DEVELOPING A LOCAL INSTRUCTION THEORY FOR INFORMAL STATISTICAL ESTIMATION IN MIDDLE SCHOOLS

SONGHEE LEE
Seoul National University
thd2432@gmail.com

CHANG-GEUN SONG
Cheongju National University of Education
cgsong@cje.ac.kr

KYEONG-HWA LEE
Seoul National University
khmath@snu.ac.kr

ABSTRACT

This study aimed to develop a Local Instruction Theory for Informal Statistical Estimation (ISE). Although Informal Statistical Inference (ISI) has been addressed in prior studies, ISE—defined as the cognitive process of making conjectures about the likely range of a population parameter based on random sampling outcomes and expressing the uncertainty of such conjectures using probabilistic language—has received limited attention. We conceptualized ISE as situated within the broader ISI framework but distinct in its focus on interval estimation. Grounded in design heuristics of Realistic Mathematics Education, we conducted a teaching experiment with ninth-grade students to explore how they developed informal reasoning about sampling distributions. Students engaged in a series of activities designed to foster a shift from intuitive reasoning to the development of models of and for statistical estimation. Data sources included student artifacts, classroom discourse, and simulation outputs. Using Vergnaud's conceptual field theory, we analyzed students' emergent modeling processes by identifying concepts-in-action and theorems-in-action. Findings illustrate how instructional design can support students in understanding variability, uncertainty, and the meaning of confidence levels.

Keywords: *Informal statistical estimation; Confidence interval; Realistic Mathematics Education; Local instruction theory; Conceptual field*

1. INTRODUCTION

Fostering students' statistical inference competence is widely regarded as a primary goal of statistics education (Bargagliotti et al., 2020; Makar & Rubin, 2018). However, teaching and learning statistical inference continues to pose significant challenges for both teachers and students (Dolor & Noll, 2015; Pfannkuch et al., 2012). These challenges may be attributed to two major causes. First, due to the inherent conceptual complexity of statistical inference, students struggle to understand the relationships among statistical concepts, such as random processes, confidence intervals, and sampling variability (Pfannkuch et al., 2012). Second, students experience difficulties in integrating multiple statistical concepts when first introduced to *formal statistical inference* (*FSI*), such as statistical estimation or hypothesis testing, through traditional instructional methods (Chance et al., 2004; Garfield & Ben-Zvi, 2008).

Given students' difficulties in coordinating multiple statistical concepts and in connecting descriptive statistics with formal inferential procedures, Zieffler et al. (2008) proposed informal inferential reasoning as a research framework for examining how learners make data-based claims beyond the observed data without using formal statistical methods. Building on this perspective, Makar and Rubin (2009) conceptualized *informal statistical inference (ISI)* in classroom contexts as involving generalization beyond the data, the use of data as evidence, and the expression of uncertainty through

Statistics Education Research Journal (early access), https://doi.org/10.52041/serj.891

probabilistic language. Later, Makar and Rubin (2018) reviewed ISI research as part of broader efforts to understand how students can learn statistical inference before and alongside formal inferential methods. Taken together, the work of these scholars positions ISI as a way to create a conceptual bridge between students' work with data in descriptive statistics and their later learning of FSI.

Prior research on supporting students' learning of ISI can be broadly categorized into two strands (Nilsson, 2023). Results from the first strand broaden the scope of statistical inference by examining how students generate probabilistic generalizations from data without relying on formal inferential methods. Researchers in the first strand have explored how elementary and middle school students can develop foundational ideas for later FSI—ideas such as sample, variability, distribution, and uncertainty—through informal reasoning activities (e.g., Estrella et al., 2023; van Dijke-Droogers et al., 2020). The second strand focuses more directly on how students' informal inferential reasoning can be organized around the core logic of particular formal methods. For example, researchers recently designed learning trajectories that support students' informal engagement with the logic of hypothesis testing before students encounter formal hypothesis testing procedures (e.g., Nilsson, 2023; Nilsson & Eckert, 2024). Despite advances in ISI research, relatively little attention has been given to how statistical estimation, another central component of FSI, can be introduced and developed progressively from an informal level. Several researchers have designed learning trajectories in which students engage in informal inferential reasoning to develop concepts, such as sampling, variability, distribution, and uncertainty, that are foundational to statistical inference (e.g., van Dijke-Droogers et al., 2020; Nilsson & Eckert, 2024). Although such concepts are also foundational for statistical estimation, prior learning trajectories have not explicitly addressed how students can progressively develop informal reasoning about estimation. Given the central role of statistical estimation in inferential statistics (Bargagliotti et al., 2020; Budgett & Rose, 2017; Pfannkuch et al., 2012), the limited attention to estimation-oriented learning trajectories represents an important gap in the literature. Among few exceptions, Budgett and Rose (2017) and Inzunza (2018) showed that informal approaches using simulations and visualizations can support students' participation in statistical estimation. However, their findings also suggest that participation in estimation activities does not necessarily lead to an integrated understanding of the fundamental concepts involved in statistical estimation. Therefore, research is needed to articulate learning trajectories that introduce statistical estimation progressively from an informal level while supporting students in developing a conceptually robust understanding of estimation.

To address the limited attention to estimation-oriented learning trajectories, we propose *Informal Statistical Estimation (ISE)* as a specific cognitive process within ISI focused on statistical estimation. Whereas informal hypothesis testing (Nilsson, 2023) concerns reasoning about whether data provide evidence for a claim or difference, ISE concerns reasoning about the likely range of a population parameter based on random sampling outcomes and about the uncertainty associated with such reasoning. In this study, we defined ISE as "the cognitive process of making conjectures about the likely range of a population parameter based on random sampling outcomes and expressing the uncertainty of such conjectures using probabilistic language." This definition aligns with the key principles of ISI (Makar & Rubin, 2009) but emphasizes the cognitive process of interval estimation without invoking formal probability distribution theory. We positioned ISE as developmentally foundational for later FSI involving confidence intervals, given that a focus on ISE can support students in grappling with sampling variability and uncertainty.

For the present study, we adopted a design-oriented perspective grounded in *Realistic Mathematics Education* (*RME*). RME provides design heuristics for supporting students' progressive development of mathematical ideas through carefully sequenced emergent modeling activities, emphasizing the transition from contextually situated activity (*'model of'*) to more generalizable forms of reasoning (*'model for'*). Drawing on these heuristics, our goal was to develop a *Local Instruction Theory (LIT)* for ISE and to investigate how students' reasoning develops through a teaching experiment with ninth-grade students (ages 14-15). We focused on ninth graders because students at this level can participate meaningfully in repeated sampling and simulation-based activities while building on prior frequentist experiences with probability. The resulting LIT details key learning goals, instructional activities, and means of support that can help students coordinate repeated sampling, sampling variability, and uncertainty to develop a coherent understanding of confidence levels before formal instruction on confidence intervals.

To analyze students' reasoning at a micro level, we drew on Vergnaud's (1998) theory of conceptual fields (CF). In CF, concepts-in-action refer to aspects of a situation that students treat as relevant for reasoning, such as whether a sample proportion yields an interval that contains the population proportion. Theorems-in-action refer to propositions that students treat as true in reasoning, such as the idea that the confidence level can be approximated by the relative frequency of sample proportions whose intervals contain the population proportion. Accordingly, the research questions were as follows:

1. In the emergent modeling process guided by the LIT for ISE, what concepts-in-action and theorems-in-action emerge, and how do they evolve?
2. What instructional supports facilitate students' development of concepts-in-action and theorems-in-action during the emergent modeling process?

2. THEORETICAL BACKGROUND

2.1 LOCAL INSTRUCTION THEORY BASED ON REALISTIC MATHEMATICS EDUCATION DESIGN HEURISTICS

In this study, we drew on design heuristics from RME to develop an LIT for ISE. Rooted in Freudenthal's view of mathematics as a human activity, RME emphasizes that students should actively construct mathematical knowledge through meaningful activity rather than receive it as a set of ready-made systems (Gravemeijer, 1999). At the same time, the mathematics constructed by students needs to be aligned with conventional mathematical knowledge. Thus, teaching in RME involves supporting students in developing their own reasoning while guiding this development toward intended learning goals (Gravemeijer & Prediger, 2019). The goal of RME design research in supporting this process is to develop an LIT that encompasses both theories about students' learning processes and the means to support that learning (e.g., learning activities, tasks and tools, the role of teachers, classroom culture, etc.) (Ibid).

In RME design research, three instructional design heuristics are employed: *guided reinvention, didactic phenomenology*, and *emergent modeling* (Gravemeijer, 1999). In this study, guided reinvention informed the design of learning trajectories through which students progressively developed key statistical ideas underlying ISE. Within such trajectories, students are given opportunities to construct mathematical knowledge through progressive mathematization, in which they mathematize context-based problem situations that are realistic and meaningful to them and reorganize their own activity into more formal mathematical concepts, structures, or ideas (Gravemeijer & Doorman, 1999). From this perspective, context problems in the present study were designed to support students in mathematizing estimation situations that were "experientially real" (Ibid, p. 111) to them.

Didactic phenomenology emphasizes the selection of phenomena that can support students' progressive mathematization of intended mathematical ideas (Freudenthal, 1991; Larsen, 2018). In RME, contexts are not merely used to motivate students, but serve as starting points for knowledge construction through progressive mathematization (Gravemeijer & Doorman, 1999). From this perspective, the estimation situations in the present study were selected and organized to support students in progressively mathematizing variability and uncertainty in the context of ISE.

Emergent modeling brings structure to an LIT by providing a framework that organizes the trajectory of the guided reinvention process (Larsen, 2018). Models are intended to support students in modeling their own informal mathematical activity rather than merely concretizing abstract mathematical knowledge (Doorman & Gravemeijer, 2009). Within this framework, models initially function as *'models of'* students' activity, representing and structuring their informal reasoning within specific contexts, and gradually transition into *'models for'* more formal mathematical reasoning that can be applied across situations (Gravemeijer, 1999). In the present study, emergent modeling informed the design of instructional activities through which students developed descriptive models of estimation situations and gradually shifted toward using them as models for reasoning about variability and uncertainty in ISE.

The above heuristics of RME have been utilized to design learning trajectories through which elementary or middle school students engage in ISI (e.g., Büscher & Schnell, 2017; van Dijke-Droogers et al., 2022). For example, van Dijke-Droogers et al. (2022) developed an eight-step learning trajectory in which ninth-grade students explored sampling variability and uncertainty through repeated sampling in a black box context. Although their trajectory supported students' development of key ideas for statistical inference, it did not explicitly organize students' reasoning around interval estimation or the likelihood that an interval constructed from a sample would contain a population parameter. Building on this line of research, the present study drew on these RME heuristics to design an LIT for ISE at the middle school level.

To clarify the theoretical framing, we distinguish between an LIT and a *Hypothetical Learning Trajectory* (*HLT*) (Gravemeijer, 2004). An HLT refers to the planning of instructional goals, activities, and anticipated student learning processes within a specific classroom context (Simon, 1995). In contrast, an LIT provides a more general account of an envisioned learning route for a particular topic, including both a description of students' learning processes and a rationale for supporting them (Gravemeijer, 2004). In this sense, an LIT serves as a 'framework of reference' that guides teachers in designing their own context-specific HLTs (Gravemeijer, 2004; Nickerson & Whitacre, 2010).

2.2. PEDAGOGICAL APPROACHES TO TEACHING INFORMAL STATISTICAL ESTIMATION

Teaching statistical estimation has traditionally followed a formal approach by primarily focusing on the algebraic derivation of confidence intervals based on normal theory (Pfannkuch et al., 2012). This has led to a limited understanding of statistical estimation among students (Henriques, 2016; Reaburn, 2014). In response, several researchers have proposed new approaches aimed at fostering students' understanding of (informal) statistical estimation or its foundational concepts. Three interrelated strategies have received significant attention.

