trends in higher education

MDPI

Article

Development of Statistical Reasoning in Pharmacy Students Through a Hypothetical Learning Trajectory Using a Dissolution Profile Simulator

Martín Hernández-Rivera [1], Claudia Orozco-Rodríguez [2,*] and Nelson Bruno de Almeida-Cunha [3]

[1] Maestría en Enseñanza de las Matemáticas, UCEI, Universidad de Guadalajara, Guadalajara 44100, Mexico; martin.hernandez7434@alumnos.udg.mx
[2] Departamento de Matemáticas, CUCEI, Universidad de Guadalajara, Guadalajara 44100, Mexico
[3] Departamento de Farmacobiología, CUCEI, Centro de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Guadalajara 44430, Mexico; bruno.dealmeida@academicos.udg.mx
* Correspondence: claudia.orozcor@academicos.udg.mx

Abstract

Given the challenges in statistics education, such as limited view of variability, memorization-based approach, and the use of decontextualized problems, among others, this study aimed to analyze the development of statistical reasoning on variability among undergraduate students enrolled in the Bachelor's program in Pharmaceutical Chemistry and Biology through a didactic sequence based on conjecture map and a retrospective analysis using the SOLO (Structure of the Observed Learning Outcome) Taxonomy. In this sense, the conjecture map and the hypothetical learning trajectory proposed that interpreting variability through statistical measures (measures of central tendency and dispersion), together with verbal and statistical argumentation and graphical representations, would enable students to attain the Multistructural and Relational levels of understanding. Initially, the students' statistical reasoning was algorithmic and decontextualized, placing them at the Prestructural, Unistructural and Multistructural levels. Nonetheless, following the hypothetical learning trajectory, most students progressed to the Relational and Multistructural levels, demonstrating the ability to apply quantitative variability in their solutions to the proposed problems. These findings support the conjecture map underlying this design research. It is concluded that prioritizing understanding variability and contextualized argumentation, rather than memorizing formulas, fostered the development of students' statistical reasoning.

Keywords: statistical reasoning; variability; SOLO Taxonomy; simulation data

Academic Editor: Ricarda Micallef

Received: 8 September 2026
Revised: 13 November 2026
Accepted: 16 November 2026
Published: 25 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Descriptive statistics plays a fundamental role in higher education, and it is relevant in fields such as engineering and pharmaceutical sciences, as it enables students to interpret large amounts of data and make informed decisions based on statistical measures. However, the way statistics is taught presents several problems that limit students' understanding and meaningful application of the subject. Statistics is often taught through a rote-learning approach in which students are required to memorize formulas and concepts, limiting their understanding [1,2] and leading students, including

future engineers, to view statistics as a mere set of rules, without understanding its practical application.

Furthermore, research by Shaughnessy [3], and Orta and Sánchez [4] notes that statistics is often taught with a narrow view of dispersion, limiting it to measures such as standard deviation or variance. These authors propose that it is necessary to address the multiple interpretations of variability in different contexts so that students can make sense of variability, rather than learn how to calculate measures of dispersion, which, according to Garfield and Ben-Zvi [5], rarely understand what these measures represent, either numerically or graphically, and do not recognize their importance or their connection to other estimators.

In this sense, a central element in exploring statistical reasoning is the types of tasks and contexts in which it takes place. Orta and Delgado [6] define risk context as a situation of uncertainty in which data variability is analyzed to assess potential consequences and support informed decision-making. It highlights that statistical dispersion is not merely a mathematical concept but can have significant practical implications for decision-making, though. The integration of risk context helps students understand variability and link greater dispersion to higher risk when making decisions. Therefore, analyzing students using the SOLO Taxonomy, they found that 95% remained at the Prestructural, Unistructural, and Multistructural levels of understanding, showing difficulty interpreting dispersion meaningfully. This aligns with Koklu and Kaplan [7], who observed that undergraduate students often rely on primitive notions of variability, focusing their attention solely on maximum and minimum values or equating variability exclusively with the range. Understanding variability in this way limits students' statistical reasoning, leading them to create restricted notions of variability and to disregard it in problem-solving. This study suggests using more data, diverse graphical representations, and richer contexts to develop an understanding of variability.

Addressing this need for contextualization, various studies show that task design directly affects which concepts students evoke. For example, in their systematic review, Boels et al. [8] found that students have significant difficulty linking the problem's context to graphs such as histograms, tending to omit variability and focus almost exclusively on the shape or center of the data. Similarly, Kaplan et al. [9] showed that subtle variations in question formulation drastically alter undergraduates' responses; if the task design does not intentionally guide attention toward dispersion, the concept goes unnoticed. This highlights that working with a large amount of real data is not enough: students must interpret shape, center, and variability [9].

This directly grounds the design of the hypothetical learning trajectory in the present work. Through this, the aim is for students to move beyond isolated algorithmic calculation and consider variability as a determining factor in problem-solving. To ensure students advance toward higher levels of reasoning and understanding, as noted by Chaphalkar and Wu [10], technology is indispensable. In this regard, García-García et al. [11] demonstrated that educational simulators help students reach the Multistructural and Relational levels in the SOLO Taxonomy. Thus, articulating risk contexts, focusing tasks on dispersion, and using interactive tools provide the means for students to evaluate the behavior of the data as a whole.

This article presents the design and implementation of a teaching sequence on variability, with a particular focus on the analysis of measures of dispersion, for learning descriptive statistics. The students' responses across the different stages of the hypothetical learning trajectory were evaluated using the SOLO Taxonomy, which made it possible to identify the level of complexity and depth of their answers and, based on their characteristics, to analyze the statistical reasoning they demonstrated. To this end, worksheets, a problem set featuring real-life scenarios involving risks, a simulator called

BioSimuLab, and Excel were used to help students develop the ability to interpret, compare, and evaluate drug dissolution profiles. Students first completed the activity on paper and then used BioSimuLab data, so that it made it possible to apply statistical estimators to real-world, meaningful situations.

Interactive tools and simulations in statistics education enhance conceptual understanding and data-driven reasoning [12], and thus, according to Orozco-Rodríguez et al. [13], digital simulators and laboratories recreate realistic environments that facilitate students' active participation in all stages of the statistical process: from problem formulation, data generation, and collection to analysis, interpretation, and decision-making. In this study, the BioSimuLab simulator generated data and replicated a laboratory experimentation process in the classroom. This approach contextualized statistical estimators in real, meaningful situations [14]. The Materials section provides further details on the simulator.

Therefore, the objective was to analyze the development of statistical reasoning regarding variability in a group of students in the Bachelor's program in Pharmaceutical Chemistry and Biology. The analysis focused on students' responses to the activities implemented as part of a hypothetical learning trajectory on descriptive statistics, as well as on the characteristics of these responses, which were classified according to the SOLO Taxonomy, which led to a learning trajectory developed in accordance with the design conjectures formulated for this study.

In general, the sequence covers descriptive statistics concepts such as the mean, standard deviation, coefficient of variation, quartiles, maximum and minimum values, box-and-whisker plots, and scatter plots, in the context of risk scenarios. In this regard, the research question guiding this study is: How does statistical reasoning about variability develop among students in the Bachelor's program in Pharmaceutical Chemistry and Biology through the implementation of a hypothetical learning trajectory on descriptive statistics, informed by the research conjectures?

1.1. Conceptual Framework

Because this design study analyzes statistical reasoning about variability, we will distinguish between variability and variation and define the subject of this design study.

1.1.1. Variability and Statistical Reasoning

Variability is defined as an observable characteristic or property of an object or phenomenon that makes it susceptible to change. Under this conceptualization, variation refers to the process of describing and quantifying this characteristic of variability through cognitive and statistical procedures, which represents the property of changing, while variation constitutes its formal description and measurement based on observed data [15].

Garfield and Ben-Zvi [16] note that variability gives statistics its meaning and essence as a discipline, underscoring the importance of studying and understanding it. Therefore, according to Garfield and Ben-Zvi [16] (p. 24), "Statistical reasoning is the way people reason with statistical concepts and make sense of statistical information, which may involve making connections between concepts (for example, means and dispersion) or combining ideas about data and chance. Statistical reasoning also means understanding and explaining statistical processes and interpreting statistical results."

Reading and Shaughnessy [15] point out that reasoning about variability must consider several aspects. Their proposal expanded Wild and Pfannkuch's [17] original list from four to five elements by adding description and representation. Together, these elements structure statistical reasoning into a cohesive process, which begins with noticing and acknowledging variation in the data, and then describing and representing it to communicate it effectively using visual tools. Moreover, it involves measuring and

modeling to predict, explain, or control phenomena through measures of dispersion, and also requires explaining and managing variation by searching for its causes, learning to work with it, and incorporating it into decision-making. Ultimately, this reasoning includes developing investigative strategies that proactively design methods to account for variation at its source.

Teaching variability and shaping statistical reasoning require the design of specific pedagogical strategies. Sánchez et al. [18] propose three fundamental principles for designing these pedagogical strategies. First, instruction should be organized around problem-solving by structuring class sessions centered on the resolution of problematic situations. Second, promoting teamwork is essential to foster collaboration, encourage collective discussion, contrast ideas, and develop mathematical and statistical discourse among students. Finally, using applied technology is crucial, as technological tools help make abstract concepts concrete, drive statistical reasoning, and facilitate the construction of mathematical and statistical meaning.

Variability is a fundamental component of statistical reasoning, which involves understanding and explaining data through the relationships between ideas, data, and chance. It also requires assessing how variability influences outcomes in problem-solving, moving beyond the mere application of routine procedures.

Under this conceptualization, statistical reasoning serves as the unifying framework of the present study. This materializes in the study's hypothetical trajectory, which gives students a way to engage in problem-solving in risk contexts, using interactive simulators as applied technology to make sense of variability. In turn, the SOLO Taxonomy serves as a reference framework to evaluate the extent to which students navigate these aspects of reasoning.

1.1.2. SOLO Taxonomy

Statistical reasoning can be analyzed and categorized using the SOLO Taxonomy (Structure of the Observed Learning Outcome), proposed by Biggs and Collis [19], as it provides a framework for examining the quality and structural complexity of learning reflected in students' responses. The taxonomy, for its part, describes a progression from responses in which students have not yet recognized the relevant elements of a task to those in which they integrate, connect, and transfer those elements to more complex situations. This progression includes the Prestructural, Unistructural, Multistructural, Relational, and Extended Abstract levels [19], which are defined in more detail later in the paper; specifically, their definitions appear in the third column of Table 1.

In statistics education, these levels can be interpreted in terms of how students identify, use, and connect statistical information when justifying their answers. For this reason, the SOLO Taxonomy has been used as a framework to classify student responses and, based on their characteristics, analyze the statistical reasoning they demonstrate. Its use makes it possible to consider not only whether statistical concepts or procedures are present but also how they are selected, connected, and integrated to support a conclusion.

Table 1 synthesizes the basic characteristics of these levels, although it is not a direct reproduction from a single source, but a synthesis of the reviewed literature. Specifically, the first two columns (Capacity and Related Operation) paraphrase the general cognitive foundations proposed by Biggs and Collis [19]. According to these authors, capacity refers to the working memory or attentional capacity required to process different levels of understanding. Meanwhile, the related operation describes the logical way the stimulus and response interrelate at a given level of understanding. For example, at the Relational level, the related operation is structured and coherent, whereas at a Prestructural level, it manifests through superficial or confusing mechanisms. On the other hand, the third column of the table integrates the general definition of the comprehension levels

structured by Orta and Sánchez [4]. This unified framework served as the analytical basis for classifying the students' responses.

Table 1. Levels of the SOLO Taxonomy.

Level	Capability [16]	Related Operation [16]	Statistical Characterization [3]
Prestructural	Minimal: characterized by signals and responses that lack clarity or are disorganized.	Negation, tautology, and transduction, reflecting a limited comprehension bound to specific details, without making broader or more meaningful connections.	The students' response does not present any relevant elements.
Unistructural	Low: a single relevant fact accompanied by a cue, but without making connections to other contextual elements.	Generalization is limited and is made solely in terms of an isolated aspect, without integrating other elements or relationships.	The students' response presents only one relevant element.
Multistructural	Intermediate: Multiple relevant facts are identified, but they remain isolated and accompanied by cues, without making meaningful connections between them.	Generalization is limited to a few independent and isolated aspects, without integrating a broader or more complex view of the problem.	The students' response includes two or more relevant elements, but without establishing clear links between them.
Relational	High: the student is able to identify key elements along with relevant data and establish interrelationships between them, demonstrating a deeper and more organized understanding of the problem or topic.	The student is able to perform induction, generalizing within a given or experienced context, using related aspects and establishing connections between them to reach broader conclusions.	The student's response integrates several elements in a coherent manner.
Extended Abstract	Maximum: the student is able to integrate cues, relevant data, and relationships between them, in addition to formulating hypotheses to apply knowledge to new contexts or situations not previously experienced, demonstrating highly complex and creative reasoning.	The student is able to apply deductive and inductive reasoning to generalize to previously unexperienced situations, utilizing learned principles and relationships to make predictions or formulate solutions in new or abstract contexts.	The student's response achieves a Relational level, but it is extended and applied to other situations or tasks; to identify this level, it is generally necessary to extend the given tasks.

