

# Framework for Integrating Generative AI Into Statistical Training in Doctor of Education Programs

**Christine Eith**  
Johns Hopkins University  
ceith@jhu.edu

**Denise Zawada**   
University of New Mexico  
dzawada@unm.edu

## ABSTRACT

This paper proposes a framework for integrating generative artificial intelligence (AI) tools into statistical training for Doctor of Education (EdD) students. The rigorous demands of doctoral education, coupled with the challenges of learning complex statistical software and coding language, often lead to anxiety and frustration among students, particularly those in part-time or online programs. This article explores how generative AI can serve as a scaffold for learning, potentially mitigating statistics anxiety and enhancing students' abilities to focus on core statistical concepts rather than software intricacies. The proposed framework, grounded in constructivist learning theory, outlines a process for faculty to facilitate dialogues using generative AI tools that support students in developing research questions, selecting appropriate statistical tests, checking assumptions, and conducting statistical analyses. By leveraging AI as a dialogic partner, students can engage in self-regulated learning and enhance critical thinking skills essential for practitioner-scholars. This approach has the potential to improve statistical training in EdD programs, producing more competent translators of research who can effectively apply and interpret statistical methods in their professional practice. The article concludes by discussing implications for EdD programs and suggestions for improving the curriculum that includes statistical training.

## KEYWORDS

*generative AI, statistics anxiety, EdD programs, constructivism, self-regulated learning*

## FRAMEWORK FOR GENERATIVE AI IN RESEARCH METHODS AND STATISTICAL ANALYSIS PREPARATION

The rigorous demands of doctoral education have uniquely challenged students in part-time degree programs, such as those entering the Doctor of Education (EdD). Students have been expected to acclimate to the pace of doctoral curriculum, rarely leaving much time to catch up on skills missed during coursework or not part of previous training. Students have also faced a *hidden curriculum*—a set of unspoken norms and expectations related to the doctoral journey—that contributed to a sense of anxiety, isolation, and confusion, especially for students who have matriculated into an asynchronous online program. Anxiety around academic preparation has been further amplified for individuals who joined doctoral programs as career professionals and entered their degree programs with expertise in practice yet had little or no experience in research and statistical analysis, the latter a common skillset required for degree attainment. EdD students must be able to read, interpret, translate, and evaluate research, which has been crucial preparation as they also must develop and conduct a study within their professional context by examining a problem of practice (Schulman

et al., 2006).

Preparing EdD students to use social science research methods has often been coupled with training in statistics. Students must be able to conduct their own data analysis, which requires knowledge of research design, methods, and analysis tools—often software that uses a unique coding language – to complete their research studies. For most EdD students taking a research or statistics course where data analysis was required, it was their first time using code-based statistical software. The inclusion of statistical software packages has had the potential to hinder the development of applied statistical knowledge because students were required to learn the software necessary to run the analysis in a constrained timeframe and for a unique purpose. This process has often been quite distinct from their professional experiences and practices.

Notably, statistical analysis software was typically not required outside of a small number of courses and the dissertation. It is rarely used in a student's professional context. According to student feedback, the outcome of these challenges has often resulted in high levels of frustration and anxiety and low levels of efficacy and authentic learning. Consistent student feedback in course evaluations indicate that they have experienced challenges using the



New articles in this journal are licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by Pitt Open Library Publishing.



This journal is supported by the Carnegie Project on the Education Doctorate: A Knowledge Forum on the EdD (CPED) [cpedinitiative.org](http://cpedinitiative.org)

[impactinged.pitt.edu](http://impactinged.pitt.edu)  
Vol.10 No.1 (2025)

ISSN 2472-5889 (online)  
DOI 10.5195/ie.2025.518

statistical software, often requiring them to invest substantial time in finding additional learning materials and reducing the time available to master the statistical concepts. As novice learners in this area, there is often frustration and anxiety around identifying reliable supplemental resources then spending the time to review and digest the information. This is further exacerbated when students are working outside of hours where faculty are readily available to respond to questions. Students report leaving the course unsure of what they learned with the majority of their time spent learning the software rather than statistical analysis and interpretation.

Ultimately, the research requirements of the program have affected students' progress and have led to attrition, coursework repetition, and extended time to degree completion.

Further, the introduction of statistical formulas and code-based statistical software such as STATA, in which students must be able to write simple lines of code, may have created additional distance between the application of coursework to practice, engendering anxiety and frustration. Therefore, there must be an alignment of the learning in the practice-based doctorate, which builds on and extends their professional expertise and facilitates students' abilities to better engage in their scholarly work and examine their problems of practice. By facilitating the alignment between research methods and data analysis training, this discussion explores how students can leverage generative AI tools to enhance their confidence and skill development in addressing problems of practice within EdD programs.

Generative AI approaches leverage tools that employ learning algorithms to produce prompted responses based on underlying data. Developers have created these tools to generate text, code, and engage in conversations, among other applications. Research has demonstrated that generative AI tools can support learning by providing diverse perspectives, relevant examples, constructive feedback on writing, and guidance in coding language acquisition (Bozkurt, 2023; Kuznetsov, 2024; Radford & Narasimhan, 2018). These tools have offered opportunities for students to learn new concepts or software language that facilitates data analysis alongside a consistent, real-time scaffold of support where students were accountable for evaluating the accuracy of AI-generated responses.