The first strategy involves using empirical sampling distributions to help students construct an intuitive understanding of variability and uncertainty. Grounded in the idea that students can build mental models of how sample statistics vary by engaging in repeated sampling activities and visualizing empirical sampling distributions (Cobb, 2007; Dvir & Ben-Zvi, 2023; Pfannkuch et al., 2012; van Dijke-Droogers et al., 2020), the strategy focuses on making the invisible structure of variability visible by aggregating and interpreting data from multiple random samples. Through these experiences, students can distinguish between more and less frequently occurring sample statistics based on repeated random sampling and further develop an image of the range within which more frequently occurring values tend to fall. In this regard, Pfannkuch et al. (2012) referred to a cognitive structure representing the range of values within most sample statistics fall as a "sampling variability band image." Such an image of the range within which likely sample statistics tend to fall provides the basis for informal understanding of confidence levels and the variability inherent in statistical estimation (Shaughnessy, 2006).

The second strategy involves using simulations as an instructional tool for teaching and learning ISI (Bakker, 2004; Chance et al., 2004; Konold & Kazak, 2008; Pfannkuch et al., 2012; Rossman, 2008; Song et al., 2023; van Dijke-Droogers et al., 2020). Using simulations to easily repeat random sampling processes and discuss the outcomes can provide a foundation for the objectification of the repeated random sampling process (Song et al., 2023; van Dijke-Droogers et al., 2020). van Dijke-Droogers et al. (2020) showed that this process can play an important role in the transition of empirical sampling distributions from functioning as a 'model of' the result of repeated random sampling to serving as a 'model for' interpreting sampling variability and uncertainty. Additionally, in statistics classes that use simulations, asking students what occurs when the process is repeated (Rossman, 2008) and encouraging them to predict the outcomes (Bakker, 2004) can help them develop reasoning and imagery related to statistical inference (van Dijke-Droogers et al., 2020). Incorporating informal and dynamic visualizations into these simulations is also regarded as an effective strategy for fostering the development of mental images related to variability (Pfannkuch et al., 2012; Song et al., 2023).

The third strategy involves engaging students in reflective reasoning about the range and consistency of sample statistics, which promotes meta-cognitive awareness and helps students articulate and refine their emerging understanding of statistical uncertainty (Bakker, 2004; Chance et al., 2004; Makar & Confrey, 2005). Although this may seem similar to the first strategy, the emphasis here is not on the content or representation of sampling distributions per se, but on fostering students' ability to reason about why certain outcomes occur more frequently than others and to explain their thinking in relation to probabilistic ideas. Reflection prompts learners to evaluate the plausibility of their inferences, consider variability in outcomes, and develop coherent explanations grounded in empirical evidence (Makar & Confrey, 2005). Such reflective practices are especially important for developing the habits of mind necessary for informal inferential reasoning (Chance et al., 2004).

In summary, the three strategies aim to (a) develop students' intuitive model of the foundational ideas of ISE, (b) transform these initial models into more elaborated models for statistical reasoning, and (c) promote meta-level reflection that refines and extends their conceptual understanding. This progression enables students to move from concrete encounters with variability and uncertainty to principled inference, building a stable bridge to subsequent formal methods.

2.3. THE MEANINGS AND ROLES OF 'MODEL OF' AND 'MODEL FOR' IN INFORMAL STATISTICAL ESTIMATION

As previously mentioned, in RME, the relationship between mathematical 'thought thing' (concepts, structures, or ideas) and phenomena is analyzed from a didactical point of view, with a focus on how this relationship is acquired in the teaching-learning process (Gravemeijer, 2004). In the context of teaching and learning ISE, 'model of' and 'model for' can hold the following meanings and roles.

The focal phenomenon in this study—and a key idea for students to make sense of in learning ISE—is the success rate of estimate intervals capturing the parameter as experienced through repeated random sampling simulations. A confidence level can be interpreted as a frequentist probability grounded in repeated random sampling (Moore, 1992). Accordingly, a coherent understanding of confidence intervals includes recognizing that the confidence level refers to the long-run proportion of intervals, constructed by the same estimation procedure across repeated random samples, that capture the population parameter. In simulation-based learning, students can empirically approximate this long-run coverage by examining the relative frequency of intervals that contain the population parameter, thereby interpreting the uncertainty associated with the estimation procedure (Pfannkuch et al., 2012). To develop such an understanding, students should recognize a repeatable long-run random process in which repeated samples generate a stable distribution of estimation outcomes. They also should understand that any observed estimation outcome from a single sample is one realization of this long-run process (Pfannkuch et al., 2012). Consequently, students' pre-existing images of a sampling variability band—for example, 'a range of sample statistics that are likely to occur' or 'a range of frequently occurring sample statistics'—can be transformed into an image of 'the range of sample statistics whose corresponding intervals contain the population parameter' [1] (see Figure 1). This transition represents the development of a descriptive model of the outcomes of repeated statistical estimation, providing a promising starting point for the reinvention of ISE.

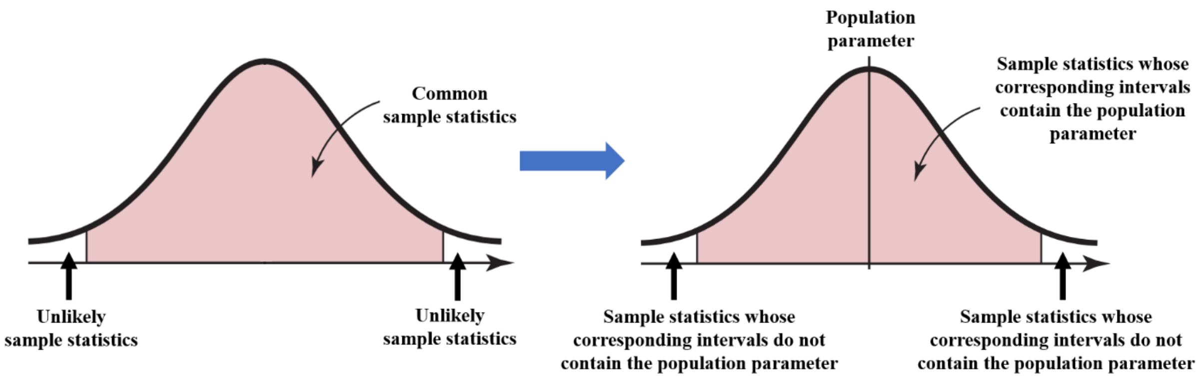


Figure 1. Expected transition of images about range of sample statistics in developing descriptive 'model of' the outcomes of repeated statistical estimation

[1]Throughout this paper, the term *range* refers to an interval of values rather than the statistical range (i.e., the difference between the maximum and minimum values).

The purpose of using confidence intervals is to represent both the estimate of a population parameter and the uncertainty stemming from variability in the data (Moore, 1990). Students should understand the limitations of conclusions based on sample surveys and be able to quantify the uncertainty associated with these conclusions by using the margin of error and related characteristics of the sampling distribution (Bargagliotti et al., 2020). This quantification entails reasoning about the range of sample statistics with an integrated understanding of the margin of error. Students who have developed a descriptive 'model of' the outcomes of repeated statistical estimation can (a) determine the range of sample statistics whose corresponding intervals contain the population parameter from an empirical sampling distribution (e.g., Inzunza, 2018) and (b) interpret the relative frequency of sample statistics within this range as an approximation of the confidence level. In this process of repeated estimation, the empirical sampling distribution shifts from serving as a descriptive 'model of' the outcomes of repeated statistical estimation to a 'model for' quantifying sampling variability and uncertainty in the context of statistical estimation. Through students' emergent modeling process, they come to understand that 'the likelihood of sample statistics within a specific range centered on the parameter' can be interpreted as 'the confidence level of a confidence interval with a width equal to that range' (Pfannkuch et al., 2012; Thompson & Liu, 2005), which enables them to organize both familiar and novel situations related to statistical estimation.

2.4. INTERPRETIVE FRAMEWORK: THE THEORY OF CONCEPTUAL FIELD

Design-based research requires researchers to explain how particular forms of reasoning emerge through students' participation in a designed learning environment and how instructional supports contribute to such emergence (Cobb et al., 2015). Accordingly, an interpretive framework is needed to support theoretically grounded interpretations of students' activity and of the means of support embedded in the learning environment (Cobb et al., 2015; Gravemeijer & Prediger, 2019). As briefly introduced in Section 1, the present study adopted Vergnaud's theory of conceptual fields (CF) as an interpretive framework for analyzing students' reasoning during the emergent modeling process.

In this study, CF was used to conduct a micro-level and ongoing analysis of the emergent modeling processes of participants. CF is appropriate for this purpose because it explains how students' implicit knowledge develops through activity in specific situations (Vergnaud, 1996, 1998). Students' reasoning is not always expressed as formal mathematical knowledge; rather, students may rely on ideas that remain implicit, are situation-bound, and are gradually reorganized through experience and learning (Vergnaud, 1996; Büscher & Prediger, 2019). To analyze students' implicit and situation-bound reasoning in estimation tasks, the present study drew on two closely related constructs from CF: concepts-in-action and theorems-in-action (Vergnaud, 1998).

Concepts-in-action indicate what students notice, select, or regard as relevant in a task situation (Vergnaud, 1998). In the context of the present study, concepts-in-action refer to aspects of an estimation situation that students treat as relevant for reasoning. Tracking students' concepts-in-action helps explain why students attend to particular aspects of a task, such as the frequency of sample proportions or whether a sample proportion yields an interval that contains the population proportion (Schnell & Büscher, 2015). Theorems-in-action indicate how students relate those aspects when making judgments or inferences (Vergnaud, 1998). In the context of this study, theorems-in-action refer to propositions that students treat as true when coordinating aspects attended to by students. For example, a student may reason that there is a range of sample proportions whose corresponding intervals contain the population proportion, or that frequently occurring sample proportions yield intervals that contain the population proportion. Concepts-in-action and theorems-in-action reflect learners' informal and developing reasoning and thus do not necessarily align with formal mathematical concepts (Büscher & Schnell, 2017). Because both constructs may remain implicit in students' activity, they are useful for identifying forms of reasoning that are not yet expressed as formal mathematical knowledge (Vergnaud, 1996, 1998; Büscher & Schnell, 2017).

In the context of ISE, students' reasoning about empirical sampling distributions and interval estimates illustrates the relationship between concepts-in-action and theorems-in-action. When students attend to whether a sample proportion yields an interval that contains the population proportion, the relevant concept-in-action can be represented as, "Success or failure of interval estimation based on a sample proportion." When students use this distinction to infer that there is a range of sample proportions whose corresponding intervals contain the population proportion, the corresponding theorem-in-action can be represented as, “There is a range of sample proportions whose corresponding intervals contain the population proportion.” In contrast, when students attend mainly to how frequently sample proportions occur in an empirical sampling distribution, the relevant concept-in-action can be represented as “The frequency with which a sample proportion occurs.” Such a focus may support the theorem-in-action, “Frequently occurring sample proportions yield intervals that contain the population proportion.” Although this theorem-in-action reflects meaningful attention to empirical sampling distributions and variability, it does not yet capture the logic of interval estimation, which requires coordinating the sample proportion, the population proportion, and the margin of error (Pfannkuch et al., 2012; Thompson & Liu, 2005).

CF is particularly useful for RME-based design research because it can help trace how students' reasoning is reorganized through mathematizing activity. RME design research emphasizes students' progressive mathematization and the transition from models of situated activity to models for more general reasoning (Gravemeijer, 1999, 2004; Doorman & Gravemeijer, 2009; Gravemeijer & Prediger, 2019). CF complements this perspective by identifying which aspects of the modeled situation become salient to students and how students coordinate those aspects into increasingly general propositions. Previous RME-based design studies have used CF to analyze students' mathematization processes and the development of models from situated tools into resources for more formal reasoning (Büscher & Schnell, 2017; Büscher & Prediger, 2019). Following this line of work, the present study used CF to analyze how students' reasoning about empirical sampling distributions develops from a descriptive model of repeated estimation outcomes into a model for quantifying sampling variability and uncertainty in ISE and the instructional supports that appeared to facilitate students' development.

