Int J of Sci and Math Educ (2026) 24:19
https://doi.org/10.1007/s10763-025-10634-y

Examining Pre-service Mathematics Teachers' Statistical Investigations: Challenges and Instructional Support

Muhammet Şahal[1] · **Oguz Koklu**[2]

Received: 2 December 2024 / Accepted: 13 November 2025
© The Author(s) 2026

Abstract

The ability to use data effectively has become increasingly essential, and statistical investigations are central to developing this ability. A statistical investigation involves formulating questions, collecting and analyzing data, and interpreting results to draw conclusions. However, pre-service mathematics teachers (PSTs) often face challenges in engaging with the full cycle of statistical investigations, and research examining their progress is still emerging. This study examines the engagement of 26 PSTs who participated in 6 weeks of group-based statistical investigations. Data were collected through audio recordings of in-class group work, classroom discussions, and weekly submissions, and analyzed within a design-based research perspective. Findings show that the problem stage of the statistical investigations was the most demanding for PSTs. In terms of the other stages, formulating investigative questions, forming relevant variables and aligning them to a question, and selecting representative samples required considerable time and influenced subsequent stages of the cycle. Additional difficulties that PSTs experienced included planning data collection and analysis, as well as a tendency to prioritize categorical over quantitative variables. Future studies could focus on designing structured interventions that systematically target the stages of the statistical investigative cycle–especially the earlier ones–by incorporating diverse statistical investigative questions and variable types to foster foundational skills for statistical investigations.

Keywords Statistical investigations · Challenges in the investigative cycle · PPDAC cycle · Pre-service mathematics teachers

✉ Muhammet Şahal
msahal@yildiz.edu.tr

Oguz Koklu
oguz.koklu@bogazici.edu.tr

[1] Department of Mathematics and Science Education, Yildiz Technical University, Istanbul, Türkiye

[2] Department of Mathematics and Science Education, Bogazici University, Istanbul, Türkiye

Published online: 07 February 2026

Introduction

Every day, people encounter situations that necessitate using data to make decisions. Making such decisions accurately requires statistical knowledge. As statistical thinking and problem-solving skills have become increasingly crucial (Bargagliotti et al., 2020), teachers must be well-prepared with statistical knowledge to teach statistics (Groth, 2011). According to Franklin et al. (2015), the successful implementation of statistics "depends to a large extent on the teachers who will bring them to life in the classroom" (p. 1). Research in teacher education indicates that engagement with the statistical investigation cycle is linked to the development of students' foundational statistical skills and data-related reasoning (Burrill, 2023; da Ponte & Noll, 2018; Franklin et al., 2015; International Data Science in Schools Project [IDSSP], 2019; Suh et al., 2020).

In a statistical investigation, the first step is to transform the problem at hand into a statistical question. After data are obtained, they are analyzed using statistical methods, and ultimately, a conclusion is drawn regarding the original problem that initiated the investigation (Bargagliotti et al., 2020). The literature on preservice mathematics teachers' (PSTs) progress in statistical investigative stages remains limited, particularly in understanding how they formulate questions, collect and analyze data, and interpret results (Biehler et al., 2017, 2018; Leavy & Frischemeier, 2022; MacGillivray & Pereira-Mendoza, 2011; Zapata-Cardona & Martínez-Castro, 2021). Engaging in statistical investigations has been argued to provide PSTs with experience that enhances their teaching of statistics (Makar & Fielding-Wells, 2011). Given that PSTs are preparing to teach statistical concepts to students, examining their experiences in statistical investigations is essential. Accordingly, this study aims to explore how PSTs engage with the statistical investigation cycle during a six-week activity by analyzing their progress across all stages. The findings offer insights for enhancing the integration of statistical investigations into mathematics teacher education.

Literature Review

Statistical Investigations

Statistical investigations are central to the discipline of statistics (Watson et al., 2018; Wild et al., 2018), and are recognized as an essential component of statistics instruction (da Ponte & Noll, 2018). Statisticians and statistics educators recommend that students experience and develop proficiency in conducting statistical investigations (Arnold, 2013; MacGillivray & Pereira-Mendoza, 2011). In the literature, various terms have been used to describe student-centered data explorations, including project-based lessons (Suh et al., 2020), statistical (data) modeling (Kazak et al., 2021; Mulligan et al., 2024; Zapata-Cardona & Martínez-Castro, 2021), investigative projects (MacGillivray & Pereira-Mendoza, 2011),

statistical (data) investigations (Forster & MacGillivray, 2010; Zapata-Cardona & Martínez-Castro, 2021), and statistical investigation projects (González et al., 2022).

Research highlights several challenges encountered by students, as well as by pre-service and in-service teachers, when conducting statistical investigations (Heaton & Mickelson, 2002; Makar & Fielding-Wells, 2011; Metz, 2010; Schubatzky et al., 2020). In this context, a *challenge* refers to any barrier that hinders the explorer from advancing in a statistical investigation (MacGillivray & Pereira-Mendoza, 2011). These challenges appear at various stages of the investigative cycle—from posing questions to analyzing data and drawing conclusions. Of these, the formulation of statistical questions represents the first and one of the most crucial stages in developing a statistical investigation (Arnold, 2013; Gould et al., 2017). The Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report (Bargagliotti et al., 2020) stresses the necessity of creating statistical investigative questions (SIQs) and distinguishing them from deterministic questions. While a deterministic question produces a single, fixed response, an SIQ anticipates variability in data within a real-world context (Cobb & Moore, 1997). Thus, initiating an SIQ helps learners recognize the need for data (Estrella, 2017). In addition, an SIQ establishes the focus of a statistical investigation by identifying its parameters, features of interest, and the population, which in turn, informs the data to be collected, the analyses to be conducted, and the conclusions to be drawn (Arnold, 2013; Arnold & Franklin, 2021; Frischemeier & Biehler, 2018). In their examination of statistical investigation tasks designed by PSTs, Casey et al. (2021) found that the tasks designed by many PSTs failed to include an SIQ to drive students' data explorations. Similar to the students' documented difficulties in posing questions (e.g., Arnold, 2008, 2013; Budgett & Puloka, 2019), many PSTs also struggle to construct SIQs, demonstrating limited proficiency in this crucial phase of statistical investigation (Frischemeier & Biehler, 2018; Schubatzky et al., 2020; Yilmaz & Yetkin-Ozdemir, 2024).

Data collection, which usually follows the formulation of SIQs, involves designing surveys and experiments, identifying appropriate sampling methods, and implementing them in real settings. This stage has the potential to form the foundation for analyzing various data types and connecting the statistical question to the conclusions drawn (Estrella, 2017). However, identifying and justifying appropriate samples necessitates understanding representativeness, bias, and variability—concepts that many students find abstract and challenging to apply in real-world situations. Students often rely on intuition and prioritize convenience over representativeness when selecting samples (Jacobs, 1999; Meletiou-Mavrotheris & Paparistodemou, 2015).

The next stage in statistical investigations is data analysis, followed by the conclusion stage, where students are expected to interpret results in context and reflect on the implications and limitations of the investigation. Neglecting the initial stages or lacking experience with data exploration (Hannigan et al., 2013; Makar, 2010; Schreiter et al., 2024) can hinder students' ability to analyze and interpret data effectively (da Ponte & Noll, 2018; Leavy, 2004; Schubatzky et al., 2020). Recognizing that such challenges are not limited to students, many studies have extended this line of inquiry to PSTs. Studies focusing on PSTs' data analysis explorations

demonstrated improved learning outcomes such as recognizing the importance of data exploration (Leavy, 2006), understanding how to use statistics in a real-world context (Confrey et al., 2004), and developing ideas and tools for informal inferential methods (Dolor & Noll, 2015).

Researchers (e.g., Hall, 2011; Makar & Fielding-Wells, 2011; Shaughnessy, 2007) argue that statistics education curricula often emphasize the later stages of the investigative cycle. Instructors frequently perceive the formulation of SIQs and the use of real, mainly primary, data to be time-consuming tasks (da Ponte & Noll, 2018; Hall, 2011; Makar, 2010), leading these stages to be partially or entirely omitted from classroom practice. Such perceptions may also extend to teacher education contexts, where similar priorities and constraints can limit PSTs to engage with initial stages of statistical investigations to follow a similar trend (da Ponte & Noll, 2018; Heaton & Mickelson, 2002; Santos & da Ponte, 2014; Shaughnessy, 2007), indicating a need for structured support for PSTs in posing questions (Leavy & Frischemeier, 2022).

Researchers revealed a similar trend in their analysis of how textbooks present statistical content. Little attention is given to posing a statistical question or considering what data to analyze in order to answer a posed question (Bargagliotti et al., 2020), and textbook tasks commonly overemphasize data analysis (Jones et al., 2015). Tran and Tarr (2018) found a similar trend in their analysis of U.S. high school textbooks. The study revealed the following patterns: (a) the textbooks provide all the statistical questions ready for the investigations; (b) they rarely allow students to collect data; and (c) nearly all tasks require students to analyze data, with most also involving interpretation of the results. Hall (2011) argues that textbook-driven tasks using synthetic, decontextualized data often reduce statistics to disconnected procedures, leaving little room for students to frame meaningful questions or design statistical investigations. Overall, Shaughnessy's (2007) observation appears to remain valid:

"Most of the current statistics education in the United States places a heavy emphasis on the DAC [data, analysis, conclusion] parts of the investigative cycle, but precious little time is devoted in classrooms to the PP [problem, plan] parts" (p. 963).

The various facets of study about the formulation of questions in statistical investigations can also be examined via the perspective of problem posing (PP) literature in mathematics education. Kontorovich (2024) argues that PP occurs whenever students engage with a mathematical situation, which is defined as "circumstances in which learners consider themselves bound to act mathematically" (Kontorovich, 2024, p. 4). Prior research suggests that engaging students in PP tasks could yield substantial learning benefits such as improved motivation, creativity, and problem-solving skills (Kontorovich, 2016). While many studies highlight the difficulties learners encounter when creating well-formulated problems, there are also "success stories" in which students produce problems regarded by themselves or their peers as interesting, meaningful, and cognitively demanding (Koichu & Kontorovich, 2013).

Kontorovich (2024) noted that "despite its almost four-decade history, research remains in the early stages of understanding the phenomenon of mathematical

problem posing" (p. 1). For example, in his seminal work in 1987, Jeremy Kilpatrick asked, "Where do good problems come from?" (Kilpatrick, 1987). More recently, Cai and Leikin (2020) identified future directions for PP research, emphasizing the need to investigate "the processes involved when learners generate their own problems, the instructional strategies that can effectively promote productive problem posing, and the effectiveness of engaging students in problem-posing activities" (p. 228). Given that the research base in statistics education—particularly studies of PSTs' statistical investigations—is considerably less established than that of PP in mathematics education, PSTs' engagement in statistical investigations and the effectiveness of instructional support require further exploration.