1.1.3. Design Research

Design research is the systematic study of the planning, development, and evaluation of educational interventions, with the aim of providing solutions to complex problems in educational practice. Furthermore, it seeks to generate knowledge about the characteristics of these interventions, and their design and development processes; therefore, it is a methodological approach, not a methodology [20–22].

According to Bakker [22] and Cobb et al. [23], the transversal characteristics of design research are, first and foremost, that its purpose is not only to develop theories about learning, with an emphasis on designing media to support student or teacher learning, which is interventionist in nature, but also incorporate both reflective and prospective components. In this case, the reflective components involve basing design research on conjectures, which are organized through Conjecture Mapping [24] and used for the initial design. Prospective components refer to the way in which designs are implemented through hypothetical learning trajectories. Because design research includes prospective and reflective components, the process becomes cyclical, consisting of three phases:

preparation and design, the teaching experiment, and retrospective analysis. Thus, beyond applying developed theories to different contexts, theory must do real work, which relates not only to a domain's learning processes but also to design activity.

1.1.4. High-Level, Design, and Theoretical Conjectures

The overarching hypothesis is that integrating risk contexts and using dynamic or static real-world data, facilitated by collaborative discursive practices, supports learning. The BioSimuLab simulator promotes statistical reasoning and enables students to solve problems by considering variability through measures of central tendency and/or dispersion, as well as their graphical representations.

The design conjectures for this research are as follows:

- SOLO Taxonomy: If students achieve visualization and interpretation of the aspects inherent to variability when answering questions designed based on the levels of comprehension (described in the Data Analysis), then the process of visualizing, describing, analyzing, and interpreting variability will emerge. This will enable students to integrate and coordinate estimators, such as measures of central tendency and/or dispersion, to support data-based decision-making in risk-related contextual problems, which will in turn foster the development of levels of comprehension among students.
- Production of artifacts: If students prepare statistical reports in their field (where they can apply their interdisciplinary knowledge) and create graphical representations, both individually and collaboratively, to solve risk-related problems using data that exhibit variability, then a process of quantification and reasoning based on estimators will emerge, in addition, relating different estimators (measures of central tendency and/or dispersion) to support conclusions.
- Discursive practice: If students take an active role in group discussions, in creating statistical reports related to their field, and in the various activities of the instructional sequence involving risk scenarios, justifying their actions through the use of statistical estimators, then levels of comprehension will emerge.

The theoretical conjectures for this mapping are as follows:

- Cognitive development: If students are able to identify and interpret aspects of variability in response to questions designed based on the levels of statistical reasoning categorized by the SOLO Taxonomy as described in the Data Analysis, they will progress to higher levels of statistical reasoning.
- Quantification and Reasoning: If students' responses demonstrate quantification and the use of different measures of central tendency and dispersion, as well as different graphical representations in risk contexts, they will reach higher levels of comprehension.
- Competence and Reasoning: If students use verbal and statistical language to reason through problem-solving throughout the instructional sequence, they will develop the ability to make informed, data-driven decisions in risk contexts, taking variability into account as a determining factor.

An important point is that without quantification and reasoning, students will not develop the skills or the ability to construct arguments.

2. Materials and Methods

This study used a qualitative, exploratory approach through a hypothetical learning trajectory to develop statistical reasoning on measures of central tendency and dispersion. This study analyzed students' work, strategies, and arguments during the activities to identify changes in their reasoning processes.

2.1. Participants

Twenty-eight students, aged 19 to 22 (9 men and 13 women), enrolled in the Bachelor's program in Pharmaceutical Chemistry and Biology at a public university in Mexico, participated in a hypothetical learning trajectory.

2.2. Materials

2.2.1. Infographic

Infographic (Figure 1), a tool for presenting the Mexican Official Standard NOM-177-SSA1 in an engaging, accessible way for students, provides context on the dissolution testing needed to complete the first activity. It explains that 12 units of each drug, reference, and test are required at least 5 times, and notes that a drug can be labeled as an interchangeable generic if at least 85% dissolves within 15 min [25]. In addition, the infographic outlines the characteristics of generic, brand-name, and similar drugs.

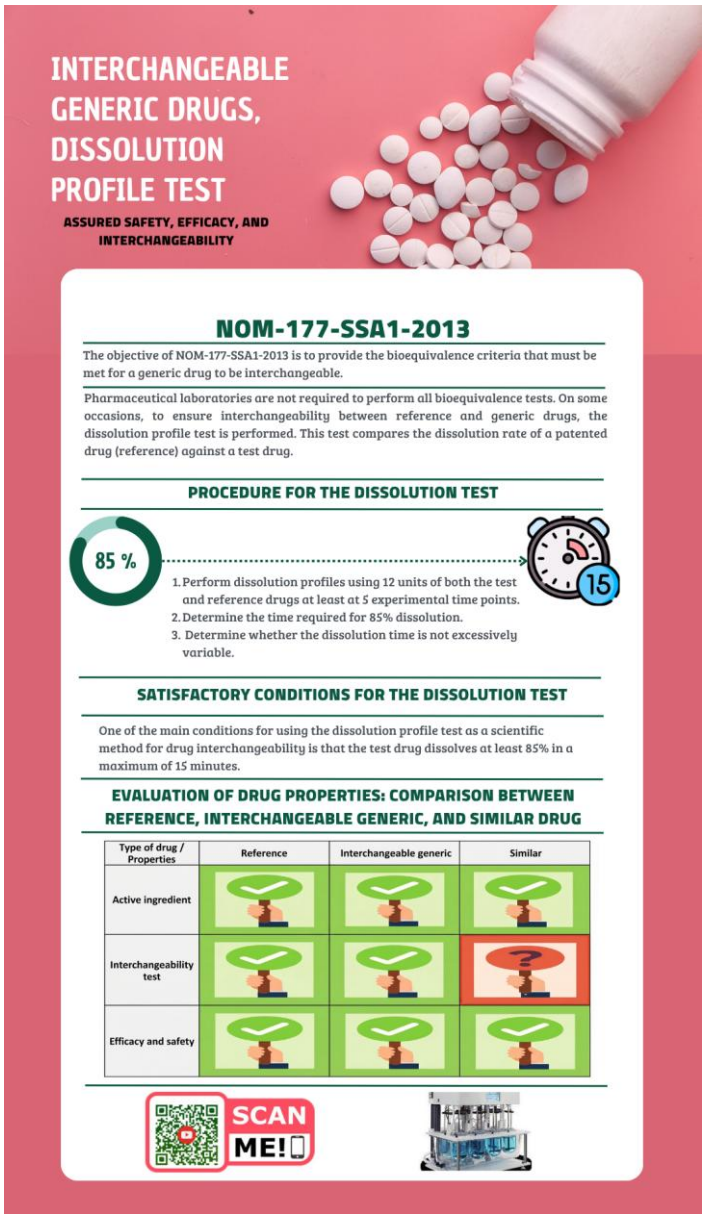


Figure 1. Infographic on NOM-177-SSA1-2013.
Source: Figure created by the authors.

2.2.2. Worksheet for Activity 1

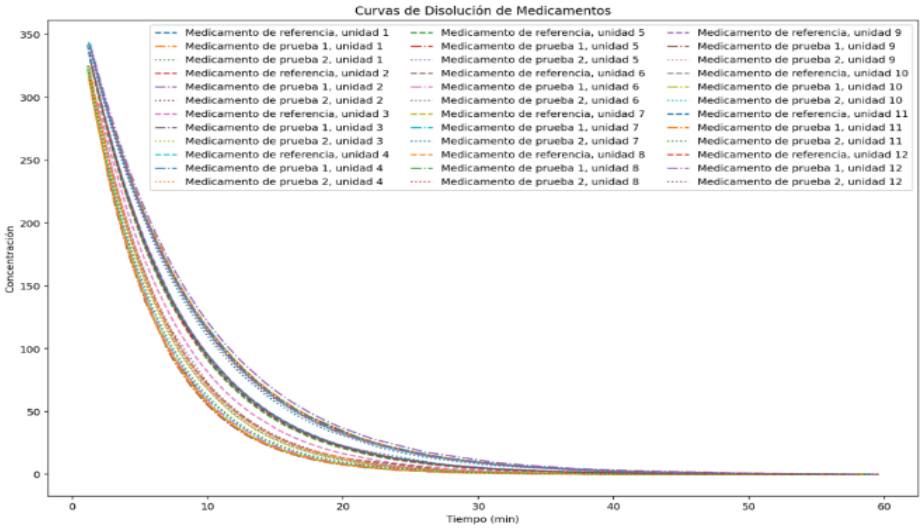
The worksheet (Appendix A.1) presents the context of drug dissolution profiles used, where a table was provided to show students the time it took for 12 units of three drugs (one reference drug and two test drugs) to reach at least 85% dissolution, and, based on NOM-177-SSA1 and the use of measures of central tendency and/or dispersion, to determine which of the two test drugs can be labeled as an interchangeable generic.

The hypothesis regarding the learning process for Activity 1 is that, as students complete the activity, they will demonstrate an overreliance on measures of central tendency, using statistical reasoning in an algorithmic and decontextualized manner. When using measures of dispersion, students will apply them qualitatively, interpreting them within context, that is, they will perceive variability but not quantify it. It was expected that students' responses would fall primarily within the Unistructural and Multistructural levels.

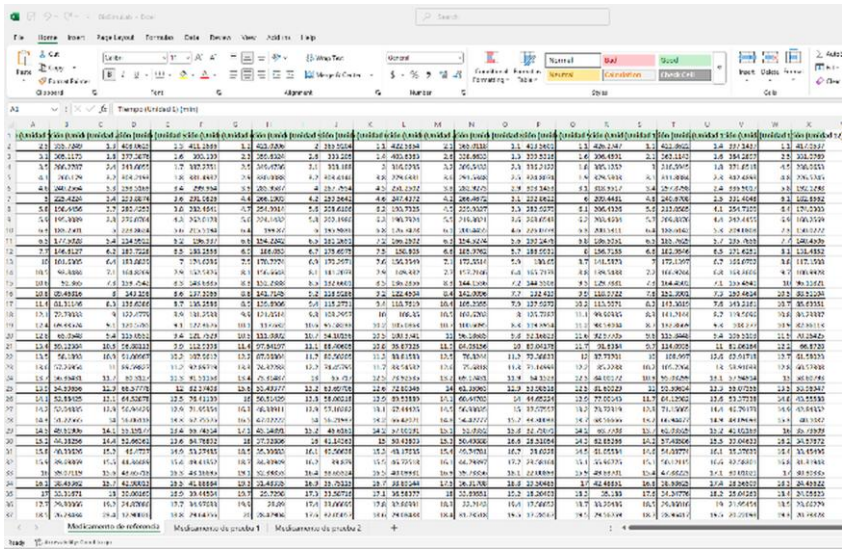
2.2.3. BioSimuLab

BioSimuLab is a simulator that was developed based on NOM-177-SSA1 and the SOLO Taxonomy, using an exponential model to simulate the dissolution of a reference drug and two test drugs. For more information on the interface simulator, see Hernández-Rivera et al. [14].

BioSimuLab creates an Excel file with three worksheets, one for each drug. Test drug 1 has noise variables introduced to create variability that prevents it from being labeled as an interchangeable generic because it does not meet the dissolution profile. In contrast, test drug 2 meets the dissolution profile. Figure 2 illustrates this most clearly.



(**a**)



(**b**)

Figure 2. Variability in the dissolution data for the reference drugs generated by BioSimuLab. (a) Plot of 12 units for the three types of drugs: reference, test 1, and test 2; (**b**) Example of the Excel file provided by BioSimuLab.

2.2.4. Worksheet for Exam

The exam consisted of two problems (see Appendix A.3) and was designed based on the learning process hypothesis that, as students solved the exercises, they would engage in a quantification process involving the use of various measures of central tendency and/or dispersion, as well as diverse graphical representations in risk contexts. Likewise, students were expected to use verbal language to support problem-solving in these contexts, considering variability as a determining factor in decision-making.

In addition, the responses were expected to fall within the Multistructural and Relational levels as follows: the first problem concerned four groups of physical therapy patients who underwent different treatments, and the number of days taken for them to recover was recorded. Wayne [26] posed this problem, and the data were presented so students can relate the estimators to the box-and-whisker plot and make decisions that account for variability and risk.

Problem 2, however, focused on evaluating the quality of four types of tequila, for which students had to analyze their electrical conductivity, as shown in a data table in millisiemens per centimeter (mS/cm). The relevance of analyzing electrical conductivity lay in its relationship with the amount of dissolved solids present in the product (such as minerals, salts, or impurities). This problem was reported by Carreon-Alvarez et al. [27].