3. METHODS

We adopted a small-scale design experiment grounded in the methodological principles of design-based research, which is characterized by a reflexive relationship between instructional design and classroom-based empirical inquiry (Stephan et al., 2003). This approach was chosen because it allows researchers to iteratively refine a conjectured LIT by examining how it plays out in actual classroom settings (Gravemeijer, 2004; Nickerson & Whitacre, 2010). The small-scale design experiment enabled us to examine in depth the learning trajectories of a small group of ninth-grade students and the role of instructional supports within the LIT. Through this format, the instructional design and its theoretical underpinnings were continuously tested and adjusted based on students' reasoning patterns. To explore the students' concepts-in-action and theorems-in-action during their emergent modeling process, we employed a classroom design study in which researchers collaborated with teachers to co-implement and refine the instructional sequence (Cobb et al., 2015). This approach has been found particularly effective for analyzing learning processes and instructional activity systems in situated classroom contexts (Cobb & Jackson, 2015) and was therefore considered appropriate for our purpose of developing and testing a conjectured LIT for ISE. This section begins by presenting the characteristics of the participants and the context of the teaching experiment. Next, we outline the conjectured LIT and teaching experiment. Finally, we describe the data collection and analysis procedures.

3.1. PARTICIPANTS AND CONTEXT

The teaching experiment was conducted in an after-school class at a middle school in Seoul, South Korea, involving ten ninth-grade students (S1-S10) and one mathematics teacher with seven years of teaching experience. The teacher had taught these students for two years in their regular classes and had established a close rapport with them; classroom norms also supported students' active expression of ideas and peer discussion. Prior to the teaching experiment, students had learned probability from a frequentist perspective, which was a key prerequisite for our LIT involving predictions of success rates in statistical estimation for a fixed confidence interval length. Meanwhile, they had no prior experience with formal concepts such as samples, probability distributions, sampling distributions, or confidence intervals. According to the teacher, three students showed high mathematics achievement (S3, S5, S9) and seven showed medium achievement (S1, S2, S4, S6, S7, S8, S10).

Although the Korean national curriculum aims to foster students' development of statistical reasoning and their ability to interpret diverse real-world contexts through a statistical lens, these goals are often undermined by formal, exam-oriented teaching practices (Lee et al., 2016; Lee et al., 2023; Tak & Jee, 2025). In this regard, the present study was situated in a particularly relevant context to address this persistent educational challenge—middle school, where students are expected to engage with foundational statistical ideas that support later learning (Bakker & Gravemeijer, 2004). Addressing challenges related to statistics instruction requires approaches that strengthen coherence across grades, particularly by connecting elementary experiences with variability and data to the development of ISE and, ultimately, ISI in middle school. In response to a broader call for coherence, recent research has advocated for instructional approaches grounded in a data-driven paradigm that foregrounds the development of ISI (Lee et al., 2021). Although Korea's 2022 revised national curriculum and recent textbooks have begun to incorporate inquiry-based tasks to promote ISI (Ministry of Education, 2022), concrete instructional strategies for supporting middle school students' ISE—and for positioning ISE as a cognitive process within ISI—remain scarce. Accordingly, this study explored pedagogical strategies for teaching ISE to middle school students as a bridge to fostering later FSI.

3.2. CONJECTURED LOCAL INSTRUCTION THEORY FOR INFORMAL STATISTICAL ESTIMATION

In the initial phase of the teaching experiment, we constructed a conjectured LIT. The conjectured LIT broadly presented a four-stage pathway for learning ISE, specifying the core concepts to learn, and, for each stage, the goals and the activities that support them. This pathway was initially proposed based on our analysis of prior research findings and theoretical considerations, reflecting the researchers' conjectures about how students' reasoning in ISE might develop. Although the conjectured LIT was primarily grounded in interpretations of prior studies and thought experiments, our broader aim was to establish a more robust and validated LIT by empirically testing and iteratively refining this pathway through the teaching experiment. In turn, the refined LIT can serve as a reference for teachers, enabling them to design and enact classroom-specific HLTs to enhance the likelihood of students' ISE development.

Table 1 summarizes the conjectured LIT and the intended learning pathway. Below, we outline the anticipated four-step learning progression, detailing goals, underlying assumptions, and instructional activities along with theoretical and empirical rationales.

Step 1 positioned the empirical sampling distribution as a descriptive 'model of' the outcomes of repeated random sampling (van Dijke-Droogers et al., 2020), supporting students' initial encounters with ISE-related concepts such as sample, sampling distribution, representativeness, and variability. Designed for students without prior exposure to these concepts, activities include: (a) collecting and interpreting data through inquiry-based estimation tasks (Franklin et al., 2007; Makar & Rubin, 2009), (b) anticipating and comparing the shape of a simulated sampling distribution with observed outcomes (Chance et al., 2004; van Dijke-Droogers et al., 2020), and (c) identifying 'frequent' and 'surprising' outcomes to engage in early inferential reasoning (Zieffler et al., 2008). This step functioned as the entry point in the LIT, establishing foundational images such as the sampling variability band image (Pfannkuch et al., 2012).

Step 2 reframed the empirical sampling distribution as a 'model for' interpreting sampling variability and uncertainty (van Dijke-Droogers et al., 2020). Activities included having students examine the characteristics of sampling distributions, articulate patterns in their own words to encourage sense-making (Bakker, 2004; Makar & Confrey, 2005), and predict simulation outcomes before observing them (Chance et al., 2004). These activities were intended to deepen reasoning about likelihood and variability, marking a conceptual shift in the LIT from descriptive familiarity to interpretive use. In Step 2, students deepen their understanding of ISE-related concepts, which in turn supports the estimation practices in Steps 3 and 4.

Step 3 involved using the empirical sampling distribution as a descriptive 'model of' repeated statistical estimation, prompting students to view parameter estimation outcomes as binary events (success or failure) defined relative to a fixed confidence interval length (Pfannkuch et al., 2012). A key pedagogical feature introduced here was the fixed-width interval estimation method (later presented to students as 'Junyoung's method'). This step began with Activity 1, in which students observed the outcomes of repeated statistical estimation simulations (Bakker, 2004; Rossman, 2008) to predict the success rate of the fixed-width interval estimation method for a given confidence interval length, with a trial defined as successful if the interval includes the population proportion. Building on this, Activity 2 was designed to guide students to determine, using an empirical sampling distribution, the range of sample proportions whose corresponding intervals contain the population proportion. The goals for the activities were for students to identify this range of sample proportions within the distribution and approximate the confidence level corresponding to the fixed interval length. By consolidating procedural imagery with probabilistic reasoning, Step 3 was designed to bridge students' informal observations toward a quantified understanding of statistical inference.

Table 1. Summary of the conjectured LIT for ISE and the intended learning pathway

Step	Concepts	Key learning goals	Instructional activities
1	Sample Repeated sampling Sampling variability Frequency distribution Empirical sampling distribution	• Developing repeatable process images for random sampling (Liu & Thompson, 2007) • Developing sampling variability band images (Pfannkuch et al., 2012) • Using frequency distribution as a 'model of' visualizing repeated random sampling outcomes (van Dijke-Droogers et al., 2020)	• Observing repeated random sampling simulation outcomes and guessing the population proportion • Conjecturing 10,000 repeated random sampling outcomes and representing the anticipated distribution as a frequency distribution • Expressing a range of frequently observed sample proportions in empirical sampling distributions generated by repeated random sampling
2	Sampling variability Uncertainty Empirical sampling distribution	• Developing reasoning about the range of sample proportions (Song et al., 2023) • Using empirical sampling distribution as a 'model for' interpreting sampling variability and uncertainty (van Dijke-Droogers et al., 2020)	• Observing the characteristics of empirical sampling distribution and interpreting the possibility of sampling variability and sample proportions within a specific range • Repeating a random sampling simulation for a given number of times to observe sample proportions that fall within a specific range and discussing the characteristics of the relative frequency of the sample proportions within that range
3	Empirical sampling distribution Confidence interval Confidence level	• Developing the way of thinking that views sample statistics as binary events of success or failure in interval estimation (Pfannkuch et al., 2012) • Using empirical sampling distribution as a descriptive 'model of' the outcomes of repeated interval estimation	• Observing and describing the success rate of a fixed-width interval estimation method based on sampling simulation • Determining the range of sample proportions whose corresponding intervals contain the population proportion under a fixed-width estimation method using an empirical sampling distribution
4	Empirical sampling distribution Confidence interval Confidence level Margin of error	• Developing reasoning about the range of sample proportions whose corresponding intervals contain the population proportion • Using empirical sampling distribution as a 'model for' quantifying sampling variability and uncertainty	• Determining the approximate confidence level of a fixed-width interval estimation method using an empirical sampling distribution • Determining the confidence interval length required to achieve a target confidence level using an empirical sampling distribution • Reflection on the reasoning process: Using an empirical sampling distribution to (1) compare confidence levels across different confidence interval lengths, (2) explain the relationship between the confidence interval length and the confidence level

Step 4 advanced to using the empirical sampling distribution as a 'model for' quantifying sampling variability and uncertainty in the context of ISE. This step was designed to develop reasoning about the range of sample proportions whose corresponding intervals contain the population proportion through three targeted activities: determining the approximate confidence level of a fixed-width interval estimation method using an empirical sampling distribution (Activity 3); determining the confidence interval length required to achieve a target confidence level (Activity 4); and reflecting on the reasoning process by comparing confidence levels for different interval lengths and explaining their relationship (Activity 5). In the design, Activity 3 incorporated a planned direct explanation by the teacher, intended to address potential persistent misconceptions and anticipated difficulties in connecting emergent models to formal statistical ideas. This planned intervention was not conceived as a conventional teacher-centered lecture, but as a deliberately timed scaffold within the instructional sequence, consistent with the RME principle of guided reinvention.

By structuring the LIT in this way, each step served as both a design conjecture and a developmental milestone within the broader HLT, aligning instructional actions with hypothesized student learning progressions.

3.3. TEACHING EXPERIMENT

The teaching experiment was conducted in an environment developed using STEAMuP (www.steamup.academy/ko), an educational authoring tool that supports inquiry-based mathematics learning. The experiment consisted of four sessions, each corresponding to a specific step of the conjectured LIT (90 minutes per session). In this study, we focused our analysis on students' emergent modeling activities during the third and fourth sessions (steps 3 and 4 of the conjectured LIT), where students engaged explicitly in ISE. The tasks designed based on the learning goals and instructional activities presented in steps 3 and 4 of our conjectured LIT are summarized as follows. (See Appendix 1 for additional information about the tasks.)

The learning goal of session 3 (step 3 of the conjectured LIT) was for students to develop a descriptive 'model of' the outcomes of repeated population proportion estimation. In RME instruction, students engage in experientially real instructional activities. Although the activities may involve mathematical modeling of real-world context, problems devoid of a story context can also serve as experientially real for students (Nickerson & Whitacre, 2010). For this session, students evaluated the success rate of the method referred to as 'Junyoung's guessing method,' which was characterized by the use of fixed-width interval estimation. The method comprised three phases. First, a random sample of colored balls was drawn from a "random box," and the sample proportion of red balls was computed. Second, a fixed-width confidence interval centered on this proportion was constructed. Third, the observations from the first two phases were used to form a statement about the interval that was expected to capture the population proportion. The simulation was performed repeatedly to obtain an empirical success rate (see Figure 2).

