A CURRICULUM PERSPECTIVE ON STATISTICAL ARGUMENTATION: EVIDENCE FROM EGYPT'S REFORMED GRADE 7 MATHEMATICS TEXTBOOK

SAMAH GAMAL AHMED ELBEHARY
Tanta University
dr.samahelbe7ary@edu.tanta.edu.eg

ABSTRACT

Grounded in the significance of textbooks and the need to engage students in statistical argumentation, this study examined the extent to which the statistics tasks in Egypt's reformed Grade 7 mathematics textbook afford opportunities for students to construct statistical arguments. To this end, it developed a framework and coding protocol to specify statistics tasks by their embedded data and warrants, and by the backings, rebuttals, and qualifiers they may prompt through contextual, variability, and uncertainty circumstances, respectively. Only 7% of the scrutinized content was highly conducive to constructing statistical arguments, particularly through the inclusion of rebuttals and qualifiers. While 21% remained at an acceptable level, 72% required significant adaptation to reduce their deterministic characteristics, which limit such opportunities.

Keywords: *Statistics education research; Statistical argumentation; Task analysis*

1. BACKGROUND AND RATIONALE

As data-driven evidence increasingly shapes decisions across diverse fields and everyday contexts, learning statistics extends beyond memorizing facts and procedures. It fundamentally entails understanding statistical concepts, developing effective reasoning strategies, and constructing critical arguments (Ben-Zvi et al., 2018). Thus, educating students to reason and argue with data has become indispensable—not only for professionals but also for consumers of statistical information who participate in data-informed decision-making processes (Burrill & Pfannkuch, 2023; Weiland, 2019a). Accordingly, cultivating statistical argumentation practices among students has emerged as a foundational goal in statistics education. This imperative has been highlighted in Pfannkuch's (2018) call for research on "what constitutes good argumentation practices and how to grow and scaffold students' argumentation" (p. 400), highlighting a critical contemporary research theme.

One way to respond to Pfannkuch's (2018) call is to examine the school curriculum, particularly textbooks, which represent the potentially enacted curriculum (Valverde et al., 2002) and serve as a key resource for both teachers and students. On the one hand, textbooks are a vital resource for teachers, especially when teaching statistics, where most teachers lack sufficient preparation in delivering statistical content (Burrill & Biehler, 2011); on the other hand, they provide crucial learning opportunities for students, enhancing the capabilities intended for them to acquire (Büscher, 2022; Kuş, 2022; Landtblom, 2023). Because incorporating evidence-based argumentation into the statistics curriculum is a promising approach to enriching students' learning experiences (Pfannkuch, 2018), this study adopts the lens of statistical argumentation, that is, the type of argumentation that relies on data and statistical concepts as evidence (Pardue, 2017), to scrutinize the statistics tasks presented in school textbooks. Ultimately, insights into the extent to which our local curriculum enables students to construct and negotiate statistical arguments, along with recommendations for improvement, can be presented.

In 2018, the Egyptian Ministry of Education and Technical Education (MoETE) launched a reformed curriculum series as part of a modern educational framework, known as Education 2.0 (EDU 2.0). This

© International Association for Statistical Education (IASE/ISI), 2026

initiative sought to raise the quality of K-12 education to meet international standards and to transform the entire pre-university system by 2030, in line with Egypt's Vision 2030 (Moustafa et al., 2022). As an integral component of this reform, a new series of school textbooks was launched, with Grade 7 mathematics textbooks released merely a few months before the present study period (September 2024) and scheduled for implementation in the ongoing academic year (2024-2025). In this regard, the present study examines specific aspects of the reformed textbooks, particularly for Grade 7, which marks the onset of the preparatory phase as outlined by the Egyptian education system, through an analytical lens, that is, statistical argumentation.

Statistical argumentation depicts a relatively recent area of research (Fielding, 2024). That is, the scarcity of research on argumentation in statistics education compared to mathematics has lately been spotlighted (Conner & Peters, 2023; Pfannkuch, 2018), particularly when weighing the epistemological differences between these two disciplines, wherein the findings of research on mathematics education cannot be directly transferred to the field of statistics (Krummenauer & Kuntze, 2019). Although most literature on proof and argumentation originates from studies carried out in developed settings (e.g., European and North American nations), and "systematic research on proof and argumentation across international contexts remains limited" (Shongwe, 2022, p. 86), it has been regarded in the Egyptian context due to:

Connection to local conditions: Acknowledging that the EDU 2.0 initiative aims to raise the quality of K-12 education in Egypt to satisfy international benchmarks, argumentation offers an authentic, globally recognized lens for critically examining the newly developed statistics content, thus promoting a shared basis for multinational comparison. In particular, the Grade 7 textbooks emphasize decision-making as a principal outcome of learning statistics (MoETE, Mathematics: First Preparatory, First Term, 2024–2025, Student's Book, p. 46), while statistical argumentation is conceived as a pedagogical and epistemic practice that could enhance this competency.

Connection to the regional research gap: Although existing literature highlights the scarcity of studies on argumentation in the Middle East and Africa (Kartika et al., 2021; Petocz et al., 2018), a few investigations do exist (e.g., Ajani, 2023; Shongwe, 2022). These investigations do not focus explicitly on statistical argumentation, but they identify argumentation as a promising avenue for research in developing countries. Therefore, extending this line of research to the Egyptian context can help reconcile findings from the Western community with local perspectives, fostering a more balanced dialogue between indigenous and exogenous languages of argumentation. In turn, it responds to discussions of social factors, presented by classroom norms and communities of practice, that are critical in argumentation research (e.g., Balacheff, 2024; Elbehary, 2026).

In light of these points, the study responds to the need to evaluate Egypt's reformed curriculum (Hussein, 2023), especially in content areas where students perform poorly on international assessments (Elbehary, 2020; Moustafa et al., 2022). Furthermore, it helps to enhance the global reputation of African statistics education research within the broader research community (Petocz et al., 2018). Based on this rationale, the present study addresses the following research question: *To what extent do the statistics tasks in Egypt's reformed Grade 7 mathematics textbook afford opportunities for students to construct statistical arguments?* In this study, a statistics task is defined as any learning activity classified within the statistics content of the school textbook.

2. PREVIOUS RESEARCH AND STUDY PERSPECTIVE

2.1. STATISTICS CONTENT IN THE EGYPTIAN MATHEMATICS TEXTBOOKS

In Egypt, the school academic year is divided into two terms. The first term usually starts around mid-September and ends in late December, while the second term begins in early February and concludes at the end of May. For each term, students should study the content using the designated textbook assigned to that term. All public schools use a single official series of school textbooks, available in both Arabic and English versions, since certain institutions (e.g., official language schools and private schools) use English as the

medium for mathematics instruction. Each textbook is available in both printed and digital formats, with the digital version accessible on an E-learning platform.

In broad terms, the status of statistics education in Egypt mirrors the international situation, in which statistics is routinely incorporated into the school mathematics curriculum. This reflects the perception that statistics is merely a branch of mathematics and consequently should be taught deterministically (Burrill & Biehler, 2011; Elbehary, 2020; Weiland, 2019a). Egyptian students begin their statistics education in kindergarten with lessons such as 'Data and Graph' which continue into Grade 1. Starting in Grade 2 and continuing through subsequent grades, the textbooks adopt a systematically organized structure. Although the present study focuses on Grade 7 textbooks, Table 1 provides an overview of the statistical topics covered in the preceding grades. Grade 7 marks the beginning of the preparatory stage, which comprises grades 7, 8, and 9; further details on statistics content at this grade are presented in the Method section.

Table 1. Progression of statistics content in the Egyptian reformed G7 mathematics textbooks

	T1	***T2***
G2	L1: Reading data L2: Collecting and representing data L3: Comparing data L4: Representing and interpreting Data L5: Representing data with a scale of 1 L6: Representing data with a scale of 2 L7: Representing data with a scale of 10 L8: Bar graph L9: Pictograph L10: Graph elements	L1: A scale for representing data using the bar graph L2: A scale for representing data using the pictograph L3: Applications on data representations
G3	L1: More of bar graphs L2: Line plot	L1: Applications on data representations
G4		L1: Different graphs L2: Plotting along L3: Breaking the bar
G5		L1: Introduction to pie charts L2: Understanding pie charts L3: Making pie charts
G6	L1: Data and statistical questions L2: Exploring the histogram L3: Representing data using histograms L4: Exploring box plot L5: Applications on data representations. L6: Exploring the balance of data sets L7: Interpreting mean L8: Exploring median, mode, and outliers L9: Exploring the range	
G7	L1: Organizing data L2: Arithmetic mean L3: Pie charts	L1: Random experiment - Sample space – Events L2: Theoretical & experimental probability

G refers to Grade; T refers to Term; L refers to Lesson

2.2. APPROACHES TO TEXTBOOK ANALYSIS AND STUDY POSITION

Textbooks play a pivotal role in teaching and learning processes, embodying national educational goals and shaping curriculum implementation (Valverde et al., 2002). They remain the primary resource through which teachers determine the knowledge that students are expected to acquire, particularly in statistics, where teachers have limited knowledge and little experience (Elbehary, 2020; Weiland, 2019a). Thus,

reflecting on textbook content is crucial to scrutinizing the quality of learning opportunities afforded to students.

Jones and Jacobbe (2014) argued that the curriculum afforded in textbooks influences the curriculum delivered to students (i.e., the content actually taught), which in turn shapes the curriculum that students ultimately attain (i.e., the attained curriculum). In essence, there is global consensus that textbooks, as embodiments of the potentially implemented curriculum, shape classroom practices and students' competencies (Kuş, 2022; Robitaille & Travers, 1992).

Considering the role of textbooks in equipping students with essential learning experiences, textbook analysis conveys critical ongoing research in the field of statistics education, wherein studies continue to be conducted (Büscher, 2022; Tran & Tarr, 2018). This body of research continues to emphasize the use of real contexts and fostering conceptual understanding of statistical ideas rather than solely concentrating on the computational aspects (Jones et al., 2015; Weiland, 2019a). Regardless, the emphasis and approaches undertaken to clarify textbook content have altered, with little known about this content from the perspective of statistical argumentation. That is the viewpoint embraced in the present study.

Research on analyzing statistical content has commonly adopted one of three approaches: (a) broad content analysis (e.g., Jones et al., 2015), (b) focused analysis of specific statistical concepts, such as measures of center (e.g., Kuş, 2022; Landtblom, 2023), or (c) directed analysis informed by a particular theoretical or ideological perspective (e.g., Büscher, 2022; Weiland, 2019b), including studies that employed comparative, discursive, or statistical literacy perspectives as analytical lenses. Still, some investigations integrated two of these approaches. For instance, Tran and Tarr (2018) examined bivariate data tasks (approach b) through the dual lenses of statistical investigation and cognitive demand (approach c).

The GAISE framework was central to investigations of the first research strand. For instance, Jones et al. (2015) examined the statistical content in mathematics textbooks for grades 1 to 5, focusing on tasks aligned with the statistical problem-solving process, a related version of the PPDAC cycle (Franklin et al., 2007). The study concluded that most statistical tasks emphasized data analysis, particularly reading displays and performing calculations; conversely, opportunities to construct representations and employ forms of statistical reasoning remained marginal. These findings align with Tran and Tarr (2018) and Weiland (2023), indicating that data analysis constituted the most heavily emphasized statistical practice in high school mathematics textbooks. In contrast, data interpretation and data collection conventions were less prioritized. Furthermore, opportunities for students to formulate their own statistical questions were minimal, as such questions were predominantly predefined by the textbook authors.

Landtblom's (2023) study exemplifies the second research strand. It explored the opportunities provided to students while engaging with measures of central tendency in Swedish textbooks by advancing a framework organized around four themes: context (contextual vs. non-contextual settings), input objects (the variable level in the data set), transformations (the procedure employed to obtain the result), and output objects (the result of transforming the input). Through this framework, the study highlighted that Swedish textbooks prioritize procedural knowledge, particularly in calculating the mean, which limits students' opportunities to develop conceptual understanding of measures of central tendency.