2.2.5. Opinion Survey

At the end of the instructional sequence, an opinion survey was conducted to obtain students' perceptions of their experience with the instructional sequence, taking into account its contribution to the development of statistical reasoning regarding data variability; the usefulness of the BioSimuLab simulator; and collaborative work in understanding, interpreting, and making data-driven decisions. The survey consisted of an electronic questionnaire with 23 closed-ended items and two open-ended questions (Appendix B). The survey used a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The questionnaire covered five dimensions:

1. **Understanding and Description of Statistical Variability:** Recognition, comprehension, and use of statistics to describe and explain data variability.

2. **Analysis and Interpretation of Variability:** Analysis of graphs and relationships among estimators to interpret and justify variability, taking context into account.
3. **Contextualization and Decision-Making**: Consideration of context, variability, and risk to reflect on and provide a rationale for decisions.
4. **BioSimuLab Simulator Activity**: Evaluation of the simulator's usefulness for visualizing and interpreting variability and for connecting theory with practice in bioequivalence.
5. **Collaborative work and group discussion:** Assessment of the exchange of ideas, feedback, and collaboration in understanding variability.

2.3. Procedure

Data collection was carried out through the implementation of a Hypothetical Learning Path, structured in four phases (Figure 3) and designed to develop statistical reasoning:

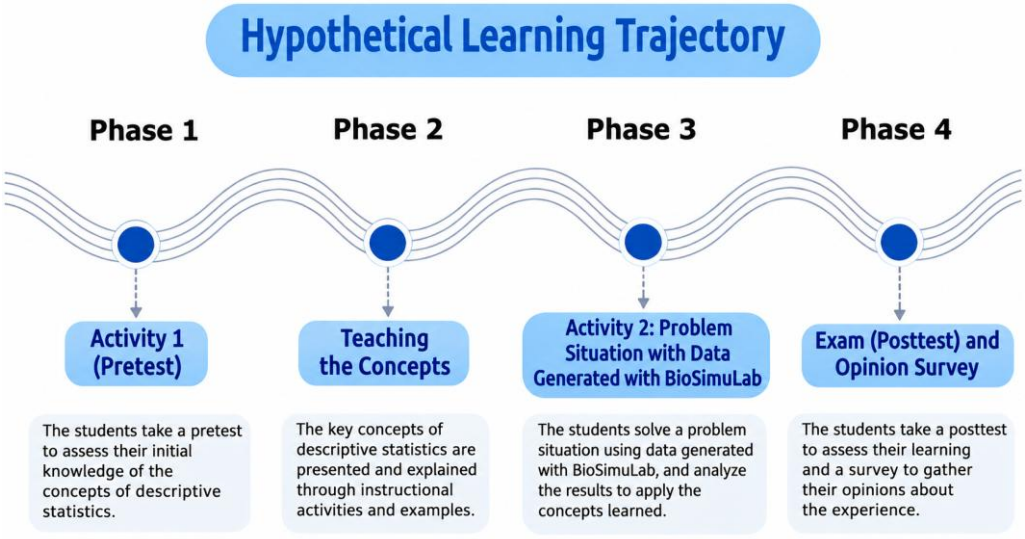


Figure 3. Sessions in the hypothetical learning trajectory.

Phase 1. Initially, the students' prior knowledge of dissolution testing and the classification of drugs into reference products, interchangeable generics, and similar products was assessed. Next, an explanatory video was shown on how the dissolution equipment works and the experimental procedure, along with an infographic on the methodology for evaluating generic and interchangeable drugs in accordance with NOM-177-SSA1. Next, Activity 1 (Pre-test) was presented, in which a problem situation was posed involving three medications: a reference product, considered the standard of comparison, and two test medications (1 and 2), which were generic alternatives whose equivalence was to be evaluated. Based on the provided data, students were required to determine which test drug met the criteria for an interchangeable generic, in accordance with NOM-177-SSA1, and to justify their choice using descriptive statistical tools. Students were given 30 min to complete the activity, followed by a group discussion in which they compared the procedures, results, and arguments used, with guidance from the teacher.

Phase 2. Teaching concepts of descriptive statistics. The teacher formally addressed measures of central tendency and dispersion through theoretical explanations, contextualized exercises, and Excel-based activities. This phase aimed to move students from the Unistructural level—identifying isolated concepts—to the Multistructural

level—using various statistical measures—and to promote conceptual integration as a step toward the Relational level.

Phase 3. BioSimuLab activity. The problem scenario from Activity 1 was revisited (Appendix A.2), focusing on comparing a reference drug with two test drugs to determine which of the former met the interchangeability criteria established in NOM-177-SSA1. In contrast to the first phase, in which the data were provided in a table, the students used BioSimuLab to generate the data through simulated experiments, sampling, and organizing the information into databases. They then applied descriptive statistics to compare the results, assess variability, and justify the selection of the test drug.

Phase 4. Final assessment with an exam (post-test) and an opinion survey. An exam based on problem-solving and data interpretation was administered to analyze participants' responses, identify the level of statistical reasoning achieved in accordance with the SOLO Taxonomy, and compare it with the level identified in Activity 1. Additionally, an opinion survey explored participants' perceptions of the teaching strategy implemented and the use of BioSimuLab as a learning support resource.

2.4. Data Analysis

The students' responses were analyzed using the SOLO Taxonomy with ATLAS.ti software(V.9). The codification was based on the statistical elements present in the responses to the problem situations in the pre-test and post-test, and the relationships established between them, in accordance with the criteria defined for each level of comprehension. This classification was based on the proposals by Brabrand and Dahl [28], Orta and Sánchez [29], Orta and Sánchez [4], Inzunza and Enciso [30], Vidal-Szabó and Estrella [31], Hernández-Rivera et al. [14], and Biggs and Collis [19].

- No Level. This category included student responses that did not address the questions posed, provided quantitatively without using estimators in their responses, or could not interpret the estimators and merely calculated them.
- Prestructural. The response addressed statistical ideas and concepts in a qualitative manner relevant to the problem, where students recognized variability but did not quantify it.
- Unistructural. The response involved selecting an estimator (mean, mode, median, maximums or minimums, or others) to compare the samples in the problem and could include measures of dispersion. The variables were described quantitatively.
- Multistructural. The response used two or more measures of central tendency and/or dispersion, without relating them to one another. The variables were described quantitatively.
- Relational. The response involved relating measures of central tendency and/or dispersion to support comparisons that take risk into account. The variables were described quantitatively.
- Extended Abstract. The response involved extrapolating statistical measures and their relationships to new contexts, transferring knowledge to unfamiliar situations or different fields of application. Inductive and deductive reasoning were used to generate hypotheses, formulate predictions, and theorize. The student was capable of reflecting, critiquing, and evaluating the results from multiple perspectives.

For the analysis of Activity 1 (pre-test) and the exam (post-test), two sets of complementary codes were established. The first corresponded to the levels of understanding defined by the SOLO Taxonomy and allowed us to characterize the structure of statistical reasoning expressed in the responses. The second group identified the estimators and statistical measures students used to solve the problems. This approach

examined reasoning levels, the statistical resources used, and the relationships between them.

The coding process was carried out in two stages. First, a researcher independently classified the responses, assigning the appropriate SOLO codes and identifying the estimators and statistics used. The research team then reviewed the codes jointly. The team analyzed and discussed potentially ambiguous responses and discrepancies in code assignment, considering the statistical elements present, their relationships, and the criteria defined for each SOLO level. The team reached final decisions by consensus to ensure consistent application of the classification criteria.

Once this process was complete, we conducted a code analysis for each document, cross-referencing SOLO Taxonomy levels with the pre-test and post-test documents. This comparison permitted an examination of changes in the distribution of comprehension levels between the two time points and an analysis of the possible evolution of statistical reasoning following the instructional intervention. This procedure was based on the premise that the SOLO Taxonomy allows the structural complexity of students' responses to be characterized. Thus, applying it at two assessment points provided a consistent framework for identifying changes in the organization and integration of statistical knowledge.

In addition, we coded students' responses to the problem scenario developed during the BioSimulator activity to identify the levels of statistical reasoning demonstrated throughout the activity. This analysis made it possible to assess an intermediate stage of the learning process and complement the pre-test and post-test comparison with evidence of the reasoning expressed during the intervention.

Furthermore, we analyzed the co-occurrence of SOLO levels and the estimators and statistical measures used by students. These relationships were visualized using a Sankey diagram, which enabled us to explore associations between the statistical resources employed and the structural complexity of the reasoning reflected in students' responses.

The opinion survey was analyzed using descriptive statistics based on responses from 28 students to a five-point Likert scale. The 23 items were organized into five categories, and individual item means and overall mean scores were calculated for each category. The results were visualized using a radar chart to compare students' ratings and identify strengths and potential areas for improvement in the instructional sequence. The analysis was purely descriptive, and no inferential statistical tests were performed.

3. Results

In this section, a retrospective analysis of the results from the classification of responses on Activity 1 (pre-test), the BioSimuLab activity, and the exam (post-test) are presented. This study addresses the following research question: How does statistical reasoning about variability develop among students in the Bachelor's program in Pharmaceutical Chemistry and Biology through the implementation of a hypothetical learning trajectory on descriptive statistics, informed by the research conjectures? To address this question, examples of students' responses are presented to illustrate and characterize their statistical reasoning. Subsequently, Sankey diagrams are presented to visualize the statistical reasoning levels identified in students' responses to the three worksheets corresponding to each phase of the hypothetical learning trajectory, with the aim of illustrating the distribution of reasoning levels and identifying changes in students' reasoning throughout the instructional sequence. Finally, the results of a student opinion survey are presented to explore students' perceptions of their learning of statistics, teamwork, and the contribution of the simulator to their learning.

3.1. Analysis of Levels

3.1.1. No Level

The No Level category includes responses from students who do not answer the questions posed, provide quantitative answers but do not use estimators, or are unable to interpret the estimators and merely calculate them. The examples shown in Figure 4 were taken from the responses to the first activity, which presented the following problem scenario: the students were part of a pharmaceutical team tasked with evaluating two different brands of the same medication, which were test samples.

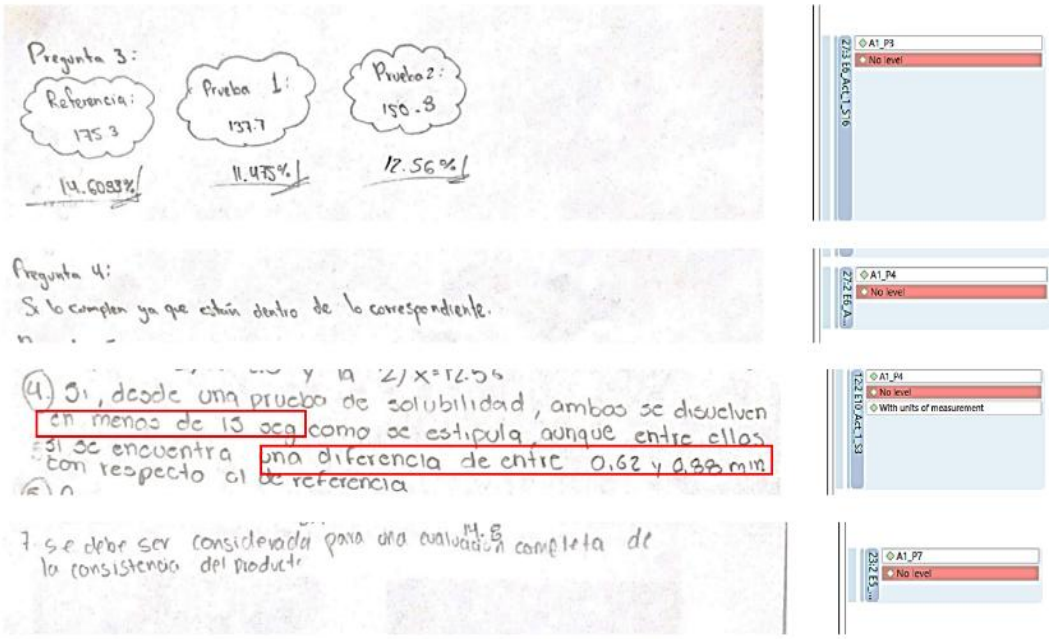


Figure 4. Examples of responses coded at No Level. Note: The first image shows that subject S_{16} states, "Reference 175.3 14.6033%, Test 1 137.7 11.475%, Test 2 150.8 12.56%". The second image shows the response from subject S_3, who states: "Yes, based on a solubility test, both dissolve in less than 15 s as stipulated, although there is a difference between them of 0.62 and 0.88 min compared to the reference drug". The third image shows the response from subject S_{16}, "Reference 175.3 14.6033%, Test 1 137.7 11.475%, Test 2 150.8 12.56%". The fourth image shows the response from subject S_{13}, who states, "This should be considered for a complete evaluation of the product's consistency".

In this activity, students had to create dissolution profiles for each medication. As a first example of the "No Level" category, we considered Question 3 of Activity 1, which asked, "How long does it take for each drug (reference and test drugs) to reach 85% dissolution?" Here, subject S_{16} stated, "Reference 175.3 14.6033%, Test 1 137.7 11.475%, Test 2 150.8 12.56%" (S_{16}, see Figure 4), which fell into the No Level category, since the subject answers the question quantitatively but does not use estimates in the response.

