding Summary, and Appendix D - Interview Transcripts are openly available at https://osf.io/za9ds/.

AI Statement

No AI tools were used to write this manuscript, nor in any other part of this project, except for the inherent incorporation of AI in Rtutor.AI which is the object of study.

References

Adams, B., Baller, D., Jonas, B., Joseph, A. C., and Cummiskey, K. (2021), “Computational Skills for Multivariable Thinking in Introductory Statistics,” *Journal of Statistics and Data Science Education, 29*(sup1), S123–S131, DOI: 10.1080/10691898.2020.1852139

Arnold, P. (2007), “What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation,” *Education*, 55.

Biehler, R., Frischemeier, D., and Podworny, S. (2015), “Preservice Teachers’ Reasoning About Uncertainty in the Context of Randomization Tests,” In A. S. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning*, Minnesota: Catalyst Press, pp. 129–162.

Bien, J., and Mukherjee, G. (2025), “Generative AI for Data Science 101: Coding Without Learning to Code,” *Journal of Statistics and Data Science Education*, 1–14, DOI: 10.1080/26939169.2024.2432397

Bray, R. (2024), “Lessons Learned When Teaching Data Analytics with ChatGPT to MBAs in Spring 2023,” Available at SSRN 4484395.

Clarke, V., and Braun, V. (2017), “Thematic Analysis,” *The journal of positive psychology*, 12(3), 297-298, DOI: 10.1080/17439760.2016.1262613

Ellis, A. R., and 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, DOI: 10.1080/26939169.2023.2223609

Franklin, C. A., and Garfield, J. B. (2006), “The GAISE Project. Developing Statistics Education Guidelines for Grades pre-K-12 and College Courses,” In G. F Burrill & P. C. Elliot (Eds.), *Thinking and reasoning with data and chance*, Reston, VA: National Council of Teachers of Mathematics, pp. 345-375.

GAISE College Report ASA Revision Committee (2016), “Guidelines for Assessment and Instruction in Statistics Education: College Report,” Alexandria, VA: American Statistical Association. Available at amstat.org/education

Habib, S., Vogel, T., Anli, X., and Thorne, E. (2024), “How Does Generative Artificial Intelligence Impact Student Creativity?,” *Journal of Creativity*, 34(1), 100072, DOI: 10.1016/j.yjoc.2023.100072

Lewis, D. G. R., Gorson, J., Maliakal, L. V., Carlson, S. E., Gerber, E. M., Riesbeck, C. K., and Easterday, M. W. (2018), “Planning to Iterate: Supporting Iterative Practices for Real-world Ill-structured Problem-solving,” in Rethinking Learning in the Digital Age: Making the Learning Sciences Count, eds. J. Kay and R. Luckin, Volume 1, London, UK: International Society of the Learning Sciences, pp. 9-16.

Liao, S. M. (2022), “SCRATCH to R: Toward an Inclusive Pedagogy in Teaching Coding,” *Journal of Statistics and Data Science Education*, 31(1), 45–56, DOI: 10.1080/26939169.2022.2090467

Merriam, S. B. (1988), “Case Study Research in Education: A Qualitative Approach”. JosseyBass.

Moshman, D., and Tarricone, P. (2016), “Logical and Causal Reasoning,”. In Handbook of epistemic cognition, eds. J. A. Greene, W. A. Sandoval, I. Bråten, Routledge, pp. 54-67.

Nickerson, R. S. (2000), “Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy,” *Psychological Methods*, 5(2), 241-301, DOI: 10.1037/1082-989x.5.2.241

Nolan, D., and Temple Lang, D. (2010), “Computing in the Statistics Curricula,” *The American Statistician*, 64(2), 97-107.

Pelletier, K., McCormack, M., Muscanell, N., Reeves, J., Robert, J., and Arbino, N. (2024), “EDUCAUSE Horizon Report, Teaching and Learning Edition,” Boulder, CO: EDUCAUSE. Available at educause.edu

Priemer, B., Eilerts, K., Filler, A., Pinkwart, N., Rösken-Winter, B., Tiemann, R., and Zu Belzen, A. U. (2020), “A Framework to Foster Problem-Solving in STEM and Computing Education,” *Research in Science & Technological Education*, 38(1), 105-130, DOI: 10.1080/02635143.2019.1600490

Rao, C. R. (1975), “Teaching of Statistics at the Secondary Level An Interdisciplinary Approach,” *International Journal of Mathematical Education in Science and Technology*, 6(2), 151–162.

Sardareh, S. A., Brown, G. T. L., and Denny, P. (2021), “Comparing Four Contemporary Statistical Software Tools for Introductory Data Science and Statistics in the Social Sciences,” *Teaching Statistics*, 43(S1), S157-S172.