Additionally, the study by Büscher (2022) presented an investigation situated within the third strand of research, wherein statistical literacy was employed as a lens to examine Grade 6 German textbooks, revealing a lack of tasks that engage students in interpreting statistical models or constructing statistical arguments, a finding related to the present study. Moreover, considering the importance of context, Weiland (2019b) scrutinized the statistics lessons embedded in school mathematics textbooks from a discursive perspective. The study uncovered a scarcity of controversial situations, such as those raised in media and societal discussions; in contrast, the statistical scenarios frequently presented were often fictional and trivial, and the data embedded within these scenarios lacked authenticity, leading students to assume their validity. This would negatively influence students' disposition to be critical citizens in modern society.

The present study aligns with the third strand. It positions argumentation as a lens to scrutinize statistics tasks. Specifically, the study examines the extent to which statistics tasks in Egypt's reformed Grade 7 mathematics textbook afford opportunities for students to construct statistical arguments. To achieve this

aim, an analytical framework, grounded in statistical argumentation theory, has been developed to scrutinize textbook tasks. In doing so, the study aims to contribute to statistics education research by offering a curriculum perspective on statistical argumentation, specified through task analysis.

2.3. ARGUMENTATION IN STATISTICS EDUCATION RESEARCH

While argumentation describes a form of discourse in which an individual seeks to convince audiences of the validity of a particular claim, statistical argumentation is a specific type of argumentation in which the employed evidence and reasons are grounded in data and statistical concepts (Pardue, 2017). Statistical argumentation is often debated under the umbrella of statistical literacy. This is evident in the work of Schield (1999) and Gal (2002), who characterized statistical literacy as the study of arguments that use statistics as evidence; it involves the capacity to critically analyze and evaluate statistical data while conveying a well-reasoned standpoint on information. Similarly, Büscher (2022) identified the construction of concise statistical arguments, which involves selective reading, as an aspect of statistical literacy.

Fielding (2024) conceptualized argument as having complementary structural and functional facets, in which an argument constitutes a structured product of reasoning presented through discourse. This reflects Pardue's (2017) distinction between argument and argumentation: the former refers to the constructed discourse/product, and the latter pertains to the process of developing arguments. In statistics education, several initiatives are underway to examine issues of argumentation; some of these are discussed below:

The structural aspect of argumentation: argumentation analyzed through Toulmin's model This body of research primarily draws on Toulmin's model, a seminal framework for studying arguments. It consists of six components: three main components, namely Data, Warrants, and Claims, along with three supplementary components (Backings, Rebuttals, and Qualifiers). These are defined as follows (Conner & Peters, 2023; Gómez-Blancarte & Tobías-Lara, 2018; Groth & Barry, 2025; Toulmin, 2003):

- Data: the information that underpins a claim, providing evidence such as the facts that demonstrate the reasoning behind a conclusion.
- Warrants: the rationales or deductions that connect data to conclusions. This may include rules or principles that serve as a validating connection between evidence and conclusions.
- Claims/ Conclusions: the statements for which validity is being determined.
- Backings: the discourse that supports or reinforces the warrant employed.
- Rebuttals: the circumstances under which warrants may not be valid.
- Qualifiers: the degree of uncertainty with which claims are made (e.g., probabilistic language).

Although all constructs of Toulmin's model are crucial, many researchers have highlighted rebuttals and qualifiers as essential for specifying the limitations of statistical arguments. This is because statistical questions are typically answered inductively, yielding a degree of uncertainty rather than definitive conclusions (Groth & Barry, 2025). Conner and Peters (2023) emphasized that rebuttals highlight potentially unexplored aspects of a warrant, while qualifiers indicate that decisions made from sample data are inherently uncertain; both remain influential in constructing statistical arguments.

The study by Gómez-Blancarte and Tobías-Lara (2018) symbolizes an exemplary model for this line of research. Drawing on Toulmin's (2003) framework, they examined the validity of students' inferential reasoning by proposing an analogy between Toulmin's model constructs and features of inferential reasoning. Their analysis further highlighted the role of rebuttals, which delimit the conditions under which statistical principles or contextual assumptions may not hold, and qualifiers, which express the uncertainty inherent in inferential claims, in statistical inference.

In a similar vein, Krummenauer and Kuntze (2019) studied primary students' construction of data-based arguments across various tasks. They conceptualized data-based argumentation as a type of reasoning in which statistical data is employed to persuade individuals regarding the truthfulness or fallacy of a claim, taking into account the empirical nature of data collection and the inherent variability of phenomena; as they expressed it, even interpretations that appear contradictory can be justified using the same dataset.

The functional aspect of argumentation: argumentation in teaching practices and classroom discourse This line of research emphasized classroom interactions, including teachers' practices and students' negotiations, aligning with the definition of argumentation as "a critical discussion during which points of view are constructed, negotiated and transformed" (Corrêa & Lopes, 2018, p. 1). Most studies in this area have associated statistical argumentation with models rooted in the statistical investigative cycle, the PPDAC (Pardue, 2017; Wild & Pfannkuch, 1999).

For example, Conner and Peters (2023) explored how to engage students in collective argumentation, i.e., reaching group consensus, to develop statistical reasoning. Drawing on Conner et al.'s (2014) framework, Conner and Peters (2023) identified three ways in which teachers sustain collective argumentation during mathematical activities: (a) by directly delivering argument components (e.g., claims, warrants, data), (b) by asking questions that elicit parts of arguments (e.g., requesting fact, an idea, a method, elaboration, and evaluation), and (c) by using supportive actions (e.g., directing, promoting, evaluating, informing, and repeating actions). This closely resembles the DIVINE framework (Chua, 2017); it relies on mathematical proof, reasoning, and justification studies to outline the nature (i.e., the cognitive processes used) and purpose (i.e., the rationale) of argumentative tasks. Accordingly, the study identified four argumentative tasks (decision-making, inference, validation, and elaboration), corresponding to the explanatory questions of whether/which, what, why, and how, respectively.

Fielding's (2024) study provides another example, demonstrating how argumentation could be employed to facilitate statistical investigations, thereby improving students' data-ing practices (i.e., posing statistical problems, drawing on context to make decisions, selecting and justifying statistical methods, and making decisions, based on data, to support or reject claims). Fielding (2024) utilized the PPDAC cycle (Wild & Pfannkuch, 1999) and the argument-based statistical inquiry (ABSI) framework (Fielding-Wells, 2016) to identify the types of knowledge (i.e., contextual, statistical, and argumentation knowledge) that students draw upon while engaging in ABSI. Therefore, argumentation knowledge has been characterized through four dimensions: generic structure, discourse, quality, and disposition, referring to knowledge of components of the argument (Toulmin's components), understanding the language of persuading others with the argument, determining the argument quality (e.g., validity, accuracy), and willingness to engage in argumentation practices, respectively.

Based on this category of research, and mirroring Weiland's (2019a) overview of the key differences between mathematics and statistics, including issues of context, variability, uncertainty, and contrast between inductive and deductive reasoning, the present study assumes that studying argumentation in the field of statistics should weigh the following factors:

Role of context (contextual knowledge): Admitting the essential premise that "data are not just numbers, they are numbers with a context" (Cobb & Moore, 1997, p. 801), contextual knowledge remains critical for practicing statistical argumentation (Pardue, 2017; Pfannkuch, 2018) since it provides a broader picture of the data (Fielding, 2024). Unlike mathematics, data analysis and interpretation in statistics must consistently consider the context surrounding the data (Cobb & Moore, 1997), as interactions between the statistical problem and its contextual background are always present (Wild & Pfannkuch, 1999). This is evident in Büscher's (2022) study; it highlighted students' reliance on contextual knowledge when engaging in statistical argumentation, primarily to reflect on the limitations, relevance, and consequences of arguments. Similarly, Fielding (2024) distinguished contextual knowledge as an essential domain of knowledge students draw on when practicing argument-based statistical inquiry; it describes "knowledge of the context in which the statistics is situated and which may be derived from any prior experiences the student may have had, whether in school or not" (p. 5)

Essence of statistical reasoning: Statistical reasoning is characterized by the recognition of variability and assumes that certain inexplicable factors might yield answers with a level of uncertainty. Mathematical reasoning, by contrast, is deterministic in nature and seeks definitive answers while disregarding variability. Put differently, while mathematics is objective, statistics hold a degree of subjectivity (Pardue, 2017); therefore, answers to statistical investigations are often associated with appropriate levels of uncertainty; this is different from mathematical proof, wherein valid deductive reasoning leads to certainty in assertions (Groth & Barry, 2025; Groth & Choi, 2023; Weiland, 2019a). This explains Conner and Peters' (2023)

emphasis on qualifiers in responses to statistical claims, given that "statistical reasoning will often contain qualifiers that reflect a level of uncertainty for the claim, such as probably, possibly, or even likely" (p. 269).

2.4. RESEARCH GAP AND ANALYTICAL FRAMEWORK

As discussed earlier, research on argumentation has recently gained attention in statistics education; however, little focus has been given to how textbooks may support students in constructing statistical arguments. This perspective recognizes arguments as an implicit component embedded in the curriculum, precisely in textbook tasks, and thus aligns with initiatives to reimagine the curriculum (Pfannkuch, 2018). In other words, the statistics curriculum should immerse students in data-rich environments and statistical investigations to help them critically construct arguments in diverse contexts. Building on these ideas, which provide the rationale for the present study, the following framework was developed to scrutinize the statistics tasks presented in mathematics textbooks.

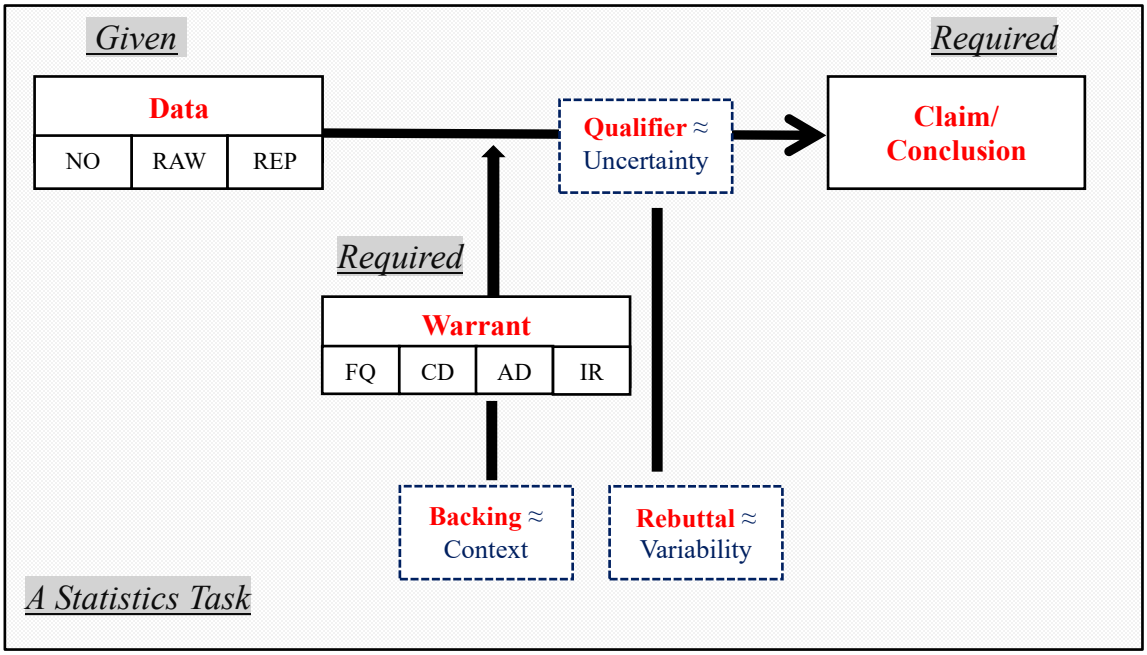


Figure 1. The study framework

Figure 1 outlines the statistics task by categorizing it into the 'Given' information and the 'Required' queries that prompt responses; the former was used to determine the *Data* incorporated, whereas the latter provides the basis for identifying the *Warrant* and *Claim*. Specifically, Data refers to the information provided in the task and may be organized in various possible formats (Kazak, 2023), including numbers (NO: with no context), raw data (RAW: quantitative or qualitative), and representations (REP: tables, figures). Warrant refers to the anticipated reasoning through which students are expected to respond to required queries; it is classified according to the statistical problem-solving process expected to be practiced in responding to such queries (i.e., FQ- Formulate Investigative Questions, CD- Collect Data, AD- Analyze the Data, IR- Interpret the Results). The claim clarifies the response, conclusion, judgment, or decision that the task directs students to establish eventually based on the available data. To situate Toulmin's model in statistics and emphasize the process of statistical reasoning, the three constructs of *Backing*, *Rebuttal*, and *Qualifier* were revisited in terms of context, variability, and uncertainty, respectively (Fielding, 2024; Weiland, 2019a). This analogy is explained as follows:

Gómez-Blancarte and Tobías-Lara (2018) identified backings within the realm of inferential reasoning as the discourse that supports the warrants employed, and "this discourse can come from either the field of statistics or the context of the problem" (p. 3). Hence, the present study interprets backings as other domains of knowledge, rather than as statistics, that students may rely on to respond to claims; i.e., contextual knowledge (Fielding, 2024). In other words, backings determine whether responding to claims prompts revisiting the problem context or deliberating possible related personal experiences to validate the response.