In this article, we have proposed a framework for faculty members to better facilitate generative AI as a collaborative tool for doctoral students to build their research knowledge and data analysis skills while promoting self-regulated learning, which is essential in doctoral education. For this framework, the goal was to address the barriers of software language that challenge students' ability to effectively use statistical software to analyze data and translate findings that affect their professional contexts.

## REVIEW OF THE LITERATURE

The preparation of EdD students in research methods has been an ongoing discussion, especially within CPED institutions. A central challenge in the preparation of scholar-practitioners is bridging the gap between theoretical knowledge and practical application, especially around statistical data analysis. This section explores the context of research methods preparation within EdD programs, highlighting the shift from formulaic instruction to a more practice-oriented model. We then look to the role of Generative AI tools as a

potential scaffold for learners and move to situate this understanding in the theoretical grounding of constructivism.

## Research Preparation and the Doctor of Education

A decades-long debate about research methods for preparing EdD students has led to numerous discussions and program improvements (Page, 2001). Barnett and Muth (2008) noted that research preparation of doctoral students was intertwined with student retention, as national trends indicated that nearly half of doctoral students did not complete their dissertations. Bates and Goff (2012) claimed that students entering doctoral programs without adequate research skills faced challenges in completing their degrees. This is often the case because doctoral students required a clear understanding of how to apply research knowledge to their dissertations. Labaree (2003) suggested incorporating students' professional expertise into doctoral programs as without connections to practical application, research skills alone were insufficient to prepare practitioner-scholars. Haller (1979) has historically emphasized that research should not be merely a means to an end. Education professionals increasingly needed to analyze and interpret data but lacked adequate training (Jimerson & Wayman, 2015). The traditional focus on formula centered statistical instruction and reliance on code-based statistical software packages for quantitative data analysis has hindered the development of practitioner-scholars by disconnecting coursework from real-world applications, often leading to statistics anxiety.

More recently, faculty members leading EdD programs have focused on making research and methods courses more relevant to EdD research (Acton, 2023; Pautler, 2023). By increasing course relevance to students' research and professional context, EdD faculty have supported the development of practitioner-scholars. They have provided opportunities for students to connect issues in their professional context programs with academic program requirements. This more recent approach contrasts sharply with traditional approaches that emphasized learning through statistical formulas and practice problems from textbooks. Such approaches hindered the development of practitioner-scholars by distancing coursework from relevant application and increasing statistics anxiety.

## Statistics Anxiety

Many instructors teaching research methods or statistics courses have observed student anxiety. Research on statistics anxiety, especially related to the use of statistical software, indicated a negative impact on learning and academic outcomes (Alnofaie, 2018; Onwuegbuzie & Wilson, 2010). Online course delivery exacerbated these challenges (Devaney, 2010).

Numerous studies have confirmed high rates of statistics anxiety among students in various disciplines (Chew & Dillon, 2014; Onwuegbuzie & Wilson, 2003; Siew et al., 2019). Students often experienced discomfort with required statistics or methods courses that included statistical analysis due to negative past experiences and low levels of self-confidence (Macher et al., 2012; Zeidner, 1991). Research focused on apprehension towards new situations and problem-solving involving fear of statistical languages, computing anxiety, and seeking help indicated how students manifested these concerns in emotional and physical ways (Cruise et al., 1985; Onwuegbuzie et al., 1997). Tutkun's (2019) research underscored computer skills and software knowledge as key

challenges for students learning statistics, highlighting their anxiety when using unfamiliar software. Students expressed concern about learning new software, and as Alnofaie (2018) adds, there is a greater increase in anxiety when statistical software used a code-based language.

### Generative AI as a Scaffold

Nolan and Temple Lang (2010) advocated for a statistics pedagogy that harnessed computational technology to prepare learners for a data-intensive world. Generative AI offers a promising avenue to meet this vision while also fostering students' critical thinking, problem-solving, and statistical reasoning. By developing these conceptual skills, students, particularly those with limited experience with statistics, can more effectively analyze and interpret educational data. The opportunity presented through the use of AI tools, such as ChatGPT and Claude AI, allows for the creation of an inclusive, level learning environment that does not privilege previous knowledge and experience with statistical software on students' paths to degree attainment.

In an exploratory study where ChatGPT was evaluated on three tasks, Megahed et al. (2024) found that although generative AI could produce statistical code and perform other structured tasks as directed by a prompt, there was still a need for students to use existing knowledge to review the results for accuracy. This finding highlighted the value in using this AI tool resided in the dialogue between humans and the tool, rather than merely seeking a definitive answer from a direct question. This dialogic approach to generative AI afforded a scaffold whereby an AI tool facilitated and co-created learning with the student, which offered students an opportunity to manage statistics anxiety (Wu et al., 2022). This exemplified learning through the constructivist lens where the AI tool serves as an interactive collaborator.

### Theoretical Grounding

This proposed framework for integrating generative AI into statistical training within the professional doctorate was situated in constructivism. This approach required learners to be actively engaged in developing their knowledge of data, analysis, and statistical procedures. Many EdD students entered the program with limited statistics experience. Thus, it was crucial to approach learning as novel to foster knowledge building. This approach is supported by Li et al. (2024) that highlighted how effective scaffolding promoting dialogue, engagement, and reflection improved academic outcomes. Further, critical attention was needed to identify adaptable scaffolds for online learning (Azevedo & Hadwin, 2005).