For the following example from the No Level category, we considered Question 4 of Activity 1, which asked, "Do the test drugs comply with NOM-177-SSA1 to be considered interchangeable generics?" Here, subject S_3 stated, "Yes, based on a solubility test, both dissolve in less than 15 s as stipulated, although there is a difference between them of 0.62 and 0.88 min compared to the reference drug" (S_3, see Figure 4), which fell into the No Level category, since the subject answered the question quantitatively but did not use estimators in the response.

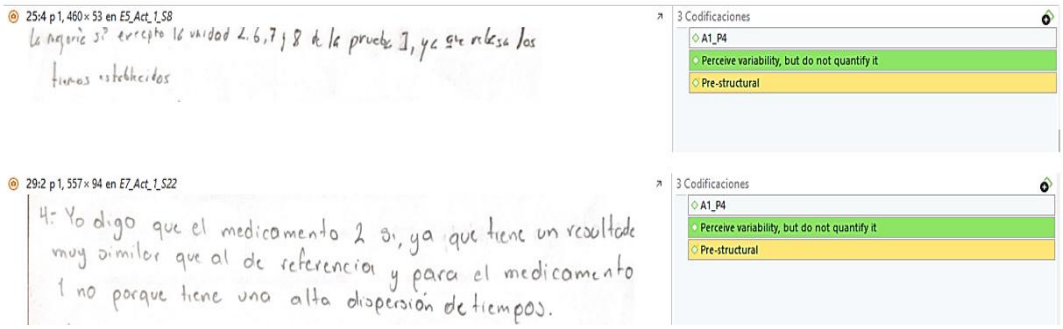


Figure 5. Examples of responses coded at the Prestructural level in Activity 1. Note: The first image shows S_8's answer to Question 4 of the first activity. S_8 states: "Most of them can, except for items 2, 6, 7, and 8 in Test 1, since they exceed the established time limits", while the second image shows S_{22}'s answer to the same question; S_{22} states: "I say that Drug 2 does, since it has a result very similar to that of the reference, and Drug 1 does not because it has a high dispersion of times".

For the following example from the No Level category, we consider Question 4 of Activity 1, in which subject S_{16} stated, "Yes, they do comply, since they fall within the applicable range" (S_{16}; see Figure 4), which was classified as No Level, since the subject did not answer the question using statistical ideas or concepts.

As a final example of the No Level category, we considered Question 7 of Activity 1, which stated: "Given the dispersion of the data, how long does it take for each medication (reference and test medications) to reach 85% dissolution?" Here, subject S_{13} stated, "This should be considered for a complete evaluation of the product's consistency" (S_{13}, see Figure 5), which fell under the No Level category because the subject did not answer the question using statistical ideas or concepts and evaded it.

3.1.2. Prestructural

At the Prestructural level, the students' responses did not use measures of central tendency or dispersion quantitatively. However, they drew conclusions or recognized variability, but they did not quantify it, and variables were described qualitatively.

The examples for the Prestructural level shown in Figure 6 were taken from the responses to the first activity, in which the problem situation was described for the examples at the No Level. For these examples, we considered Question 4, which asked, "Do the medications comply with NOM-177-SSA1 to be considered interchangeable generics?"

In Activity 1, Question 4, when deciding which of the two test medications could be labeled as an interchangeable generic, S_8 stated, "Most of them can, except for items 2, 6, 7, and 8 in Test 1, since they exceed the established time limits" (S_8, see Figure 6), which was qualitative, since it was not based on statistical estimators that would justify compliance with or noncompliance with NOM-177-SSA1; the subject merely mentioned the units that exceeded the 15 min time limit specified by NOM-177-SSA1. Thus, the subject acknowledged the variability but did not quantify it, so that the response fell at the Prestructural level.

In Activity 1, Question 4, S_{22} stated, "I say that Drug 2 does, since it has a result very similar to that of the reference, and Drug 1 did not because it has a high dispersion of times" (S_{22}, see Figure 6). The respondent acknowledged variation but does not quantify or explain how it affects compliance with NOM-177-SSA1; therefore, the response was classified at the Prestructural level.

3.1.3. Unistructural

The Unistructural level was characterized by responses that used a single estimator, such as a measure of central tendency or dispersion. It also used graphical representations, such as box-and-whisker plots or scatter plots, but without integrating them into an interpretation or relating them to risk. Therefore, the variables are described quantitatively. Some examples of responses coded at the Unistructural level in the individual activities are shown in Figure 6.

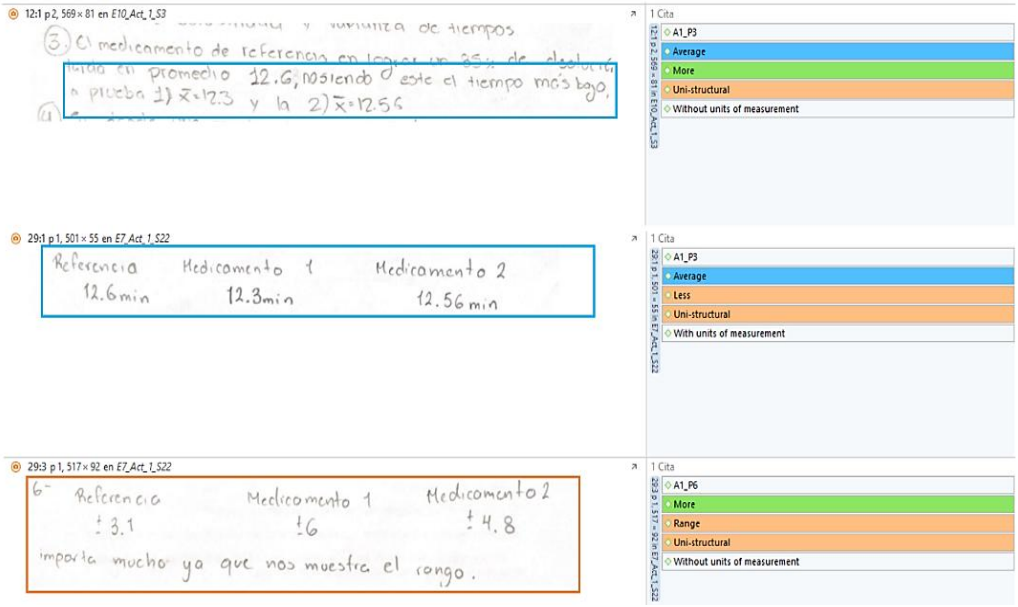


Figure 6. Examples of responses coded at the Unistructural level in Activity 1. Note: The first image shows that subject S_3 states, "The reference drug takes an average of 12.6 to achieve 85% dissolution; this is not the shortest time, as Test 1) $\bar{x}$ = 12.3 and Test 2) $\bar{x}$ = 12.56". The second image shows the response from subject S_{22}: "Reference 12.6 min, Drug 1 12.3 min, Drug 2 12.56 min". The third image shows the response from subject S_{22}: "Reference ± 3.1, Drug 1 ± 6, Drug 2 ± 4.8; this is very important because it shows us the range".

The examples for the Unistructural level were taken from the responses to the first activity, which asked participants to determine which drug met the dissolution profile using measures of central tendency and/or dispersion, and to apply the infographic to data on the dissolution of two test drugs and one reference drug. For example, Question 3 asked "How long does it take for each drug (reference and test drugs) to reach 85% dissolution?", and Question 6 asked "What are the maximum and minimum times observed in the tests, and how do these affect dissolution?"

S_3 in Activity 1, Question 3, stated: "The reference drug takes an average of 12.6 to achieve 85% dissolution; this is not the shortest time, as Test 1) $\bar{x}$ = 12.3 and Test 2) $\bar{x}$ = 12.56" (S_3, see Figure 6). The answer relied exclusively on the mean for each drug, without incorporating measures of dispersion, establishing relationships, or considering variability; therefore, the answer was at the Unistructural level.

Meanwhile, S_{22} in Activity 1, Question 3, stated: "Reference 12.6 min, Drug 1 12.3 min, Drug 2 12.56 min" (S_{22}, see Figure 6). The mean was used as the sole estimator, without accounting for variability or justifying reliability; therefore, the response was at a Unistructural level.

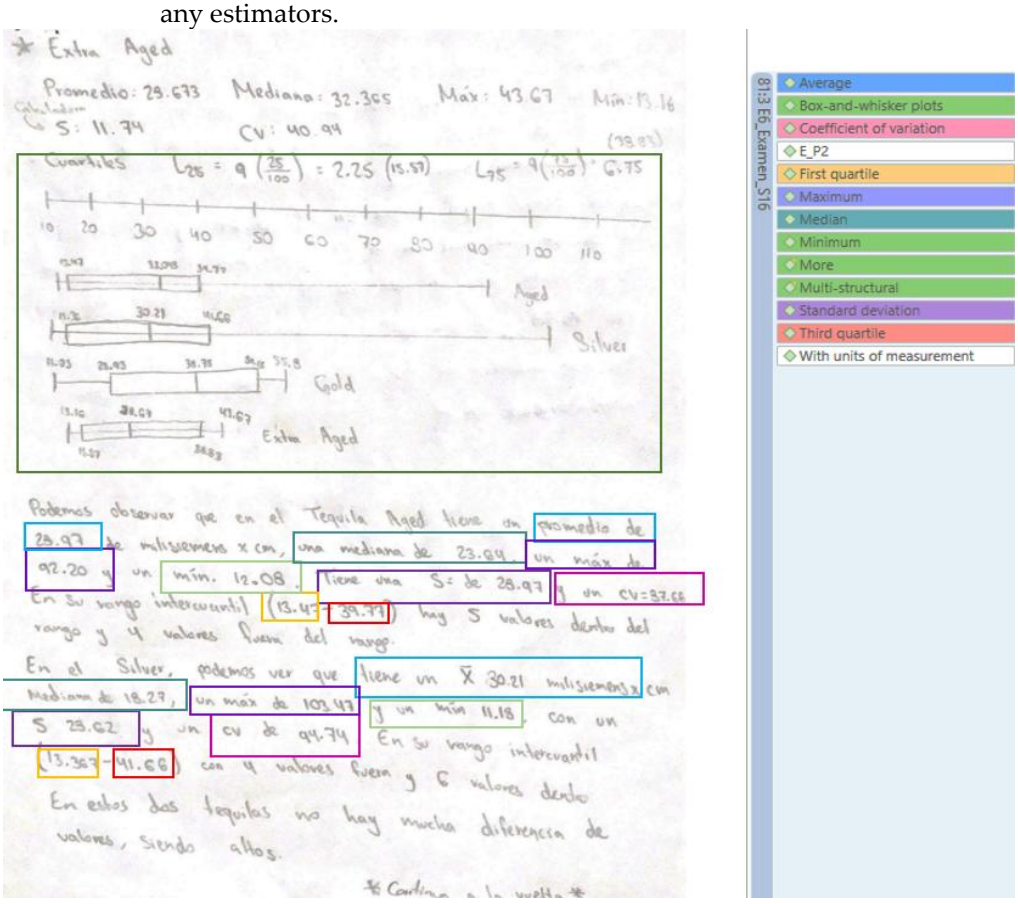
Furthermore, S_{22} in Activity 1, Question 6, stated, "reference ± 3.1, Drug 1 ± 6, Drug 2 ± 4.8, this is very important because it shows us the range" (S_{22}, see Figure 6). Although it quantifies dispersion using the range, it did not explain how this dispersion affected compliance with NOM-177-SSA1 and relied only on this measure of dispersion; therefore, the response was at the Unistructural level.

3.1.4. Multistructural

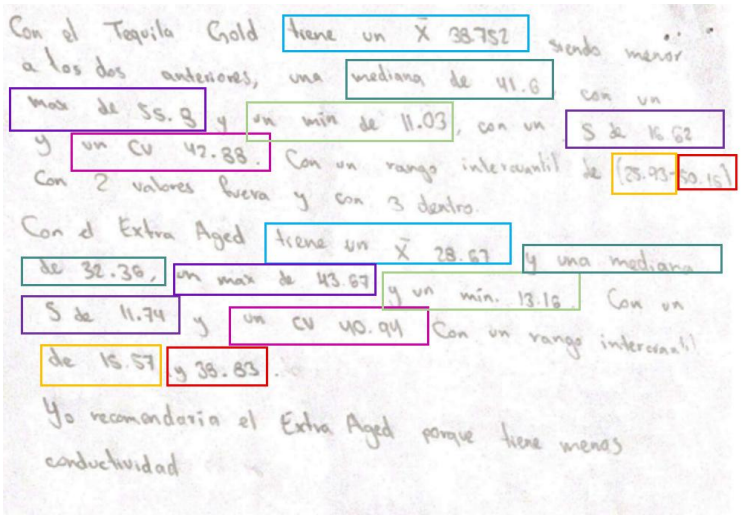
The Multistructural level was characterized by responses that included multiple statistical measures, whether measures of central tendency and/or measures of dispersion used independently, as well as graphical representations such as box-and-whisker plots or scatter plots, but without establishing a relationship between the estimators and the graphical representations. The variables were described quantitatively, though.