While rebuttals establish conditions under which warrants may be deemed inapplicable (Gómez-Blancarte & Tobías-Lara, 2018), this study assumes that variability can provide empirical grounds for identifying such conditions; variability refers to how attributes may differ "from individual to individual or object to object," encompassing measured, natural, induced, and sampling variation (Weiland, 2019a, p. 398). Such variability becomes relevant to rebuttal when differences in the data, observations, measures, or samples inform circumstances under which a warrant does not sufficiently support the claim. For example, in a comparative scenario wherein an analyst evaluates two pain-relief medications and finds that patients receiving one medication recover, on average, more quickly than those receiving the other. Based on a warrant that treats the average as an indicator of overall treatment effectiveness, the analyst may infer that the first medication is better. However, examining how recovery times differ among patients may stimulate reasonable rebuttals. In this sense, variability does not constitute the rebuttal itself; rather, it may reveal possibilities that make a warrant less adequate for supporting a particular claim.

Accordingly, the study assumes that opportunities to consider potential rebuttals arise when statistics tasks require students to attend to how variation within or across possible data sets, measurements, or samples may affect the inferential pathway from data to claims, thereby revealing conditions under which a warrant may no longer sufficiently support a claim. That is, the presence of variability in a task is considered analytically related to rebuttal insofar as they create opportunities for students to deliberate on or reconsider the sufficiency of the warrant. Importantly, variability does not constitute a rebuttal in itself; rather, it may provide empirical grounds for identifying conditions that may weaken the warrant. Such an association between rebuttals and variability is evident in Bargagliotti et al.'s (2020) view on responses to statistical queries, wherein these responses should be stochastic in nature to recognize the "limitations regarding generalizability" due to several factors such as sample-to-sample variability, absent data, bias, and other statistical or contextual factors. From this perspective, the study framework connects rebuttals to reasoning about probable variations that may limit the warrants linking data to claims.

Makar and Rubin (2009) emphasized the importance of using probabilistic language to express varying degrees of confidence in conclusions drawn through statistical reasoning, as documented: "[b]ecause generalizations go beyond the given data, they cannot be stated in absolute terms" (p. 87). This corresponds to the description of qualifiers as a possible emergent construct in arguments. Moreover, Fielding (2024) stressed the reciprocal relationship between qualifiers and expressions of uncertainty since drawing inferences from a sample to a population necessarily acknowledges uncertainty, and "argument structure supports expression of that uncertainty through the use of qualifiers" (p. 41). Similarly, Groth and Choi (2023) stated that statistical interpretations should not be conveyed with certainty; uncertainty remains a critical factor in qualifying claims. Thus, from the study perspective, qualifiers reflect the extent to which possible responses to claims might include probabilistic terminology to express the degree of uncertainty surrounding conclusions drawn.

Again, the study framework (see Fig. 1) follows Toulmin's basic path from data to claims, with warrants serving as the intermediary link. It also preserves the placement of backings, rebuttals, and qualifiers to qualify conclusions; however, these three constructs were rethought in terms of context, variability, and uncertainty, respectively, to anchor argument, a domain-specific practice, in statistics.

3. METHOD

3.1. CONTENT ASSIGNED

As previously outlined, the preparatory phase of the Egyptian education system comprises three grade levels (7, 8, and 9), during which students learn mathematics and statistics using formal textbooks. The seventh-grade mathematics textbook, published by the Egyptian Ministry of Education and Technical Education, was developed by a committee of mathematics teachers, researchers, and university professors (MoETE, Mathematics: First Preparatory, First Term, 2024–2025, Student's Book). It is accessible via the ministry's online platform: https://studentbooks.moe.gov.eg/

Statistics (including probability lessons) is primarily taught at the end of each term and assessed alongside numbers and algebra units, aligning with Tran and Tarr (2018), who noted that statistics is typically offered as separate chapters within mathematics courses. Specifically, the Grade 7 mathematics textbooks for the academic year 2024–2025 cover four primary content areas: Geometry and Measurement (35%), Numbers and Operations (24.5%), Algebra (22%), and Statistics and Probability (18.5%); the percentage of each domain was determined by the number of related pages in the textbook (see Table 2).

Table 2. Content of Mathematics and Statistics in the Grade 7 Egyptian Mathematics Textbooks

Content	***Number of pages***	***Percentage***	***Content of Statistics Including Probability "Study focus"***
First term: (21st Sep. to 28th Dec.)			
Unit 1: Numbers and Their Operations	30	15%	
Unit 2: Algebra	15	7.5%	
Unit 3: Statistics	22	11%	▪ *L1: Organizing Data* ▪ *L2: Arithmetic Mean* ▪ *L3: Pie Charts*
Unit 4: Geometry & Measurement	44	22%	
Second term: (8th Feb. to 10th May)			
Unit 1: Powers, Exponents, and Roots	19	9.5%	
Unit 2: Algebra	29	14.5%	
Unit 3: Geometry & Measurement	26	13%	
Unit 4: Probabilities	15	7.5%	▪ *L4: Random Experiment- Sample Space- Events* ▪ *L5: Theoretical & Experimental Probability*
Eight units (Over the whole school year)	**200**	**100%**	**Two units (Five lessons) 18.5%**

As highlighted in blue in Table 2, five lessons were specified. These lessons belong to the statistics and probability units, which account for 18.5% of the total curriculum to be covered throughout the school academic year. The lesson begins with a presentation of the learning outcomes, a list of relevant vocabulary, and a QR code linking to a related video. Each lesson revolves around three main themes: "Get Ready!", "Think & Discuss!", and "Learn!". The "Learn" section is further segmented into several examples (each with an accompanying answer) that correspond to an equal number of self-assessment tasks (without answers). After each lesson, supplementary tasks are offered under the heading "Lesson Assessment". Figure 2 shows an example of a lesson structure from the 7th-grade mathematics textbook.

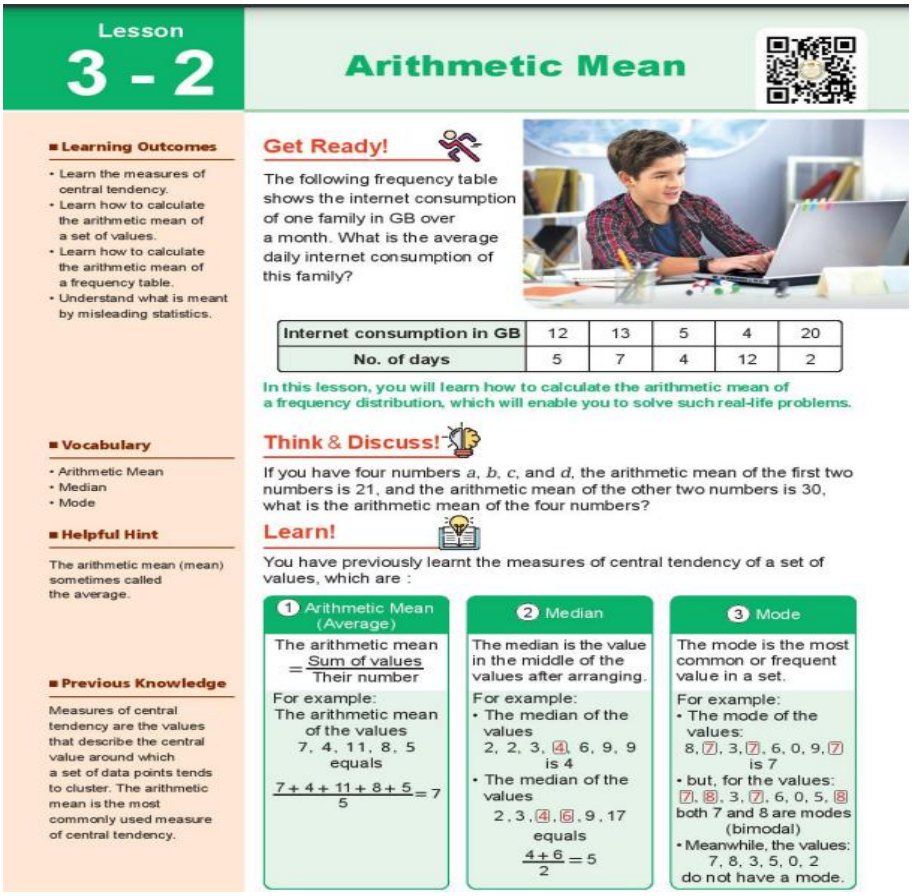


Figure 2. A statistics lesson offered in the G7 mathematics textbook in Egypt

3.2. TASK SELECTION AND UNIT OF ANALYSIS

The study focused on scrutinizing the statistics tasks presented within the three core components of each lesson: "Get Ready!", "Think & Discuss!", and "Learn!". Conversely, other tasks included in the "Lesson Assessment" section were excluded from the analysis. This is because Egyptian teachers frequently engage students in essential tasks; however, due to class time constraints, other tasks are mainly left to students as homework.

According to this study, the textbook task, outlined by the given information and required inquiries, determined the unit of analysis. Regardless, due to the absence of solutions for many tasks, particularly those found in the Get Ready and Self-evaluation sections, the researcher (a) revised the discussion regarding task solutions available in the accompanying teacher guidebook, (b) reflected on her own experience to consider multiple approaches likely to emerge in the Egyptian classroom context, (c) engaged in discussions with another researcher about her interpretation, and (d) conducted a pilot study to validate some cases. These procedures helped verify the coding decisions, particularly those in relation to the three advanced constructs of argument (i.e., backings, rebuttals, and qualifiers), as detailed in the subsequent sections.

Also, in contrast to Jones et al.'s (2015) perspective, the present study included all probability-related tasks (see Table 2) in the analysis. While Jones et al. (2015) distinguish between probability tasks situated within a statistical context (such as data collection for an experiment) and those necessitating mathematical computation (for instance, determining the probability of rolling a 4), this study took both types into account and judged them in terms of constructs of statistical argument, as articulated in the coding protocol. This

viewpoint is strengthened by Makar and Rubin's (2009) account of statistical ideas, such as average, variation, and probability, as both objects of instruction and tools for inference. Moreover, it is consistent with Jones and Tarr's (2007, p.13) identification of a probability task as any task that "contained probability," regardless of whether its primary focus was on "another content area, such as geometry, combinatorics, or statistics." The present study therefore adopted a broader inclusion criterion, as it aimed to examine the opportunities that all statistics tasks afford students to construct statistical arguments. This decision also reflects the view that understanding basic notions of probability is a core component of statistical literacy (Gal, 2002), the broader competence within which statistical argumentation is situated. Furthermore, considering the Egyptian context, the concept of probability is frequently discussed in relation to statistics. This explains the study's determination of a statistics task as any learning activity classified within the statistics content of the school textbook, as indicated earlier.