Instructional support, defined as a sequence of tasks or teaching interventions implemented within a course or research setting, can enhance PSTs' confidence and ability to conduct data investigations (Leavy & Frischemeier, 2022; Makar, 2010). Incorporating statistical investigative frameworks into instruction can help PSTs navigate these complex tasks by providing a structured guide. Explicit guidance, such as illustrating how to ask a statistical question or form and interpret graphs, is also a common practice. Technological tools play an important role in supporting data exploration, and instructors may adapt these instructional interventions in various ways to address students' specific challenges.

Statistical investigations are essential in teacher education (da Ponte & Noll, 2018; Franklin et al., 2015; IDSSP, 2019) as they provide opportunities for active engagement with data (Hannigan et al., 2013). However, previous research has often fallen short in examining how PSTs engage in these investigations – specifically, how they formulate statistical questions, explore, analyze, and interpret data (Hannigan et al., 2013; Leavy & Frischemeier, 2022; MacGillivray & Pereira-Mendoza, 2011; Schreiter et al., 2024; Suh et al., 2020; Zapata-Cardona & Martínez-Castro, 2021). Therefore, the present study examines PSTs' engagement in statistical investigations within an authentic context by addressing the following research questions:

1. What challenges do pre-service mathematics teachers encounter at each stage of the statistical investigation process when working collaboratively on a real-life problem?
2. How does instructional intervention assist pre-service mathematics teachers in overcoming the challenges that arise at each stage of their statistical investigations?

Analytical Framework

Given that numerous frameworks exist for characterizing and interpreting statistical investigations in statistics education, and they share common essential principles (Davidson, 2023), the present study adopts the Wild and Pfannkuch (1999) framework, which describes the investigative cycle in five stages: Problem, Plan, Data, Analysis, Conclusion (PPDAC).

In the PPDAC statistical investigative cycle, the first stage involves identifying and understanding the problem (the first P in PPDAC). The problem is transformed into

a state by reformulating it as a statistical question, which can then be addressed and answered with data. Franklin and Garfield (2006, p. 350) state that "the formulation of a statistics question requires an understanding of the difference between a question that anticipates a deterministic answer and a question that anticipates an answer based on data that vary." The second stage, planning (the second P in PPDAC), includes developing hypotheses about the problem and identifying the potential data sources. This stage also includes decisions on appropriate measures, variables, and a sampling design. It is important to address how to measure, collect, and record data at this stage. For example, designing a questionnaire and planning the subsequent analysis are tasks typically addressed during this stage.

The next two stages are data (D in PPDAC) and analysis (A in PPDAC). The data stage involves data collection, management, and cleaning data for analysis. The analysis stage includes data exploration, planned analyses, and necessary unplanned analyses relevant to the dataset and the statistical questions. This stage may facilitate hypothesis generation and often relies on data visualization tools to identify patterns or relationships in the data.

In the final stage, conclusion (C in PPDAC), context-based findings are developed to synthesize the investigative cycle. This involves interpreting the results in context, making inferences, communicating findings, and generating new ideas. The stage emphasizes evaluating the validity of explanations based on the data, including assessing whether the analyses address the initial statistical question and acknowledging the study's limitations.

The PPDAC framework is compatible with alternative approaches for characterizing statistical investigations (Arnold, 2008). For example, GAISE I (Franklin et al., 2007) and GAISE II (Bargagliotti et al., 2020) reports for Pre-K–12 grade levels and the Statistical Education of Teachers (SET) report (Franklin et al., 2015) proposed four components of a statistical problem-solving process: (a) formulating a statistical question, (b) collecting or considering data, (c) analyzing data, and (d) interpreting results. Similarly, Lee et al. (2022) listed the steps of an investigation as framing a problem, considering and gathering data, exploring and visualizing data, considering models, and communicating and proposing possible actions. Watson et al. (2018) suggested problem posing, planning for and collecting data, analyzing the data (devising and presenting visual representations and summarizing and reducing data), and drawing conclusions as the dimensions of the practice of statistics. IDSSP (2019) outlined a cycle of learning from data, which includes problem elicitation and formulation, securing the data, exploring the data, analyzing the data, communicating the results, and cycling back to addressing the problem at the beginning. Wild and Pfannkuch (1999) emphasized the importance of planning within the PPDAC statistical investigative cycle, positioning it as a central element of the investigation process (Watson et al., 2023). Similar to other types of inquiry, the PPDAC stages form an iterative cycle, in which results may lead to further questions and subsequent analyses (Wolff et al., 2016).

Methodology

Research Design

We implemented a guided learning environment based on the PPDAC statistical investigative cycle (Wild & Pfannkuch, 1999), revising our instructional plan according to PSTs' progress in their statistical investigations. The study followed a design-based approach (Cobb et al., 2003), in which data were collected and used iteratively to inform and improve the planning and execution of the tasks and the instructions. Each subsequent lesson was adapted based on PSTs' advancement through the PPDAC cycle, ensuring continuous progress in their statistical investigations.

Participants

Twenty-six senior students (20 female and six male) enrolled in a mathematics education program at a large public university participated in the study. These PSTs had completed a compulsory Statistics course in their third year of undergraduate education. The course covered traditional descriptive and inferential statistics, including calculating and interpreting descriptive measures, visualizing distributions using bar graphs, dot plots, and histograms, understanding the normal distribution and the Central Limit Theorem, analyzing bivariate data, and performing confidence interval estimation and hypothesis testing. The PSTs had little to no exposure to statistical software in their prior coursework in statistics. Furthermore, the prior statistics course did not provide students with opportunities such as conducting a survey, hands-on data investigation, or realistic statistical inquiry. In brief, they had no opportunity to develop statistical questions, collect and analyze data, and interpret results in meaningful settings, all of which are critical in this study.

The Implementation and Data Collection

The study took place in a classroom setting, and the participants met weekly for two hours to develop their investigations over six weeks. All PSTs gave their written consent to participate, as required by national legislation and our institution. Kilpatrick (1987) indicated that group collaboration can have a positive, synergistic effect on the hurdles that learners confront when learning to pose problems, though contrary findings have also been recorded (e.g., Koichu & Kontorovich, 2013). Thus, we asked the PSTs to be divided into groups of 3–4, which remained the same throughout the study (hereafter referred to as G1, G2, etc.). We facilitated the discussions during the course by requesting PSTs to clarify or elaborate on their ideas and offering instructional support to the groups or the whole class. We evaluated each week's work, provided feedback (during breaks in class hours and at the end of each week), and finalized the planning of the following week's intervention.

Data were collected in multiple forms. First, groups submitted weekly reports. They also audio-recorded their in-class group work, and the entire class discussion was video-recorded. Table 1 presents each week's in-class work, interventions, assignment, and email feedback. For example, the first week of in-class work began with an introduction to statistical investigations. As statistical investigations require using technology, we introduced the web-based interactive statistical data analysis platforms CODAP (http://codap.concord.org) and iNZight (https://inzight.nz/). These platforms facilitate exploring data thanks to their user-friendly drag-and-drop interface and the rapid generation of statistical measures, tables, and graphs (Casey et al., 2021).

The second week included two different problems: the first served as a warm-up activity, and the second was investigated throughout the study. Following Crespo and Sinclair's (2008) recommendations on PP, we introduced the queue in the dining hall during lunchtime as the problem context for the PSTs' statistical investigation (see Appendix A). Crespo and Sinclair (2008) note that for a student to pose problems, "she or he must have the opportunity to appreciate what might be problematic in a particular mathematical situation. This requires gaining familiarity with the situation and its constraints" (p. 411). These guiding principles are also consistent with what Kontorovich (2024) suggested in his PP studies. According to Kontorovich (2024), an initial engagement with a problematic situation is necessary in ultimate problem posing. In addition, Kontorovich claimed that it is essential to immerse learners in socio-mathematical environments that promote "good" problems. The long queue during lunchtime represented a real-world situation that students had personally experienced during the study period.

Data Analysis

The data analysis followed a two-phase process, serving as an instructional design and analytical lens. As shown in Table 1, we used each week's written reports and classroom interactions to provide group feedback and generate or revise the following week's intervention. We conducted all analyses at the group level, allowing us to record each group's developmental trajectory throughout the intervention while synthesizing performance patterns across groups to inform targeted instructional support.

In the first phase, we systematically reviewed weekly written reports and classroom interactions to identify challenges encountered by the groups and to guide the design or revision of the subsequent week's instructional materials. We traced each group's progress weekly and documented it in short memos. For instance, difficulties in formulating SIQs emerged as a recurring challenge among several groups during the early stages of the implementation. Recognizing this recurring challenge, we synthesized patterns across groups and refined the instructional design by allocating additional time to the problem stage and incorporating supplementary examples and prompts (e.g., the sleep deprivation scenario in Appendix B).

In the second phase, we analyzed the groups' written reports, group audios, and whole-class video recordings using Powell et al.'s (2003) seven-step method. These

Table 1 Weekly Intervention Process

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
In-class work	Briefing about statistical investigation projects. Introducing CODAP and iNZight (import data, create graphs, calculate statistical measures, manipulate variables, and export outputs), and the PPDAC statistical investigative cycle.	Campus Problem as the first/warm-up activity: walking around the campus to observe and investigate issues (30 min.) Brainstorming SIQs (30 min.) and submitting written responses. Introducing the Dining Hall Problem (Appendix A).	Working on the Dining Hall Problem (20 min.). Planning (30 min.) and submitting the project proposal.	Recording and cleaning the data. Analyzing the data (60 min.) Continuing to work on data analysis based on feedback.	Working on the final report (Appendix C) based on feedback after a 30-minute break.	Presenting the group project (15 minutes for each group).
Intervention	No intervention.	Providing examples of SIQs, comparing and contrasting survey questions and SIQs.	Discussing when correlation is possible: numerical bivariate data is needed.	Providing instruction on how to reveal patterns of association in bivariate categorical data and technical support for data analysis.	No intervention.	No intervention.
Assignment	Forming groups of 3–4 students.	Observing the dining hall and posing an SIQ about the dining hall.	Collecting data according to the plan.	Continuing to analyze data.	Asking the groups to provide the surveys in their reports. Preparing the first draft of conclusions and interpretations.	Preparing the final report and revising the final form to resubmit.