Sardi, J., Candra, O., Yuliana, D. F., Yanto, D. T. P., and Eliza, F. (2025), “How Generative AI Influences Students' Self-Regulated Learning and Critical Thinking Skills? A Systematic Review,” *International Journal of Engineering Pedagogy*, 15(1), 94-108, DOI: 10.3991/ijep.v15i1.53379

Schraw, G., Crippen, K. J., and Hartley, K. (2006), “Promoting Self-regulation in Science Education: Metacognition as Part of a Broader Perspective on Learning," *Research in science education*, 36(1), 111-139, DOI: 10.1007/s11165-005-3917-8

Tucker, M. C., Shaw, S. T., Son, J. Y., and Stigler, J. W. (2022), “Teaching Statistics and Data Analysis with R,” *Journal of Statistics and Data Science Education*, 31(1), 18–32, DOI: 10.1080/26939169.2022.2089410

Wickham, H. (2010), “A Layered Grammar of Graphics,” *Journal of computational and graphical statistics*, 19(1), 3-28, DOI: 10.1198/jcgs.2009.07098

Wild, C. J., and Pfannkuch, M. (1999), “Statistical Thinking in Empirical Enquiry,” *International statistical review*, 67(3), 223-248, DOI: 10.1111/j.1751-5823.1999.tb00442.x

Woodard, V., and Lee, H. (2021), “How Students Use Statistical Computing in Problem Solving,” *Journal of Statistics and Data Science Education*, 29(sup1), S145–S156, DOI: 10.1080/10691898.2020.1847007

Yilmaz, R., and Yilmaz, F. G. K. (2023), “The Effect of Generative Artificial Intelligence (AI)-based Tool Use on Students' Computational Thinking Skills, Programming Self-efficacy and Motivation,” *Computers and Education: Artificial Intelligence*, 4, 100147.

Zavez, K., and Harel, O. (2025), “Teaching Statistical Concepts Using Computing Tools: A Review of the Literature”, *Journal of Statistics and Data Science Education*, 1–21, DOI: 10.1080/26939169.2024.2445541

Zhou, X., Teng, D., and Al-Samarraie, H. (2024), “The Mediating Role of Generative AI Self-Regulation on Students’ Critical Thinking and Problem-Solving,” *Education Sciences*, 14(12), 1302, DOI: 10.3390/educsci14121302

**Figure 1.** *Framework for statistical thinking integrating the unique aspects of context, statistics, and technology with the statistical investigative cycle*

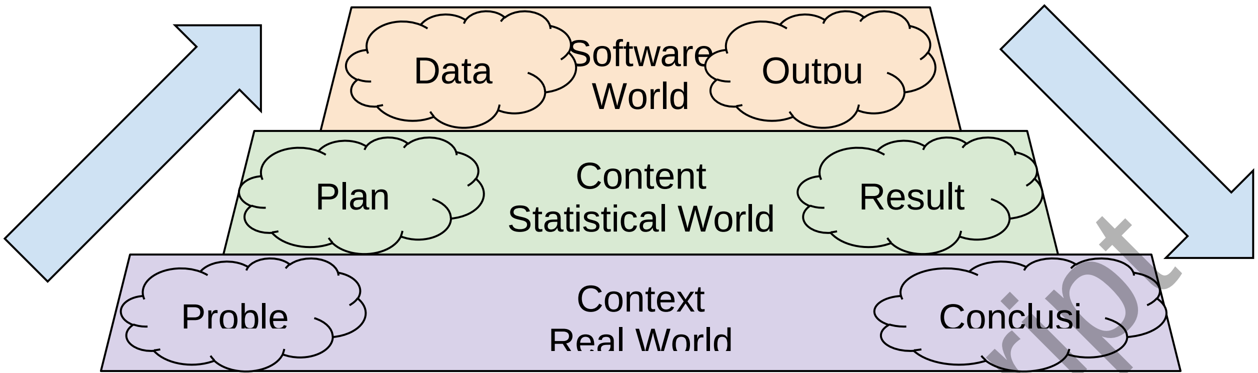

**Table 1.** *Examples of Students' Planning, Monitoring, and Evaluating in the Software World*

| Planning | Monitoring | Evaluating |
|---|---|---|
| "[for] the ACT scores. You want to say like where [it goes]? So I guess [you have to say] 'Represent act scores on the Y-axis'" | "What I've been trying to do is, I've been trying to read it [the prompts] back and see if it makes sense." | "Looks like it's labeled and scaled fine. This … makes sense. Use distinct colors and shapes. Okay, it did that. … Yeah, I think this is fine .. I think this compared with what I had in my head." |

**Table 2.** *Examples of Students' Thinking about Context, Content, and Software*

| Real World Context | Statistical Content World | Software World |
|---|---|---|
| "If these are the best top rated public high schools in Chicago, they're probably not getting an ACT score of like 10" | "I'm trying to like, you know, on the graph where it [the y-axis] starts … decide that. And we'll just start at 20, I guess" | "Represent act_scores on the Y-axis from 25 to 36." |

**Table 3.** *Preliminary and Initial Coding Example*

| Interview Transcript | Preliminary Codes | Initial Codes |
|---|---|---|
| **Interviewer:** So before you hit submit, what is your mindset? Why did you go with trying to copy and paste it there? | | |
| **Participant:** It's the easiest way to put it all together at once. Or I could do … well, I could do just "create a multivariate scatter plot" first and then do that later. Maybe I'll do that: the rest later. | • Copy-and-pasting = easiest way to put it all together at once<br>• Decides partway through to just create a multivariate scatter plot first | • Step-by-step approach<br>• "I could do just 'create a multivariate scatter plot' first"<br>• "...the rest later" |
| **Interviewer:** What do you think is the advantage of doing one maybe versus the other? | | |
| **Participant:** I think doing it one at a time can show you, as the student, kind of how it works—like the steps in creating the graph—whereas, all at once just gets you the end result. So I kind of want to see the steps, so I'll just do one at a time. | • Doing instructions one part at a time shows students how it works<br>• All at once gets you end result<br>• "I kind of want to see the steps" | • Step-by-step shows how graphing works<br>• Prompting all at once "gets you the end result"<br>• "I kind of want to see the steps" |