3.3. CODING PROTOCOL

Guided by the study framework depicted in Figure 1, a coding protocol was developed to identify argument constructs embedded in the statistics tasks across the five designated lessons. Figure 3 presents the coding procedures, followed by a description of the criteria used to judge each construct. The section concludes with an example illustrating how the task introduced previously in Figure 2 was coded.

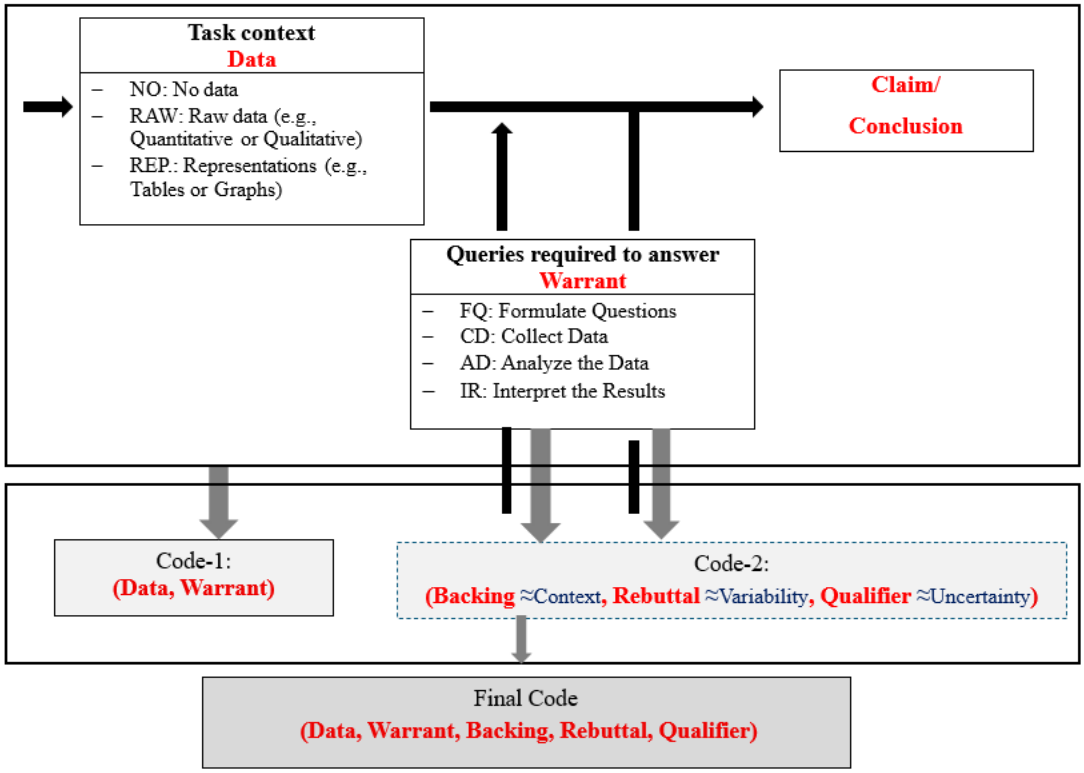


Figure 3. The coding protocol

The criteria for judging code-1 and code-2 constructs

The criteria for judging Code-1 constructs: Data and Warrant

Data defines the information embedded in the task context that serves as the grounds for addressing the inquiry posed by the task. It is classified as follows:

- NO (No data): The task provides no data; that is, the data needed to address the inquiry should be generated, measured, or collected by the students themselves.
- RAW (Raw data): The data is presented in a raw format, either quantitative (numerical) or qualitative (categorical). Additionally, numerical values (i.e., data presented without a contextual referent) are also classified in this category.
- REP (Representations): The data is presented through tables or graphs (e.g., bar graphs, pie charts).

Warrant defines the anticipated statistical reasoning that connects the data (i.e., available or generated data) to the conclusion; because tasks rarely state the warrant explicitly, warrant affordances are operationalized according to the statistical problem-solving demand through which this conclusion is expected to be established, wherein:

- FQ (Formulate Questions): The task inquiry directs students to formulate pertinent questions for data collection.
- CD (Collect Data): The task inquiry directs students to determine a strategy to gather relevant data.
- AD (Analyze the Data): The task inquiry directs students to read a display, organize and represent it (e.g., tables or graphs), or calculate related measures (e.g., measures of center, probabilities, or relevant calculations).
- IR (Interpret the Results): The task inquiry directs students to evaluate the given, collected, or generated data to decide on or predict the phenomena, considering that predictions do not require simply reading the data, but rather provoke students to look beyond this data.

The criteria for judging Code- 2 constructs: Backing, Rebuttal, and Qualifier

Backing determines whether the statistics task allows students to draw on relevant contextual or experiential knowledge to support the warrant connecting data to the claim. It is assessed dichotomously using a score of 0 or 1.

- Score 0: The expected response to the task inquiry can be developed without drawing on contextual knowledge or relevant personal experiences.
- Score 1: The expected response to the task inquiry cannot be fully addressed without considering the context of the problem, stimulating students' reflection on relevant prior knowledge or personal experiences to strengthen their reasoning.

Rebuttal determines whether the statistics task allows students to consider conditions under which variation within or across possible data sets, measurements, or samples may weaken or limit the warrant connecting data to the claim. It received a score of 0 or 1.

- Score 0: The task inquiry provides no opportunity to consider limitations, exceptions, counterevidence, alternative conditions, or variability that could weaken or restrict the applicability of the warrant or the conclusion.
- Score 1: The task inquiry provides an explicit or implicit opportunity to consider conditions under which the claim may need to be challenged, qualified, or restricted.

Qualifier determines whether the statistics task allows students to qualify the claim by expressing its degree of certainty or strength. It received a score of 0 or 1.

- Score 0: The task inquiry calls for a deterministic conclusion without requiring qualification of its degree of certainty or scope.
- Score 1: The task inquiry leads to a conclusion that prompts probabilistic or uncertain expressions to qualify the extent to which the claim can be generalized.

Accordingly, the final coding scheme comprises five constructs: Data, Warrant, Backing, Rebuttal, and Qualifier (see Figure 3), which operationalizes the potential affordances of textbook tasks for students to construct statistical arguments, rather than assessing the actual arguments produced when engaging with these tasks.

An example of employing the coding protocol to analyze statistics tasks The *Get Ready* task, shown in Figure 2, presents a family's internet usage over a month, which characterizes the task context. Because the data are presented in a frequency table, the task was coded as Rep. Following this, the inquiry, *"What is the*

average daily internet consumption of this family?", was coded as AD because it requires students to calculate a measure of center to answer it. The task also affords *Backing (1)* because students may draw on contextual information to appropriately interpret the variables and the resulting measure, for example, by identifying the relevant unit of measurement and understanding what the calculated average represents in the context of the family's internet usage. In contrast, the task affords neither *Rebuttal (0)* nor *Qualifier (0)*. The inquiry concerns a single dataset and requires an exact numerical result; thus, it does not invite consideration of alternative conditions or limitations, nor does it demand students qualify the certainty and extent of the conclusion. Accordingly, as detailed in Table 4, the task was coded as (Rep., AD, 1, 0, 0).

3.4. RELIABILITY

Coding reliability was established through both inter-rater and intra-rater procedures. First, an Egyptian researcher with a master's degree in mathematics education independently coded the sampled tasks using an initial version of the coding protocol developed by the author. The initial inter-rater agreement was approximately 72%, with the two coders agreeing on 63 of the 88 task inquiries. At that point, it was crucial to investigate the sources of the discrepancies and resolve them by discussing the coding rationale for each designated task. This process improved the coding protocol, particularly for the criteria used to determine the presence of backings, rebuttals, and qualifiers across the coded tasks.

More specifically, inconsistencies in coding the backing construct arose from differing interpretations of the extent to which contextual knowledge is necessary for responding to an inquiry. For instance, the author initially judged that an inquiry such as *"Represent this data using a dot plot and a bar graph"* did not interrogate contextual knowledge, as, in general, neither Egyptian teachers nor students pay attention to the meaning of the data visualized; then, the task could be completed by attending only to the numerical values. However, discussion with the second researcher led to the recognition that even a minimal degree of contextual interpretation remains necessary to consider in the coding process. Thus, although contextual recognition may be limited to labeling the axes when constructing the bar graph, such labeling nevertheless requires students to identify what the variables and values represent and, consequently, what the data convey. Therefore, the inquiry was ultimately coded as 1 for the presence of backing.

Regarding rebuttal, after consulting with the researcher about the anticipated responses from Egyptian students, alongside revising the teacher guidebook and the insights from the pilot study conducted (see the Results section), rebuttal was evaluated as present whenever students may draw on diverse possible data sets, resources, concepts, or measurements to respond to the inquiry. At its core, this indicates that the characteristics of a conveyed phenomenon may vary across individuals, shaped by how they measure, sample, or derive those attributes. For instance, an inquiry such as *"Represent this data in a suitable graphical chart"* received a code of (1) for the rebuttal construct since it was expected that students might debate in deciding which is the suitable representation to convey attributes of a phenomenon; this, in turn, prompts a discussion on the circumstances for presenting and analyzing different data sets using various graphs.

It is also necessary to note that the abbreviation MD, i.e., Making a Decision, was sometimes used in conjunction with a statistical problem-solving process to highlight the inquiries that prompt students to make a decision (i.e., Which). Item 9 in Table 3 offers an example of this code, in which the code IR & MD was attached to the inquiry: *Which graph could be misleading and why?"*

After the second coding process using the revised protocol, the author's discussion with the researcher increased their agreement from 72% to 94%. Additionally, the author repeated the coding process three weeks after her conversation with the researcher to assess agreement between the results (intra-coder reliability), achieving a rate of 96% agreement.

In alignment with the preceding procedures and to ensure the trustworthiness of the data analysis and coding techniques, a pilot study was conducted with a sample of 50 eighth-grade students during the 2025–2026 academic year; that is, the pilot study could be used to test a research protocol and enhance the credibility of the qualitative analysis (Kim, 2010). In detail, a survey comprising selected statistics tasks,

on which the researcher and the author had disagreed during the coding process, was administered to the student sample, with each student attempting a different subset of these tasks.

4. RESULTS AND DISCUSSION

This section addresses the study question by determining *the extent to which the statistics tasks in Egypt's reformed Grade 7 mathematics textbook afford opportunities for students to construct statistical arguments*, using the coding protocol outlined earlier. First, the Data Presentation section highlights the statistical argument constructs defined in each task across all lessons; afterward, the subsequent section (Discussion) characterizes these findings in relation to the study question.

4.1. DATA PRESENTATION

Tables 3–7 present the results of applying the coding protocol to the statistics tasks across the five statistics lessons offered in Egypt's reformed Grade 7 mathematics textbook. Organized by lesson, the tables document the statistical argument constructs identified in tasks addressing Organizing data (L1), Arithmetic mean (L2), Pie charts (L3), Random Experiments, Sample Spaces, and Events (L4), and Theoretical and Experimental Probability (L5). Together, these tables provide the basis for the subsequent synthesis of findings presented in Section 4.2.