Learning statistics as an active process involving conceptual and computational elements required time for students to engage with both aspects. Vygotsky's (1978) work highlighted the importance of social interaction among faculty, students, and technology; that learning should be collaborative to extend one's knowledge. Yet, Vygotsky (1978) noted that constructed knowledge required an internal motivation to acknowledge and advance this learning process. Recent studies suggest that generative AI can serve the function as a more knowledgeable other and engage in socially constructed knowledge (Robertson et al., 2024; Stojanov, 2023). Collaborative learning was central to doctoral classrooms but required self-regulated learning through independent scholarship and research projects like the dissertation.

Self-regulated learning has been critical in guiding learning as individuals acquired new information. These learning processes included metacognition, motivation, and strategic action (Winnie & Perry, 2000; Zimmerman, 2008), known as essential for authentic data analysis. Metacognitive processes involved understanding learning preferences, demonstrating resilience, and strategically tackling problems. Students needed to control their learning through initial motivation, goal setting, strategy implementation, and self-observation (Pintrich, 2000). Self-regulated learning is crucial when using generative AI, especially conversational AI tools. Early research on interactive scaffolding showed promise in fostering self-regulated learning (Song & Kim, 2021), but its impact on participation and achievement was unclear. Generative AI, particularly as a dialogic agent, can support and scaffold self-regulated learning (Broadbent & Poon, 2015) and be used to socially construct knowledge (Robertson et al., 2024).

## A FRAMEWORK FOR FACULTY FACILITATION OF GENERATIVE AI IN STATISTICAL TRAINING IN DOCTOR OF EDUCATION PROGRAMS

As generative AI tools continue to permeate our classrooms, educators are confronted with the challenge of integrating this technology into their classrooms responsibly and effectively. The following section examines how faculty can leverage AI tools for learning and the development of critical thinking, especially around statistical concepts. We examine diverse approaches to integrating AI dialogue into coursework and highlight strategies that foster student engagement and model appropriate use of AI tools as a scaffold for learning research methods and statistics. Finally, we bring together our learning around the use of generative AI to offer a proposed framework on using AI tools to bridge the learning gap for EdD students new to learning statistics and using statistical software.

### Faculty Member Facilitation of Effective Generative AI Dialogue

Faculty members can leverage generative AI to help students overcome the challenges of learning statistical software while fostering a deeper understanding of statistical concepts. By emphasizing the use of software as a tool for data analysis rather than a standalone skill, faculty can empower students to focus on the underlying statistical principles. This shift in emphasis can significantly enhance students' learning experiences, especially for those in practitioner-centered programs. Generative AI tools can serve as valuable scaffolds for students as they engage in self-regulated learning. By providing opportunities for students to interact with large language models (LLMs) through conversation and prompt engineering, these tools can shift the focus from mastering coding intricacies to a deeper understanding of statistical concepts. This allows students to concentrate on the core ideas underlying data analysis rather than getting bogged down in the mechanics of software implementation. Generative AI tools have demonstrated the ability to produce accurate code for popular statistical software (R, STATA, SPSS, SAS) (Ellis & Slade, 2023). Therefore, the use of these tools has the potential to mitigate statistics anxiety. This is especially true when students struggle to see the utility or value of statistics in their own work (Chiou et al., 2014; Sloomtaeckers, et al., 2014; Siew et al., 2019).

## Faculty Member Facilitation in Constructing Learning

Faculty members play a critical role in equipping students with the necessary knowledge to effectively engage with generative AI tools for data analysis. Students require a foundational understanding of research design, data structure, and statistical analysis to make informed decisions when interacting with AI tools. To successfully utilize generative AI, students must be able to articulate research questions, identify appropriate variables, understand data structure, and select suitable statistical tests (Ellis & Slade, 2023). Faculty can create course activities that encourage students to interact with generative AI tools. Through this process, students can evaluate the tools' responses to deepen their conceptual understanding.

While statistical software can generate results, students must possess the foundational knowledge to critically evaluate these outputs against established statistical principles and assumptions. We suggest incorporating the following questions into course activities or discussions to help students initiate their data analysis.

- What is my research question?
  - Is this question aligned with my problem of practice?
  - Are my variables appropriate for my research question?
- What is the structure of my data?
  - What is the level of measurement of the variables I want to use?
  - What is my sample size and are missing data a concern?
- Based on my research question, do I know what analysis is needed?
  - Do I know the assumptions of this test?
  - Do I have the appropriate data to perform this test?
  - Do the variables (data) I am using in my analysis meet the assumptions of this test?

Faculty members can effectively integrate generative AI into research methods and statistics curricula by focusing on the critical knowledge needed for students to engage in a dialogue with the technology tool. This approach empowers students to become critical thinkers and creative problem solvers. By grounding students in a solid understanding of their data and appropriate statistical tests, faculty can leverage generative AI to quickly and accurately produce the necessary code. Consequently, generative AI can serve as a thought partner, assisting students in navigating statistical software from data importation through analysis. As students gain statistical confidence, generative AI can support self-regulated learning, particularly metacognition and problem-solving, as they engage in data analysis and interpretation.

Faculty members should model the use of generative AI as an integral part of the learning process. By openly demonstrating how they employ generative AI, such as crafting code for statistical software procedures or developing effective prompts to engage in dialogue, instructors can exemplify the metacognitive skills essential for effective AI interaction. This includes not only showcasing desired outcomes but also the iterative process of refining prompts and evaluating outputs. By modeling these practices, faculty members

can effectively communicate the concept of generative AI as a learning partner rather than a question-answering tool. Students should understand that the tool's value lies in its ability to support their learning process through ongoing dialogue and experimentation.