This environment was designed to guide students in reinventing the confidence level corresponding with a particular confidence interval length based on repeated random sampling simulations. In the preliminary Task of session 3, students examined an empirical sampling distribution obtained from 10,000 runs of a random sampling simulation. They used the distribution to determine a range where they believed the proportion of red balls (sample proportions) would 'almost certainly be' if the simulation were run one more time. Additionally, they estimated the probability that the proportion of red balls among 50 newly drawn balls would actually fall within that determined range.

In Task 1 (Activity 1 of the conjectured LIT), students adopted the length of the interval that they had determined as a group during the preliminary task as the fixed width for Junyoung's method. They were asked to predict the success rate of the estimation method corresponding to the confidence interval length by running the simulation one trial at a time and observing whether each sample proportion yielded an interval that contained the population proportion. In Task 2 (Activity 2 of conjectured LIT), students examined the empirical sampling distribution of 200 simulated samples along with the success rate of estimation over 200 trials. Based on the empirical sampling distribution and their observations from 200 simulation trials, they were asked to determine the range of sample proportions whose corresponding intervals contained the population proportion.

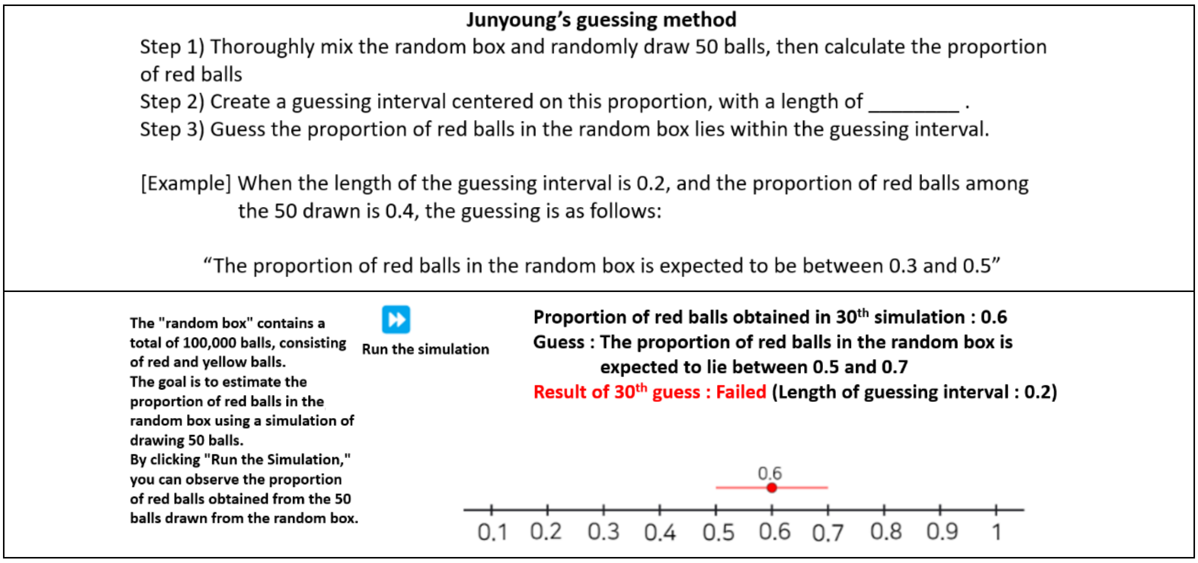


Figure 2. Junyoung's guessing method (top) and the interval estimation simulation (bottom)

The learning goal of session 4 (step 4 of the conjectured LIT) was for students to utilize the empirical sampling distribution as a 'model for' quantifying sampling variability and uncertainty. Students were expected to identify which sample proportions yielded confidence intervals that contained the population proportion and to express the uncertainty of the interval estimation method. To support students' reasoning, we designed an empirical sampling distribution in which the numerical values of individual sample proportions were hidden, but the population proportion was visually indicated by a vertical line (see Figure 3). This empirical sampling distribution was generated from 10,000 random samples from a population with a population proportion of 0.7 and sample size 50. With the sample size of 50, increments between adjacent sample proportions were fixed at 0.02. Students could also identify the frequency of each sample proportion from the numbers displayed above the bars in the graph.

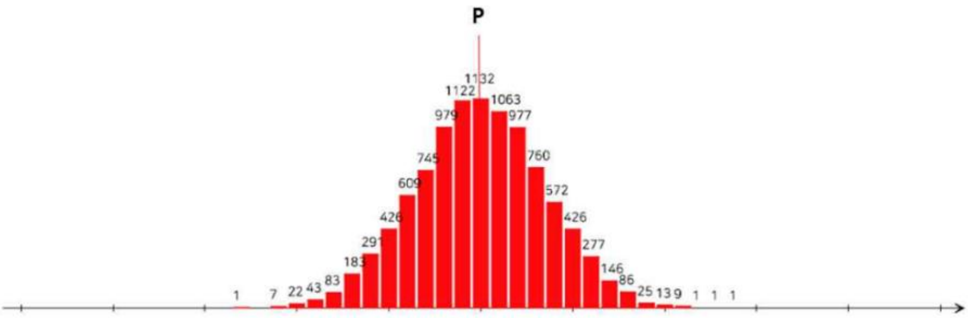


Figure 3. Empirical sampling distribution used in Task 3 and Task 4

In Task 3 (Activity 3 of the conjectured LIT), students were asked to determine the confidence level corresponding to a confidence interval length of 0.16 using the empirical sampling distribution. Conversely, in Task 4 (Activity 4 of the conjectured LIT), the direction of reasoning was reversed. Students were asked to determine the confidence interval length that would result in a confidence level of approximately 90%. Through these activities, students were expected to determine the range of sample proportions whose corresponding intervals contained the population proportion by comparing the differences between individual sample proportions and the population proportion relative to the margin of error, and to interpret the relative frequency of sample proportions within this range as an approximate confidence level. In Task 5 (Activity 5 of the conjectured LIT), students were encouraged to reflect on their reasoning about the range of sample proportions whose confidence intervals contained the population proportion. Using an empirical sampling distribution, students expressed the approximate confidence levels as probabilities for different confidence interval lengths assigned to each group (Task 5-1) and discussed the relationship between confidence interval length and confidence level (Task 5-2).

After the group work on each task, the teacher orchestrated a whole-class discussion, guiding students to share their group responses. In addition, during the whole-class discussions, the teacher noticed pedagogically critical events (Leatham et al., 2015) in students' responses and responded to them in order to encourage student reflection and advance the mathematical thinking of the entire class.

3.4. DATA COLLECTION AND ANALYSIS

The teaching experiment was conducted in an environment where each group of students used a single laptop to encourage reasoning and communication (Group 1: S1–S2, Group 2: S3–S4, Group 3: S5–S7, Group 4: S8–S10). We collected classroom video recordings, group audio recordings, screen recordings from the laptops, individual students' worksheets, and field notes taken by the authors during lesson observations. Additionally, to gather further information about students' thinking, interviews were conducted with selected students during breaks or immediately after class. Based on the field notes, interview participants were purposefully selected to include those who demonstrated the intended forms of reasoning as well as those whose responses deviated from the researchers' intentions. All audio data were transcribed.

Regarding research question 1, we analyzed the concepts-in-action and theorems-in-action that emerged and evolved during students' emergent modeling process. To ensure analytical transparency, we differentiated the interpretation of students' responses by examining how concepts-in-action and theorems-in-action were manifested in the 'model of' phase (Step 3) and how they were reorganized as students transitioned to a 'model for' phase (Step 4), respectively. In the development of a 'model of,' students' concepts-in-action reflected the identification of new categories in a situation-specific way, whereas their theorems-in-action reflected the formation of situation-bound inferential rules for using those categories. As students moved toward a 'model for,' these categories were abstracted for generalized use, and the inferential rules were generalized across situations. In our research context, distinguishing between concepts-in-action and theorems-in-action provided a comprehensive account of both students' developing statistical concepts and their inferential reasoning as they move from using an empirical sampling distribution as a 'model of' to using it as a 'model for.' This distinction allowed us to trace students' conceptual development in a more precise and fine-grained way.

To identify students' concepts-in-action and theorems-in-action as reflected in their actions and verbal explanations, we applied qualitative content analysis (Mayring, 2015). Qualitative content analysis is a methodological approach that incorporates both the hermeneutical position and the positivistic position of text analysis. It was an appropriate analytical method for the purpose of this study, which aimed to identify and trace changes in students' concepts-in-action and theorems-in-action within our conjectured LIT. The analysis procedure is as follows.

First, we collected data related to each task from student utterance transcripts and worksheet responses. We then divided the data into segments, with each segment representing a single idea or reasoning step in students' responses, and summarized the content of each segment. For example, when a student explained that a sample proportion yielded an interval that contained the population proportion, or that another sample proportion did not, we treated the explanation as one segment because it expressed a single criterion for distinguishing successful and unsuccessful interval estimates. Following Mayring's (2015) qualitative content analysis, we interpreted unclear segments by examining both the immediate textual context and additional contextual materials. The immediate textual context included the surrounding utterances in the transcript, whereas additional contextual materials included video recordings, screen recordings, diagrams in students' worksheets, and follow-up interview data. This process allowed us to clarify students' intended meanings while preserving the connection between their written or verbal responses and the broader task context.

Second, based on the mathematical concepts and properties expected to be used by students in each task, we deductively derived codes for the 'intended concepts-in-action' and 'intended theorems-in-action.' These codes were derived from the prior studies and formed the basis for the first round of coding. Concepts-in-action and theorems-in-action were denoted using the symbolic notations ‖...‖ and <…>, respectively. For example, students with stochastic conceptions of confidence interval regard individual sample proportions as binary events of success or failure in interval estimation (Pfannkuch et al., 2012). In Step 3 of the instructional activities in our conjectured LIT, students were expected to develop such a way of thinking. Accordingly, we extracted the concept-in-action code ||*Success or failure of the interval estimation based on a sample proportion*|| and the intended theorem-in-action code <*There is a range of sample proportions whose corresponding intervals contain the population proportion.*>.

Third, we conducted a second round of coding, in which we inductively assigned codes to recurring responses in text segments that had not been coded with the intended concepts-in-action or intended theorems-in-action during the first coding phase. Additionally, we organized the inductively derived codes into broader categories that encompassed similar or related codes, thereby identifying overarching concepts-in-action and theorems-in-action. For example, in solving Task 1, S1 and S2 focused on 'Values of sample proportions that rarely appear' as they observed the outcomes of repeated interval estimation. The responses "Proportion is not below 0.6" and "It does not go below 0.6" were coded as 'Rarely occurring sample proportions.' In addition, utterances such as "It seems like the proportion doesn't go beyond 0.6 to 0.9, right? Even though we keep trying, it never goes outside this range." were assigned the code 'Range of rarely occurring sample proportions.' This suggests that the students divided sample proportions into frequently or rarely occurring values based on a sampling variability band image (Pfannkuch et al., 2012). Consequently, the overarching concept-in-action embedded in their responses was represented as ||*The frequency with which a sample proportion occurs*||. Based on this conceptual distinction between frequent and rare values, the students recognized the existence of a specific boundary for these occurrences. This led to the identification of the theorem-in-action <*There is a range of rarely occurring sample proportions.*>, representing the proposition they held to be true regarding the distribution of sample proportions.

Fourth, to ensure the reliability of the coding results, a codebook comprising definitions for each task-specific code, anchoring samples, and demarcation rules was developed (Mayring, 2015). The second author then independently coded the same dataset using this shared codebook. Inter-rater reliability, calculated using Cohen's κ, was .98, indicating a high level of agreement. Any discrepancies were subsequently discussed and resolved through consensus.