The example for the Multistructural level shown in Figure 8 was taken from the answers to the second exam question, which presented the following scenario: the students were part of a team from the Tequila Regulatory Council (TRC) tasked with analyzing four types of tequila, Aged, Silver, Gold, and Extra Aged. To do this, they had to measure the electrical conductivity in millisiemens per centimeter (mS/cm) whose physicochemical property in tequila related to the amount of dissolved solids; low conductivity, however, suggested good quality control, while high conductivity indicated more dissolved solids, which could compromise tequila quality. Students were given a table showing data on the electrical conductivity of the four types of tequila, and were asked to determine which tequila was of the highest quality.

Respondent S_{16} answered at the Multistructural level for Question 2 on the exam (see Figure 7) and described the variables quantitatively. The student used multiple measures: mean, median, maximum, minimum, standard deviation, coefficient of variation, and interquartile range; then, they listed the data for each group as isolated blocks of information. The conclusion is "I would recommend the Extra Aged because it has lower conductivity" (S_{16}); however, in reaching this conclusion, this student did not relate it to any estimators.



(**a**)



(**b**)

Figure 7. Example of a response at a Multistructural level. Note: The first and second images show the response of subject S3, "We can observe that the Aged Tequila has an average of 23.97 millisiemens × cm, a median of 23.84, a maximum of 92.20, and a minimum of 12.08. It has an *S* of 28.97 and a $cv = 37.68$. In its interquartile range (13.43–39.77), 5 values fall within the range and 4 fall outside them. In the Silver, we see that it has a $\underline{x}$ of 30.21 millisiemens × cm, a median of 18.27, a maximum of 103.47, and a minimum of 11.18, with an *S* of 23.62 and a *cv* of 94.74. In its interquartile range (13.367–41.66), there are 4 values outside and 6 values inside. In these two tequilas, there is not much difference in values, both being high. The Gold Tequila has a $\underline{x}$ of 38.752, being lower than the previous two, a median of 41.6, with a maximum of 55.9 and a minimum of 11.03, with an *S* of 16.62 and a *cv* of 42.88. It has an interquartile range of (28.93–50.15) with 2 values outside and 3 inside. The Extra Aged has a $\underline{x}$ of 28.67 and a median of 32.36, a maximum of 43.67, and a minimum of 13.16. With an *S* of 11.74 and a *cv* of 40.94. With an interquartile range of 16.57 and 38.83. I would recommend the Extra Aged because it has lower conductivity".

3.1.5. Relational

The Relational level was achieved when students' responses meaningfully integrated measures of central tendency and dispersion, interpreted variability as a fundamental component of risk, and used this information to justify context-specific decisions. S18 provides an answer to question one at the Relational level (see Figure 8), since the participant relates "…less than the average for those who received no treatment. The box-and-whisker plot also shows that, although the third box in the massage data has a wider range of values, it is noticeably shifted downward, indicating a shorter recovery time, and the deviation is not too high" (S18: Ex2_7-10). In addition, they relate the maximum to the mean, as expressed in "…Although it is also true that massage results showed greater variation, with a standard deviation of 9.39, even the person with the longest recovery time, 82 days, took about the same amount of time as the average for those who did water exercises…" (S18: Ex2_4-7).

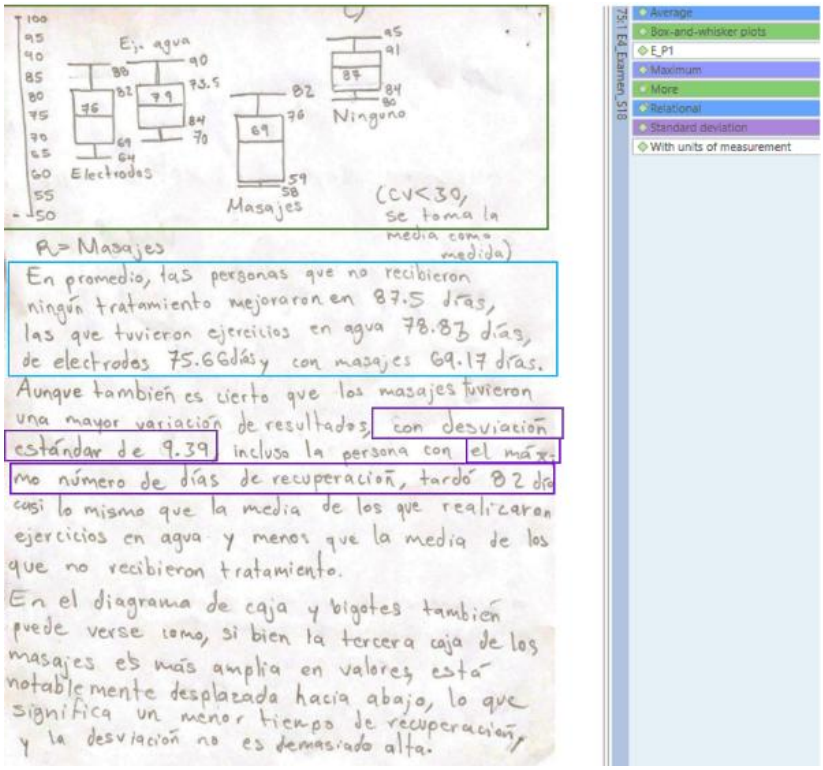


Figure 8. Example of a response at the Relational level from subject S_{18}. Note: The image shows the response of subject S_{18}, "On average, people who received no treatment improved in 87.5 days, those who had water exercises in 78.83 days, electrode therapy in 75.66 days, and massage therapy in 69.17 days. Although it is also true that massages had a greater variation in results, with a standard deviation of 9.39, even the person with the maximum number of recovery days took 82 days, almost the same as the mean of those who performed water exercises and less than the mean of those who received no treatment. In the box and whisker plot, it can also be seen that, although the third box for massages is wider in values, it is notably shifted downward, which means a shorter recovery time, and the deviation is not too high".

In Question 2 of the exam, S_3 provides an answer at the Relational level (see (S_3) in Figure 9). This student uses several measures of central tendency: mean, range, interquartile range, CV%, minimum, and maximum. The response demonstrates, however, an understanding of variability and its role in differentiating the quality of the various types of tequila. Also, the respondent recognized that a high CV% indicates greater variability and a less desirable option. Furthermore, the student justified why one option was preferable despite its rank and CV%, noting that lower possible values ranging from 13.16 to 43.67 ms/cm, compared to the vast range of data and possible values for the Silver category, which range from 103.47 to 11.18, with a CV% close to 95%, an absurd and undoubtedly manipulated value.

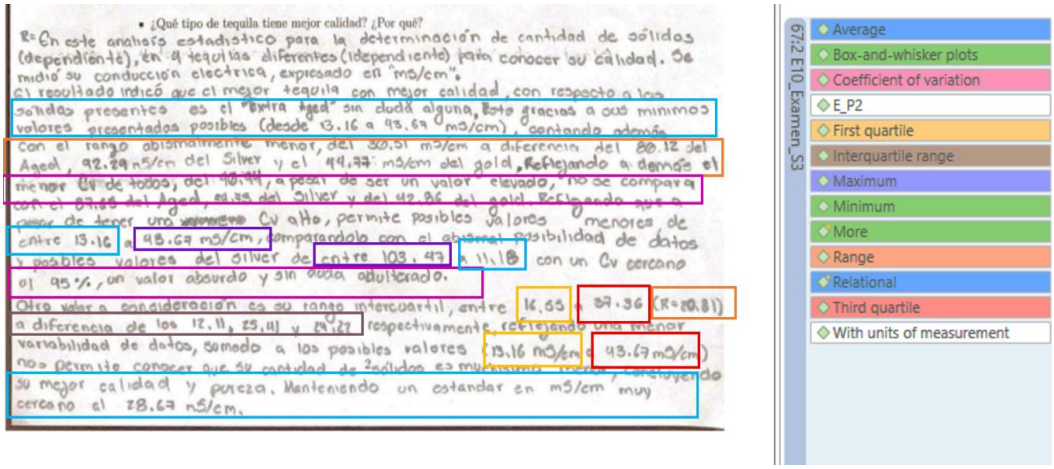


Figure 9. Example of a response at the Relational level from subject S_3. Note: The image shows the response of subject S_3, "R. In this statistical analysis to determine the amount of solids (dependent variable) across 4 different tequilas (independent variable) to determine their quality, electrical conductivity was measured and expressed in mS/cm. The result indicated that the best tequila with the highest quality regarding the solids present is the 'Extra Aged' without a doubt. This is thanks to its minimum possible values (from 13.16 to 43.67 mS/cm), and its much smaller range of 30.51 mS/cm compared to 80.12 for the Aged, 90.29 mS/cm for the Silver, and 44.44 mS/cm for the Gold. It also reflects the lowest CV of all at 40.94; despite being a high value, it does not compare to the 37.68 for the Aged, 94.74 for the Silver, and 42.88 for the Gold, reflecting that despite having an apparent high CV, it allows for possible lower values between 13.16 and 43.67 mS/cm, when compared to the abysmal data spread and possible values of the Silver between 103.47 and 11.18 with a CV close to 95%, an absurd and undoubtedly adulterated value. Another value to consider is the interquartile range, between 16.57 and 37.36 ($R = 20.81$) compared to 12.11, 25.41, and 29.27, respectively, reflecting lower data variability. Combined with the possible values of 13.16 mS/cm and 43.67 mS/cm, this allows us to know that its amount of solids is much lower, concluding its superior quality and purity, while maintaining a standard in mS/cm very close to 28.67 mS/cm".

Here, the student uses statistics to formulate a hypothesis about the nature of the sample, linking the range and CV% values to possible adulteration of the Silver tequila. This places subject S_3 at the Relational level.

3.2. Analysis of the Effect of a Hypothetical Learning Trajectory on Levels of Comprehension

Comprehension levels showed a positive trend after implementing the instructional sequence, as shown in Figure 10. In Activity 1, students' responses were distributed primarily at the Unistructural level (32%) and the Multistructural level (35%), while 18% fell at the Prestructural level and 15% were classified as "No Level". On the one hand, it should be noted that no responses were coded at the Relational level in this activity, which, unlike in previous exams, showed no student responses at the "No Level" or "Prestructural levels", and only 2% of responses remained at the "Unistructural level". On the other hand, the remaining 27% of responses were classified at the Multistructural level, and 71% reached the Relational level, which reflected development in statistical reasoning and refinement in their interpretations. Figures 10 and 11, in turn, corroborate this by showing the estimators used at each student's comprehension level during the individual activities.

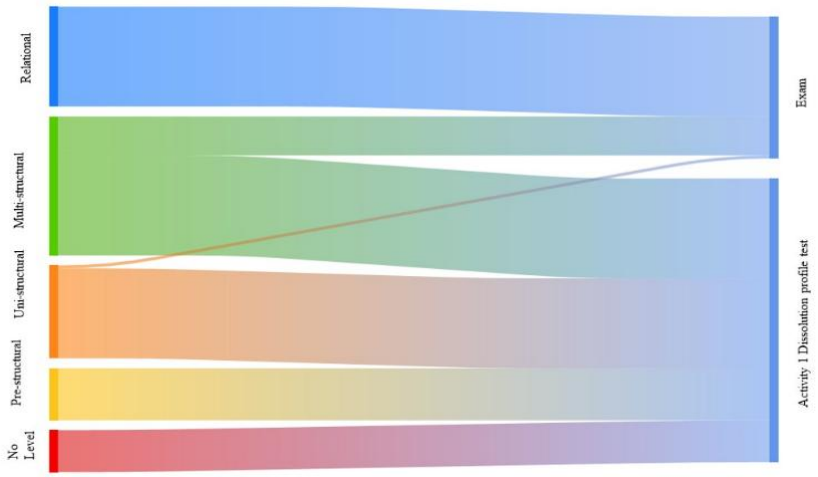


Figure 10. The Sankey diagram illustrates the distribution of levels of statistical reasoning among students during Activity 1 and the exam.

The comparative analysis between Activity 1 (pre-test) and the exam (post-test) reveals significant development in students' statistical reasoning, as reflected in SOLO Taxonomy levels. Statistical reasoning evolved from algorithmic to contextualized in the following way: In the pre-test (Activity 1), most reasoning was algorithmic and decontextualized. A marked overvaluation of measures of central tendency was identified (using only the mean in 15 responses), alongside a qualitative use of dispersion in which students perceived variability but failed to quantify it, placing them mainly at the Unistructural and Multistructural levels. In contrast, in the post-test (exam), 100% of the responses considered the context and argued based on it, and likewise, most students integrated measures of central tendency and dispersion simultaneously and quantitatively (18 out of 22 responses in both exam problems).

Variability became a determining factor, while dispersion was initially used in an isolated or superficial way; the exam results show that the intervention favored students who considered variability a key factor in making informed decisions in risk contexts.