Table 1 (continued)

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Email	No email feedback.	Sending a review document (Appendix B) to assist in overcoming the difficulties of posing SIQs and providing feedback on the identified issues by sending emails tailored to each group.	No email feedback.	No email feedback.	Sending emails to each group with feedback about the first draft (Appendix D).	No email feedback.

steps included (a) viewing the recordings attentively, (b) describing the recordings, (c) identifying critical instances, (d) transcribing selectively, (e) coding, (f) building a storyline of the issue under investigation, and finally (g) composing a narrative. Our goal was to obtain a profile summarizing each group's performance and developmental trajectory across the stages of the PPDAC statistical investigation cycle. These profiles enabled us to track progress within each group while facilitating cross-group comparisons. Adopting the PPDAC as a framework helped us identify each group's challenges and determine the types of instructional support needed to address them. The first researcher conducted a second review of the recordings, annotating them with notes and timestamps to highlight the most critical segments illustrating the main issues under investigation. These segments were subsequently transcribed and coded for analysis.

In the initial two weeks of implementation—the warm-up and problem stages—we analyzed how each group conceptualized the problem and constructed SIQs, drawing on Arnold and Franklin's (2021) analytical framework. Within this framework, a good SIQ meets the six criteria in Table 2.

In addition to the criteria for a good SIQ, Arnold and Franklin (2021) suggested three types of SIQs: descriptive, comparison, and association. Descriptive questions summarize a single data set, comparison questions evaluate differences between subsets of data across a common categorical variable (e.g., male and female), and association questions explore relationships between paired variables.

In the plan stage, the analysis focused on each group's strategies for sampling, defining variables, and designing data collection procedures, drawing on both written reports and audio recordings. During the data stage, each group's decisions concerning data collection, cleaning, and preparation for analysis were examined. In the analysis stage, each group's use (and selection) of statistical software and representations (e.g., graphs and statistical tests to be employed) were evaluated based on their SIQs. Patterns in misunderstanding were aggregated across groups to inform whole-class reviews, including clarifying distinctions between bivariate quantitative data and the Chi-squared test, as well as verifying the assumptions underlying these tests. In the conclusion stage, we focused on how well each group synthesized the results and related them to their SIQs, data collection, and analysis.

The coding was open-ended after the first stage, as no rubric or checklist was available to guide the subsequent stages. An open coding approach was applied first, and then we compiled codes into a list and checked each group's data against these

Table 2 The Criteria for a Good SIQ

Criteria
The variable(s) of interest is/are clear and/or available.
The population or interest group is clear.
The intent is clear, e.g., summary/descriptive, comparison, association.
The answer to the question(s) is possible using the data.
The question(s) is worth investigating, as it is interesting and has a purpose.
The question(s) do/does allow for analysis of the whole group(s).

emerging codes. We then tabulated the codes and their occurrences across groups and constructed a storyline following Powell et al.'s (2003) seven-step method in conjunction with the PPDAC statistical investigative cycle. This approach enabled us to trace how PSTs engaged with statistical investigations as a group over time, identify challenges at each stage, and assess how targeted instructional interventions supported their development of statistical investigations.

Findings

In this study, we examined the challenges PSTs encountered during their statistical investigations and how targeted instructional support addressed these challenges at each stage of the PPDAC cycle. Accordingly, we present the findings based on the research questions and the PPDAC statistical investigative cycle.

The PSTs' Challenges in Statistical Investigations

Problem

In the problem stage, we observed challenges that could be grouped under two themes: difficulties with variables and issues in selecting the population and a representative sample. Table 3 summarizes the deficiencies observed in each group's SIQs.

As shown in Table 3, most groups experienced difficulties in identifying or appropriately incorporating the variables of interest into their SIQs. For example, G3 proposed including gender, year at university, and monthly income as the variables in their investigation. These variables, however, were absent in their SIQ, "How do the undergraduate students at X University use their spare time?" A similar issue was evident in G4's SIQ, "What alternative methods can help X University students reduce their accommodation costs and make housing more affordable?" Although the group identified variables such as type of residence (with levels: students' shared house, parents' house, and dormitory), rental cost, district, and gender, they failed to incorporate them into the SIQ. This challenge forced some groups to change the

Table 3 The Criteria and the Group Performances

Criteria	The group(s) with deficiencies
The variable(s) of interest is/are clear and/or available.	1, 3, 4, 6, 7
The population or interest group is clear.	2, 5
The intent is clear, e.g., summary/descriptive, comparison, association.	4, 6
The answer to the question(s) is possible using the data.	4
The question(s) is worth investigating, as it is interesting and has a purpose.	1, 3, 6
The question(s) do/does allow for analysis of the whole group(s).	-

problem context. For instance, G1 decided to alter their problem context to investigate the relationship between students' vegetarian and non-vegetarian menu preferences and the faculties they are enrolled in. The group ultimately abandoned the second scenario and returned to their initial context as one member reported during the class discussion: "We cannot merge the variables here [referring to their SIQ]. We will have to deal with a lot more stuff. I think we should go back to the first scenario."

In the problem stage, challenges primarily centered on defining the population, selecting a representative sample, and ensuring the question's worthiness. For instance, G3's SIQ during the warm-up activity focused on students' year at university, preferred lunch time slots, and dining hall floor during lunchtime. However, for the "students on campus" population, the group suggested "students who use the dining hall" as a sample. Although all groups revised their SIQs and variables for the dining hall problem after completing the task, G3 and G5 struggled to identify the population and a representative sample. G3 used the exact phrase to describe both the population and the sample; G5 defined the sample as "Faculty of Education students who use the dining hall at noon," which could be the population, as they were only interested in students at this faculty. It should be noted that students using the dining hall do not represent the entire campus population, and that Faculty of Education students dining at noon do not represent all students who dine at that time. In other words, we concluded that the groups' proposed samples were not unbiased, which we also address in the Plan stage below. In addition, some of the questions were unlikely to be worth investigating, as they lacked a clear purpose or did not anticipate variability. For instance, G6's "Are you happy with the university's cafeterias?" was not an SIQ but could only be a survey question. To sum, the problem stage indicated two common challenges: (i) coordinating variables with SIQs, and (ii) identifying an appropriate sample and population, ensuring that the sample represents the population. The disconnect between the variables and the SIQs made it difficult for researchers to derive the purpose of the posed questions and the investigations.

Plan

In the planning stage, the PSTs needed to determine how to collect and analyze data for the dining hall context, which we summarize in Table 4. As the table shows, the groups aimed to collect data using in-person and online questionnaires from 50 to 500 students (for populations ranging from approximately 15,000 to 30,000). The rationale for the sample sizes chosen by the groups was based on practical considerations, such as their willingness or capability to achieve those numbers. G7, for example, initially planned to collect data from 150 students, but reduced it to 100 due to time constraints and the group members' availability. One group member explained their rationale regarding sample size expectations:

"I thought we needed to aim for a minimum sample size of 100. We could go up to 200 or even 300 if we want, but let's be realistic, we may not have the time to gather that much data. We might have other stuff going on that week. So, let's stick with 100 for now."

Table 4 The Group Proposals for the Investigations

Group	Sample Size	Sampling Method	Data collection	Data analysis	Software
G1	300	Systematic	Online questionnaire & database	Descriptive statistics	CODAP
G2	200	Systematic	Observation & interview	Descriptive statistics & chi-squared test	CODAP
G3	50	Convenience	Questionnaire & interview	Descriptive statistics	CODAP, iNZight, & Excel
G4	500	Convenience	Online & in-person questionnaire	Descriptive statistics	CODAP
G5	150	Cluster	Online questionnaire	Descriptive statistics & chi-squared test	iNZight & Excel
G6	50	Simple random	Online questionnaire & Observation	Chi-squared test	iNZight
G7	100	Convenience	Questionnaire	Descriptive statistics	iNZight & Excel

The groups aimed to collect data on categorical variables such as students' gender, faculty enrolled, day of the week, and time slot for lunch. Five groups identified the preferred lunch menu as contributing to the lunchtime queue length. Notably, variables employed by each group were categorical; no group could devise numerical variables. While several groups intended to conduct correlational studies by exploring relationships between variables, this was likely due to a misunderstanding regarding which statistical analyses were suitable for the specific variable types. As a result, they encountered difficulties in selecting suitable analysis methods for their SIQs. The following interaction between G2 and one of the researchers illustrates the group's confusion:

G2: "Can we establish a correlation between the faculty to which students belong and the number of students?"

R2: "To do a correlation, you need to collect data on two variables from the same participants, I mean, you need bivariate data, and they [the variables] both need to be numerical. Since yours sounds like categorical data, a descriptive statistical approach or chi-squared test might be more suitable. Besides, I do not see a second variable in your inquiry. For correlation, please secure bivariate data that are both quantitative."

The above dialogue, along with similar inquiries from other groups early in their investigation, indicated that participants struggled to select appropriate data analysis methods. The presence of frequencies (i.e., the counts in each level of the categorical variable faculty) appeared to prompt some groups to contemplate conducting a correlational study. Upon realizing this approach was inappropriate for the data they collected, the groups opted for descriptive statistical methods or chi-squared tests for data analysis. The following dialogue from G2 illustrates the difficulty the group experienced in identifying the design of their study:

S1: We cannot do this based on gender, right? However, we can do correlation by faculties, can't we?

S2: The direction of the linear relationship between two variables. You know, we were saying that correlation approaches 1 or 0.

S1: Is that with binary variables?

S2: Yes.

S1: Gender?

S3: Isn't it just male and female?

S2: I do not think ours would be a correlation anyway.

S3: That is true.

S2: Hmm, if we want to do something with faculties and gender, but doing it with faculties and numbers is a correlation. Faculty and the number of students is correlated.

S3: The teacher told me that it falls into descriptive statistics.

S2: Yes.

S3: Look, descriptive statistics is about collecting, summarizing, and analyzing numerical data.

S1: Aren't we already looking at that? However, our data will be so small that the mode, median...

S2: Are those used in correlation?

S1: No, that is descriptive statistics. We will arrange it accordingly.

S2: Is the mode or the median important for us?

S3: We are actually looking at distribution.

S1: What type of data analysis is suitable for gender? Write that down. [Asks the group member to online search for the appropriate data analysis method for the case at hand] For gender distribution? I want to say bar chart, is that okay? I guess those also count under distribution.

S3: I am searching online now. Hımm, only the categorical variable comes up.

S1: Professor, we are a bit stuck with data analysis. Can we say distribution by faculty, for example?

As the dialogue shows, the group aimed for a correlation but could not proceed as they had chosen categorical variables. In addition to a mismatch between variables and the intended study design, several groups reported difficulties in selecting an appropriate sampling method aligned with their data collection plan. The following exchange with G7 illustrates this challenge:

G7: "Is this sampling stratified?"