Regarding research question 2, we identified the instructional supports that seemingly contributed to the development of students' concepts- and theorems-in-action. To identify the instructional supports that likely facilitated students' emergent modeling, we tracked how their concepts-in-action and theorems-in-action evolved through each activity. We compared each student's evolution in concepts-in-action and theorems-in-action with the intended learning pathway of the LIT (Prediger et al., 2015), while systematically contrasting these changes across students to discern both the heterogeneity and recurring patterns in learning pathways (Büscher & Schnell, 2017). Based on this, the instructional supports that led to the progress of students' concepts-in-action and theorems-in-action were derived. For example, in Activity 3, the classroom discussion initiated by the teacher in response to errors in the majority of students' reasoning strategies, facilitated the shift of students' concepts-in-action from ||*Sample proportions near the population proportion*|| to ||*Sample proportions within the margin of error from the population proportion*|| that conforms to a more formal statistical concept. In tandem with this conceptual shift, their theorems-in-action transitioned from a situated observation—<*An approximate confidence level to a given confidence interval length can be obtained by calculating the relative frequency of sample proportions that fall within a range near the population proportion, where the width of the range matches the given confidence interval length.*>—to a more formal and generalized rule: <*An approximate confidence level corresponding to a given confidence interval length can be obtained by calculating the relative frequency of the sample proportions within the margin of error from the population proportion.*>. This transition demonstrates how specific instructional supports within the LIT enabled students to refine their informal reasoning into mathematically robust concepts and theorems-in-action.

4. RESULTS

This section presents students' concepts-in-action and theorems-in-action for each task, along with the instructional supports that facilitated their advancement.

4.1. DEVELOPMENT OF 'MODEL OF'

Activity 1: Observing and describing the success rate of a fixed-width interval estimation method based on sampling simulation In Activity 1, students observed repeated interval estimation outcomes and predicted the success rate of the estimation method corresponding to their confidence interval length. Table 2 summarizes the concepts-in-action and theorems-in-action identified in Task 1.

Table 2. Concepts-in-action and theorems-in-action identified in Task 1

Concepts-in-action and Theorems-in-action		Description
Intended	‖*Success or failure of the interval estimation based on a sample proportion*‖ (n = 8; S3–S10) <*There is a range of sample proportions whose corresponding intervals contain the population proportion.*> (n = 8; S3–S10)	Students classify sample proportions as either successful or not in interval estimation, determining the range of sample proportions whose corresponding intervals contain the population proportion. (e.g., "Success at 0.82, failure at 0.84. Success at 0.62, failure at 0.6.")
Not intended	‖*The frequency with which a sample proportion occurs*‖ (n = 2; S1, S2) <*There is a range of rarely occurring sample proportions.*> (n = 2; S1, S2)	Students distinguish between frequently and rarely occurring sample proportions, recognizing that there is a range of sample proportions that rarely occur. (e.g., "It seems like the proportions don't go outside the range from 0.6 to 0.9. Even as we keep going, they just keep fluctuating within this range.")
	‖*The difference between the mean of sample proportions and individual sample proportions*‖ (n = 1, S8) <*As the confidence interval length increases, the confidence level also increases.*> (n = 3; S8–S10)	Students conjecture that as the confidence interval length increases, the range of sample proportions whose corresponding intervals contain the population proportion becomes wider, thereby increasing the confidence level. (e.g., "As the proportions of red balls deviates further from the average proportion, the guess interval must widen to increase the probability of success.")

This simulation-based activity served as a key instructional support that led most students to a conceptual shift about the range of sample proportions. The intended concept-in-action and theorem-in-action were identified in the majority of students' reasoning (n = 8; S3, S4, S5, S6, S7, S8, S9, S10). In response to "What is the success rate of your guessing method?", they observed the simulation outcomes, determined the range of sample proportions whose corresponding intervals contained the population proportion, and predicted the success rate of the method. They categorized sample proportions based on the ‖*Success or failure of interval estimation based on a sample proportion*‖. Furthermore, they recognized that <*There is a range of sample proportions whose corresponding intervals contain the population proportion.*>. For example, S3 and S4 identified and recorded the sample proportions whose corresponding intervals contained the population proportion. Based on these observations from the simulation, S3 determined the range of sample proportions whose confidence intervals contained the population proportion, with confidence interval length 0.22, to be between 0.62 and 0.82 (see Figure 4).

> It succeeded at 0.78, while it failed at 0.6. It succeeded at 0.8 but failed at 0.86. It also succeeded at 0.66. At 0.6, it failed again. It also failed at 0.62, wait, no. It succeeded at 0.62. It was successful at 0.62. Therefore, the failure occurred between 0.6 and 0.62. At 0.8 … this clearly indicates that the range has been narrowed. It failed at 0.86 but succeeded at 0.8. It also failed at 0.84. … Based on the results, it was successful at 0.82 but failed at 0.84. Similarly, it succeeded at 0.62 but failed at 0.6. This means that a failure occurred with just a 0.02 difference, an extremely narrow margin.

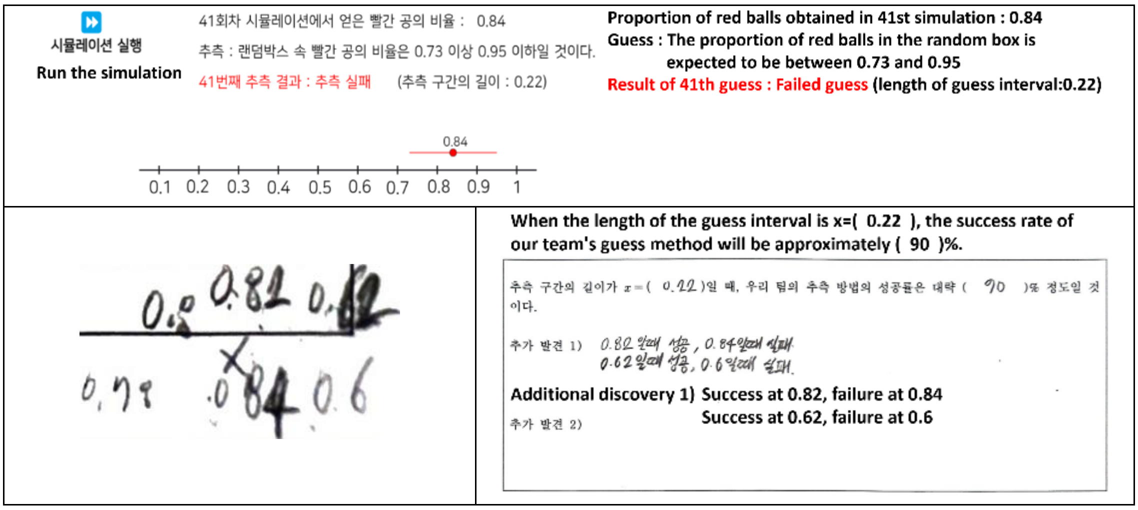


Figure 4. Screenshot of group 2's laptop (top) and S3's record and response to Task 1 (bottom)

Additionally, unintended concepts-in-action and theorems-in-action were identified in this activity. Students in group 4 (n = 3; S8, S9, S10) conjectured that, as the confidence interval length increases, the range of sample proportions whose intervals contained the population proportion becomes wider, thereby increasing the confidence level. The theorem-in-action they used was represented as <*As the confidence interval length increases, the confidence level also increases.*>. In particular, S8 inferred, from her perspective, the farther a sample proportion is from the mean of sample proportions, the longer the confidence interval needs to be in order to achieve a high success rate (see Figure 5). This inference indicates that the concept-in-action ||*The difference between the mean of sample proportions and individual sample proportions*|| was used as a basis for distinguishing among sample proportions.

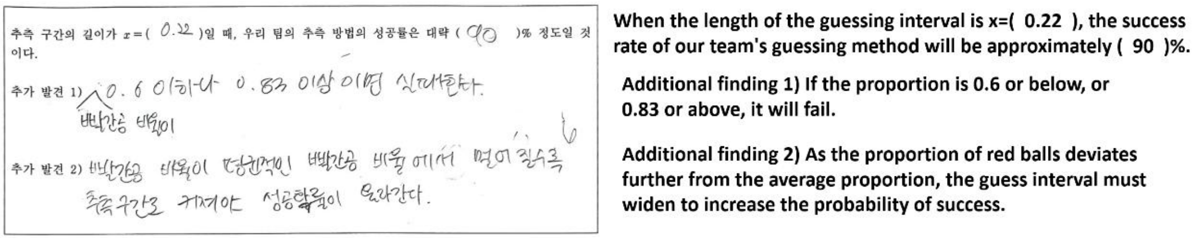


Figure 5. S8's response to Task 1

Meanwhile, S1 and S2 distinguished between frequently and rarely occurring sample proportions based on the concept-in-action ||*The frequency with which a sample proportion occurs*||. Initially, they set the confidence interval length to 0.08 and predicted the success rate of the estimation method by observing the first 67 simulation outcomes, focusing on the frequency of sample proportions whose confidence intervals contained the population proportion. However, during the review of the remaining 133 simulation outcomes, their attention shifted from success or failure in interval estimation to the range of rare sample proportions. In their discussion, they described the sample proportions as fluctuating within the range from 0.6 to 0.9 and suggested that values outside this range might occur only very infrequently. This view was also reflected in Figure 6, which shows that they did not expect the sample proportion to fall below 0.6. Taken together, their discussion suggests that they were beginning to distinguish a range of sample proportions that they regarded as very unlikely to occur; accordingly, their theorem-in-action was represented as <*There is a range of rarely occurring sample proportions.*>.

S2: Wait a second. Let's try again. It seems like the proportions don't go outside the range from 0.6 to 0.9. Even as we kept going, they just keep fluctuating within this range. At most, they go a little above 0.8, but they never exceed 0.9.

S1: Then what values does it usually take?
S2: They just fall between 0.6 and 0.9. They don't seem to go beyond that.
S1: But it might if we keep doing it.
S2: I don't think it ever will. Keep running it and see. It just won't happen.
S1: Yeah, it keeps moving back and forth within that range.
S2: Well, if you kept doing it for an hour, maybe it would happen once.

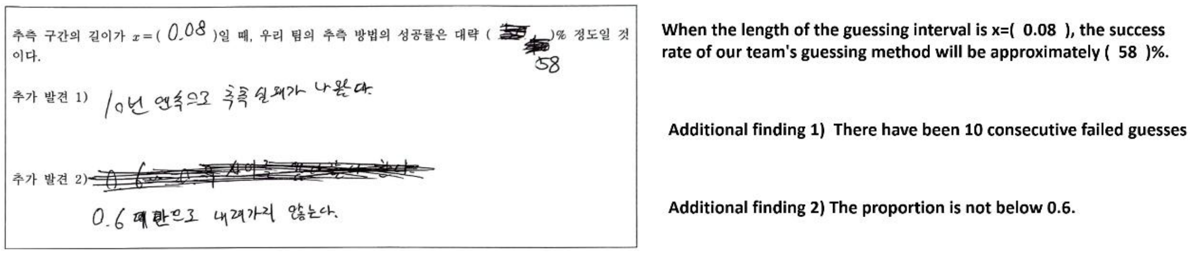


Figure 6. S2's response to Task 1

Activity 2: Determining the range of sample proportions whose corresponding intervals contain the population proportion under a fixed-width estimation method using an empirical sampling distribution In Activity 2, students examined the empirical sampling distribution obtained from 200 simulation trials they had observed in Activity 1 and confirmed the success rate of the estimation method up to 200 trials. They also indicated the range of sample proportions whose corresponding intervals contained the population proportion on a number line and explained the reason. Table 3 summarizes the concepts-in-action and theorems-in-action identified in Task 2.