Table 3. Results of coding tasks of L1: Organizing Data

The problem context	**Data embedded**	**Resultant codes**
1. Goals scored by a football team in 34 matches in an Egyptian league.	Numerical data	• 2- AD (1, 1, 0)
2. Relationship between the number of children and the number of families.	A single bar graph	• CD (1, 1, 0) • CD (1, 1, 1) • AD (1, 0, 0)
3. Marks of 25 students in a quiz.	Numerical data	• 3- AD (1, 0, 0) • IR & MD (1, 1, 1)
4. Laptop sales in a computer shop in six months.	Table of categorical data	• AD (1, 1, 0)
5. Masses of a group of 40 judo players in kilograms.	Numerical data	• 2- AD (1, 0, 0)
6. Masses of a group of 30 students in kilograms.	Numerical data	• 3- AD (1, 0, 0) • AD (0, 0, 0)
7. Number of weekly hours spent by 27 students using the Internet.	Numerical data	• 3- AD (1, 0, 0)
8. Temperatures of a city over three weeks.	Numerical data	• 2- AD (1, 0, 0)
9. Number of post offices in four Egyptian governors for a year.	Two bar graphs	• IR & MD (0, 0, 0)
10. Masses of Sarah, Hind, and Amal in kilograms.	Two bar graphs	• IR & MD (0, 0, 0)
11. Collaborate with your classmates to answer the following question: what means of transportation is used to go to school?	No data	• CD (1, 1, 1) • AD (1, 1, 0) • AD (1, 1, 0) • IR (1, 1, 1)
11 Tasks		**27 inquiries**

Note: The numeral preceding each code denotes the frequency with which that code occurred within the same task. Accordingly, "3-AD (1, 0, 0)" indicates that the code AD (1, 0, 0) emerged three times, once in response to each of the three inquiries.

Table 4. Results of coding tasks of L2: Arithmetic Mean

The problem context	Data embedded	Resultant codes
12. The internet consumption of one family over a month in GB over a month.	Frequency table	• AD (1, 0, 0)
13. The arithmetic mean of four numbers.	Numerical data (with no context)	• AD (0, 0, 0)
14. Sara's marks in five tests.	Numerical data	• AD (1, 0, 0)
15. Number of children who took the infantile paralysis vaccine over five days.	Numerical data	• AD & MD (1, 1, 0)
16. Daily pocket money of a student over week.	Frequency table	• AD (1, 0, 0)
17. Masses of the school's Karate team players in KG.	Frequency table	• AD (1, 0, 0)
18. number of working hours for a group of workers	Frequency table	• AD (1, 0, 0)
19. Number of children in each family for a group of families.	Frequency table	• AD (1, 0, 0)
20. Menu of lunch meals in a restaurant display prices in L.E.	Categorical data	• AD (1, 1, 0)
21. A student's marks in five tests.	Numerical data	• AD & MD (1, 1, 0)
22. Using Excel to Find Arithmetic Mean, Median, and Mode	Numerical data (with no context)	• AD (0, 0, 0)
23. Technological Evaluation	No data	• CD & AD (1, 1, 1)
12 Tasks		**12 Inquiries**

Table 5. Results of coding tasks of L3: Pie Charts

No. The problem context	Data embedded	Resultant codes
24. Percentages of water and land that cover the Earth's surface	Categorical data	• AD (1, 0, 0)
25. Results of a survey on the favorite school subject for a group of preparatory students	Pie chart	• 2- AD (1, 0, 0)
26. Elections: Number of votes each of Khaled, Anas, and Hamza gained in the classroom election.	Frequency table	• AD (1, 0, 0)
27. Contributions of three friends (in thousands) to a project.	Dot plots	• AD (1, 0, 0)
28. Sports: In a survey, a group of people were asked about their favorite sport	Table with percentages	• AD (1, 0, 0)
29. A shopping market observed the selling activity for one day.	Table with percentages	• AD (1, 0, 0)
30. Daily Life: how Bassem spends his entire day, represented by a figure.	Horizontal single bar graph with percentages	• AD (1, 0, 0)
31. The preferred season for a group of people.	Bar graph with percentage	• AD (1, 0, 0)
32. Nutrition: The percentage of calories Sally had in her meals in one day.	Pie chart	• 2- AD (1, 0, 0)
33. The broadcasting of a TV channel for 10 hours in one day.	Pie chart	• 2- AD (1, 0, 0)
34. Social Life: The number of children and youth members of one club according to their ages.	Frequency table with intervals	• AD (1, 0, 0)
35. Marks of 30 students in the first-term exam.	Table	• AD (1, 0, 0)
36. Technology: A student's weekly study hours for each school subject.	Table	• AD (1, 0, 0)
13 Tasks		**16 Inquiries**

Table 6. Results of coding tasks of L4: Random Experiment- Sample Space- Events

No. The problem context	Data embedded	Resultant codes
37. Available types of school activities to participate.	Table with categorical data	• AD (1, 0, 1)
38. Experiment of drawing a ball	Categorical data	• 2- AD & MD (1, 0, 0)
39. Multiple experiments of drawing a ball or card	Categorical & Numerical data	• AD & MD (1, 0, 0) • AD (1, 0, 1)
40. Multiple experiments of drawing a ball or a card and rolling a fair cube.	Categorical & Numerical data	• AD & MD (1, 0, 0) • AD (1, 0, 1)
41. Random experiments of tossing a fair coin twice and rolling a fair die twice.	Categorical & Numerical data	• AD (1, 0, 1)
42. Types of juice offered by a restaurant.	Categorical data	• AD (1, 0, 1)
43. Random experiment of rolling a fair die	Numerical data	• AD (1, 0, 1) • AD & MD (1, 0, 0)
44. Random experiment of selecting a number from numbers 2 to 11.	Numerical data	• AD (1, 0, 1) • AD & MD (1, 0, 0)
45. Random experiment of forming 2-digit number of a set of numbers.	Numerical data	• 2- AD (1, 0, 1)
46. Random experiment of rolling a fair die twice.	Numerical data	• AD (1, 0, 0)
10 Tasks		**16 Inquiries**

Table 7. Results of coding tasks of L5: Theoretical & Experimental Probability

No. The problem context	Data embedded	Resultant codes
47. Two football two players taking penalty kicks.	Numerical data	• AD & MD (1, 1, 1)
48. Designing a spinning disc for a game.	Multiple discs, each is divided into circular sectors	• AD & MD (1, 0, 0)
49. Random experiment of rolling a fair die.	Numerical data	• AD (1, 0, 0)
50. Random experiment of drawing a card.	Numerical data	• AD (1, 0, 0)
51. Random experiment of tossing a fair coin twice.	Categorical data	• AD (1, 0, 0)
52. Random experiment of forming 2-digit number of a set of numbers.	Numerical data	• AD (1, 0, 0)
53. Multiple experiment of drawing a ball.	Numerical data	• AD (1, 0, 0)
54. Selection of films in a cinema.	Categorical data	• AD (1, 0, 0)
55. Experiment of tossing a fair coin 100 times.	Numerical data	• AD (1, 0, 0)
56. Experiment of rolling a fair dice 150 times.	Frequency table	• 2- AD (1, 0, 0)
57. Experiment of spinning disc of colored sectors equal in area.	Frequency table	• 3- AD (1, 0, 0)
58. Experiment of drawing a ball.	Table with percentages	• 3- AD (1, 0, 0)
12 Tasks		**17 Inquiries**

4.2. DISCUSSION

As detailed in the previous section, the statistics tasks presented in the five designated lessons were coded according to the embedded constructs of argument: Data, Warrant, Backing, Rebuttal, and Qualifier.

In total, the eighty-eight codes generated were organized into five categories, as illustrated in Figure 4, with each category comprising codes that share the same argumentative constructs.



Figure 4. Distribution of constructs of argument embedded in statistics tasks in the seventh-grade Egyptian mathematics textbook

This discussion begins with Group 4, which accounted for 7% of the analyzed content; it categorized the most effective tasks, in terms of the opportunities afforded for students to construct statistical arguments, as all constructs of statistical arguments were evident in these tasks. An exemplary case of this category is the task presented in Figure 5. In addition, Figure 6 visualizes selected samples of students' responses to this task, collected during the pilot study.

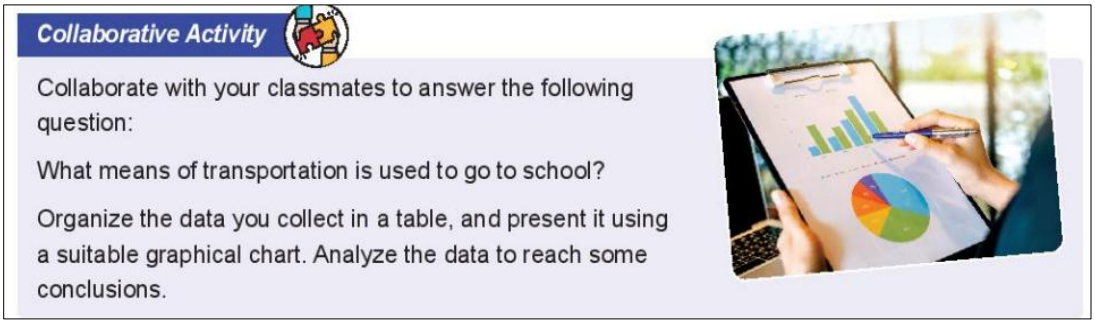
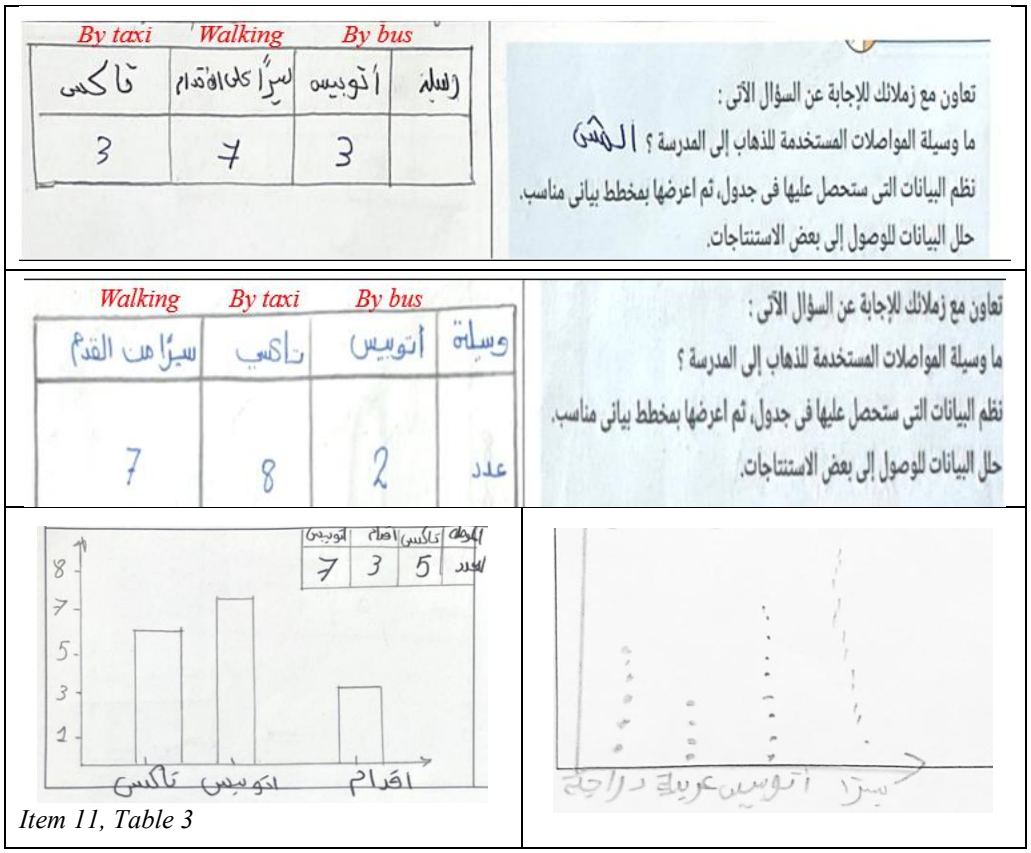


Figure 5. An example for the most promising category of tasks in terms of the opportunities afforded for students to construct statistical arguments



Note: The words colored in red represent the English translation for the Arabic words proposed by the students

Figure 6. Samples of students' responses to the task visualized in Figure 5

The task in Figure 5 was situated in an authentic context, clarifying the means of transportation students use to travel to school; thereby, it prompted them to collect relevant data, analyze it, and draw conclusions. Given the open-ended nature of the task, students' responses were expected to vary depending on the sample they selected for data collection. This variation could, in turn, create opportunities to consider the limitations associated with different samples and the conclusions derived from them. This expectation was ascertained by reflecting on students' responses to the task shown in Figure 5 during the pilot phase. As illustrated in Figure 6, students differed in the transportation modes they identified and considered in responding to the task (e.g., bus, taxi, bicycle, and walking), the forms they used to represent the data they collected (e.g., tables, bar graphs, and dot plots), and the number of cases they selected for investigation (e.g., 13, 15, and 17 surveyed cases). Although these differences primarily reflect variation in students' choices of context, representation, and sampling procedures, they also create opportunities to examine how conclusions might differ across samples and, consequently, to consider the extent to which such conclusions could be generalized beyond the cases investigated.