R1: "No, here you preferred convenience sampling. You have selected one (out of many) department."

R2: "This would be a biased sampling. First, you need to define the population correctly. Simple random sampling requires a list of all students, and the selection is random, i.e., each selection is equally likely and independent."

In conclusion, some groups chose sample sizes and sampling techniques arbitrarily without offering statistical justification. They also encountered difficulties distinguishing between variable types, particularly in differentiating frequencies in categorical variables from quantitative variables. The confusion—especially the challenge in generating appropriate quantitative variables—appeared to hinder the selection of suitable data analysis methods for their investigations.

Data

During class discussions and end-of-study presentations, the groups reported that data collection was generally challenging. Since students in the dining hall were reluctant to be surveyed, the groups often failed to reach their intended sample size. Recognizing the challenges of in-person data collection, G4 opted for an online survey. However, this decision introduced another complication: the risk of duplicate data arising from both online and in-person surveys. Detailing their data collection process, the group provided the following justification:

"We originally wanted to get data from 500 students using the easiest way, which is called convenience sampling. We sent a link to an online survey, but only 200 students responded. We decided to get data from 300 students in person to reach our target sample. This could have led to some students taking the survey multiple times, so we decided to collect their student ID numbers to prevent that. However, some students abstained or did not want to give us their ID numbers because they thought we could identify them."

During the data collection stage, Group 3 showed some awareness of the challenges of obtaining 'good' data and made compromises accordingly. The group acknowledged this challenge as a reflection:

"During the data collection process, it was sometimes difficult to communicate with people one-on-one. We encountered situations in which surveyed students thought we were people the university administration assigned. As a result, they were telling us about the different kinds of problems they had related to the campus or university."

G1 also reported making some adjustments to their sampling method in order to complete data collection:

"We planned to use systematic sampling during the data collection, but we had great difficulty. We came across students who did not want to participate in the study, were in a hurry, or did not have spare time because they came with a large group of friends. According to our systematic sampling rule, we were to select one student every minute. However, this method did not allow us to include all students who arrived simultaneously as a group. Despite these difficulties, we somehow managed to collect data from one student per minute as a group."

Analysis

The challenges encountered at the analysis stage varied in type and severity across the groups. For G1, G2, and G7, the analyses aligned with their data collection and organization, ensuring consistency with their SIQs and plans. G3 and G4, however, needed additional work to prepare their data for analysis. By contrast, G5 and G6 deviated substantially from their original plans and, as a result, addressed different SIQs than they had initially intended. For example, G5 intended to apply a chi-squared test for the categorical variables time slot and grade level. Despite prior advice from researchers to increase the sample size to meet the assumptions of a chi-squared test, G5 attempted to analyze data from only 80 subjects. The group could not run a chi-squared test as they had categorical variables with several levels. Moreover, they could not conduct a cluster sampling as they had planned. Collecting data from only three out of 17 departments within the faculty may have resulted in sampling bias and reduced the generalizability of the results. G6 encountered a similar challenge after raising their sample size from 20 to 55 students. The group decided to use descriptive statistics. One group member provided a rationale: "We use descriptive analysis to reflect the views of the individuals interviewed or observed strikingly and to interpret whether the answers directly obtained through the questionnaire are suitable for our question." The group's rationale for using descriptive statistics was not compatible with the statistical question that the group initially formulated: "Do the menu selections and chosen time slots for Elementary Mathematics Teaching 3rd and 4th year students at the university's central campus, Faculty of Education contribute to the long student queues in front of the dining hall?" As a result, G5 and G6 attempted to modify the investigation process to accommodate the data they had collected, rather than adhering to the original plan. The challenges encountered by G5 and G6 provide evidence of their difficulties in understanding

statistical investigations, particularly in deciding on an appropriate sample size for a predetermined statistical test.

Because their data were not ready for analysis in their current form, G3 and G4 needed to reorganize the data. Although the groups were able to construct rectangular datasets, where rows represent cases or entities, and columns represent variables, they needed to create new levels (categorical values) for some of the variables (the day variable for G4 and the preferred main dish for G3). For instance, G3 anticipated only vegetables, meat, and chicken as levels for the preferred primary dish variable. However, after data collection, the group realized that the levels for the preferred main dish were insufficient in covering the varied responses in the dataset. As given in Figure 1, some of the surveyed students from whom G3 collected data listed more than one main dish. When G3 began to analyze their data in a CODAP spreadsheet file, the software treated each response as a distinct level. CODAP automatically determines the type of graph to generate based on the attributes selected for visualization. As a result, the group increased the number of levels to include answers such as meat-chicken, chicken-vegetable, meat-chicken-desert, and all. One group

	A	B	C	D	E	F
1	**İsim**	**Cinsiyet**	**Menü Tercihi**	**Gittiği gün**		
2		Erkek	Et	Her gün		
3		Erkek	Tavuk	Her gün		
4		Erkek	Et, tavuk	Her gün		
5		Erkek	Hepsi	3 gün	Pazartesi Salı Perşembe	
6		Erkek	Hepsi	Her gün		
7		Erkek	Hepsi	2 gün		
8		Erkek	Tavuk	Her gün		
9		Erkek	Et, tavuk	Her gün		
10		Erkek	Et, tavuk	2 gün	Salı Perşembe	
11		Erkek	Hepsi	Her gün		
12		Erkek	Et, tavuk	2 gün	Salı Cuma	
13		Erkek	Et, tavuk	2 gün	Salı cuma	
14		Erkek	Tavuk, tatlı	3 gün	Salı Çarşamba Perşembe	
15		Erkek	Et, tavuk, tatlı	3 gün		

	B	C	D
1	Okul Numarası	Yemekhaneyi hangi günler kullanıyorsunuz?	Aşağıdaki ana yemeklerden hangisi ya da hangileri yemekhaneyi tercih etmenizi sağlıyor?
2		Pazartesi, Salı, Çarşamba, Perşembe	Baklagil, Sebze, Tavuk, Hamur işi
3		Pazartesi, Çarşamba, Perşembe	Kırmızı Et, Sebze, Tavuk
4		Pazartesi, Salı, Çarşamba, Perşembe	Kırmızı Et, Baklagil, Sebze, Tavuk, Hamur işi
5		Pazartesi, Salı, Çarşamba, Perşembe	Tavuk
6		Pazartesi, Salı, Perşembe	Tavuk, Hamur işi
7		Pazartesi, Çarşamba, Perşembe, Cuma	Kırmızı Et, Tavuk
8		Pazartesi, Salı	Tavuk
9		Salı, Çarşamba, Perşembe	Sebze, Tavuk, Hamur işi
10		Pazartesi, Salı, Çarşamba, Perşembe, Cuma	Kırmızı Et, Baklagil, Sebze, Tavuk, Hamur işi
11		Pazartesi, Salı, Çarşamba, Perşembe, Cuma	Tavuk, Hamur işi
12		Pazartesi, Çarşamba, Perşembe	Sebze, Tavuk
13		Pazartesi, Salı, Çarşamba, Perşembe, Cuma	Kırmızı Et, Tavuk, Hamur işi

Figure 1 Rectangular datasets submitted by Group 3 (upper panel) and Group 4 (lower panel). *The information in the first column was removed as it provided identifying information about the surveyed students. Column C in the upper and column D in the lower spreadsheet illustrate the issue

member suggested revising the initial levels of the variable based on the distribution of responses observed in CODAP: "CODAP displayed the levels in different colors, but I think there are numerous combinations such as meat-chicken-vegetable, meat-vegetable-chicken, and meat-vegetable. We should separate them accordingly." This modification enabled a more robust representation of their collected data.

To conclude, groups that performed better at the problem and plan stages successfully collected data and then prepared their data for analysis, whereas other groups encountered several challenges at this stage. Some groups recognized that their data required modifications and created new levels for their categorical variables to perform appropriate data analysis. Additionally, some groups adapted their data organization and investigation processes according to the data they obtained. Although the groups demonstrated some proficiency in gathering and arranging data, they generally struggled to conduct analyses that were consistent with their SIQs and planned procedures.

Conclusion

The groups organized their final reports around the problem, plan, and data sections; however, the conclusion stage frequently lacked coherence. For example, G2 presented graphs and their interpretations separately, making it difficult to discern the relevance of the graphs to their statistical investigation. Additionally, G5 and G6, who experienced challenges in data analysis, struggled to explain what their graphs showed. Furthermore, the conclusions of G5 were largely unrelated to their SIQ and the data analysis. Contrary to their inability to conduct a chi-squared test, the group attempted to offer interpretations regarding the independence of variables. One group member explained:

"We studied the association between students' undergraduate programs and menu preferences. We found that very few people preferred vegetarian menus. Looking at these two variables and their rates, we can easily see that they are independent and that there is no relationship between students' undergraduate program and their menu preference."

Similarly, G2, after applying the chi-squared tests, concluded that there were no significant relationships between day-gender and faculty-gender variables. Nevertheless, they speculated about potential gender-related dependencies across faculties: "According to the graph, we saw that the percentages for girls in the Faculty of Science and Literature are high. There might be a dependency compared to other faculties."

Instructional Intervention Provided to the PSTs During Statistical Investigations

Problem

Following the warm-up activity, several shortcomings were observed in the PSTs' SIQs, particularly concerning variables, sample, and population (see Table 2). To address these issues, we provided group-specific feedback on the

questions they formulated in preparation for the dining hall investigation. In response to requests from some groups for a good SIQ example, we emailed all groups an SIQ formulation for a sleep deprivation context (see Appendix B). This example included clearly defined SIQs, relevant variables, the target population, and the sample. Table 5 shows the groups' SIQs before and after the instructional support.

As shown in Table 5, some groups' initial questions in the warm-up activity lacked relevant variables, a clear purpose, or a population. Furthermore, G5, G6, and G7 were unable to formulate questions that were in the proper question format. After receiving feedback via email and observing the dining hall, the groups' questions demonstrated noticeable improvement. As shown in Table 5, each group developed their SIQs further by incorporating relevant variables and specifying the population.