Finally, a consolidation of the higher levels (Multistructural and Relational) can be seen, given that during the exam, a large group of students (14 in Problem 1 and 12 in Problem 2) integrated measures of central tendency and dispersion quantitatively and relationally, correctly interpreting graphical representations (box plots) and linking the study variables. This confirms that reaching a Relational level is directly associated with the ability to incorporate data dispersion into argumentation, unlike lower levels focused exclusively on the static structure of the data (Multistructural).

The findings above directly answer the research question about how statistical reasoning about variability develops following implementation of the hypothetical learning pathway. This evidence also supports the learning process hypotheses established for both Activity 1 and the exam.

Figure 11 compares the initial activity and the exam, suggesting development in students' statistical reasoning, reflected in both the number and type of estimators connected within their responses. In the initial activity, though, responses classified at the Unistructural level relied mainly on a single estimator, with the mean being the most frequently used. At the Multistructural level, students incorporated several estimators, particularly measures of central tendency and extreme values, such as the mean, maximum, and minimum. However, using several estimators did not necessarily indicate a higher level of comprehension, since they could still be applied independently without establishing meaningful relationships among them.

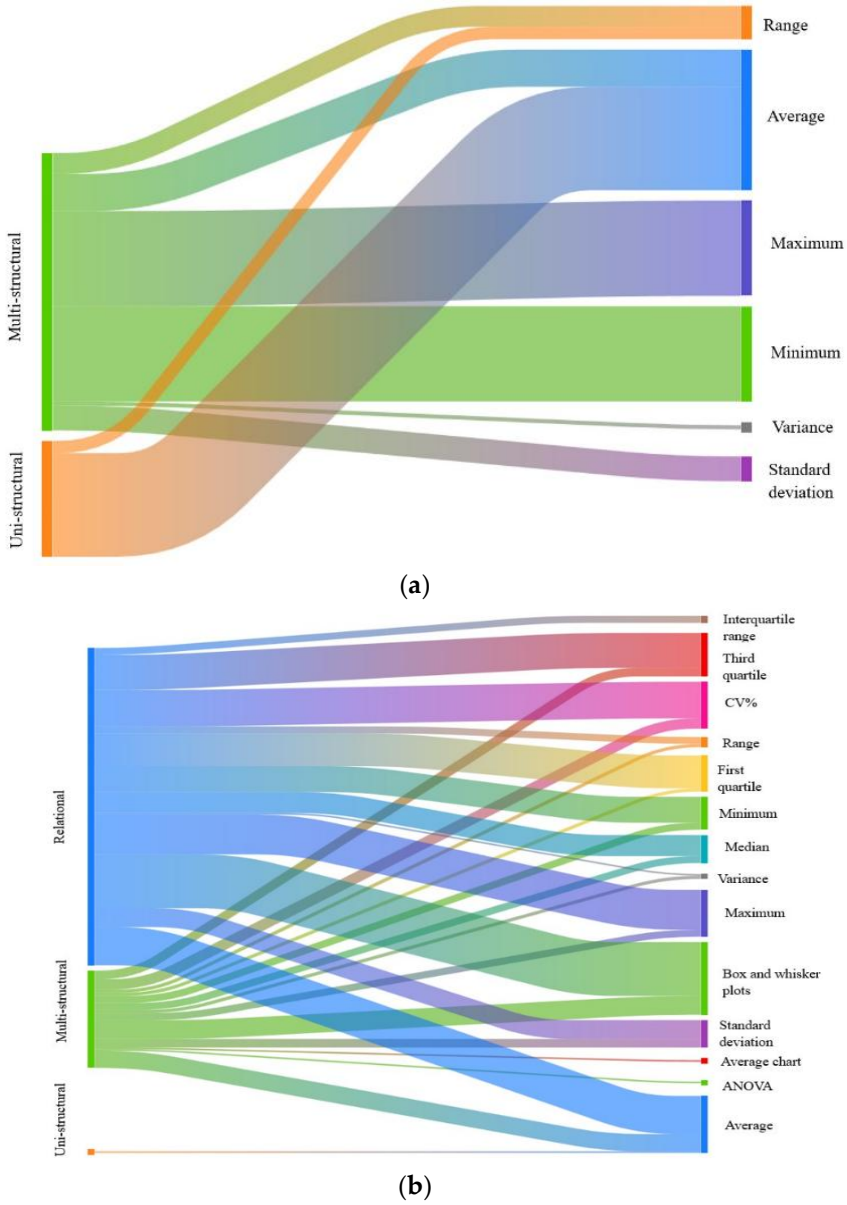


Figure 11. Estimators by levels of statistical reasoning for (**a**) Activity 1 and (**b**) the exam.

In the exam, a broader range of estimators was observed, including measures of position and variability such as quartiles, the median, standard deviation, coefficient of variation, and interquartile range, together with graphical representations such as the box-and-whisker plot. Also, it coincided with the emergence of responses classified at the Relational level, which had not been identified in the initial activity, which from the perspective of the SOLO Taxonomy, suggests that the transition from the Multistructural to Relational levels is not determined simply by the use of a larger number of estimators, but by the selection of estimators with different statistical purposes and the establishment of relationships among them to interpret a distribution, compare data, or support a conclusion. Thus, the assessment conducted after the implementation of the hypothetical learning trajectory revealed more structurally complex responses, in which the combined use of measures of central tendency, position, variability, and graphical representations became an important criterion for classifying higher levels of understanding; furthermore, the qualitative analysis revealed responses in which students incorporated measures of dispersion, related different estimators, interpreted graphical representations, and used elements of context and variability to support their decisions. Taken together, these results

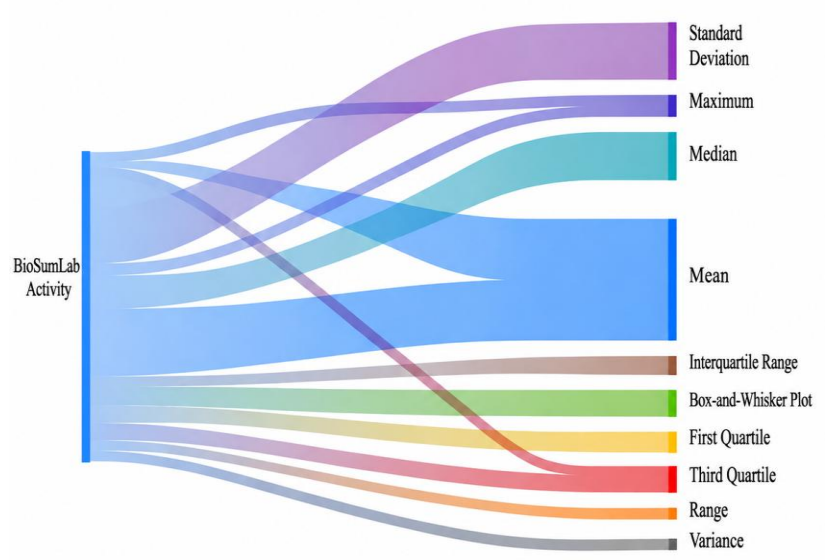
are consistent with the development of more integrated forms of statistical reasoning about variability during the instructional sequence.

3.3. BioSimuLab Activity Analysis

An analysis of the Sankey diagrams, based on the SOLO Taxonomy, reveals a predominance of the Relational level, followed by the Multistructural, Unistructural, and Prestructural levels, with a smaller proportion of unclassified responses (Figure 12). The predominance of the Relational level suggests that students were able to integrate different statistical concepts; establish meaningful connections among them; and build a deeper, more structured understanding of the subject matter. The second diagram, meanwhile, shows a higher concentration of connections related to the mean, standard deviation, and median, while the other statistical measures are less represented. Taken together, these results suggest that the BioSimuLab activity offers opportunities to develop skills in analysis, interpretation, and conceptual integration, thereby fostering more complex cognitive processes.



(**a**)



(**b**)

Figure 12. Analysis of the statistical reasoning present in the responses to the BioSimuLab activity based on the SOLO Taxonomy. (**a**) Sankey diagram showing the distribution of comprehension levels. (**b**) Sankey diagram of the distribution of statistical concepts.

3.4. Opinion Survey Assessment

The radar chart shows that students had a favorable perception of the instructional sequence, with averages ranging from 3.69 to 4.03 (Figure 13). The category "Contextualization of Variability and Decision-Making" received the highest average (4.03), reflecting a positive assessment of the ability to relate variability to risk and decision-making. This was followed by "Collaborative Work and Group Discussion" (3.96), highlighting the usefulness of exchanging ideas. The categories "Understanding and Description of Statistical Variability" and "Analysis and Interpretation of Statistical Variability" received identical average scores (3.94), suggesting similar evaluations of the understanding and application of statistical concepts. Meanwhile, the "BioSimuLab Simulator Activity" had the lowest average score (3.69), although it still received a favorable evaluation.

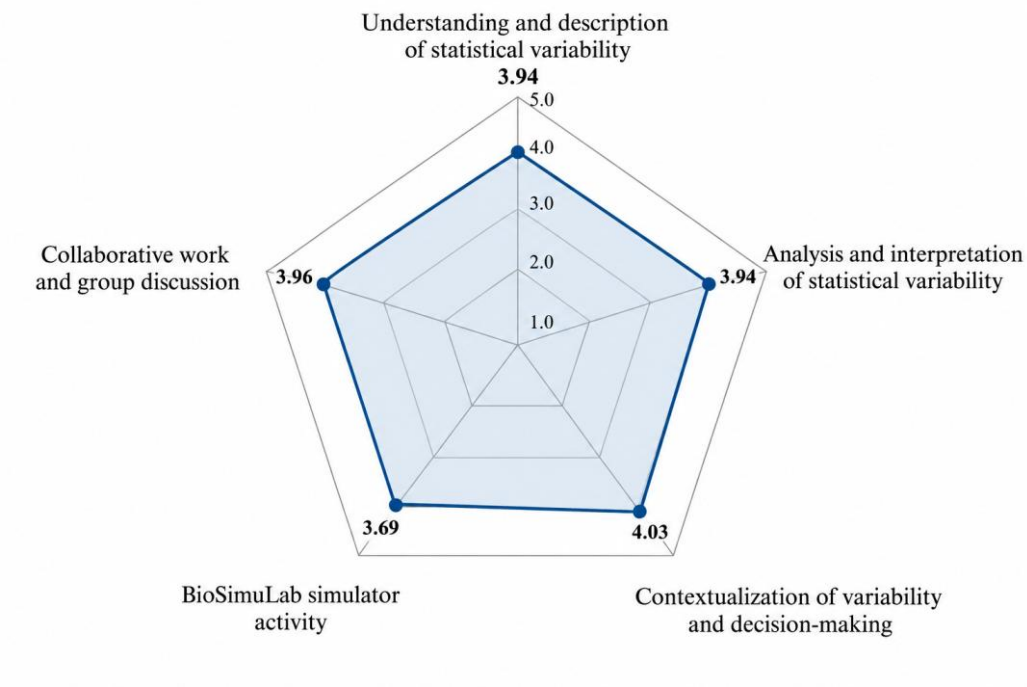


Figure 13. Radar chart showing the evaluation of the dimensions of the opinion survey.

The radar chart shows relatively consistent performance across the five categories, with a maximum difference of 0.34 points between their averages. There is a slight upward trend in terms of contextualization and decision-making, while the simulator activity received a lower rating than the other categories. These results suggest that students particularly recognized the value of applying statistical knowledge to real-world situations and sharing their interpretations with peers. Furthermore, the lower score on the simulator indicates a potential area for improvement in data visualization and interpretation activities. Overall, the results reflect a positive perception of the instructional sequence and provide information to strengthen future implementations. However, the differences observed are descriptive and do not allow identification of statistically significant differences or direct verification of improvements in statistical reasoning.

4. Discussion

The results of implementing the instructional sequence show that teaching descriptive statistics, focused on understanding variability and the connection between estimators and their numerical and graphical representations in various contexts, rather

than on memorizing formulas and concepts, promoted the development of statistical reasoning among students, as proposed by Garfield and Ben-Zvi [16].

For this reason, the process of quantification and the use of different estimators, along with the use of verbal and statistical language to justify problem-solving by students, fostered the development of their statistical reasoning, as outlined in Orta and Sánchez [29], Orta and Sánchez [4], Inzunza and Enciso [30], and Vidal-Szabó and Estrella [31]. Furthermore, these are the two phases (quantitative and qualitative) identified by Biggs and Collis [19], into which the levels of understanding in the SOLO Taxonomy are grouped.

This phenomenon confirms the findings of Shaughnessy [3], who identifies variability as the central axis for understanding and explaining data and chance. On the other hand, statistical reasoning evaluates how variability influences results during problem-solving, rather than merely applying standard procedures.

The design of the activities, which presented data with similar structures but different dispersion, as well as problems with different structures and dispersion, reduced the exclusive use of measures of central tendency, as reported in Orta and Sánchez [4]. This strategy led students to seek answers in variability, facilitating transition to the Relational level, which converges to what Orta and Delgado [6] point out in studies of medical contexts, which states that the ability to articulate data and integrate variability defines higher levels of statistical reasoning, as shown by this study's results. Moreover, following the suggestion by Orta and Sánchez [4], problems with different means were included to prepare the students prior to the analysis of variation.