For instance, when introducing students to non-parametric statistical tests such as the Mann-Whitney, we modeled how to generate initial code using a generative AI tool. We demonstrated how to refine the STATA code based on the tool's output and our understanding of the statistical procedure. This process allowed students to observe how we validated the code's accuracy and made necessary adjustments. By sharing these steps, we aimed to illustrate the iterative nature of working with generative AI and the importance of critical thinking in interpreting and utilizing its outputs.

This example highlights how faculty can transparently demonstrate the use of generative AI as a tool to enhance student learning rather than a substitute for statistical knowledge. Effective use of generative AI necessitates skill in prompt engineering, which must be developed through practice. Faculty should position AI as a collaborative learning tool rather than a source of ready-made answers. By interacting with the AI, students can enhance their critical thinking and problem-solving abilities (Wang & Yin, 2021).

Faculty members have played a critical role in developing the foundational knowledge necessary for successfully integrating generative AI into students' methods and statistical training. By establishing learning objectives that illuminated students' self-efficacy regarding core concepts, faculty empowered students to engage meaningfully with the technology. A part of this scaffolding requires the instructor to model an appropriate dialogue that reflects a set of expectations around the practice of inquiry to foster discursive engagement (Engle & Conant, 2002). Moreover, to facilitate students' learning, faculty may have set up activities to include learning in small groups, which has been shown to be effective (Barron & Darling-Hammond, 2010).

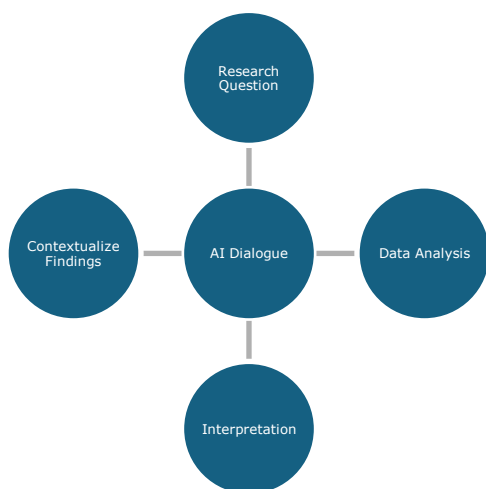
## Generative AI as a Bridge to Learning

A critical factor in successfully using generative AI as a learning bridge is teaching students to engage with it dialogically rather than viewing it as a simple answer provider. Much of the research on generative AI as a collaborator or dialogic peer is grounded in Bakhtin's (1981) work, which proposes that language is inherently dialogic, suggesting generative AI serves as a collaborator or co-creator of shared knowledge (Tang et al., 2024). This dialogic approach has particular benefits for non-traditional students. For those engaged in full-time employment, with families and other responsibilities, having a dialogic tool available on their schedule mitigates feelings of isolation and helplessness that often arise when working independently without immediate access to traditional support structures. Furthermore, for students experiencing statistics anxiety—a common fear or apprehension about working with statistical concepts—this tool provides real-time feedback and support as they engage in new tasks. Appropriate scaffolds, such as guided prompts or reflective exercises, help students realize the affordances of generative AI tools, build understanding and efficacy about their base knowledge, and foster long-term skills in critical thinking and self-directed learning.

### Creating Dialogic Scaffolds

Faculty members can support student learning by guiding them through the utility of generative AI tools for their dissertation work. Integrating AI dialogue into the curriculum through discussions, group work, and assignments provides the most valuable approach. Instructors should utilize both the course and laboratory to train students in the best ways to use generative AI. In Figure 1, we present examples of how instructors can use AI tools in the classroom to scaffold students' understanding: (a) writing and refining research questions, (b) planning data analysis, (c) interpreting statistics, and (d) contextualizing findings. This wide-ranging process offers various opportunities to incorporate AI activities that enhance students' understanding and development throughout the data analysis and interpretation process.

Figure 1. Centering AI Dialogue for Students Examining a Problem of Practice



As students prepare to design their research study, course instructors should scaffold and model effective dialogue with generative AI tools. One of the first challenges involves awareness of the different available tools and choosing among them. Faculty members should clarify that many generative AI tools exist, and the field evolves quickly. Much like choosing between Mac or PC, individuals will have preferences; however, each tool has strengths and weaknesses. Students should identify the tool(s) they use or compare responses from multiple tools to evaluate the outputs.<sup>1</sup>

Prompt engineering, or structuring conversations with the tool, should develop a conceptual understanding that facilitates computational knowledge with AI assistance (Zamfirescu-Pereira, et al., 2023). Users must guide the dialogue, which may include asking for clarification or correcting miscommunicated information. Students must rely on their base knowledge to design appropriate prompts

that produce the most suitable output. With each exercise, students can build their efficacy as prompt engineers and learn how to best engage the AI tool in dialogue around their area of inquiry.