Plan

During the planning week, researchers primarily focused on providing feedback to each group. For example, G1 asked how to collect data about the variable "days of the week" for their statistical investigation. In response to the group's data collection plan, the researchers offered a practical recommendation to improve validity: "Either you need to go and observe every weekday, or if you are going to collect data only one day, a question can be added to your questionnaire to learn which days the surveyed students typically visit the dining hall." In pursuit of their goal to examine the distribution of students based on day and menu variables, G4 sought guidance on a suitable data analysis method. The researchers advised them to start with descriptive statistics for this purpose. G2, G3, and G7 wondered if they could examine a correlation between categorical and quantitative variables. To clarify this common source of confusion, the researchers offered instruction by contrasting designs appropriate for the chi-squared test and correlation analyses. Because several groups demonstrated confusion about examining relationships between predetermined categorical variables in their SIQs, the researchers addressed the entire class with the following clarification:

"You seem to think there needs to be a correlation when we collect multivariate data. Please note that these data need to be paired and quantitative. Although not valid for each group, you might have the following situation: Let's say you have gender and menu preference information from the subjects. These are not numerical variables, right? These variables are categorical, and the ones that appear as numbers in your dataset are just frequencies of these categorical variables. [Having] Frequencies do not make these data quantitative! What we can do with this data is, maybe, a chi-squared test of independence."

In summary, instructional support during the planning week was mainly about analyzing the association between categorical variables and how PSTs can succeed in the overall data collection process. However, no discussion arose on sampling methods and sample size as the groups did not raise any issues.

Table 5 Statistical Questions Before and After the Instructional Support

Group	Before	After
1	Do students studying at the faculty of political sciences prefer the ivy-mantled road?	On which days of the week do students enroll in different faculties studying at X University X Campus and use the dining hall between 12-1 pm?
2	What are the problems experienced by people using the library?	How are the students who prefer a vegetarian menu in the X University X Campus dining hall distributed according to gender and faculty?
3	How do the undergraduate students at X University use their spare time?	How does the daily changing menu content of the campus dining hall affect the dining hall preferences of male and female students and crowdedness?
4	What alternative methods can help X University students reduce their accommodation costs and make housing more affordable?	Which days and menus do X University X Campus students prefer for lunch?
5	The grade level of the students studying at X campus of X University and the time slot, floor, and food menu they prefer during lunch break*	In which time interval, which menu do the students of X University Faculty of Education who use the dining hall at lunchtime prefer, and at which grade level are they?
6	Are you satisfied with X University cafes?**	How does the type of food served in the dining hall of X University, X campus, and the working hours of the dining hall affect the number of students waiting in the queue?
7	The relationship between X University students' clothing expenditures and the type of accommodation*	How many days a week and due to which menu do male and female students who use the X University X campus dining hall between 11–14 prefer the dining hall?

*The groups did not use a question format.

**The group posed a survey question instead of an SIQ.

Data

During the data stage, we supported the groups by demonstrating the capabilities of statistical software. This included assisting with data cleaning, reformatting the data for analysis, displaying distributions using graphs, and reformatting the graphs, such as organizing axes, adjusting colors, and splitting bars. Since all groups worked with categorical variables, we introduced contingency tables, marginal distributions, and the formulation of hypotheses in a chi-squared test situation. Then, we presented graphical representations and reviewed how to interpret the graphs, tables, and statistical test results in context. The instructional support was helpful in the groups' work in the analysis stage. Figure 2 shows G2's graphical representations.

G2's question was "How are the students who prefer a vegetarian menu in the dining hall at the university campus distributed according to gender and faculty?" The group used stacked and side-by-side graphs for the relationships, although they had no previous experience using this type of graph or conducting a chi-squared test. Their null hypothesis was "There is no dependency between the faculty and gender of students who prefer a vegetarian menu." The p-value of 0.5429 showed that the two variables are independent.

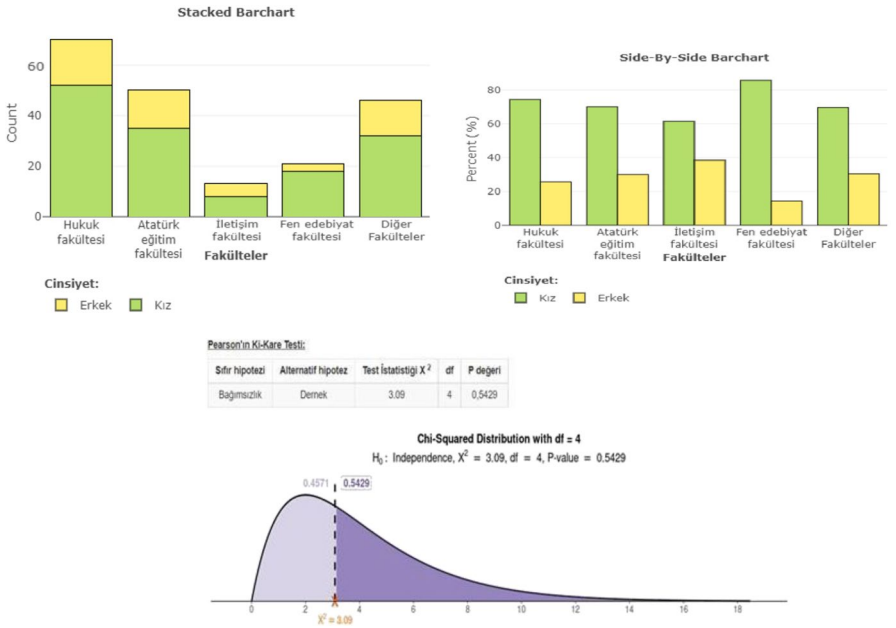


Figure 2 The Graphical Representations by Group 2

Analysis

We provided immediate feedback to the groups as they conducted their analyses. For instance, G1's analysis was based on the difference between the faculty-to-university and dining hall-to-student ratios, with a 5% threshold value as shown in Figure 3. They asked the researchers whether the 5% threshold value was appropriate. In response to the group's question about determining the meaningfulness of differences in descriptive statistics, the researchers provided the following clarification:

"Such a number may not exist in descriptive statistics. However, if we analyze data using inferential statistical methods such as hypothesis testing, we can determine whether there is a significant difference. Suppose you are not going to analyze data using inferential statistical methods. In that case, you may consider taking a value such as 5% difference and state that those with more than 5% difference are taken as significant differences specific to your study; those with less than 5% difference are not."

The group aimed to determine whether students from a faculty use the dining hall remarkably more than any other faculty. Given that the total number of students is not equal across the faculties on campus, the group needed to generate a measure to weight each faculty's contribution based on its total number of students. In Figure 3, the blue bars represent the ratio of students using the dining hall relative to the total number of students surveyed. In contrast, the orange bars represent the proportion of students in each faculty relative to the total number of students at the university. According to the group's analysis, the faculties circled in red contributed most to the congestion in the dining hall. For example, the blue bar for the Faculty of Education represents 40%, while the orange bar represents 22.92%. Because the difference between the blue bar and the orange bar (17.02) was greater than the group's predetermined threshold value (5%), the group concluded that the Faculty of Education was one of the faculties that contributed most to the dining hall's crowdedness. Regarding the challenges G5 and G6 encountered, there was no longer any possibility of returning to the investigation process. Therefore, we told them that even though

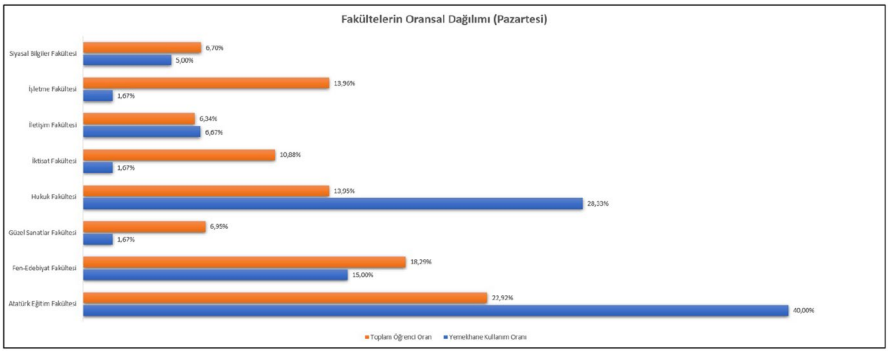


Figure 3 G1's Graphical Representation Based on the Ratios

they could not apply the chi-squared test due to small sample sizes, they could include some graphical representations.

In summary, groups that devoted sufficient time to thoroughly complete the problem, plan, and data stages were able to analyze the data with minimal modifications. Conversely, the analysis stage proved more challenging for groups that had struggled in the earlier stages.

Conclusion

We sought to provide feedback to enhance the clarity and coherence of the groups' reports in the conclusion stage of the PPDAC statistical investigative cycle. We asked the groups to align their graphs with their interpretations to improve their reports' readability for both the intended audience and the researchers, thereby facilitating more precise conclusions. The groups incorporated graphs, tables, and interpretations into their work. We also asked them to include surveys they used to strengthen the credibility of their investigations.

Additionally, we provided feedback on a group basis. For instance, at the initial stage of the investigation, G1 claimed that newly relocated faculties on campus might explain the dining hall congestion; however, they later rejected the claim based on their findings. However, they failed to provide alternative reasons for the long queues. In addition, the formulation method for the ratios they used was unclear. We asked them to discuss the reasons for the congestion and clarify their formula by giving a mathematical equation. G4 did not specify how many of the 525 data points were collected in person versus online. G7's report revealed inconsistencies between their plan and sampling method. Similarly, G2 had discrepancies between their test results and interpretations. After feedback, each group addressed these issues and made corrections or added more explanations to their reports. In summary, while the groups effectively presented the problem, plan, and data collection stages in their reports, the analysis and conclusion required revisions. With the help of instructional support, the PSTs clarified the essential points in their investigation in the final two stages.

Discussion

This study examined PSTs' engagement in statistical investigations through the PPDAC statistical investigative cycle. At the outset, the groups struggled to transform problems into statistical investigative questions, but with feedback and review of worked examples, they gradually refined their question formulations to some extent. However, they often overlooked statistical theory when determining sample size and sampling methods during the planning stage. Data collection and the organization of datasets for analysis also posed difficulties, highlighting the critical role of careful planning in statistical investigations. We observed deviations from the groups' initial SIQs and planning as they progressed in their investigations. The findings also suggested that, with the support of researchers, PSTs could progress through the PPDAC statistical investigative cycle, albeit not linearly (Wild

& Pfannkuch, 1999). Some groups needed more iteration, especially until the data collection process was completed. Additionally, the groups progressed at varying speeds based on their performance at the initial stages of the investigations.