Table 3. Concepts-in-action and theorems-in-action identified in Task 2

Concepts-in-action and Theorems-in-action		Description
Intended	||*Success or failure of the interval estimation based on a sample proportion*|| (n = 8; S3–S10) <*There is a range of sample proportions whose corresponding intervals contain the population proportion.*> (n = 8; S3–S10)	Students classify sample proportions as either successful or not in interval estimation, determining the range of sample proportions whose corresponding intervals contain the population proportion. (e.g., "If it goes below 0.6 or above around 0.83, the estimation fails.")
Not intended	||*The frequency with which a sample proportion occurs*|| (n = 2; S1, S2) <*Frequently occurring sample proportions yield intervals that contain the population proportion.*> (n = 2; S1, S2)	Students perceive the range of frequently occurring sample proportions from an empirical sampling distribution as the range that yields intervals that contain the population proportion. (e.g., "They matched it exactly to the range that appeared most frequently.")

In Task 2, students' strategies for determining the range of sample proportions whose corresponding intervals contain the population proportion were categorized into two groups. Among these, the intended concept-in-action and theorem-in-action were identified from students in groups 2, 3, and 4 (n = 8, S3–S10). They determined the range of sample proportions whose intervals contained the population proportion from the empirical sampling distribution corresponding to the 200 simulation outcomes, taking into account their observations from the simulation in Task 1. For example, S10 determined the range to be between 0.61 and 0.82 (see Figure 7). This can be regarded as a descriptive 'model of' the outcomes of repeated statistical estimation, which indicates that S10's sampling variability band image shifted into an image of the range of sample proportions whose intervals contain the population proportion. Among these students, S3 showed a noteworthy response by placing a dot near the center of the identified range (see Figure 8), suggesting an emergent awareness of the structural relationship between the population proportion and the range of sample proportions whose corresponding intervals contain the population proportion. However, despite these individual variations, the reasoning of all eight students at this stage remained primarily situated within their simulation observations. They determined the range without yet explicitly considering the underlying relationship between the population proportion and the confidence interval length. Accordingly, their concept-in-action and theorem-in-action can still be represented as ||*Success or failure of the interval estimation based on a sample proportion*|| and <*There is a range of sample proportions whose corresponding intervals contain the population proportion.*>, respectively.

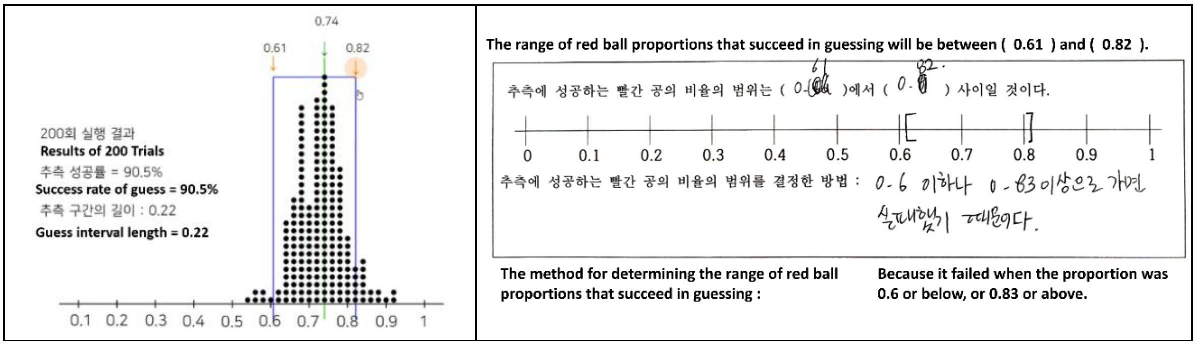


Figure 7. Screenshot of group 4's empirical sampling distribution and S10's response to Task 2

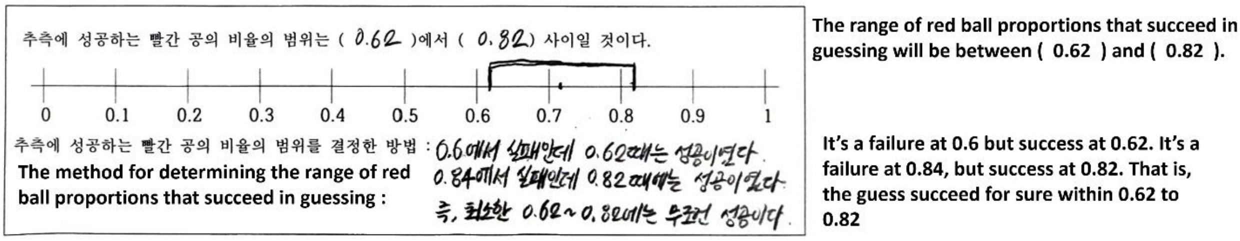


Figure 8. S3's response to Task 2

By contrast, in the case of Group 1, students (n = 2; S1, S2) did not make a conceptual shift to the range of sample proportions in Activity 1. They relied primarily on the sampling variability band image in determining the range of sample proportions whose corresponding intervals contained the population proportion. They set the range of 'the most frequently occurring outcomes' as the range of whose intervals contained the population proportion (see Figure 9). Their concept-in-action and theorem-in-action were represented by ||*The frequency with which a sample proportion occurs*|| and <*Frequently occurring sample proportions yield intervals that contain the population proportion.*>, respectively.

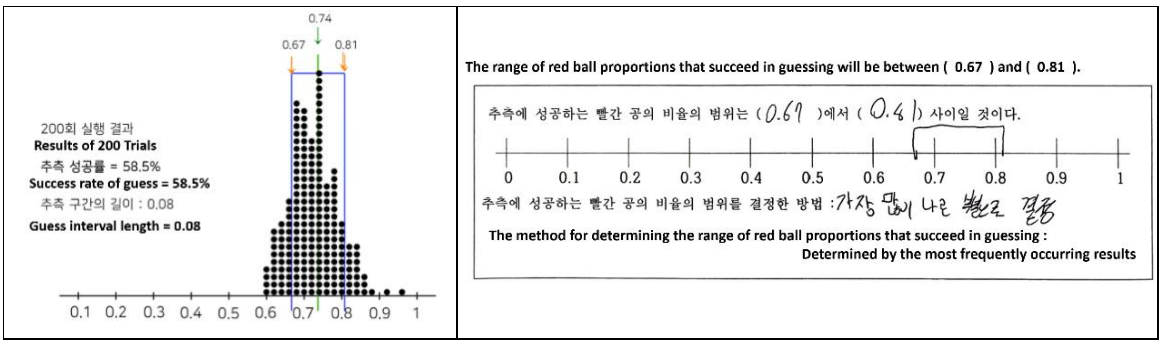


Figure 9. Screenshot of Group 1's empirical sampling distribution and S1's response to Task 2

4.2. TRANSITION TO 'MODEL FOR'

Activity 3: Determining the approximate confidence level of a fixed-width interval estimation method using an empirical sampling distribution In Activity 3, students examined the empirical sampling distribution (see Figure 3) and determined the confidence level corresponding to a confidence interval length of 0.16. Table 4 summarizes the concepts-in-action and theorems-in-action identified in Task 3.

Table 4. Concepts-in-action and theorems-in-action identified in Task 3

Concepts-in-action and Theorems-in-action		Description
Intended	||*Sample proportions within the margin of error from the population proportion*|| (n = 1; S6) <*An approximate confidence level corresponding to a given confidence interval length can be obtained by calculating the relative frequency of the sample proportions within the margin of error from the population proportion*> (n = 1; S6)	Students determine the range of sample proportions whose corresponding intervals contain the population proportion as those falling within the margin of error from the population proportion. In addition, students interpret the relative frequency of the sample proportion within that range as an approximate confidence level corresponding to the confidence interval length. (e.g., "Since it's symmetric around P, I used that as a reference. Adding up everything in this range gave me 7,959. And since the total is actually 10,000, it's 7,959 out of 10,000, which is 79.59%.")
Not intended	||*Sample proportions near the population proportion*|| (n = 7; S3–S5, S7–S10) <*An approximate confidence level corresponding to a given confidence interval length can be obtained by calculating the relative frequency of sample proportions that fall within a range near the population proportion, where the width of the range matches the given confidence interval length*> (n = 7; S3–S5, S7–S10)	Students determine the range of sample proportions whose corresponding intervals contain the population proportion as a range near the population proportion, not symmetric around it but equal in width to the confidence interval. In addition, students interpret the relative frequency of the sample proportion within that range as an approximate confidence level corresponding to the confidence interval length. (e.g., "Since each block is 0.02, that's about 8 blocks. Because 609 (The frequency of the sample proportion P – 0.08) is higher than 572 (The frequency of the sample proportion P + 0.08), I decided on the range from 609 to 760 (The frequency of the sample proportion P + 0.06).")
	<*If the confidence interval length is doubled, the confidence level is also doubled.*> (n = 1; S1)	Students conjecture that there is a directly proportional relationship between the confidence interval length and the confidence level. (e.g., "Since it's twice as much as 0.08, I doubled the success rate.")
	<*Considering the empirical sampling distribution, it is possible to estimate the change in the confidence level as the confidence interval length doubles.*> (n = 1; S2)	Based on the empirical sampling distribution, students conjecture that the confidence level will be about 1.6 times when the confidence interval length is doubled. (e.g., "When the confidence interval length was previously set to 0.08, the success rate was 58.5% (47.2%). Therefore, I predicted that it would be about 1.6 times greater.")

In Task 3, the strategies students used to determine the range of sample proportions whose intervals contained the population proportion were divided into four categories. The intended concept-in-action and theorem-in-action were identified from only one student, S6. S6 determined the range by focusing on the difference between the population proportion and individual sample proportions. In a brief interview with the researcher immediately after completing Task 3, S6 responded as follows.

R: From 609 to 572? How did you get it?

S6: Since it's symmetrical around *P*, I used that as a reference. Adding up everything in this range (from *P* – 0.08 to *P* + 0.08) gave me 7,959. And since the total is actually 10,000, it's 7,959 out of 10,000, which is 79.59%.

S6 calculated the relative frequency of sample proportions that fell within 0.08 of the population proportion, *P*, and interpreted this as the confidence level corresponding to the confidence interval length of 0.16. This transition represented a pivotal shift where the empirical sampling distribution no longer merely described repeated statistical estimation outcomes but functioned as a 'model for' quantifying the uncertainty of the estimation method.

By contrast, most students (n = 7; S3, S4, S5, S7, S8, S9, S10) aimed to determine the range of sample proportions near the population parameter so that the difference between the maximum and minimum values in this range would be 0.16. However, they misinterpreted that one bar in the empirical sampling distribution, which represented individual sample proportions, corresponded to a value of 0.02. Consequently, they decided to include eight neighboring sample proportions near the population proportion (which actually made the interval length 0.14), resulting in an error that led them to estimate the approximate confidence level of 73% instead of 79% (see Figure 10). These students interpreted the relative frequency of sample proportions within a certain range as an approximate confidence level, which indicates proportional reasoning about the range of sample proportions (Song et al., 2023). However, their approach did not align with formal statistical concepts because they did not consider the difference between the population proportion and individual sample proportions when determining the range of sample proportions whose intervals contained the population proportion.