We illustrate the process using an example of dialogue about writing research questions. Often the first step in the research process, developing research questions, frequently challenges novice research students. Creating a question appropriate in scope, clearly articulated, and researchable often requires iteration. Encouraging students to engage in iterative work on their questions with AI tool support provides opportunities to practice dialogue and engage in metacognitive processes to monitor progress. Students must explain the context and purpose of their study to the tool, often requiring clear articulation of their POP and the ability to clarify murky points. As they develop their research questions, students must state what they know and how they think about the concepts. Class discussions about using AI to iterate research questions help students explain their thinking. To aid students in this work, faculty members should consider the following questions to guide class discussion:

1. What did you learn about writing research questions?
2. How did your research question(s) evolve in dialogue with the AI tool?
3. What process did you use to assess or verify the information provided by the AI tool?
4. How would you guide another student through your process?
5. Considering your research questions, what would be the best approach for your data analysis?

As students familiarize themselves with the dialogic process, faculty members can guide discussions toward data analysis and how to use generative AI in learning statistical software. Current generative AI tools, such as Claude AI, ChatGPT, and Gemini, can quickly and accurately provide code or procedures for statistical analyses across most common statistical software packages, including R, Python, SPSS, STATA, and SAS.

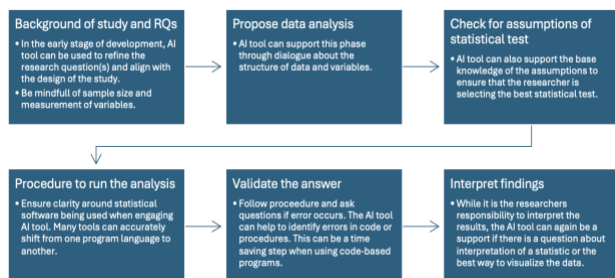
### Bridging the Learning Gap for Using Statistical Software

Figure 2 illustrates a framework for bridging the software learning gap for statistical data analysis. Using this framework, novice researchers can use generative AI to easily navigate statistical analysis in any commonly used statistical software package. The articulation of this process, which often becomes rote memory through repetition, highlights the affordances AI tools provide for novices in the learning process. By beginning with a clear articulation of one or more research questions, students employ metacognitive processes as they engage in a dialogue about the strategic method of analyzing their data with a purpose. The process then moves into identifying the most appropriate statistical test, requiring students to draw upon their base knowledge about statistical tests as they choose. Although the AI tool provides real-time support, students must make the decision and subsequently justify it in their dissertations. We prompt students to make the best choice by checking that the data meet all the assumptions of the statistical test, a step sometimes overlooked, followed by directions on how to conduct the test using their chosen statistical software package. Critically, all students now have access to a tool that enables them to perform statistical analyses in any standard statistical software package. Students still must learn the course

<sup>1</sup> At the time of this publication, Claude is known for its performance in mathematics and coding responses; however, it has only been available in the US and UK.

content, but through the support of generative AI tools, they have real-time assistance when needed for running statistical analyses. They are no longer limited to watching videos and searching for resources. The resource is available and can be used as support until students have enough practice and feel comfortable running the analyses independently.

**Figure 2. Framework for Bridging Software Learning Gap for Statistical Data Analysis with an Example of Generative AI Dialogue for Statistical Analysis.**



**Note.** Dialogue with Claude AI and ChatGPT occurred between April 24 and April 25, 2024.

We produced results from this inquiry using the noted prompts and substantially truncated AI responses for space to illustrate this approach. It is important to note at the time of this article, Claude AI cannot produce tables or graphics. ChatGPT 4.0 would be a better tool for data visualization. In the following example, we highlight the efficacy of the framework for bridging the software learning gap. The dialogue follows the steps outlined in the framework and emphasizes the constructivist roots, grounding this process. We acknowledge that many students entering the dissertation phase of doctoral study are still novices in their statistical knowledge and computational skills and may experience statistics anxiety. This approach aims to reduce anxiety and remove barriers students experience in feeling nervous, anxious, or uncomfortable using or learning new statistical software. This model offers a basic structure of how to guide students through the use of generative AI to refine a research question, explore possible statistical tests appropriate for their research question(s) and data, examine the statistical software to run the statistical test, and serve as a dialogic partner when interpreting the findings.

### Step 1—Study Purpose and Research Questions.

To assist students in developing well-defined research questions, we employed an AI tool (Claude AI) to engage in simulated dialogues. Students were encouraged to input their initial research questions and receive feedback on clarity, specificity, and focus. For example, when presented with a vague research question about bullying, the AI suggested specifying the type of bullying, target population, and desired outcome. This iterative process facilitated the development of more precise and actionable research inquiries.

To illustrate the interactive nature of the AI tool, a sample dialogue between a student and Claude AI is presented below.

*Student:* "I am working on my dissertation study examining the relationship between student absenteeism and bullying. Can you help me refine my research question?" *Claude AI:* "A well-defined research question is crucial in guiding the entire research process and ensuring that the study remains focused and addresses a specific gap in the existing literature."

The initial interaction highlights the importance of a well-crafted research question in guiding the overall research process. To further refine the research question, the student provided additional context:

*Student:* "My research question states 'what is the relationship between student absenteeism and bullying?' How can this question be improved?" *Claude AI:* "Here are some suggestions for improving this research question:

- Specify the direction of the relationship: What is the effect of bullying on student absenteeism? or How does student absenteeism relate to experiences of bullying?
- Define the types of bullying and absenteeism: What is the relationship between physical/verbal/cyber bullying and excused/unexcused absenteeism among students?
- Identify the target population:"

The AI's response demonstrates its capacity to provide specific guidance on refining research questions by suggesting key elements to consider. By breaking down the research question into its component parts, the AI assists students in developing more focused and answerable inquiries.