Formulating SIQs proved the most time-demanding task. As Watson et al. (2018) suggest, posing good SIQs requires considerable time and effort. This outcome resonates with the findings in problem-posing literature in mathematics education. For example, Kontorovich (2024) argued that initial outputs in newcomers' problem-posing activities fail to meet the expected standards. The difficulty we observed in formulating questions, especially during the initial stage, also aligns with findings from previous statistics education research (Arnold, 2008, 2013; Burgess, 2007; Frischemeier & Biehler, 2018; Gal & Ograjenšek, 2016; Leavy & Frischemeier, 2022). Our study, however, underscores the crucial role of devising variables in formulating SIQs. The disconnect between variables and SIQs appears to make it difficult for the PSTs to derive the purpose of the questions posed and the investigations. Similar difficulties have been raised in the PP literature, with ambiguity and missing givens being identified as sources contributing to poor quality problem-posing (Kontorovich, 2016). In statistics, data collection (e.g., sampling plan), analysis, and reporting of results become ambiguous and confusing if an SIQ is not well-formed (Gal & Ograjenšek, 2016). Our study showed that the deficiencies not comprehensively addressed in posing SIQs and planning stages can severely affect the later stages of statistical investigations. In addition, focusing on variables during the problem stage could help PSTs formulate SIQs and, thus, align survey questions and SIQs.

Our findings suggest that the challenges with generating SIQs may be linked to PSTs' challenges in identifying appropriate variables, samples, and populations in context, thereby affecting their ability to make accurate sample-to-population inferences. Moreover, the absence of statistical justification for sample size and sampling methodology – as noted by Jacobs (1999), Lovett and Lee (2018), and Meletiou-Mavrotheris and Paparistodemou (2015) – may have further impeded their success in statistical inference. Although PSTs may have been exposed to statistical inference in prior coursework, they need practical support in applying these concepts when formulating SIQs. The challenges encountered by G5 and G6 reveal a gap in their statistical investigations: deciding on the appropriate sample size for the predetermined statistical test. One possible explanation for these findings is that the PSTs lacked a solid understanding of the purpose of random selection and the role of sampling variation in statistical inference (Hannigan et al., 2013; Lovett & Lee, 2018), limiting their ability to critically evaluate whether their results could be generalized beyond the sample. This gap may also account for their failure to consider conducting an appropriate statistical test to determine whether the observed patterns in the data could be attributed to chance.

The planning stage was crucial for identifying variable types and analysis methods to answer SIQs. Some groups, however, lacked an understanding of the critical role of planning. Although our initial assumption was that the dining hall context could accommodate various variable types, the PSTs generated only categorical variables. The absence of quantitative data in the groups' investigations may be explained by the fact that categorical data are generally more intuitive to students

(Koklu, 2017; Koklu & Kaplan, 2023); accordingly, statistics instruction often begins with categorical variables (Bargagliotti et al., 2020; Franklin et al., 2007; MacGillivray & Pereira-Mendoza, 2011). Categorical variables can be analyzed in simple ways related to counts (e.g., frequencies, percentages, and cross-tabulations), whereas quantitative variables necessitate relatively more detailed procedures (Koklu & Kaplan, 2023; Ograjenšek & Gal, 2016). Categorical data collection is achieved through counting, while quantitative data collection is more tedious as it necessitates using a counting or measuring device, such as a stopwatch to measure the duration of time in queues. Furthermore, similar to Casey et al. (2021), PSTs in this study often misinterpreted the observed frequencies of categories as quantitative variables. Requests from PSTs to calculate the correlation between two categorical variables or between one categorical variable and one quantitative variable indicate challenges in distinguishing variable types. Overall, these findings highlight the importance of distinguishing quantitative from categorical variables (Casey et al., 2021), which may support appropriate study design and interpretation (Ograjenšek & Gal, 2016) and allow for a broader statistical investigation experience.

The data stage can present several challenges, as also noted by previous researchers (Davidson, 2023; MacGillivray & Pereira-Mendoza, 2011. Groups that excelled in the problem and plan stages found the data stage less challenging, and as a result, demonstrated confidence and ease during the analysis stage. This pattern suggests a potential link between effectively transforming a problem into a well-formulated SIQ and planning data collection and analysis. However, some PSTs arbitrarily chose sample size and sampling methods, indicating a lack of understanding of the impact of data collection in statistical investigations. Studies reported students' difficulties comprehending the sample's role and sampling method in statistical investigations (Ben-Zvi et al., 2015; Casey et al., 2021; Zieffler et al., 2008). Although we did not have an opportunity to observe the PSTs while they collected data, we noticed how groups struggled to organize the data, altered data collection methods, and even abandoned their original SIQs. Gelman and Nolan (2017) argue that having students collect their data can expose them to the practical challenges of data collection and the incidental decisions that must be made. Sole and Weinberg (2017) add that students can see firsthand how difficult it can be to obtain good data and discover the importance of inspecting and cleaning data for analysis. Statistical investigations may necessitate working with vague and often ill-defined problems and messy data (Wild & Pfannkuch, 1999). We concluded that the PSTs had an opportunity to start to understand the role of data collection, sample size, and sampling method through their work in this study.

The conclusion stage of the investigation cycle revealed that the PSTs, with the support of the researchers, were able to structure their reports and draw conclusions from the data. In response to a call by Kontorovich (2020), Watson et al. (2018), and Crespo and Sinclair (2008), we intentionally asked the PSTs to work on a familiar, everyday-life context at the university. While most of the analyses remained primarily at a descriptive level, the PSTs occasionally went beyond presenting their results by offering justification and argumentation for necessary changes based on their findings. One of the reasons why the analyses remained largely descriptive may be the PSTs' limited understanding of the inferential nature of statistics, which

prevented them from using their sample data to make generalizable claims about the population. Some groups engaged in more in-depth interpretations and supported their claims with evidence from the data. The affordances of the groups' attempt to draw inferences from the data could be attributed to the fact that contextual knowledge plays a crucial role in making statistical interpretations, as reported in previous studies (Biehler et al., 2018; Budgett & Pfannkuch, 2018; English & Watson, 2018; Kazak et al., 2021; Zapata-Cardona & Martínez-Castro, 2021).

Recommendations and Implications

The study suggests that PSTs may need more targeted support and guidance (Casey et al., 2021), especially when posing SIQs and planning a statistical investigation. Providing the PSTs with feedback and structured support –such as examples of well-formulated questions for specific scenarios, exemplar statistical investigation tasks from credible sources (e.g., Casey et al., 2021; Statistics Teacher: https://www.statisticsteacher.org), and targeted guidance on linking SIQs to variables (and their levels), samples, and populations–may help them recognize the importance of formulating questions in statistical investigations (Arnold & Franklin, 2021; Yilmaz & Yetkin-Ozdemir, 2024). These supports could enable PSTs to engage in authentic scenarios and gain first-hand experience in data collection and analysis (Davidson, 2023), to work with both quantitative and qualitative variables (Kazak, 2023), to apprehend essential statistical concepts such as sample and sampling method (Anhalt et al., 2018; MacGillivray & Pereira-Mendoza, 2011), and to discover the nonlinear and iterative nature of statistical investigations (Doerr, 2007; Wild & Pfannkuch, 1999).

While statistical investigations expose students to essential practices, the investigations may be insufficient on their own to support learning. Incorporating frameworks such as the PPDAC statistical investigative cycle can guide students in reviewing and refining their processes, drawing insights, and progressing systematically through each stage of an investigation. Future research could explore how particular stages of the PPDAC statistical investigative cycle contribute to students' statistical reasoning and where targeted instructional support is most beneficial (Santos & da Ponte, 2014). It is important to identify which stages of investigation are likely to pose greater challenges for specific groups of learners and determine the types of interventions that are most effective.

A key challenge observed in our study was the formulation of SIQs. Given student confusion between chi-squared tests and correlational designs, explicit attention to SIQs appears promising for enhancing PSTs' statistical investigations. Arnold and Franklin (2021) emphasized the potential of modeling well-structured SIQs through precise language and think-aloud strategies. Although our intervention included descriptive, comparative, and relational question types, it was not embedded within an explicit instructional framework. Presenting association questions may help learners differentiate between correlational and associative contexts, while comparison questions illustrate the link between categorical and quantitative variables. Instructional support in this area may also help students identify appropriate

variables and prepare for subsequent stages of an investigation. Future studies could more systematically address different types of questions at the problem stage and examine whether learners can better distinguish between statistical and non-statistical questions

The study explored the PSTs' engagement in statistical investigations based on in-class work and dialogues. Therefore, the study did not include how out-of-classroom factors, attitudes, dispositions toward statistics, and theoretical knowledge differences among PSTs might influence their engagement in statistical investigations. Future studies could elucidate the role of these individual characteristics.

Appendix A: Dining Hall Problem (Main Scenario)

X Campus is the main campus of X University and houses students from many faculties. Since the Fall 2022 semester, the number of students on X Campus has increased remarkably due to the transfer of students from other campuses. However, the facilities in student-serving areas such as the dormitory, gym, dining hall, library, ATM, gate, and dining hall have remained the same. As a result, long queues have formed in front of the dining hall, especially during lunch break, in recent days. We expect you to design a statistical investigative question for this problem and answer the question with data by following the project steps.

The Form for Posing a Statistical Investigative Question

Statistical Investigative Question	
Population	
Sample	Sample
	Sampling Method
	Sample Size
Variable 1*	Name & Description**
	Variable Type
	Variable Unit***
Variable 2	Name & Description**
	Variable Type
	Variable Unit***

*You can identify more than three variables./**Variable: If a particular definition is needed for the variable, please write it down./***Answer the following question: What values can the variable take?

Appendix B: Sleep Deprivation Scenario

Posing a statistical investigative question is the first and most critical step in any research study. Posing a meaningful statistical investigative question is especially challenging for people with no previous research experience. For this reason, we

will illustrate how to pose a statistical investigative question through an example scenario.

Scenario: Sleeping Deprivation (Adapted from Arnold & Franklin*,* 2021*)*

Administrators and teachers within a large US school district are concerned about the perceived lack of sleep among middle and high school students. More than ever, students seemingly struggle to stay awake and focused during the school day. Parents have expressed concerns that too much homework is assigned. The school officials believe several potential factors contribute to sleep deprivation among the students, other than academic reasons. These include the number of extra-curricular activities middle and high school students are actively involved in and the time spent on the internet and electronic devices. The school district decides to survey selected middle and high school students in the large US school district to investigate the potential factors contributing to sleep deprivation among their middle and high school students.

1. What is a possible statistical investigative question that will assist the school officials in investigating and better understanding the sleep deprivation of these middle and high school students?

Statistical Investigative Question:

What is the number of extracurricular activities that middle and high school students in the large US school district actively participate in during the school year?

We can ask ourselves some interrogative questions to understand how good this question is. For example;

How does knowing the number of extracurricular activities help us to investigate the causes of insomnia among middle and high school students?

Besides the number of extracurricular activities, is there other information about extracurricular activities that we should consider?

How can the time spent on extracurricular activities be taken into account?

Are there other variables associated with sleep deprivation and other research questions we should ask?