Furthermore, the Figure 5 task may stimulate the presence of backings, represented by students' knowledge of the various available modes of transportation to reach school, as well as rebuttals that are characterized by the constraints involved in deciding whether each mode is suitable. For example, one could anticipate that the data gathered, and thus the conclusions drawn, would consider a larger number of cars compared to buses, contingent upon whether the school is a public or private institution; furthermore, a constraint related to the distance from the respondent's home to the school may also influence decisions. Accordingly, allowing students to select the surveyed group themselves would introduce variability, personal experiences, and an examination of the limitations associated with each decision.

Additionally, because the data obtained may differ according to the sample selected for investigation, the task does not support a deterministic conclusion that can be straightforwardly generalized to the entire population. Hence, students' conclusions are expected to be inductive and consequently subject to a degree of uncertainty (Groth & Barry, 2025). For instance, a student who finds that the "bus" is the most prevalent mode of transportation among the surveyed students may conclude that using the bus is the preferred option, whereas another student, based on a different sample, may identify "private cars" as the most prevalent mode and argue in favor of them because they offer a quieter or more comfortable means of transportation. Consequently, such conclusions would be discussed as recommendations presented within a probabilistic context. For instance, students may state that "one might consider using the family car if a calm environment is preferred, whereas a school bus could offer a faster and safer alternative." Such an argument contains rebuttals and qualifiers that remain crucial for any statistical argument (Conner & Peters, 2023). This aligns with Groth and Choi's (2023) perspective on evaluating statistical interpretations, recognizing that students' interpretations should demonstrate their ability to balance explaining data-based claims with appropriate expressions of uncertainty and qualification.

Following Group 4, the tasks classified within Groups 2 and 3 were considered acceptable for supporting students' construction of statistical arguments. Collectively, these two groups accounted for approximately 21% of the analyzed content. The principal distinction between Group 2 and Group 3 is that the former primarily addressed uncertainty, whereas Group 3 foregrounded variability across potential alternatives arising during statistical investigations. Despite this distinction, both groups required students to integrate statistical data (e.g., raw data and graphical representations) with contextual knowledge to reach decisions.

Regarding Group 2 (11%), it is important to note that most codes classified within this group were associated with Lesson 4, which focused on developing students' understanding of random experiments, sample space, and events. Figure 7 depicts a task from this category, coded as (RAW, AD, 1, 0, 1) (see Table 6, item 41).

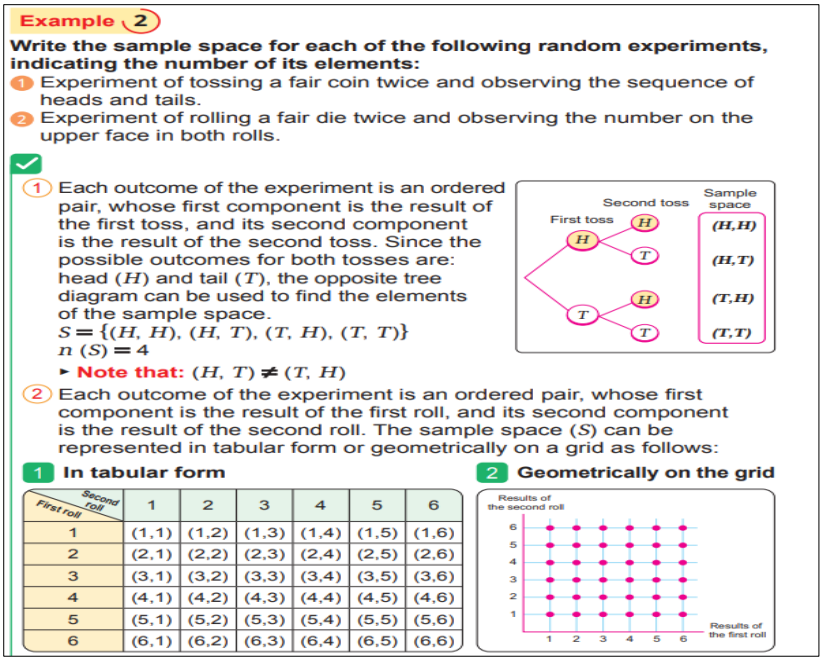


Figure 7. Example 1 for the acceptable category of tasks in terms of the opportunities afforded for students to construct statistical arguments

The task in Figure 7 required students to determine the sample space of random experiments involving tossing a fair coin and rolling a fair die twice. Although probability studies describe such tasks as quasi-pedagogical or conventional tasks, as they require a marginal understanding of the context (e.g., Elbehary, 2021), the author believes that even this seemingly trivial attention to context remains essential to addressing the assertion. In such instances, contextual knowledge may be cultivated by discussing what the elements of a sample space signify and by emphasizing the term "fair" as it appears in simulation-based experiments, which may evoke students' prior experiences with games of chance (i.e., backings). Although variability may not be evident, since all students are likely to identify the same sample space, uncertainty is expected to remain present throughout their discussions; in particular, every event in the sample space of such random experiments remains probable. Alternatively, probabilistic terms are anticipated to emerge in students' discourse, i.e., qualifiers (Gómez-Blancarte & Tobías-Lara, 2018). As shown in Figure 8, students used probabilistic language when determining the sample space of the experiment of tossing a coin twice and when listing the possible outcomes of a situation in which two individuals each select two of three juice types (M, O, and A). These responses suggest that, even when students identify the complete set of possible outcomes, they still realize that the occurrence of any particular event remains inherently uncertain.

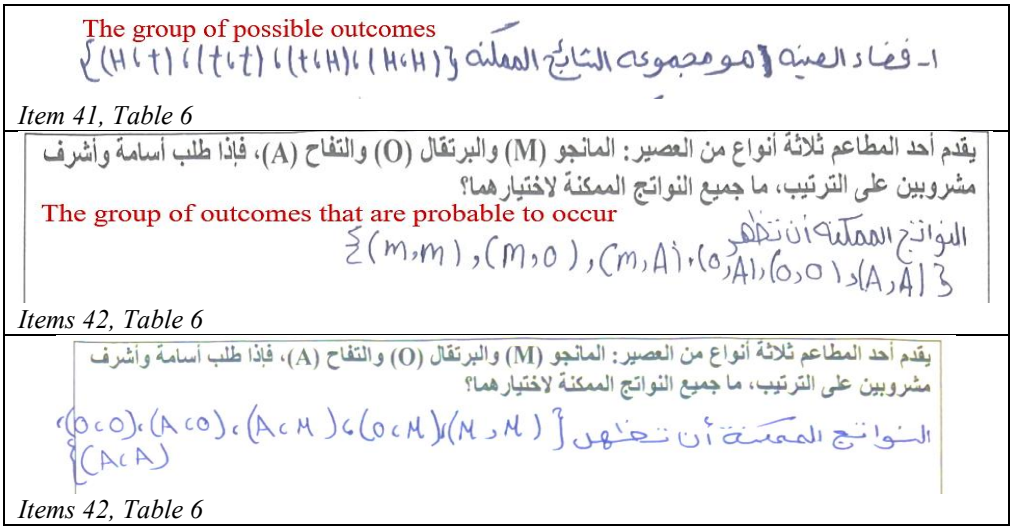


Figure 8. Samples of students' responses to the task visualized in Figure 7, and to another task with the same code

Figure 9 shows an example of Group 3 tasks (10%) that incorporate both backing and rebuttal constructs by considering context and statistical variability. This task (see Table 3, item 1) provides information on the number of goals scored by an Egyptian football team in (34) matches and prompts students to think of a suitable way to organize this data to be easily interpreted; hence, students may deem various possible approaches to organize and represent this data. This was not merely an expectation; the data collected during the pilot study made it empirically evident. Students used diverse representations, such as frequency tables, bar graphs, and dot plots, to categorize and visualize the data (see Figure 10). Interestingly, one student commented, "*Al-Ahly is the only team capable of scoring 63 goals across 34 matches*," which suggests that students retain personal experience related to the task context (i.e., backings).

When students can generate multiple representations of the same dataset, differences in their responses create opportunities to attend to variation, arising from prior experience, instructional norms, and personal heuristics. That is, these dissimilarities help recognize that the organization and representation of the dataset can vary across approaches. In this sense, the task may foreground the broader notion of variability as differences in attributes across individuals or objects (Weiland, 2019a), while also creating opportunities to consider how alternative representations can influence the interpretation of the same dataset. Unlike the

affordance of backing through contextual considerations and rebuttal through variability that appeared in Group 3 tasks, these tasks did not foreground uncertainty. That is, students were not expected to incorporate qualifiers such as 'possibly' or 'likely' into their warrants; these are linguistic indicators of uncertainty in statistical argumentation. The claims embedded in the tasks (see Figure 9) prompted students to determine, with precision and definitiveness, their criteria for selecting an appropriate representation of a dataset.

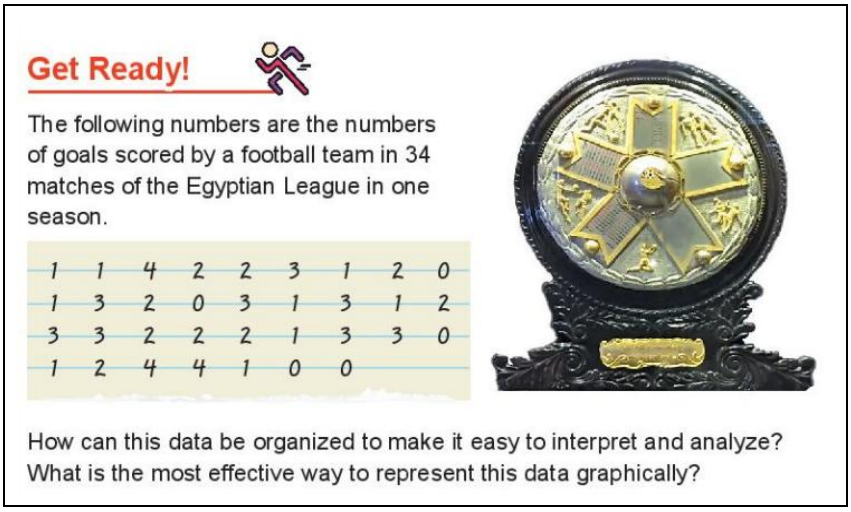
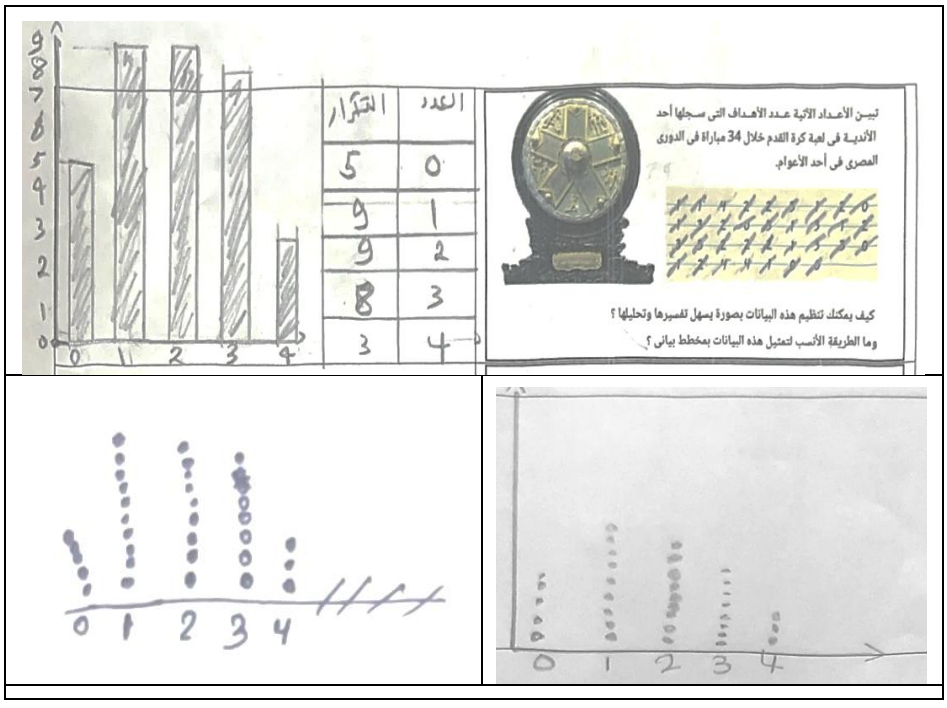
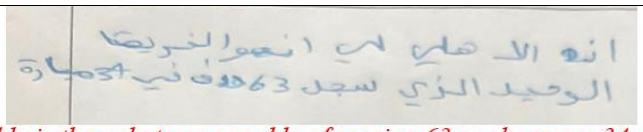