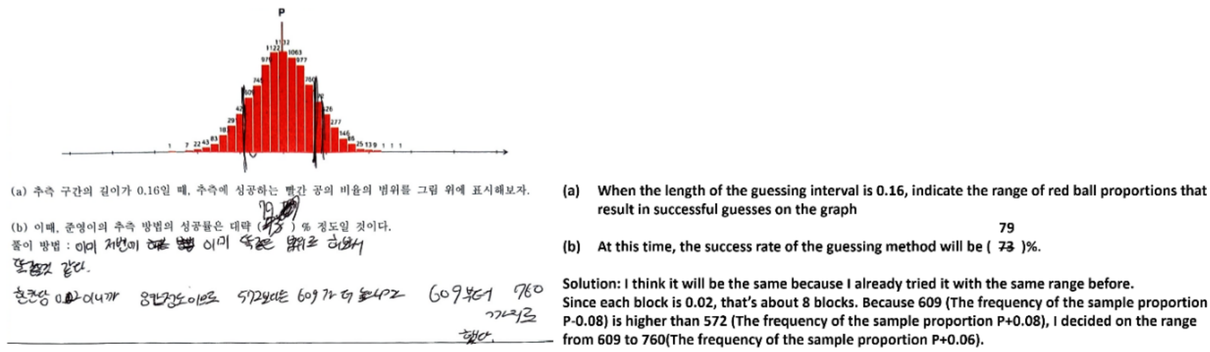


Figure 10. S7's response to Task 3

In the case of Group 1 students (n = 2; S1, S2), it was noted that the confidence interval length of 0.16 presented in Task 3 was double the length of the 0.08 interval they had previously explored in Tasks 1 and 2. During the whole-class discussion where students shared their responses to Task 3, S1 and S2 responded as follows.

T: How did you get the success rate?
S1: Earlier, when we used 0.08, it came out to forty-something.
T: We are working with 0.16.
S1: Since it's twice as much as 0.08, I doubled the success rate.
S2: I didn't do it that way.
T: How did you do it?
S2: I made a rough guess.

S1 initially determined the range of sample proportions whose corresponding intervals contained the population proportion with a confidence interval length of 0.08, ensuring that it included four bars around the population parameter (sample proportions *P* – 0.04, *P* – 0.02, *P*, *P* + 0.02), and calculated the relative frequency for this range to be 0.4296. Then, he predicted that because the confidence interval length had doubled, the confidence level would also approximately double to about 87%.

Meanwhile, S2 did not agree with S1's claim that the confidence level would double as the confidence interval length doubled. Instead, S2 predicted that the confidence level would increase by approximately 1.6 times when the confidence interval length increased from 0.08 to 0.16, based on the empirical sampling distribution (see Figure 11). The theorem-in-action applied here can be formulated as *<Considering the empirical sampling distribution, it is possible to estimate the change in the confidence level as the length of the confidence interval doubles.>*.

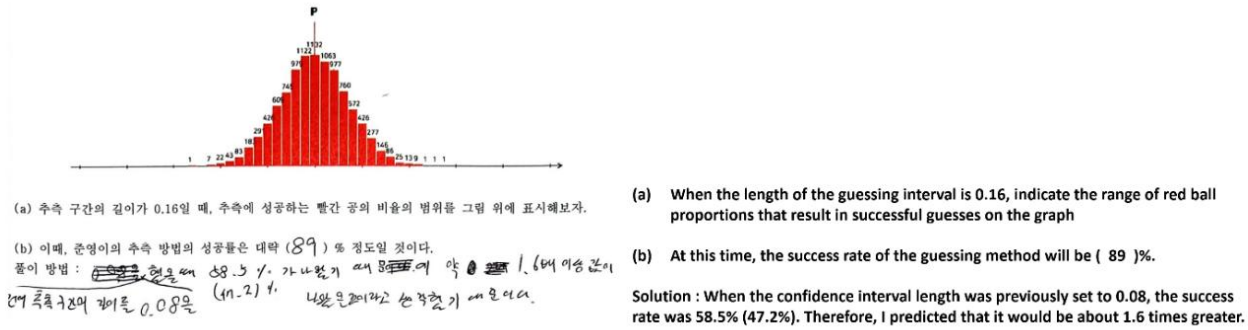


Figure 11. S2's response to Task 3

Activity 4: Determining the confidence interval length required to achieve a target confidence level using an empirical sampling distribution In Activity 4, students examined the empirical sampling distribution (see Figure 3) and determined the confidence interval length that yielded an approximate confidence level of 90%. Table 5 summarizes the concepts-in-action and theorems-in-action identified in Task 4.

Table 5. Concepts-in-action and theorems-in-action identified in Task 4

Concepts-in-action and Theorems-in-action		Description
Intended	*‖Sample proportions within the margin of error from the population proportion‖* (n = 7; S3, S4, S6–S10) *<An approximate confidence level corresponding to a given confidence interval length can be obtained by calculating the relative frequency of the sample proportions within the margin of error from the population proportion.>* (n = 7; S3, S4, S6–S10)	Students determine the range of sample proportions whose corresponding intervals contain the population proportion as those falling within the margin of error from the population proportion. In addition, students interpret the relative frequency of the sample proportions within that range as an approximate confidence level corresponding to the confidence interval length. (e.g., "It increases by 0.04 each time. It's increasing by 0.02 on each side.")
Not intended	*<Considering the empirical sampling distribution, it is possible to estimate the change in the confidence level as the confidence interval length doubles.>* (n = 3; S1, S2, S5)	Students determine the length of confidence interval that yields an approximate confidence level of 90%, using the empirical sampling distribution and the approximate confidence level of 79% for the confidence interval length of 0.16, which they confirmed from the whole-class discussion on Task 3. (e.g., "We tried 0.16 before, right? It was around 79%. So, if we go about this …")

In the whole-class discussion where groups shared their responses to Task 3, the teacher noticed that most students' reasoning about the range of sample proportions whose corresponding intervals contained the population proportion was flawed. To encourage students to reflect on their reasoning, he intervened before introducing Task 4, guiding students to focus on the difference between the population proportion and individual sample proportions.

T: Okay everyone, just a moment. There's one number right in the middle here. Do you understand what I mean? There's one right in the middle, and the ones next to it are -0.02 and +0.02. So earlier, S4 said there were 8 of them. It's not that there are a total of 8, but rather there are 4 on each side of the middle one, making the middle one the 'center.'

This teacher intervention functioned as an instructional support that corresponded with changes in the concepts-in-action and theorems-in-action of most students. The intended concepts-in-action and theorems-in-action were identified from seven students (S3, S4, S6, S7, S8, S9, S10). They gradually expanded the range symmetrically around the population proportion to identify the range of sample proportions whose corresponding intervals contained the population proportion, aiming for a confidence level close to 90%. In other words, they explicitly incorporated the margin of error into their reasoning to determine the range of sample proportions whose intervals contained the population proportion, an approach that aligns with FSI using confidence intervals. Here, the empirical sampling distribution was used as a 'model for' quantifying sampling variability and uncertainty. For example, S3 and S4 compared confidence interval lengths of 0.24, 0.26 and 0.28, ultimately selecting 0.28 as the interval length yielding a success rate closer to 90% (see Figure 12). They explained that the range should be symmetric around the population proportion as follows.

S3: When it was 0.24, it was close, but when it was 0.28, it was closer to 90%, right? It's 88%[2] now with 0.24. But if we go with 0.28, it'll be around 91%.
S4: 28? What about 26?
S3: At 0.28, it's about 90%. From 183 to 146.
S4: (Referring to *P* – 0.12 and *P* + 0.12) This is 24, right? So, 183 plus 146?
S3: So that makes 329, right? So, it'll be 91%.
S4: 91 is better than 88, right? So wouldn't 0.26 be better?
S3: It's 0.28. It increases by 0.04 each time. It's increasing by 0.02 on each side.

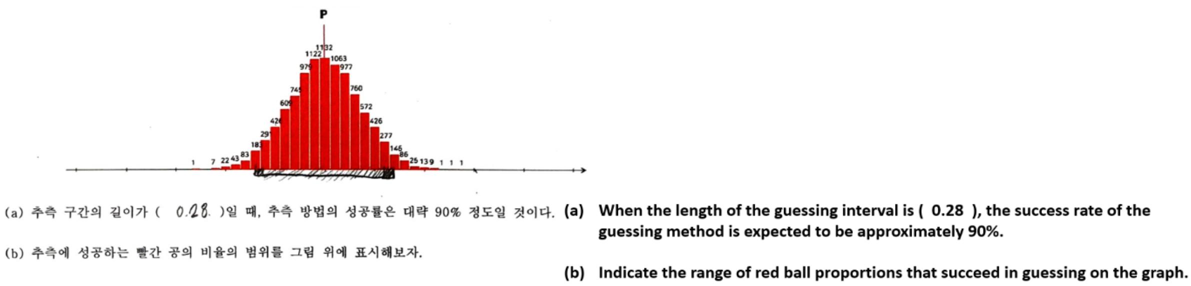


Figure 12. S3's Response to Task 4

Meanwhile, three students (S1, S2, S5) determined the range of sample proportions whose intervals contained the population proportion by using the approximate confidence level of 79% corresponding to the confidence interval length of 0.16 along with the empirical sampling distribution. This approach indicates that they employed proportional reasoning (Song et al., 2023) by using relative frequencies to estimate the probability of sample proportions within a specific range. However, they attempted to reach the 90% target by asymmetrically expanding the existing interval [*P* – 0.08, *P* + 0.08] by an additional 11% in relative frequency, prioritizing numerical accumulation over the consideration of the margin of error. For example, S1 set the range as '*P* – 0.12 or greater and *P* + 0.1 or less' (see Figure 13). S1's markings extended six bars to the left of P and five bars to the right of P, indicating an asymmetric range around the population proportion. This asymmetry suggests that S1 was accumulating relative frequencies to approach 90%, rather than using the margin of error as a symmetric criterion.

[2] When calculating the success rate for the interval length 0.24, S3 made an error by not adding the relative frequencies at $P - 0.12$ and $P + 0.12$ (0.291 and 0.277, respectively), resulting in an estimated success rate of 88%. However, this can be regarded as a simple mistake, as he correctly calculated the success rate for the interval length 0.28 by adding the relative frequencies at $P - 0.14$ and $P + 0.14$ (0.183 and 0.146, respectively) to the previous success rate of 88%, resulting in a final success rate of 91%.

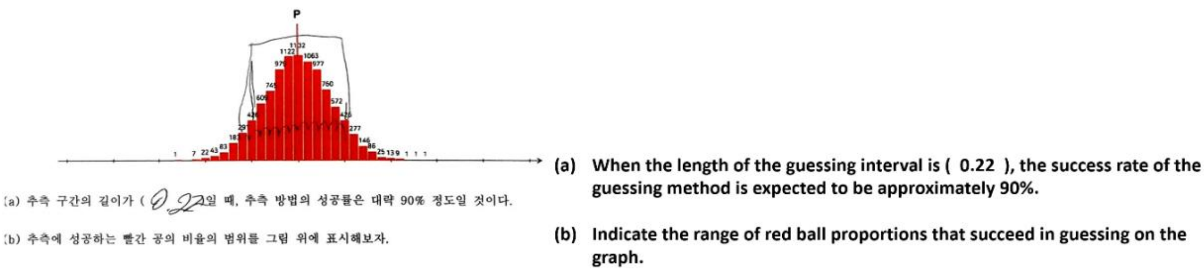


Figure 13. S1's Response to Task 4

Activity 5: Reflection on the reasoning process Activity 5 provided students with opportunities to reflect on their reasoning about the range of sample proportions whose corresponding intervals contain the population proportion. Students determined the approximate confidence level corresponding to a given confidence interval length, x (x = 0.1 for Group 1; x = 0.2 for Groups 2 and 3; x = 0.3 for Group 4), and explained why (Task 5-1). Furthermore, students explained the relationship between the confidence interval length and the confidence level (Task 5-2). Table 6 summarizes the concepts-in-action and theorems-in-action identified in Task 5.