To improve the clarity of their research questions, this student engaged in an iterative process of refinement using an AI tool. Through this interaction, the student was prompted to critically examine their research questions, receiving immediate feedback on clarity, specificity, and comprehensiveness. The AI tool's capacity to provide concrete examples and suggestions facilitated the development of well-defined research focus and a refined, parsimonious research question.

### Step 2—Identify a Data Analysis Procedure.

Once presented with the research topic and question, this student engaged in a dialogue with the AI tool to identify an appropriate data analysis procedure. The student inquired about suitable statistical tests, and the AI provided recommendations based on the research question's nature. The faculty member facilitated this process by emphasizing the role of language in shaping the research question. For instance, when a student sought a test to examine a relationship between variables, the AI correctly suggested a correlation analysis. The following dialogue highlights a common challenge student researchers encounter when selecting appropriate statistical tests.

*Student:* "Thank you for the suggestions to improve my research question. Since I am looking at a relationship between two variables, what is the best statistical test for me to use?"

In response, the AI offers guidance on correlational studies as an appropriate analytical approach for examining the relationship between student absenteeism and bullying. The AI further outlines specific statistical tests, including Pearson's Correlation Coefficient and Spearman's Rank Correlation, suitable for this type of analysis.

*Claude AI:* "Based on your research question 'What is the relationship between student absenteeism and bullying?', an appropriate statistical analysis would be a correlational study. A correlational study examines the relationship or association between two or more variables without implying causation. In your case, you want to investigate if there is a relationship between student absenteeism (one variable) and bullying (another variable). Specifically, you could use the following statistical analyses:

### 1. Pearson's Correlation Coefficient ( $r$ ):

- This measures the strength and direction of the linear relationship between two continuous variables.

This measures the strength and direction of the linear relationship between two continuous variables. You would calculate the correlation coefficient between your measures of student absenteeism and bullying.

### 2. Spearman's Rank Correlation:

- If your data violates the assumptions of Pearson's  $r$  (e.g., not normally distributed), you can use Spearman's non-parametric correlation. Before conducting the analysis, you should screen your data, check for missing values, assess assumptions (e.g., normality, linearity), and ensure your variables are properly coded and scaled. Additionally, you may want to consider controlling for potential confounding variables (e.g., socioeconomic status, school environment) using partial correlation or multiple regression.

If your data violates the assumptions of Pearson's  $r$  (e.g., not normally distributed), you can use Spearman's non-parametric correlation. Before conducting the analysis, you should screen your data, check for missing values, assess assumptions (e.g., normality, linearity), and ensure your variables are properly coded and scaled. Additionally, you may want to consider controlling for potential confounding variables (e.g., socioeconomic status, school environment) using partial correlation or multiple regression. This dialogue exemplifies the AI's capacity to provide tailored statistical advice based on a researcher's specific query. By offering concrete recommendations and explanations, the AI empowers researchers to make informed decisions about their data analysis strategies and develop a rationale for their choice of statistical test. In this instance, the response emphasized the need to examine the distribution of the data prior to selecting an appropriate test. The student is expected to use their existing knowledge to make the decision on which test is most appropriate. Additionally, the student was guided towards the subsequent step in the framework, which involves checking the assumptions of the chosen test.

### **Step 3—Check the Assumptions of the Statistical Tests.**

The following dialogue highlights the importance of assessing statistical assumptions before conducting an analysis. The student seeks clarification on the assumptions underlying Pearson's correlation coefficient.

*Student:* "I think that the Pearson's correlation is the best choice. Can you tell me what the assumptions are for that statistical test? I want to make sure I have the correct types of variables."

In response, the AI provides a detailed explanation of the key assumptions for Pearson's correlation, including linearity and the absence of outliers. The AI further emphasizes the importance of checking these assumptions before interpreting the results and offers potential solutions for addressing assumption violations.

*Claude AI:* "Certainly, here are the main assumptions for Pearson's correlation:

1. Linearity: The relationship between the two variables should be linear. If the relationship is non-linear, the Pearson correlation may underestimate the true relationship.
2. No Outliers: Outliers can have a substantial influence on the correlation coefficient, inflating or deflating it. Dealing with outliers is advisable."

It's important to check these assumptions before interpreting the Pearson correlation results. Violations can impact the accuracy and meaning of the coefficient. Various data transformations or using non-parametric alternatives may be required if assumptions are violated."

This dialogue demonstrates the AI tool's ability to provide comprehensive statistical guidance, including the crucial step of assessing assumptions. By prompting students to consider the underlying conditions for their chosen statistical test, the AI fosters critical thinking and a deeper understanding of data analysis.

### **Step 4—Procedure to Run the Analysis.**

To illustrate how AI can enhance statistical learning, consider the following interaction where a student seeks assistance in performing a Pearson's correlation using Stata.

*Student:* "I'm having trouble running a Pearson's correlation in Stata. Can you help me with the steps?"

In response, the AI provides clear and concise instructions on how to execute the analysis. The AI outlines the necessary steps, including data formatting, writing the appropriate code, and selecting the desired options.

*Claude AI:* "To run a Pearson's correlation in Stata, follow these steps: (1) Make sure your data is [sic] properly formatted and the variables you want to correlate are numeric and loaded into STATA; (2) The command for Pearson's correlation in STATA is: `pwcorr variable1 variable2, sig`. Replace "variable1" and "variable2" with the names of the variables you want to correlate. The "sig" option will display the significance levels; (3) To create a correlation matrix with all variables in your dataset: `pwcorr *, sig`; (4) If you want to display the number of observations for each correlation: `pwcorr variable1 variable2, obs sig`."