Are high school students more likely to engage in extracurricular activities than middle school students?

Now that we have created a statistical investigative question, we start thinking about how we can collect data for this statistical investigative question. Surveys are often used for this step. Survey and statistical investigative questions have different structures and are often confused. A statistical investigative question asks to be answered with data, while a survey question aims to collect data.

Survey question:

How many extra-curricular activities do you actively participate in during the school year?

To understand how good this survey question is, we can ask ourselves some interrogative questions. For example;

Will students accurately recall the extracurricular activities they actively participated in?

What does it mean to participate actively?

Which school year do you mean, my current school year, or what I did in middle school (high school student opinion)?

Will finding the number of activities help to answer the question about engagement in extra-curricular activities?

Should we ask about the time spent on extra-curricular activities?

We improved the first statistical investigative question and wrote two more, thanks to these questions.

Three different statistical investigation questions:

1. How much do middle and high school students participate in extracurricular activities during the school year in a large school district in the United States?
2. Does the amount of time high school students spend on extracurricular activities tend to be greater than the amount of time middle school students in the large US school district spend on extracurricular activities?
3. Is there a relationship between the number of extracurricular activities in which high school students participate and the typical hours of sleep on a weeknight in a large US school district?

Appendix C: The Final Report Template (Adapted from Spence et al. (2017)).

Statistical Investigation Project Conclusion Report Form

1. Problem

This section should include the statistical investigative question you have posed. The purpose of the statistical investigative question is to guide you to think about the variable (and its units), population, sample, and sampling method with justifications for each.

2. Plan

Explain how you have planned the statistical investigation process. The distribution of tasks among group members should be clearly outlined. In particular, it should be stated whether the planning you have made and the implementation you have realized match or not; if there are parts that do not match, the reasons should be stated.

3. Data

In this section, the method of data collection should be specified. It should be explained which data collection tools were used (questionnaire, interview, observation, etc.), which devices/instruments were used in the data collection process, and which software was used in the analysis (CODAP, iNZight, etc.) (Google form, computer, phone, counter, camera, etc.). It should be indicated how long the data collection process took and how tasks were shared among the group members to facilitate the process.

4. Analysis

In this section, it should be explained which statistical methods (descriptive, chi-squared, correlational, etc.) were used for the variables in the research question and which statistical models (graphs, tables, formulas, etc.) were used to

achieve the purpose of the investigation. It is also necessary to explain in detail how the raw data is organized and cleaned (sorting, outliers, handling the missing data, etc.) and how it is made ready for analysis.

5. Conclusion

This is the section that includes interpretations of the situation that emerged as a result of the analysis, what the data you obtained might imply to someone who knows the investigation, and possible stories behind the data. Whether the data you have collected on different variables confirm each other or not is also one of the issues that should be included in this section.

Appendix D: Examples of Email Exchanges

Email sent to Group 1:

20.12.2022

Hello X,

I am sending you the statistical investigative question-posing form to agree on with your group members. Dr. K. has updated this form to help you pose more qualified statistical investigative questions. I ask you to review this form with the group members until the day of the lesson and create your statistical investigative question by the next lesson. You are free to update the statistical investigative question you wrote last week.

Mail from Group 1:

20.12.2022

Hello Sir,

We made a few changes in light of what you said, but some questions remained. We wanted to consult you. Would it be appropriate to sample in this way? We aim to see which faculty students constitute the majority during lunch hours. We decided to analyze this not only by the number of students but also by analyzing what percentage of the faculty uses the dining hall, according to the number of faculty students given in the 2021 Annual Report of X University. Does specifying this percentage also require defining a variable? We would like to get your opinions on this issue as well. Thank you very much.

Have a good day.

Email sent to Group 1:

20.12.2022

Hello X,

The statistics stated in the annual report are the actual results of the population. You are trying to reach results that can represent the population through the statistical investigation process you plan to conduct. For this purpose, you will design and manage this investigation process. As for the sampling method, systematic sampling could be regarded as a random sampling method. You need to justify how you will systematize the selection you will make among all the elements of the universe (such as the ten people I get every 5 minutes, the people who come across us every 10 minutes, etc.). The percentages will emerge after the data you obtain, according to the research process you have designed. After

that, you can compare the results of your investigation and the actual report. In this respect, percentages are not defined as variables here. There is no incompatibility between the identified variables and the research question.

I wish you good work…

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Compliance with Ethical Standards

Ethics Committee Approval Ethical approval for the study was obtained from Marmara University Institute of Educational Sciences Research and Publication Ethics Committee.

Competing Interests We declare no potential conflicts of interest concerning this article's research, authorship, and publication.

Informed Consent Written informed consent was obtained from all participants to participate in this study in accordance with the national legislation and institutional requirements.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Anhalt, C. O., Cortez, R., & Bennett, A. B. (2018). The emergence of mathematical modeling competencies: An investigation of prospective secondary mathematics teachers. *Mathematical Thinking and Learning, 20*(3), 202–221. https://doi.org/10.1080/10986065.2018.1474532

Arnold, P. (2008, July). What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation. In *Paper presented at the 11th International Congress on Mathematical Education (ICME-11), Monterrey*.

Arnold, P. (2013). S*tatistical investigative questions—An enquiry into posing and answering investigative questions from existing data* [Unpublished doctoral dissertation]. The University of Auckland. Retrieved from https://researchspace.auckland.ac.nz/handle/2292/21305

Arnold, P., & Franklin, C. (2021). What makes a good statistical question? *Journal of Statistics and Data Science Education, 29*(1), 122–130. https://doi.org/10.1080/26939169.2021.1877582

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education (GAISE) report II*. American Statistical Association and the National Council of Teachers of Mathematics.

Ben-Zvi, D., Bakker, A., & Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics, 88*, 291–303. https://doi.org/10.1007/s10649-015-9593-3

Biehler, R., Frischemeier, D., & Podworny, S. (2017). Elementary preservice teachers'reasoning about modeling a "family factory" with TinkerPlots-a pilot study. *Statistics Education Research Journal, 16*(2), 244–286. https://doi.org/10.52041/serj.v16i2.192

Biehler, R., Frischemeier, D., & Podworny, S. (2018). Elementary preservice teachers' reasoning about statistical modeling in a civic statistics context. *ZDM Mathematics Education, 50*(7), 1237–1252. https://doi.org/10.1007/s11858-018-1001-x

Budgett, S., & Pfannkuch, M. (2018). Modelling and linking the Poisson and exponential distributions. *ZDM - Mathematics Education, 50*(7), 1281–1294. https://doi.org/10.1007/s11858-018-0957-x

Budgett, S., & Puloka, M. (2019). Making sense of categorical data-question confusion. In *Paper presented at the IASE 2019 Satellite Conference*. Retrieved from https://iase-web.org/documents/papers/satellite2019/IASE2019%20Satellite%20117_BUDGETT.pdf

Burgess, T. (2007). *Investigating the nature of teacher knowledge needed and used in teaching statistics* [Unpublished doctoral thesis]. Massey University.

Burrill, G. (2023). An international look at the status of statistics education. In G. F. Burrill, L. de Oliveira Souza, & E. Reston (Eds.), *Research on reasoning with data and statistical thinking: International perspectives* (pp. 11–16). Springer. https://doi.org/10.1007/978-3-031-29459-4_2

Cai, J., & Leikin, R. (2020). Affect in mathematical problem posing: Conceptualization, advances, and future directions for research. *Educational Studies in Mathematics, 105*, 287–301. https://doi.org/10.1007/s10649-020-10008-x

Casey, S. A., Harrison, T., & Hudson, R. (2021). Characteristics of statistical investigations tasks created by preservice teachers. *Investigations in Mathematics Learning, 13*(4), 303–322. https://doi.org/10.1080/19477503.2021.1990659

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801–823. https://doi.org/10.1080/00029890.1997.11990723

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13. https://doi.org/10.3102/0013189X032001009

Confrey, J., Makar, K., & Kazak, S. (2004). Undertaking data analysis of student outcomes as professional development for teachers. *ZDM - Mathematics Education, 36*(1), 32–40. https://doi.org/10.1007/BF02655755

Crespo, S., & Sinclair, N. (2008). What makes a problem mathematically interesting? Inviting prospective teachers to pose better problems. *Journal of Mathematics Teacher Education, 11*(5), 395–415. https://doi.org/10.1007/s10857-008-9081-0

da Ponte, J. P., & Noll, J. (2018). Building capacity in statistics teacher education. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 433–455). Springer. https://doi.org/10.1007/978-3-319-66195-7_14

Davidson, A. (2023). A review of the use of investigative projects in statistics and data science courses. *Journal of Statistics and Data Science Education, 32*(2), 188–201. https://doi.org/10.1080/26939169.2023.2240385

Doerr, H. M. (2007). What knowledge do teachers need for teaching mathematics through applications and modelling? In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI study* (pp. 69–78). Springer.

Dolor, J., & Noll, J. (2015). Using guided reinvention to develop teachers' understanding of hypothesis testing concepts. *Statistics Education Research Journal, 14*(1), 60–89. https://doi.org/10.52041/serj.v14i1.269

English, L. D., & Watson, J. (2018). Modelling with authentic data in sixth grade. *ZDM - Mathematics Education, 50*, 103–115. https://doi.org/10.1007/s11858-017-0896-y

Estrella, S. (2017). Enseñar estadística para alfabetizar estadísticamente y desarrollar el razonamiento estadístico [Teaching statistics to develop statistical literacy and statistical reasoning]. In A. Salcedo (Ed.), *Alternativas Pedagógicas para la Educación Matemática del Siglo XXI* (pp. 173–194). Centro de Investigaciones Educativas, Escuela de Educación. Universidad Central Venezuela.

Forster, M., & MacGillivray, H. (2010). Student discovery projects in data analysis. In C. Reading (Ed.), *The Proceedings IASE/ISI 8th International Conference on Teaching Statistics*. ISI.

Franklin, C., Bargagliotti, A. E., Case, C. A., Kader, G., Scheaffer, R., & Spangler, D. (2015). *The statistical education of teachers*. American Statistical Association. Retrieved from www.amstat.org/education/SET/SET.pdf

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A preK-12 curriculum framework*. American Statistical Association.

Franklin, C. A., & 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* (pp. 345–375). National Council of Teachers of Mathematics.