This design research demonstrated that as students' progress through the phases, moving from identifying the presence of variability, measuring and modeling it by means of measures of central tendency and dispersion, representing that variation with box plots and scatter plots, and finally explaining and managing variability for decision-making, their statistical reasoning develops, as proposed by Reading and Shaughnessy [15].

Similarly to the research by [31], this study shows that graphical representations support students' learning when interpreting variability. In addition to sharing the conclusion that participants' primitive notions of variability are incomplete, this work adds that they are also decontextualized and algorithmic.

Analysis using the SOLO Taxonomy identified that responses tend to reach the Multistructural level when participants use only the data structure without integrating dispersion. In this research, there were no cases where the Relational level was reached considering the data structure solely. This phenomenon coincides with the views of Shaughnessy [3], who places variability as the central axis for understanding and explaining data and chance. It also agrees with Garfield and Ben-Zvi [16], who point out that: "it is impossible to consider variability without also considering the center, since both ideas are necessary to find meaning in data analysis." Similarly, Kaplan et al. [9] mention that for students to interpret variability, they must know the center, given that formal measures of variation depend on the distance of the points to it, as well as identify it when solving the problem and see the necessity of its use. Furthermore, this can be justified by Garfield and Ben-Zvi [5], who state that students must employ different measures of dispersion such as range, standard deviation, and interquartile range. Still, teachers must clarify to them that these can be interpreted more comprehensively if the measures of central tendency of the data are known.

The limitations of this study indicate that the designed activities did not reach the Extended Abstract level, in which students could apply interdisciplinary knowledge to interpret the results, which was due to factors such as the academic level (third semester, whereas dissolution profiles are covered in the sixth semester in the course on Biopharmaceutics and Pharmacokinetics), time constraints, and the design of the

instructional sequence itself. Therefore, the design conjecture related to the production of artifacts is partially fulfilled, since students use verbal and statistical language in preparing their reports; however, they do not fully integrate interdisciplinary knowledge.

Similarly, the design conjecture regarding discursive practices is also partially fulfilled. In addition, the students' failure to reach the Extended Abstract level can be explained, in part, by the duration of the instructional sequence, since, as Orta and Sánchez [4] point out, identifying responses corresponding to this level requires longer instructional sequences.

A further limitation concerns the fifth cross-cutting characteristic of design research proposed by Cobb et al. [23], which states that theories developed in design research must be applied to different contexts; theory must perform real work, where these are related not only to the learning processes of a domain but also to the design activity. Additionally, the evidence generated in this study is insufficient evidence to determine whether the proposed instructional design, based on the use of simulators, is more effective than other teaching methods. Although the proposal is grounded in research indicating that simulators promote the development of statistical thinking by engaging students in the collection, organization, and analysis of data, the results obtained do not allow us to attribute the observed changes in students' reasoning exclusively to the use of simulators. Therefore, future research should include comparison groups and gather students' perceptions and reflections on the teaching strategies used.

Finally, a methodological limitation is the lack of comparability between Activity 1 (pre-test) and the exam (post-test), due to differences in their respective difficulty levels and the conditions under which they were administered. This prevented us from making direct comparisons between the two tests and from determining progress in statistical reasoning based on them. While the pre-test was administered in a learning context, the post-test was administered in a formal assessment context, where grade assignment may have influenced student performance. Nevertheless, the qualitative analysis of the students' work, together with the intermediate activities, made it possible to identify changes in their strategies, arguments, and ways of interpreting variability throughout the instructional sequence.

5. Conclusions

Students' statistical reasoning evolved from being algorithmic and decontextualized at the Prestructural, Unistructural, and Multistructural levels to a form of statistical reasoning that considers variability as a determining factor in problem-solving, in which context of variability was taken into account and represented using box-and-whisker plots or scatter plots. Students perceived, modeled, represented, explained, and considered variability in their decision-making at the Relational and Multistructural levels. Furthermore, throughout the hypothetical learning trajectory, students showed progressively more integrated responses regarding variability, evidencing transformations compatible with the development of statistical reasoning. Therefore, this reinforces the study's hypothesis that "the implementation of a hypothetical learning trajectory of this research can foster the development of statistical reasoning among students regarding variability, in risk situations and with the support of technology".

According to the retrospective analysis conducted using the SOLO Taxonomy, students' responses remained at the Multistructural level, as they relied on measures of central tendency without examining variability or considering its relationship with risk when solving the problems. Furthermore, the responses showed no evidence of reasoning at the Extended Abstract level, as students' interpretations lacked the integration of knowledge from other disciplines. This limitation may be partly explained by the relatively short duration of the instructional sequence, its focus on descriptive statistics

within the hypothetical learning trajectory, and the students' stage in the curriculum. Specifically, topics such as drug dissolution profiles and the quality control of tequila brands are addressed in later semesters of the Bachelor's program in Pharmaceutical Chemistry and Biology.

The use of design research made it possible to find evidence supporting theoretical, design, and high-level conjectures through an analysis of student responses based on the SOLO Taxonomy, which revealed the following:

- When students were able to visualize and interpret variability in response to questions designed around the levels of statistical reasoning defined by the SOLO Taxonomy, a process involving the visualization, description, analysis, and interpretation of variability emerged. This process integrated the use of estimators, including measures of central tendency and/or dispersion, to support data-driven decision-making in contextual problems involving risk. Through this process, different levels of statistical reasoning became evident in students' responses.
- When students prepared statistical reports on their subject area and created graphical representations, either individually and/or collaboratively, to solve risk problems involving data with variability, their responses showed a process of quantification and reasoning based on estimators. In addition, they related different estimators (measures of central tendency and/or dispersion) to support their conclusions. However, they did not do so at an Extended Abstract level, but only at the levels preceding it.
- Students taking an active role in group discussions and in creating statistical reports for their area during the instructional sequence activities addressing risk situations, justifying their interventions through statistical estimators, fostered the emergence of levels of statistical reasoning, although not at an Extended Abstract level.
- When the students completed the instructional sequence, which introduced quantification, various measures of central tendency and/or dispersion, and different graphical representations in risk contexts, students' responses showed higher levels of statistical reasoning.
- When students used verbal and statistical language to justify their problem-solving approaches throughout the instructional sequence, they strengthened their ability to make informed, data-driven decisions in high-risk contexts, taking variability into account as a determining factor.

Taken together, these findings suggest that the instructional sequence provided a framework for addressing some of the challenges associated with statistics education, including rote learning, the use of unrealistic problem contexts, an overemphasis on measures of central tendency, and the teaching of dispersion measures as formula-based procedures rather than concepts requiring interpretation.

Author Contributions: Conceptualization, M.H.-R.,C.O.-R. and N.B.d.A.-C.; methodology, M.H.-R., C.O.-R., and N.B.d.A.-C.; software, M.H.-R. and N.B.d.A.-C.; formal analysis, M.H.-R. and C.O.-R.; investigation, M.H.-R., C.O.-R., and N.B.d.A.-C.; writing—original draft preparation M.H.-R., C.O.-R., and N.B.d.A.-C.; writing—review and editing, M.H.-R., C.O.-R., and N.B.d.A.-C.; supervision C.O.-R. and N.B.d.A.-C.; funding acquisition, C.O.-R. and N.B.d.A.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This research was conducted as a non-interventional study involving human participants. This study was conducted in accordance with the ethical principles applicable to research involving human participants and was reviewed and approved by

Coordinación de Investigación of the Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI), Universidad de Guadalajara.

Informed Consent Statement: Written informed consent has been obtained from the participants to publish this paper.

Data Availability Statement: The datasets are not publicly available because they contain questionnaire responses that may compromise participants' privacy and confidentiality. Requests for access to anonymized data may be directed to the corresponding author.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (GPT-6) (OpenAI, San Francisco, CA, USA) to improve their English translations and to assist in adapting the manuscript to the journal template and formatting requirements. The authors reviewed and edited all outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SOLO Structure of the Observed Learning Outcome

Appendix A

This appendix presents the worksheets used in the hypothetical learning trajectory, which served as data collection instruments. These materials were developed through conjecture mapping as part of the design process.

Appendix A.1.

Activity 1: Dissolution Profile Test

Student: _____

For this activity, use a No. 2 pencil; the administrator will provide one. Answers must be written on the back of the sheets provided; if preferred, additional blank sheets may be requested. Write your name on each sheet used and properly justify each answer. You have 45 min to complete the activity. Figure 1 shows how a medication dissolution profile experiment is conducted.

You are part of a pharmaceutical team responsible for evaluating two different brands of the same 400 mg medication that is to be purchased. For these brands to be considered interchangeable generics, they must comply with the requirements established in Mexican Official Standard NOM-177-SSA1. This involves conducting bioequivalence studies to demonstrate that both brands have the same safety, efficacy, and quality profile as the reference medication.



Figure A1. Experimentation on drug dissolution.

Table 1 presents the times, in minutes, at which each of the 12 units from two different brands of the same medication, labeled Test Medication 1 and Test Medication 2, reaches 85% dissolution, allowing analysis of whether they meet the standard's requirements.

Table A1. Time in minutes required to reach 85% dissolution of the medications.

Unit	Reference Medication	Test Medication 1	Test Medication 2
1	13.2	11.5	14.2
2	12.9	15.6	10.9
3	14.3	10.4	12.6
4	13.4	11.3	13.4
5	11.6	11.6	12.7
6	12.5	15.5	12.1
7	11.6	15.7	14.0
8	12.9	15.9	10.0
9	11.2	9.9	12.6
10	13.2	9.9	11.7
11	12.2	9.9	14.8
12	12.2	10.40	11.8

Based on the information provided above, answer the following questions.

1. According to NOM-177-SSA1, how is it determined whether a medication is an interchangeable generic?

2. How was the experiment designed?

3. How long does each medication (reference and test medications) take to reach 85% dissolution?

4. Do the test medications comply with NOM-177-SSA1 to qualify as interchangeable generics?

Variability in the data plays a fundamental role in evaluating medications, since high dispersion in dissolution times may indicate inconsistencies in the formulation or manufacturing process.

5. Why do you think it is important that the data not be widely dispersed?

6. What are the maximum and minimum times for the test medications, and how do they influence dissolution?

7. Considering the dispersion of the data, how long does each medication (reference and test medications) take to reach 85% dissolution?

Appendix A.2.

BioSimuLab Activity

Student: _____

Read the following information carefully.

You have 45 min to complete the following activity:

In this activity, you will evaluate and select a medication that meets the criteria for being considered an interchangeable generic. To do so, you will use the BioSimuLab simulator, a tool that generates dissolution data to analyze equivalence between medications.

The simulator will generate an Excel file containing the dissolution profiles of three 400 mg medications:

- Reference medication: The original product against which the others are compared.
- Test medications 1 and 2: Generic alternatives whose equivalence will be evaluated.

Use the data from the simulator and the infographic to answer the questions below. Record your answers in a Word document and save the file in the same folder as the simulator, under the name "BioSimuLab#n", where n is the team number assigned by the person administering the activity. At the beginning, include the names of all team members and the recorder number assigned by the person administering the activity. Be sure to support each answer with the corresponding statistical justification.

To use the BioSimuLab simulator, wait for the instructions from the person administering the activity, as you will complete this process together.

Activity to complete

1. According to NOM-177-SSA1, what conditions must a medication meet to be labeled an interchangeable generic?
2. How was the experiment conducted?
3. Which of the two test medications would you purchase? Prepare a statistical report.

At the end of the activity, attach the Word file with your answers and the Excel file generated by the simulator to the corresponding Activity 3: BioSimuLab section in Classroom.

Appendix A.3.

Exam

Student: _____

Read the following information carefully.

For this examination, use a No. 2 pencil. Answers must be written on the back of the sheets provided; if preferred, additional blank sheets may be requested from the person administering the examination. Write your name on each sheet used and properly justify each answer. You have 120 min.

You may use your notebook, but electronic devices are not permitted, and you may not ask for or lend notes during the examination.

Problem 1

Four groups of physical therapy patients underwent different treatment regimens: electrodes, water exercises, muscle massage, and no treatment. The time in days from injury to recovery was estimated. Table 1 presents the data.

Table A2. Time in days from injury to recovery for each treatment regimen.

Electrodes	Water Exercises	Muscle Massage	No Treatment
64	76	58	95
88	70	74	90
72	90	66	80
80	80	60	87
79	75	82	88
71	82	75	85

- Which treatment regimen would you recommend? Why? Answer this question by preparing a statistical report with the following elements:

  (a) Independent variable.
  (b) Dependent variable.
  (c) Descriptive analysis; place the values on the box-and-whisker plot.

Problem 2

You are part of a team from the Tequila Regulatory Council (CRT) responsible for evaluating the quality of four types of tequila: Aged (añejo), Silver (blanco), Gold (oro), and Extra Aged (extra añejo).

One parameter analyzed is electrical conductivity, which is directly related to the amount of dissolved solids in the product (such as minerals, salts, or impurities). The more solids are present, the higher the electrical conductivity, because they facilitate the flow of electric current through the liquid. In general:

- Low conductivity suggests good filtration, quality control, and a lower risk of impurities.
- High conductivity may indicate a greater presence of dissolved solids, which could compromise quality or flavor or even imply adulteration.