This interaction underscores the potential of AI to serve as an effective learning tool by providing immediate and accessible support for statistical software. By offering clear and concise instructions, the AI empowers students to independently conduct statistical analyses. Moreover, the AI's inclusion of information about data assumptions demonstrates its capacity to support deeper statistical understanding beyond procedural knowledge.

The provided directions offer immediate and accurate information to help students navigate the steps of running a statistical test, while still requiring students to understand the process. For instance, the instructions did not guide students about the choice of a one-tailed or two-tailed test, with two-tailed being the default. This was often going to be the preferred choice, but again, use of this AI tool afforded the opportunity for students to use their existing conceptual knowledge and make a reasoned decision if one was necessary.

Use of the AI tool enabled students to seek information about the appropriate way to use the statistical software for a given test, but also continued to provide feedback and support if an error occurred in the process. Although the tool allowed for the rapid retrieval of instructions on how to utilize the statistical software, in

doing so, it reduced one of the known antecedents of statistics anxiety, learning the statistical software language, i.e., writing the code-based language. The integration of AI tools into assignments can provide faculty members the opportunity to support an important shift in students' thinking: viewing the software language as a step in the process to answering a question rather than being a barrier to overcome.

## CONCLUSION

The integration of generative AI tools into statistical training of EdD students offers the opportunity to significantly shift how we approach research methods and statistical instruction. The proposed framework provides the following implications for EdD programs who teach statistics through the curriculum: (1) enhance focus on core concepts, (2) reduce statistics anxiety, (3) develop critical thinking skills, (4) improve alignment with professional practice, and (5) increase accessibility for students.

First, this framework offers a way to enhance focus on core concepts by alleviating the cognitive load associated with learning complex statistical software languages and allowing students to dedicate more attention understanding fundamental statistical principles and their application in education research. Next, the use of AI as a dialogic partner may help mitigate the anxiety associated with statistics courses, providing the opportunity to improve student engagement and retention in courses and EdD programs. The framework, alongside faculty guidance, also encourages student to engage in metacognitive practices, fostering the development of critical thinking skills essential for practitioner-scholars. In addition, by emphasizing the application of statistical concepts and practices to examine a Problem of Practice, this approach better prepares EdD students to be translators of research in educational practice. Finally, this framework may help level the playing field for students who come into doctoral programs with varying levels of prior statistical knowledge.

This proposed framework for integrating generative AI as a scaffold for statistical training provides an opportunity to transform the EdD by producing more confident and competent practitioner-scholars. We will continue to examine and evaluate the role of AI tools in statistical training and continue to refine this approach to align with the evolving needs of the student we serve. As this is one article in the prolific body of research emerging around generative AI, we propose this framework as an opportunity for faculty to consider how AI tools can be harnessed in support of student learning in statistical training, rather than regulated and excluded from the development of scholar-practitioners.