Frischemeier, D., & Biehler, R. (2018). Stepwise development of statistical literacy and thinking in a statistics course for elementary preservice teachers. In T. Dooley & G. Gueudet (Eds.), *Proceedings of the 10th Congress of the European Society for Research in Mathematics Education* (pp. 756–763). DCU Institute of Education and ERME. https://hal.archives-ouvertes.fr/hal-01927856

Gal, I., & Ograjenšek, I. (2016). Rejoinder: More on enhancing statistics education with qualitative ideas. *International Statistical Review, 84*(2), 202–209. https://doi.org/10.1111/insr.12157

Gelman, A., & Nolan, D. (2017). *Teaching statistics: A bag of tricks*. Oxford Scholarship Online. https://doi.org/10.1093/oso/9780198785699.001.0001

González, B. A. R., Ibarra, G. N. F., Barbosa, O. G., & Muñóz, H. A. D. (2022). The use of the empirical rule in the probability class: A proposed application for university students to determine the type of statistical thinking. *Canadian Journal of Science, Mathematics, and Technology Education, 22*, 521–537. https://doi.org/10.1007/s42330-022-00237-y

Gould, R., Bargagliotti, A., & Johnson, T. (2017). An analysis of secondary teachers' reasoning with participatory sensing data. *Statistics Education Research Journal, 16*(2), 305–334. https://doi.org/10.52041/serj.v16i2.194

Groth, R. E. (2011). Improving teaching through lesson study debriefing. *The Mathematics Teacher, 104*(6), 446–451.

Hall, J. (2011). Engaging teachers and students with real data: Benefits and challenges. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education: A Joint ICMI/IASE study: The 18th ICMI study* (pp. 335–346). Springer. https://doi.org/10.1007/978-94-007-1131-0_32

Hannigan, A., Gill, O., & Leavy, A. M. (2013). An investigation of prospective secondary mathematics teachers' conceptual knowledge of and attitudes towards statistics. *Journal of Mathematics Teacher Education, 16*, 427–449. https://doi.org/10.1007/s10857-013-9246-3

Heaton, R. M., & Mickelson, W. T. (2002). The learning and teaching of statistical investigation in teaching and teacher education. *Journal of Mathematics Teacher Education, 5*(1), 35–59. https://doi.org/10.1023/a:1013886730487

International Data Science in Schools Project. (2019). *Curriculum frameworks for introductory data science.* Retrieved from http://idssp.org/files/IDSSP_Frameworks_1.0.pdf

Jacobs, V. R. (1999). How do students think about statistical sampling before instruction? *Mathematics in the Middle School, 5*(4), 240–263.

Jones, D. L., Brown, M., Dunkle, A., Hixon, L., Yoder, N., & Silbernick, Z. (2015). The statistical content of elementary school mathematics textbooks. *Journal of Statistics Education, 23*(3), Article 11889748.

Kazak, S. (2023). Pre-service mathematics teachers' statistical problem-solving process for telling a story through the use of photographs as data sources. In *Paper presented at the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)*. Alfréd Rényi Institute of Mathematics, Eötvös Loránd University of Budapest.

Kazak, S., Fujita, T., & Turmo, M. F. (2021). Students' informal statistical inferences through data modeling with a large multivariate dataset. *Mathematical Thinking and Learning, 25*(1), 23–43.

Kilpatrick, J. (1987). Problem formulating: Where do good problems come from? In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 123–147). Lawrence Erlbaum Associates.

Koichu, B., & Kontorovich, I. (2013). Dissecting success stories on mathematical problem posing: A case of the Billiard Task. *Educational Studies in Mathematics, 83*(1), 71–86. https://doi.org/10.1007/s10649-012-9431-9

Koklu, O. (2017). *Undergraduate students' informal notions of variability* [Unpublished doctoral dissertation]. University of Georgia.

Koklu, O. & Kaplan, J. J. (2023). Undergraduate students' use of primitive notions when reasoning about variability. *International Journal of Science and Mathematics Education, 21*(4), 1243–1264. https://doi.org/10.1007/s10763-022-10293-3

Kontorovich, I. (2016). Considerations of aptness in mathematical problem posing: Students, teachers and expert working on Billiard task. *Far East Journal of Mathematical Education, 16*(3), 243–260. https://doi.org/10.17654/me016030243

Kontorovich, I. (2020). Problem-posing triggers or where do mathematics competition problems come from? *Educational Studies in Mathematics, 105*, 389–406. https://doi.org/10.1007/s10649-020-09964-1

Kontorovich, I. (2024). The road to "good" problems goes through initial responses to stimulating socio-mathematical situations. *The Journal of Mathematical Behavior, 73*, Article 101125. https://doi.org/10.1016/j.jmathb.2024.101135

Leavy, A. (2006). Using data comparison to support a focus on distribution: Examining preservice teachers' understandings of distribution when engaged in statistical inquiry. *Statistics Education Research Journal, 5*(2), 89–114.

Leavy, A., & Frischemeier, D. (2022). Developing the statistical problem posing and problem refining skills of prospective teachers. *Statistics Education Research Journal, 21*(1), 10. https://doi.org/10.52041/serj.v21i1.226

Leavy, A. M. (2004). Indexing distributions of data: Preservice teachers' notions of representativeness. *School Science and Mathematics, 104*(3), 119–134.

Lee, H. S., Mojica, G. F., Thrasher, E. P., & Baumgartner, P. (2022). Investigating data like a data scientist: Key practices and processes. *Statistics Education Research Journal, 21*(2), 3. https://doi.org/10.52041/serj.v21i2.41

Lovett, J. N., & Lee, H. S. (2018). Preservice secondary mathematics teachers' statistical knowledge: A snapshot of strengths and weaknesses. *Journal of Statistics Education, 26*(3), 214–222. https://doi.org/10.1080/10691898.2018.1496806

MacGillivray, H., & Pereira-Mendoza, L. (2011). Teaching statistical thinking through investigative projects. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education. New ICMI Study Series* (Vol. 14, pp. 109–120). Springer. https://doi.org/10.1007/978-94-007-1131-0_14

Makar, K. (2010). Teaching primary teachers to teach statistical inquiry: The uniqueness of initial experiences. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society (Proceedings of the 8th International Conference on the Teaching of Statistics, Ljubljana, Slovenia, July 11–16)*. International Statistical Institute.

Makar, K., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education: A joint ICMI/IASE study: The 18th ICMI study* (pp. 347–358). Springer.

Meletiou-Mavrotheris, M., & Paparistodemou, E. (2015). Developing students' reasoning about samples and sampling in the context of informal inferences. *Educational Studies in Mathematics, 88*(3), 385–404 http://www.jstor.org/stable/43589972

Metz, M. L. (2010). Using GAISE and NCTM standards as frameworks for teaching probability and statistics to pre-service elementary and middle school mathematics teachers. *Journal of Statistics Education, 18*(3), 1–27. https://doi.org/10.1080/10691898.2010.11889585

Mulligan, J., Tytler, R., Prain, V., & Kirk, M. (2024). Implementing a pedagogical cycle to support data modelling and statistical reasoning in years 1 and 2 through the Interdisciplinary Mathematics and Science (IMS) project. *Mathematics Education Research Journal, 36*(1), 37–66. https://doi.org/10.1007/s13394-023-00454-0

Ograjenšek, I., & Gal, I. (2016). Enhancing statistics education with qualitative ideas. *International Statistical Review, 84*(2), 165–178. https://doi.org/10.1111/insr.12158

Powell, A. B., Francisco, J. M., & Maher, C. A. (2003). An analytical model for studying the development of learners' mathematical ideas and reasoning using videotape data. *The Journal of Mathematical Behavior, 22*(4), 405–435. https://doi.org/10.1016/j.jmathb.2003.09.002

Santos, R., & da Ponte, J. P. (2014). Learning and teaching statistical investigations: A case study of prospective teacher. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, ICOTS9, July 2014, Flagstaff, AZ, USA*. International Statistical Institute.

Schreiter, S., Friedrich, A., Fuhr, H., Malone, S., Brünken, R., Kuhn, J., & Vogel, M. (2024). Teaching for statistical and data literacy in K-12 STEM education: A systematic review on teacher variables, teacher education, and impacts on classroom practice. *ZDM - Mathematics Education, 56*, 31–45. https://doi.org/10.1007/s11858-023-01531-1

Schubatzky, T., Bock, B., & Haagen-Schützenhöfer, C. (2020). Physics pre-service teachers' approaches to scientific investigations by data exploration. *EURASIA Journal of Mathematics, Science and Technology Education, 16*(11), 1–18.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning. A project of the National Council of Teachers of Mathematics* (pp. 957–1009). Information Age.

Sole, M. A., & Weinberg, S. L. (2017). What's brewing? A statistics education discovery project. *Journal of Statistics Education, 25*(3), 137–144. https://doi.org/10.1080/10691898.2017.1395302

Spence, D. J., Bailey, B., & Sharp, J. L. (2017). The impact of student-directed projects in introductory statistics. *Statistics Education Research Journal, 16*(1), 240–261. https://doi.org/10.52041/serj.v16i1.230

Suh, H., Kim, S., Hwang, S., & Han, S. (2020). Enhancing preservice teachers' key competencies for promoting sustainability in a university statistics course. *Sustainability, 12*(21), 9051. https://doi.org/10.3390/su12219051

Tran, D., & Tarr, J. E. (2018). Examination of bivariate data tasks in us high school textbooks through the statistical investigation and cognitive demands frameworks. *International Journal of Science and Mathematics Education, 16*(8), 1581–1603. https://doi.org/10.1007/s10763-017-9851-1

Watson, J., Fitzallen, N., Fielding-Wells, J., & Madden, S. (2018). The practice of statistics. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 105–138). Springer.

Watson, J., Fitzallen, N., Wright, S., & Kelly, B. (2023). Characterizing the ways in which young students recognise, describe, explain, and employ variation when analysing data in a STEM context. *International Journal of Science and Mathematics Education, 21*(3), 943–968. https://doi.org/10.1007/s10763-022-10281-7

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 221–248.

Wild, C., Utts, J., & Horton, N. (2018). What is statistics? In D. BenZvi, D. K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 5–36). Springer. https://doi.org/10.1007/978-3-319-66195-7_1

Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics, 12*(3), 9–26. Retrieved from www.ci-journal.net/index.php/ciej/article/view/1286

Yilmaz, N., & Yetkin-Ozdemir, E. (2024). Pre-service teachers' knowledge of statistical questions in the context of lesson study. *Kastamonu Education Journal, 32*(3), 408–419. https://doi.org/10.24106/kefdergi.1525332

Zapata-Cardona, L., & Martínez-Castro, C. A. (2021). Statistical modeling in teacher education. *Mathematical Thinking and Learning, 25*(1), 64–78. https://doi.org/10.1080/10986065.2021.1922859

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education, 16*(2), 122. https://doi.org/10.1080/10691898.2008.11889566