Figure 9. Example 2 for the acceptable category of tasks in terms of the opportunities afforded for students to construct statistical arguments





Al-Ahly is the only team capable of scoring 63 goals across 34 matches

Item 1, Table 3

Figure 10. Samples of students' responses to the task visualized in Figure 9

As noted earlier, while the tasks in Groups 2 and 3 were deemed acceptable for sustaining students' construction of statistical arguments, pedagogical improvements that enhance them are still achievable. One potential recommendation is to restructure the task to allow students to design their own experiments or statistical investigations, thereby fostering decision-making under uncertainty through inductive approaches that may vary across students. Designing such activities would further highlight the subjectivity that arises from biases in data production, which manifest in the selection and allocation of samples and, consequently, affect students' arguments and decisions (Weiland, 2019a, 2023).

Additionally, one might consider statistical inquiries that can be addressed through various approaches or concepts to enrich classroom discourse about the rationale, circumstances, and limitations behind them. Considering the students' familiarity and the relevance of the inquiry context, they can reflect on their personal experiences to identify likely limitations (Büscher, 2022; Fielding, 2024). This highlights the significance of integrating both theoretical and practical contexts; whereas data derived from the former can be verified through simulators, data from the latter can be estimated based on students' knowledge and personal experiences (Zapata-Cardona & Marrugo Escobar, 2019). A complementary view to this argument is Weiland's (2019b) standpoint on involving students in controversial situations presented in media and societal debates to foster critical arguments rather than concentrating on fictional statistical scenarios (e.g., personal preferences or characteristics) that are prevalent in most school textbooks.

Ultimately, the tasks in Group 1 and Group 5 were assessed as distant from enabling students to construct statistical arguments. Group 5 tasks were the least prevalent category, comprising only (6%), whereas Group 1 tasks constituted the predominant type of statistics task in the Egyptian Grade 7 textbooks, accounting for (66%) of the content. Both categories similarly neglect variability and uncertainty, thus constructs of rebuttals and qualifiers, but they differ in the extent to which context is required to respond to their claims.

An example of Group 5 tasks is presented in Figure 11. Students are asked to use Excel to compute measures of central tendency for a set of numbers without any information regarding the meaning of these numbers or the context from which they were obtained (see Table 4, item 22). In that sense, it is anticipated that students will perform specific procedures in Excel to obtain certain values, regardless of what these values signify. Such types of tasks are neither suitable for learning statistical concepts nor for engaging students in constructing statistical arguments. It deviates from the fundamental principle established by Cobb and Moore (1997), which declares that "data are not just numbers, they are numbers with a context" (p. 801); that is, interpreting statistical data must always consider the contextual grounds underpinning this data (Wild & Pfannkuch, 1999).

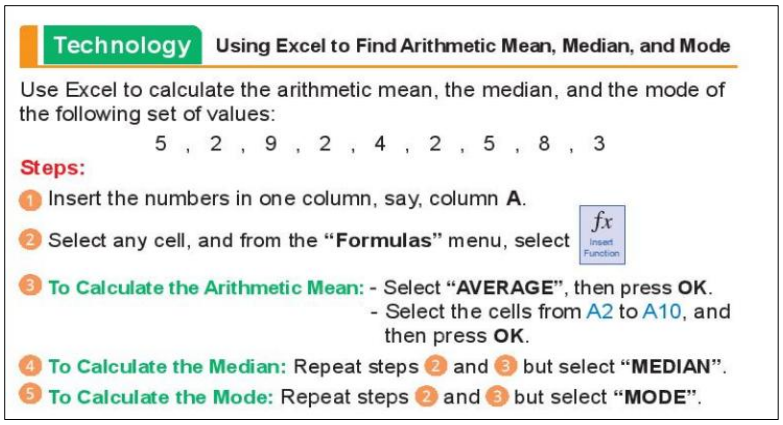


Figure 11. Example 1 for the inadequate category of tasks in terms of the opportunities afforded for students to construct statistical arguments

Unlike Group 5 tasks, Group 1 tasks treated context as relevant to the claims, yet it was deemed insufficient to scaffold students' construction of statistical arguments. These tasks were most prevalent in Lesson 3 (Pie Charts; see Table 5) and Lesson 5 (Theoretical & Experimental Probability; see Table 7). Nearly all tasks designed to develop students' comprehension of pie charts, such as the example displayed on the left side of Figure 12, require students merely to represent the given data. Accordingly, instructional attention focused primarily on computing angle measures corresponding to the percentage share of each data category; regardless, no rebuttals or qualifiers emerged. Put differently, orienting the task toward merely representing the given data using a pie chart (e.g., percentages of students' favorite sports) constrains opportunities to negotiate substantive inquiries, such as the conditions under which selecting handball might be more feasible than selecting another sport, and the extent to which these percentages can be believed reliable and stable across diverse contexts.

Similarly, the task illustrated on the right side of Figure 12 required students to simply compute the probabilities of multiple events without any discussion of either the constraints of the situation or the uncertainty of the outcomes. Such tasks differ from those of Group 2 (see Figure 7), which elicit arguments about the uncertainty of results as determined by the elements of the sample space of a random experiment. This aligns with Jones et al.'s (2015) distinction between probability tasks situated within a statistical context (Group 2) and those that demand mathematical computation (Group 1). Figure 13 displays a sample of students' responses to tasks categorized in Group 1; the upper model exemplifies a response to a task that required representing the number of votes received by three individuals in the classroom election using a pie chart, while the lower one depicts a calculation of experimental probability.

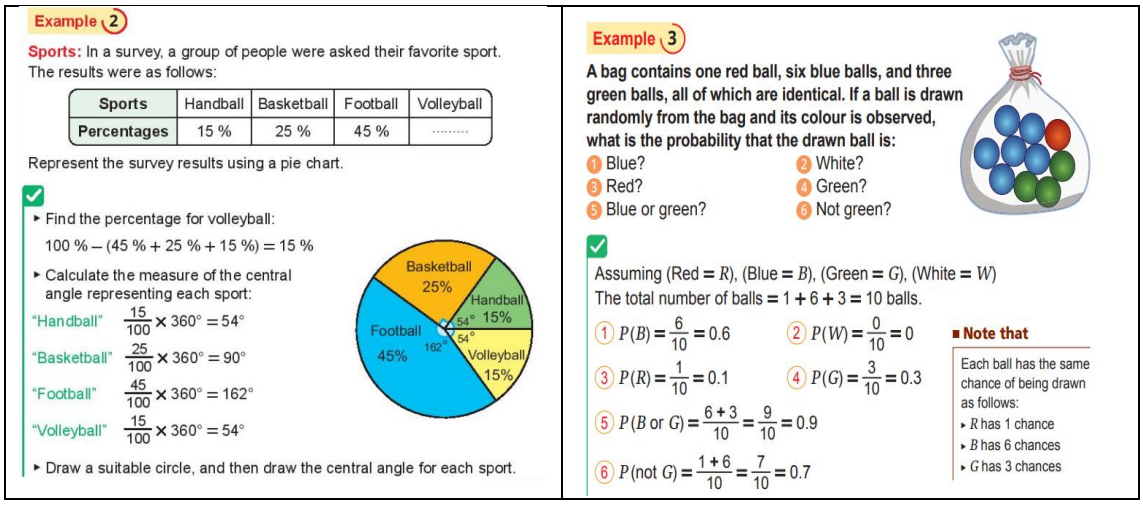


Figure 12. Examples 2 and 3 for the inadequate category of tasks in terms of the opportunities afforded for students to construct statistical arguments

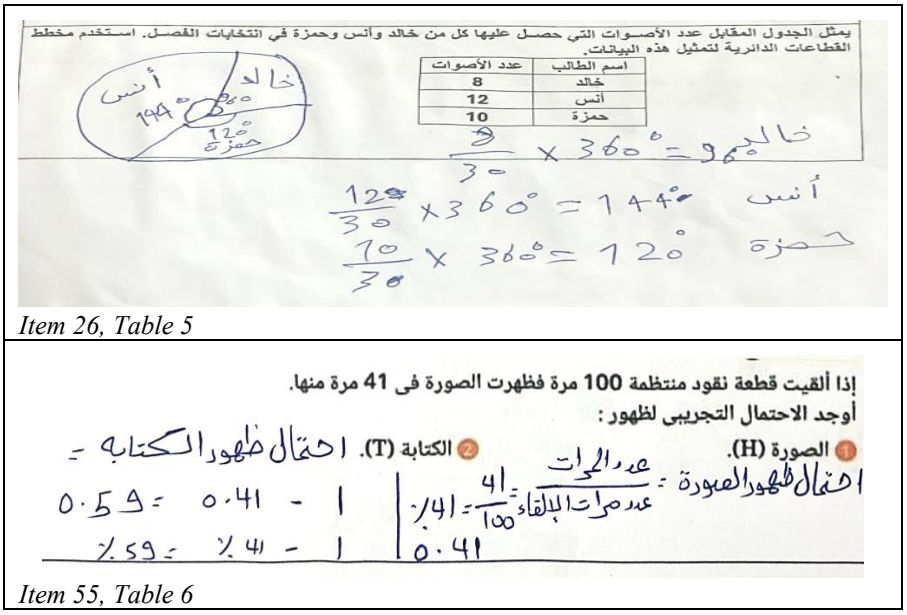


Figure 13. Samples of students' responses to the tasks categorized in Group 1

5. CONCLUSION AND RECOMMENDATIONS

The primary principle underpinning the present study is to "learn to argue" by debating the extent to which the statistics tasks afforded in the Egyptian Grade 7 mathematics textbooks may help students practice statistical argumentation. Precisely, the study sheds light on the limitations of the opportunities afforded by the Egyptian Grade 7 textbooks to sustain students' construction of statistical arguments, indicating that approximately (7%) of the content was deemed sufficient, (21%) required modifications, and (72%) remained useless for provoking argument constructs, mainly rebuttals and qualifiers.

This analysis gives teachers practical evidence on the status of the statistics tasks offered in textbooks; therefore, they must decide whether to deliver them as written or adapt them. Alternatively, teachers remain responsible for cultivating classroom statistical discourse to practice argumentation, which they can do by

posing questions that encourage students to express data and justifications, and by stimulating examination of the diversity of ideas and the constraints of selections made in uncertain situations.