Table A3 shows the conductivity results for the four types of tequila, measured with a conductivity meter, an instrument with sensors for electrical conductivity and temperature. These results were measured in millisiemens per centimeter (mS/cm). Table A3 shows the conductivity results for the four types of tequila, obtained using a conductivity meter, an instrument with sensors for measuring electrical conductivity and temperature. These results were measured in millisiemens per centimeter (mS/cm).

Table A3. Conductivity of different types of tequila, measured in millisiemens per centimeter (mS/cm).

Aged	Silver	Gold	Extra Aged
25.82	18.41	41.6	31.33
92.20	103.47	44.5	14.59
22.97	50.03	55.8	34.43
12.08	13.09	11.03	18.51
24.32	15.55	40.83	40.3
81.63	13.46		43.67
25.46	38.87		33.4
19.56	19.99		13.16
12.74	18.13		
13.71	11.18		

- Which type of tequila has the best quality? Why?

Appendix B

The following pinion survey assesses students' perceptions of their statistical reasoning about data variability, decision-making, and the use of the BioSimuLab simulator. It also explores how collaborative work and group discussions contribute to understanding and interpreting statistical variability.

Questionnaire on the Teaching Sequence

The Same, but Cheaper

Read each statement carefully and indicate the extent to which you agree with each item:

Scale: 1 = Strongly disagree; 2 = Disagree; 3 = I don't know; 4 = Agree; 5 = Strongly agree

1. Understanding and Description of Statistical Variability

Items	1	2	3	4	5
1. I identify the estimators or statistics used to describe the variability of a dataset.	□	□	□	□	□
2. I correctly explain variability in the data.	□	□	□	□	□
3. I use statistics or estimators to describe the variability of the data.	□	□	□	□	□
4. I explain the variability of the data using two or more estimators or statistics.	□	□	□	□	□

5. I explain in my own words what the estimators used to complete the activities mean.	□	□	□	□	□

2. Analysis and Interpretation of Statistical Variability

Items	**1**	**2**	**3**	**4**	**5**
6. I analyze box-and-whisker plots to interpret variability in the data.	□	□	□	□	□
7. When analyzing data, I integrate interpretations of data variability and context, rather than relying only on statistical summaries or graphs.	□	□	□	□	□
8. I consider estimators or statistics other than the mean when the context requires it.	□	□	□	□	□
9. I justify my interpretations using estimators or statistics.	□	□	□	□	□
10. I establish connections between different estimators to interpret the variability of the data.	□	□	□	□	□

3. Contextualization of Variability and Decision-Making

Items	**1**	**2**	**3**	**4**	**5**
11. When making decisions, I associate greater variability with greater risk.	□	□	□	□	□
12. I reflect on how variability affects decision-making in the problems presented in the activities for this course.	□	□	□	□	□
13. I analyze possible decisions based on the variability observed in box-and-whisker plots.	□	□	□	□	□
14. I consider the context of the problem when describing the data.	□	□	□	□	□

4. BioSimuLab Simulator Activity

Items	**1**	**2**	**3**	**4**	**5**
15. The simulator activity made it easier for me to visualize data variability.	□	□	□	□	□
16. Working with simulated data in the BioSimuLab activity improved my understanding of variability.	□	□	□	□	□
17. The simulator activity strengthened my ability to visualize and interpret variability.	□	□	□	□	□
18. The graphical representations created during the simulator activity helped me understand data variability.	□	□	□	□	□
19. The simulator helped me connect theory with practice regarding bioequivalence testing.	□	□	□	□	□

5. Collaborative Work and Group Discussion

Items	**1**	**2**	**3**	**4**	**5**
20. Collaborating with my classmates makes it easier for me to interpret data variability.	□	□	□	□	□
21. I find it useful to share different points of view to understand variability in the data.	□	□	□	□	□
22. Feedback from the group helps me improve my interpretations.	□	□	□	□	□
23. Teamwork allows me to compare ideas about how to interpret variability in the data.	□	□	□	□	□

Open-ended Questions

24. What recommendations or improvements would you suggest for future implementations or improvements of the sequence of activities?

_____

_____

_____

_____

25. What aspects do you think should be modified or improved in the BioSimuLab simulator activity?

_____

_____

_____

_____

References

1. Pomilio, C.J.; Miño, M.H.; Brignone, N.F.; Garcia Facal, G.; Telesnicki, M.C.; Fass, M.; Filloy, J.; Cueto, G.R.; Fernández, M.S.; Perez, A. Análisis de Actividades Sobre Estadística Descriptiva En Libros de Educación Media: ¿Qué Se Pretende Que Los Estudiantes Aprendan? *Rev. Estad. Educ.* **2016**, *18*, 1345–1364.
2. Sepúlveda, A.; Díaz-Levicoy, D.; Jara, D. Evaluación de la comprensión sobre Tablas Estadísticas en estudiantes de Educación Primaria. *Bolema* **2018**, *32*, 869–886. https://doi.org/10.1590/1980-4415v32n62a06.
3. Shaughnessy, J.M. Research on Statistics Learning and Reasoning. In *Second Handbook of Research on Mathematics Teaching and Learning*; Lester, F.K., Jr., Ed.; Information Age Publishing: Charlotte, NC, USA, 2007; pp. 957–1009.
4. Orta Amaro, J.A.; Sánchez Sánchez, E. Niveles de Razonamiento Sobre Variación Estadística de Estudiantes de Nivel Medio Superior al Resolver Problemas En Un Contexto de Riesgo. *Educ. Matemática* **2018**, *30*, 47–71. https://doi.org/10.24844/EM3001.02.
5. Garfield, J.; Ben-Zvi, D. A Framework for Teaching and Assessing Reasoning about Variability. *Stat. Educ. Res. J.* **2005**, *4*, 92–99.
6. Orta Amaro, J.A.; Fernández Delgado, G. Descripción de La Variabilidad Estadística de Datos En El Contexto de Una Situación de Tratamientos Médicos. *Probab. Condicionada Rev. Didáctica Estadística* **2013**, 2, 189–195.
7. Koklu, O.; Kaplan, J.J. Undergraduate Students' Use of Primitive Notions When Reasoning about Variability. *Int. J. Sci. Math. Educ.* **2022**, *21*, 1243–1264. https://doi.org/10.1007/s10763-022-10293-3.
8. Boels, L.; Bakker, A.; Van Dooren, W.; Drijvers, P. Conceptual Difficulties When Interpreting Histograms: A Review. *Educ. Res. Rev.* **2019**, *28*, 100291. https://doi.org/10.1016/j.edurev.2019.100291.
9. Kaplan, J.J.; Lyford, A.; Jennings, J.K. Effects of Question Stem on Student Descriptions of Histograms. *Stat. Educ. Res. J.* **2018**, *17*, 85–102.
10. Chaphalkar, R.; Wu, K. Students' Reasoning about Variability in Graphs during an Introductory Statistics Course. *Int. Electron. J. Math. Educ.* **2020**, *15*, em0580. https://doi.org/10.29333/iejme/7602.
11. García García, J.I.; Fernández Coronado, N.A.; Imilpán Rivera, I.A. Desarrollo Del Razonamiento Probabilístico En Profesores de Matemáticas Mediante Simulación Computacional. *Paradigma* **2020**, *41*, *404-426*. https://doi.org/10.37618/PARADIGMA.1011-2251.0.p404-426.id917.
12. Cox, F.T.; Sánchez, D.G.; Ruiz, Á.A.M.; Palma, L.O. Enseñanza de Estadística Descriptiva Mediante El Uso de Simuladores y Laboratorios Virtuales En La Etapa Universitaria. *Bordón Rev. Pedagog.* **2022**, *74*, 103–123. https://doi.org/10.13042/bordon.2022.94121.
13. Orozco-Rodriguez, C.; Palafox-González, A.; Valenzuela-García, C. A Realistic Situation, Proposed from Data Generated within the GeOrder Simulator to Elicit the Statistical Reasoning. *Heliyon* **2023**, *9*, e19330. https://doi.org/10.1016/j.heliyon.2023.e19330.
14. Hernández-Rivera, M.; Almeida-Cunha, N.B.; Cortez-Álvarez, C.R. Evaluation of Dissolution Profiles With BioSimuLab Simulator Based on NOM-177-SSA1 Mexican Standard. In *Interdisciplinary Approaches to Game Design in Healthcare and Education*; Carranza, B., Negrón, D., Pérez-Cisneros, A., Eds.; IGI Global Scientific Publishing: Hershey, PA, USA, 2026; pp. 359–398.
15. Reading, C.; Shaughnessy, J.M. Reasoning About Variation. In *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*; Ben-Zvi, D., Garfield, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 201–226.
16. Garfield, J.B.; Ben-Zvi, D. *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*; Springer: Berlin/Heidelberg, Germany, 2008.
17. Wild, C.J.; Pfannkuch, M. Statistical Thinking in Empirical Enquiry. *Int. Stat. Rev.* **1999**, *67*, 223–265.
18. Sánchez Sánchez, E.; Silvestre Castro, E.; García Ríos, V.N. Patrones de Razonamiento Sobre Pruebas de Significación de Estudiantes de Bachillerato. In *Proceedings of the Actas del Tercer Congreso Internacional Virtual de Educación Estadística*; Contreras, J.M., Gea, M.M., López-Martín, M.M., Molina-Portillo, E., Eds.; Universidad de Granada: Granada, Spain, 2019.

19. Biggs, J.B.; Collis, K.F. *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)*; Academic Press: Cambridge, MA, USA, 1982.
20. Van den Akker, J.; Bannan, B.; Kelly, A.E.; Nieveen, N.; Plomp, T. An Introduction to Educational Design Research. In *An Introduction to Educational Design Research*; Plomp, T., Nieveen, N., Eds.; SLO: Amersfoort, The Netherlands, 2010; pp. 9–16.
21. Plomp, T. Educational Design Research: An Introduction. In *An Introduction to Educational Design Research*; Plomp, T., Nieveen, N.M., Eds.; SLO: Netzodruk, Enschede, The Netherlands, 2010; pp. 9–35.
22. Bakker, A. *Design Research in Education: A Practical Guide for Early Career Researchers*; Routledge: London, UK, 2018.
23. Cobb, P.; Confrey, J.; DiSessa, A.; Lehrer, R.; Schauble, L. Design Experiments in Educational Research. *Educ. Res.* **2003**, *32*, 9–13. https://doi.org/10.3102/0013189x032001009.
24. Simon, M.A. Reconstructing Mathematics Pedagogy from a Constructivist Perspective. *J. Res. Math. Educ.* **1995**, *26*, 114–145. https://doi.org/10.2307/749205.
25. Secretaría de Salud. *Norma Oficial Mexicana NOM-177-SSA1-2013, Que establece las pruebas y procedimientos para demostrar que un medicamento es intercambiable. Requisitos a que deben sujetarse los Terceros Autorizados que realicen las pruebas de intercambiabilidad. Requisitos para realizar los estudios de biocomparabilidad. Requisitos a que deben sujetarse los Terceros Autorizados, Centros de Investigación o Instituciones Hospitalarias que realicen las pruebas de biocomparabilidad*. Mexico City: Diario Oficial de la Federación, September 20, 2013. https://dof.gob.mx/nota_detalle_popup.php?codigo=5314833.
26. Daniel, W.W. *Bioestadística: Base para el análisis de las ciencias de la salud*, 4th ed.; Editorial Limusa S.A. de C.V.: Mexico City, Mexico, 2002.
27. Carreón-Álvarez, A.; Suárez-Gómez, A.; Zurita, F.; Gómez-Salazar, S.; Soltero, J.F.A.; Barcena-Soto, M.; Casillas, N.; Porfirio-Gutierrez; Moreno-Medrano, E.D. Assessment of Physicochemical Properties of Tequila Brands: Authentication and Quality. *J. Chem.* **2016**, *2016*, 6254942. https://doi.org/10.1155/2016/6254942.
28. Brabrand, C.; Dahl, B. Using the SOLO Taxonomy to Analyze Competence Progression of University Science Curricula. *High. Educ.* **2009**, *58*, 531–549. https://doi.org/10.1007/s10734-009-9210-4.
29. Orta Amaro, J.A.; Sánchez, E.A. La Toma de Decisiones En Una Situación de Riesgo. In Proceedings of the XIV Conferencia Interamericana de Educación Matemática, Tuxtla Gutiérrez, Mexico, 3–7 May 2015.
30. Inzunza, S.; Enciso, S.S. Alfabetización y Razonamiento Estadístico de Estudiantes Mexicanos al Concluir El Bachillerato. *Rev. Chil. Educ. Matemática* **2022**, *14*, 101–117.
31. Vidal-Szabó, P.; Estrella, S. Explorando La Extensión Del Modelo MTSK al Dominio Estadístico: Características Del Aprendizaje Desde La Taxonomía SOLO. *Rev. Educ. Estad.* **2023**, *2*, 1–25. https://doi.org/10.29035/redes.2.1.3.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.