## REFERENCES

- Acton, K. S. (2023). Reinventing EdD research methods courses: Elevating traditional teaching with transformative learning. *Impacting Education: Journal on Transforming Professional Practice*, 8(4), 3–10.
- Alnofaie, S. S. (2018). Statistics anxiety towards learning new statistical software (Master's thesis). Rochester Institute of Technology.
- Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition: Implications for the design of computer-based scaffolds. *Instructional Science*, 33(5-6), 367–379. <https://doi.org/10.1007/s11251-005-1272-9>
- Barnett, B. G., & Muth, R. (2008). Using action research strategies and cohort structures to ensure research competence for practitioner-scholar leaders. *Journal of Research on Leadership Education*, 3(1), 1–42.
- Barron, B., & Darling-Hammond, L. (2010). Prospects and challenges for inquiry-based approaches to learning. In L. Darling-Hammond (Ed.), *The nature of learning: Using research to inspire practice* (pp. 199–225). Jossey-Bass.
- Bates, P. J., & Goff, L. (2012). The invisible student: Benefits and challenges of part-time doctoral studies. *Alberta Journal of Educational Research*, 58, 368–380. <http://ajer.journalhosting.ucalgary.ca/index.php/ajer/article/view/1054/894>
- Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, 27, 1–13. <https://doi.org/10.1016/j.iheduc.2015.04.007>
- Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, 18(1). <http://www.asianjde.com/ojs/index.php/AsianJDE/article/view/718>
- Chew, P. K., & Dillon, D. B. (2014). Statistics anxiety update: Refining the construct and recommendations for a new research agenda. *Perspectives on Psychological Science*, 9, 196–208. <https://doi.org/10.1177/1745691613518077>
- Chiou, C.-C., Wang, Y.-M., & Lee, L.-T. (2014). Reducing statistics anxiety and enhancing statistics learning achievement: Effectiveness of a one-minute strategy. *Psychological Reports*, 115, 297–310. <https://doi.org/10.2466/11.04.PR0.115c12x3>
- Cruise, R., Cash, R., & Bolton, D. (1985). Development and validation of instrument to measure statistical anxiety. *Proceedings of the Joint Statistical Meetings, Section on Statistical Education* (pp. 92–97). American Statistical Association.
- DeVaney, T. A. (2010). Anxiety and attitudes of graduate students in on-campus vs. online statistics courses. *Journal of Statistics Education*, 18(1). <https://doi.org/10.1080/10691898.2010.11889472>
- 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>
- Engle, R. A., & Conant, F. R. (2002). Guiding Principles for Fostering Productive Disciplinary Engagement: Explaining an Emergent Argument in a Community of Learners Classroom. *Cognition and Instruction*, 20(4), 399–483. [https://doi.org/10.1207/S1532690XCI2004\\_1](https://doi.org/10.1207/S1532690XCI2004_1)
- Haller, E. J. (1979). Questionnaires and the dissertation in educational administration. *Educational Administration Quarterly*, 15(1), 47–66.
- Jimerson, J. B., & Wayman, J. C. (2015). Professional learning for using data: Examining teacher needs and supports. *Teachers College Record*, 117(4), 1–36. <https://doi.org/10.1177/016146811511700405>
- Kuznetsov, A. (2024). The analysis of the efficiency of generative AI algorithms for creating a natural dialogue. *The American Journal of Interdisciplinary Innovations and Research*, 6(6), 26–34. <https://doi.org/10.37547/tajirr/Volume06Issue06-08>
- Labaree, D. F. (2003). The peculiar problems of preparing educational researchers. *Educational Researcher*, 32(4), 13–22. <https://doi.org/10.3102/0013189X032004013>
- Li, Z., Oon, P. T. E., & Chai, S. (2024). Examining the impact of teacher scaffolding in the knowledge building environment: Insights from students' interaction patterns, social epistemic networks, and academic performance. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12535-z>
- Macher, D., Paechter, M., Papousek, I., & Ruggeri, K. (2012). Statistics anxiety, trait anxiety, learning behavior, and academic performance. *European Journal of Psychology and Education*, 27, 483–498. <https://doi.org/10.1007/s10212-011-0090-5>
- Megahed, F. M., Chen, Y. J., Ferris, J. A., Knoth, S., & Jones-Farmer, L. A. (2024). How generative AI models such as ChatGPT can be (mis)used in SPC practice, education, and research? An exploratory study. *Quality Engineering*, 36(2), 287–315. <https://doi.org/10.1080/08982112.2023.2206479>
- Nolan, D., & Temple Lang, D. (2010). Computing in the statistics curricula. *The American Statistician*, 64, 97–107. <https://doi.org/10.1198/tast.2010.09132>
- Onwuegbuzie, A. J., & Wilson, V. A. (2010). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments: A comprehensive review of the literature. *Teaching in Higher Education*, 8(2), 195–209.





- Onwuegbuzie, A. J., & Daley, C. E. (1999). Perfectionism and statistics anxiety. *Personality and Individual Differences*, 26, 1089–1102. [https://doi.org/10.1016/S0191-8869\(98\)00214-1](https://doi.org/10.1016/S0191-8869(98)00214-1)
- Onwuegbuzie, A., DaRos, D., & Ryan, J. (1997). The components of statistics anxiety: A phenomenological study. *Focus on Learning Problems in Mathematics*, 11–35.
- Pintrich, P. R., & Zeidner, M. (2000). *Handbook of self-regulation*. Elsevier Science & Technology.
- Radford, A., & Narasimhan, K. (2018). Improving Language Understanding by Generative Pre-Training.
- Robertson, J., Ferreira, C., Botha, E., & Oosthuizen, K. (2024). Game changers: A generative AI prompt protocol to enhance human-AI knowledge co-construction. *Business Horizons*.
- Siew, C. S. Q., McCartney, M. J., & Vitevitch, M. S. (2019). Using network science to understand statistics anxiety among college students. *Scholarship of Teaching and Learning in Psychology*, 5(1), 75–89. <https://doi.org/10.1037/st10000133>
- Slootmaeckers, K., Kerremans, B., & Adriaensen, J. (2014). Too afraid to learn: Attitudes towards statistics as a barrier to learning statistics and to acquiring quantitative skills. *Politics*, 34(2), 191–200.
- Song, D., & Kim, D. (2021). Effects of self-regulation scaffolding on online participation and learning outcomes. *Journal of Research on Technology in Education*, 53(3), 249–263. <https://doi.org/10.1080/15391523.2020.1767525>
- Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20(1), 35.
- Tang, K.-S., Cooper, G., Rappa, N., Cooper, M., Sims, C., & Nonis, K. (2024, March 8). A dialogic approach to transform teaching, learning & assessment with generative AI in secondary education. SSRN. <http://dx.doi.org/10.2139/ssrn.4722537>
- Tutkun, T. (2019). Statistics anxiety of graduate students. *International Journal of Progressive Education*, 15(5), 32–41. <https://doi.org/10.29329/ijpe.2019.212.3>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Harvard University Press.
- Wang, X., & Yin, M. (2021, April). Are explanations helpful? a comparative study of the effects of explanations in ai-assisted decision-making. In *Proceedings of the 26th International Conference on Intelligent User Interfaces* (pp. 318–328).
- Wu, R., Chen, J., Li, Q., & Zhou, H. (2022). Reducing the influence of perfectionism and statistics anxiety on college student performance in statistics courses. *Frontiers in Psychology*, 13, <https://doi.org/10.3389/fpsyg.2022.1011278>.
- Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023, April). Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-21).
- Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Educational Psychology*, 61, 319–328. <https://doi.org/10.1111/j.2044-8279.1991.tb00989.x>
- Zimmernan, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45, 166–183.