The study recommends that textbook writers should prioritize tasks that engage students in all phases of the statistical investigation cycle, from identifying a problem to reaching conclusions, rather than limiting them to work with pre-existing data, with an emphasis on data analysis and interpretation. Regardless, one should realize that any changes confined to a single dimension, such as modifying specific tasks, remain inadequate for achieving sustainable change in statistics teaching and learning (Ben-Zvi et al., 2018). From this perspective, this study, aligning with Ben-Zvi et al. (2018), proposes an extended investigation that incorporates its findings into a comprehensive instructional design encompassing content, pedagogy, tasks, tools, assessment, and classroom norms to achieve the intended outcome of students' statistical arguments.

When interpreting the present study results, several limitations should be acknowledged. First, the analysis was limited to the five lessons classified under the Statistics and Probability units of the newly issued Grade 7 mathematics textbook. This limited sample restricts the scope to which the study findings can be generalized to the broader treatment of statistics in the Egyptian curriculum or to statistics education in the African context. Second, although research on argumentation emphasizes the importance of contextual and cultural norms in interpreting arguments, it is also necessary to note that the coding protocol was validated primarily through local researchers' judgments and a relatively small sample of student participants who had not yet engaged sufficiently with the relevant statistics content. Therefore, further studies examining a broader range of lessons, textbooks, grade levels, and curricular contexts are recommended to support more comprehensive conclusions. In addition, the coding protocol developed in this study should undergo further critical empirical validation across diverse educational and cultural contexts to enhance its reliability, robustness, and transferability.

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to Dr. Fatma Hamada and Ms. Esraa Elshemy for their valuable assistance with data collection and coding during the pilot study.

DISCLOSURE AND DATA AVAILABILITY STATEMENT

No potential conflict of interest was reported by the author. The author confirms that the data supporting the findings of this study are available within the article.

REFERENCES

Ajani, M.O. (2023). Reasoning and argumentation in teaching mathematics and mathematics achievement in Oyo State, Nigeria. *Global Scientific Journal, 11*(5), 2394–2413. https://www.globalscientificjournal.com/researchpaper/Reasoning_and_Argumentation_in_Teaching_Mathematics_and_Mathematics_Achievement_in_Oyo_State_Nigeria_.pdf

Balacheff, N. (2024). Mathematical argumentation, a precursor concept of mathematical proof. In J. Wang (Ed.), *Proceedings of the 14th International Congress on Mathematical Education: Volume II: Invited Lectures* (pp. 33–52). ICME. https://doi.org/10.1142/9789811287183_0003

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K-12 guidelines for assessment and instruction in statistics education II*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Ben-Zvi, D., Gravemeijer, K., & Ainley, J. (2018). Design of statistics learning environments. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 473–502). Springer. https://doi.org/10.1007/978-3-319-66195-7_16

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics:*

Challenges for teaching and teacher education—A joint ICMI/IASE study (pp. 57–69). Springer. https://doi.org/10.1007/978-94-007-1131-0_10

Burrill, G., & Pfannkuch, M. (2023). Emerging trends in statistics education. *ZDM – Mathematics Education, 56*(1), 19–29. https://doi.org/10.1007/s11858-023-01501-7

Büscher, C. (2022). Learning opportunities for statistical literacy in German middle school mathematics textbooks. In J. Hogden, E. Geraniou, G. Bolondi, & F. Ferretti (Eds.), *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education (CERME 12)* (pp. 845–852). Free University of Bozen-Bolzano and ERME. https://hal.science/CERME12/hal-03751814v1

Chua, B. L. (2017). A framework for classifying mathematical justification tasks. In T. Dooley & G. Gueudet (Eds.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education (CERME 10)* (pp. 115–122). DCU Institute of Education and ERME. https://hal.archives-ouvertes.fr/CERME10-TWG01/hal-01873071v1

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

Conner, A., & Peters, S. A. (2023). Distinctive aspects of reasoning in statistics and mathematics: Implications for classroom arguments. In G. F. Burrill, L. de Oliveira Souza, & E. Reston (Eds.), *Research on reasoning with data and statistical thinking: International perspectives* (pp. 259–278). Springer. https://doi.org/10.1007/978-3-031-29459-4_20

Conner, A., Singletary, L. M., Smith, R. C., Wagner, P. A., & Francisco, R. T. (2014). Teacher support for collective argumentation: A framework for examining how teachers support students' engagement in mathematical activities. *Educational Studies in Mathematics, 86*(3), 401–429. https://doi.org/10.1007/s10649-014-9532-8

Corrêa, S. A., & Lopes, C. E. (2018). Creative insubordination and the argumentative process in statistical education in early childhood. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward: Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July 2018, Kyoto, Japan).* Voorburg, The Netherlands: International Statistical Institute.

Egyptian Ministry of Education and Technical Education. (2024). *Mathematics: First preparatory, first term, 2024–2025 (Student's book).* https://studentbooks.moe.gov.eg/

Elbehary, S. G. A. (2020). Providing insights into the intended curriculum of probability for basic education sector in Egypt from a comparative perspective. *International Electronic Journal of Mathematics Education, 15*(2). https://doi.org/10.29333/iejme/7603

Elbehary, S. G. (2021). Reasoning under uncertainty within the context of probability education: A case study of preservice mathematics teachers. *Pythagoras, 42*(1). https://doi.org/10.4102/pythagoras.v42i1.630

Elbehary, S. G. A. (2026). Practices of mathematical argumentation among primary school students in Egypt. *Mathematical Thinking and Learning*, 1–25. https://doi.org/10.1080/10986065.2026.2623617

Fielding, J. (2024). Taking an argumentation approach to statistical investigations: Developing student data-ing practices. *ZDM – Mathematics Education*, *57*, 31–44. https://doi.org/10.1007/s11858-024-01639-y

Fielding-Wells, J. (2016). Mathematics is just 1 + 1 = 2, what is there to argue about? In B. White, M. Chinnappan, & S. Trenholm (Eds.), *Opening up mathematics education research: Proceedings of the 39th annual conference of the Mathematics Education Research Group of Australasia* (pp. 214–221). MERGA.

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 pre-K–12 curriculum framework.* American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEPreK-12_Full.pdf

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Gómez-Blancarte, A., & Tobías-Lara, M. G. (2018). Using the Toulmin model of argumentation to validate students' inferential reasoning. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward: Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July*

2018, Kyoto, Japan). International Statistical Institute. https://iase-web.org/Conference_Proceedings.php?p=ICOTS_10_2018

Groth, R. E., & Barry, J. P. (2025). Argument structures of responses to a contextually provocative hospital problem variant. *Statistics Education Research Journal, 24*(1), 5. https://doi.org/10.52041/serj.v24i1.764

Groth, R. E., & Choi, Y. (2023). A method for assessing students' interpretations of contextualized data. *Educational studies in mathematics*, 1–18. https://doi.org/10.1007/s10649-023-10234-z

Hussein, H. (2023). Trends of mathematics education research studies published in *Journal of Mathematics Education* from 2017 to 2021. *The International Journal for Research in Education, 47*(1), 49–91. https://doi.org/10.36771/ijre.47.1.23-pp49-91

Jones, D. L., & Tarr, J. E. (2007). An examination of the levels of cognitive demand required by probability tasks in middle grades mathematics textbooks. *Statistics Education Research Journal, 6*(2), 4–27. https://doi.org/10.52041/serj.v6i2.482

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). https://doi.org/10.1080/10691898.2015.11889748

Jones, D. L., & Jacobbe, T. (2014). An analysis of the statistical content in textbooks for prospective elementary teachers. *Journal of Statistics Education, 22*(3). https://doi.org/10.1080/10691898.2014.11889713

Kartika, H., Budiarto, M. T., & Fuad, Y. (2021). Argumentation in K-12 mathematics and science education: A content analysis of articles. *International Journal of Research in Education and Science (IJRES), 7*(1), 51–64. https://doi.org/10.46328/ijres.1389

Kazak, S. (2023). *Pre-service mathematics teachers' statistical problem-solving process for telling a story through the use of photographs as data sources.* In P. Drijvers, C. Csapodi, H. Palmér, K. Gosztonyi, & E.Kónya (Eds.), *Proceedings of the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME 13)*. Alfréd Rényi Institute of Mathematics and ERME. https://hal.science/hal-04410967

Kim, Y. (2010). The pilot study in qualitative inquiry: Identifying issues and learning lessons for culturally competent research. *Qualitative Social Work*, *10*(2), 190–206. https://doi.org/10.1177/1473325010362001

Krummenauer, J., & Kuntze, S. (2019). Primary students' reasoning and argumentation based on statistical data. In U. T. Jankvist, M. van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education (CERME 11)*. Freudenthal Group & Freudenthal Institute, Utrecht University and ERME. https://hal.science/hal-02398118

Kuş, M. (2022). A comparative textbook analysis on measures of center in selected school mathematics textbooks of Australia and Turkey. *International Journal for Mathematics Teaching and Learning, 23*(2), 144–173. https://doi.org/10.4256/ijmtl.v23i2.381

Landtblom, K. (2023). Opportunities to learn mean, median, and mode afforded by textbook tasks. *Statistics Education Research Journal, 22*(3), Article 6. https://doi.org/10.52041/serj.v22i3.655

Makar, K., & Rubin, A. (2009). *A framework for thinking about informal statistical inference*. *Statistics Education Research Journal*, 8 (1), 82–105. https://doi.org/10.52041/serj.v8i1.457

Moustafa, N., Elghamrawy, E., King, K., & Hao, Y. (2022). Education 2.0: A vision for educational transformation in Egypt. In Y. Hao, K. King, & N. Moustafa (Eds.), *Education to build back better* (pp. 51–74). Springer. https://doi.org/10.1007/978-3-030-93951-9_3

Pardue, T. M. (2017). *The heuristics of statistical argumentation: Scaffolding at the postsecondary level* [Doctoral dissertation, University of Georgia]. ERIC. https://eric.ed.gov/?id=ED576377

Petocz, P., Reid, A., & Gal, I. (2018). Statistics education research. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 71–99). Springer. https://doi.org/10.1007/978-3-319-66195-7_3

Pfannkuch, M. (2018). Reimagining curriculum approaches. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 387–413). Springer. https://doi.org/10.1007/978-3-319-66195-7_12

Robitaille, D. F., & Travers, K. J. (1992). International studies of achievement in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 687–709). Macmillan.

Schield, M. (1999). Statistical literacy: Thinking critically about statistics. *Of Significance, 1*(1), 15–20.

Shongwe, B. (2022). The quality of argumentation in an Euclidean Geometry context: A case study. *International Journal of Innovation in Science and Mathematics Education*, *30*(5). https://doi.org/10.30722/ijisme.30.05.004

Toulmin, S. E. (2003). *The uses of argument* (Updated ed.). Cambridge University Press. (Original work published 1958)

Tran, D., & Tarr, J. E. (2018). Examination of bivariate data tasks in U.S. 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

Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H., & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks.* Kluwer Academic Publishers. https://doi.org/10.1007/978-94-007-0844-0

Weiland, T. (2019a). Critical mathematics education and statistics education: Possibilities for transforming the school mathematics curriculum. In G. Burrill & D. Ben-Zvi (Eds.), *Topics and trends in current statistics education research* (pp. 391–411). Springer. https://doi.org/10.1007/978-3-030-03472-6_18

Weiland, T. (2019b). The contextualized situations constructed for the use of statistics by school mathematics textbooks. *Statistics Education Research Journal, 18*(2), 18–38. https://doi.org/10.52041/serj.v18i2.138

Weiland, T. (2023). High school mathematics texts construction of statistics practices. *Frontiers in Education, 8,* Article 1186522. https://doi.org/10.3389/feduc.2023.1186522

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

Zapata-Cardona, L., & Marrugo Escobar, L. M. (2019). Critical citizenship in Colombian statistics textbooks. In G. Burrill & D. Ben-Zvi (Eds.), *Topics and trends in current statistics education research* (pp. 373–389). Springer. https://doi.org/10.1007/978-3-030-03472-6_17

SAMAH GAMAL AHMED ELBEHARY
Faculty of Education
Tanta University
El-Gash Street
Tanta 31511
Egypt