Table 6. Concepts-in-action and theorems-in-action identified in Task 5

Concepts-in-action and Theorems-in-action		Description
Intended	||*Sample proportions within the margin of error from the population proportion*|| (n = 8; S3–S10) <*An approximate confidence level corresponding to a given confidence interval length can be obtained by calculating the relative frequency of the sample proportions within the margin of error from the population proportion.*> (n = 8; S3–S10)	Students determine the range of sample proportions whose corresponding intervals contain the population proportion as those falling within the margin of error from the population proportion. In addition, students interpret the relative frequency of the sample proportion within that range as an approximate confidence level corresponding to the confidence interval length. (e.g., "Adding up all the numbers within this range to calculate the proportion.")
	<*Considering the shape of the empirical sampling distribution, it is possible to estimate the change in the confidence level as the confidence interval length varies.*> (n = 10; S1–S10)	Students explain the relationship between the confidence interval length and the confidence level based on the shape of the empirical sampling distribution. (e.g., "If it becomes excessively large, the increase eventually flattens out. The larger it gets, the higher the probability becomes. As it grows, it could theoretically 0.999…., but it does not seem to be directly proportional.")
Not intended	||*Sample proportions near the population proportion*|| (n = 2; S1, S2) <*An approximate confidence level corresponding to a given confidence interval length can be obtained by calculating the relative frequency of sample proportions that fall within a range near the population proportion, where the width of the range matches the given confidence interval length.*> (n = 2; S1, S2)	Students determine the range of sample proportions whose corresponding intervals contain the population proportion as a range near the population proportion, not symmetric around it, but equal in width to the confidence interval. In addition, students interpret the relative frequency of the sample proportions within that range as an approximate confidence level corresponding to the confidence interval length. (e.g., "With 0.56 ~ 0.66 as the interval for guessing, adding the corresponding values and dividing by the total.")

In Task 5-1, students' strategies for determining the range of sample proportions whose confidence intervals contain the population proportion were categorized into two groups. The intended concept-in-action and theorem-in-action were identified in Groups 2, 3, and 4 (n = 8, S3–S10). They interpreted the relative frequency of sample proportions that fell within the margin of error from the population proportion as an approximate confidence level corresponding to the given confidence interval length. For example, S10 set 0.45 and 0.75 as the boundaries of the range, given a margin of error of 0.15 from the population proportion of 0.6 and calculated the relative frequency of sample proportions within this range (see Figure 14). This response illustrates the use of empirical sampling distribution as a 'model for' quantifying the uncertainty of the estimation method.

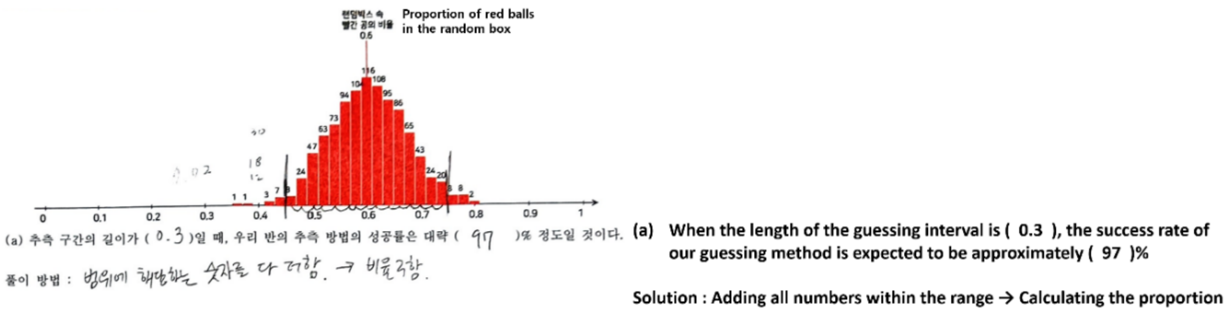


Figure 14. S10's response to Task 5-1

On the other hand, Group 1 students (n = 2; S1, S2) set the range of sample proportions whose corresponding intervals contained the population proportion from 0.56 to 0.66, aiming for a difference of 0.1 around the given population proportion of 0.6. They then calculated the relative frequency within this range to determine an approximate confidence level (see Figure 15). However, they still did not incorporate the margin of error when determining the range of sample proportions whose intervals contained the population proportion.

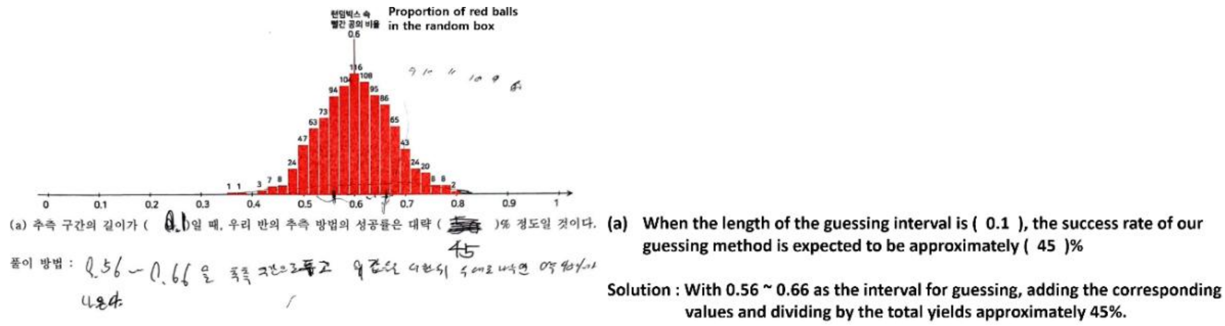


Figure 15. S2's response to Task 5-1

In Task 5-2, the intended theorem-in-action was identified across all students. Students explained the relationship between the confidence interval length and the confidence level in their own words, considering the shape of the empirical sampling distribution (see Figure 16). In this context, students used the empirical sampling distribution as a 'model for' explaining the relationship between the margin of error, which is a measure of the variability of sample statistics, and the uncertainty of the estimation method (confidence level). For example, S5 wrote that "The longer the guessing interval, the higher success rate of the guessing method. If the confidence interval length is too long, it will lose its significance." In the subsequent whole-class discussion, S5 agreed that confidence level increased as confidence interval length increased, but not in a proportional manner; rather, the rate of increase gradually leveled off. S5's written response, together with his agreement during the whole-class discussion, suggests that he was reasoning about changes in confidence level in a way that was compatible with the shape of the empirical sampling distribution, although his written response alone did not explicitly articulate this basis.

T: What is the relationship between interval length and success rate? As the interval length increases..
S5: The wider the interval, the higher the probability.

S8: Probability increases.
T: Is it a proportional relationship?
S3: No. In increases rapidly at first, but then the increase becomes more gradual.
T: What do the others think?
S9: It is not directly proportional.
T: Not directly proportional? So, if the interval length increases by 0.1, does the probability also increase by 0.1?
S5: No. That's not the case.
S3: It rises sharply at first and then gradually levels off. (Several students, including S5, nodded in agreement.)

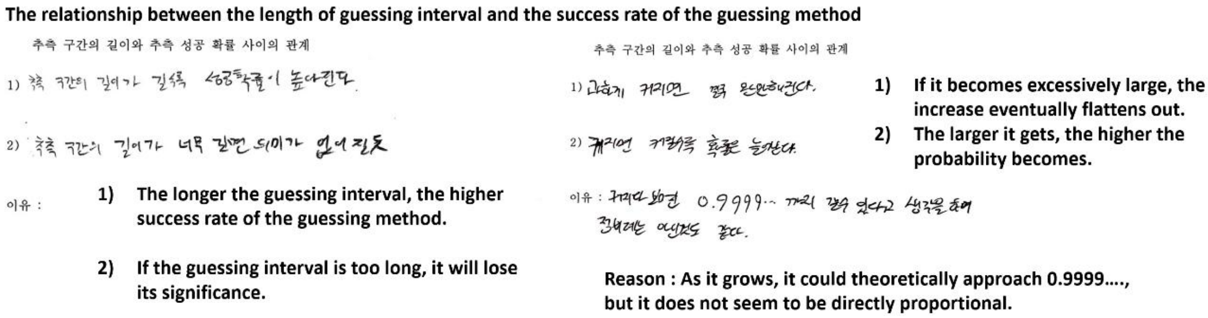


Figure 16. S5 (left) and S7 (right) responses to Task 5-2

5. DISCUSSION AND CONCLUSION

This section presents syntheses and interpretations of the key findings from the teaching experiment in relation to prior research on ISE. Guided by the conjectured LIT, the instructional sequence foregrounded (a) empirical sampling distributions, (b) simulations for repeated estimation, and (c) opportunities for students to reflect on and refine their reasoning. Across the sequence, a majority of the ninth-grade student-participants developed a descriptive model of repeated estimation outcomes and progressively transformed it into a model for quantifying sampling variability and uncertainty. Our analysis identified salient concepts-in-action and theorems-in-action, along with instructional supports that likely facilitated this progression.

We first discuss how students' concepts-in-action and theorems-in-action emerged and evolved across the instructional sequence and the instructional supports associated with the evolution. We then discuss the theoretical implications of these results for understanding students' ISE and the practical implications for designing and enacting instruction that leverages empirical sampling distributions and simulation-based reasoning. Next, we address limitations of the present study, including constraints related to the participants, setting, and scope of the teaching experiment, and we outline directions for future research to examine the transferability and refinement of the proposed instructional approach across diverse contexts. Finally, we present the conclusion of this study.

5.1. SUMMARY AND INTERPRETATION OF EMPIRICAL FINDINGS

Development of 'model of' and transition to 'model for' among students (n = 8) who achieved the key learning goals of the LIT Among the ten students, eight exhibited reasoning that was broadly aligned with the key learning goals of the conjectured LIT. This section synthesizes the development of the descriptive model of repeated estimation outcomes and the transition to the model for quantifying uncertainty and sampling variability, highlighting the instructional supports that accompanied each modeling stage.

In the 'model of' development stage, the eight students developed an image of the range of sample proportions by observing repeated estimation outcomes and predicting the confidence level corresponding to a given confidence interval length (Task 1). To describe the situation of repeated statistical estimation, they used the concept-in-action ||*Success or failure of the interval estimation based on a sample proportion*||. This categorization went beyond the sampling variability band image developed in earlier stages and served as a foundation for understanding statistical estimation as a stochastic process (Ibid). In the subsequent Task 2, students determined the range of sample proportions whose confidence intervals contained the population proportion from the empirical sampling distribution. The resulting ranges can be regarded as a descriptive 'models of' the outcomes of repeated statistical estimation.

Simulation-based activities during the 'model of' development stage likely supported students' reasoning in the following ways. The repeated estimation simulation enabled most students to conceptualize the complex processes of statistical estimation as objects of thought. This finding is consistent with prior studies suggesting that simulations can enhance students' objectification of repeated random sampling processes (Song et al., 2023; van Dijke-Droogers et al., 2020). Moreover, incorporating informal and dynamic visualizations within the simulation (Pfannkuch et al., 2012; Song et al., 2023) enabled students to distinguish individual sample proportions based on the success or failure of interval estimation. This distinction may have encouraged students to infer boundaries for sample proportions whose confidence intervals contained the population proportion and contributed to the development of an image of variability. During the 'model of' development stage, the simulation allowed students to safely explore the situation, generate and test conjectures, and engage in discussions, evaluations, and critiques of various ideas (de Beer et al., 2017; Inzunza, 2018).

In the transition to the 'model for' stage, the concepts- and theorems-in-action used by the eight students gradually evolved to become increasingly aligned with formal statistical concepts. In Task 3, we anticipated that students would determine the range of sample proportions whose confidence intervals contain the population proportion by comparing the difference between the population proportion and individual sample proportions to the margin of error 0.08. However, most students (n = 7) did not consider the margin of error as a criterion for determining whether confidence intervals corresponding to individual sample proportions contained the population proportion. The concept-in-action for these students was ||*Sample proportions near the population proportion*|| rather than the intended concept-in-action, ||*Sample proportions within the margin of error from the population proportion*||.

During the whole-class discussion on Task 3, the teacher identified a prevalent error in the students' reasoning: most students focused only on whether sample proportions appeared close to the population proportion, without considering the margin of error as a critical criterion. The teacher prompted students to revisit their reasoning by focusing on the explicit difference between the population proportion and individual sample proportions. The students' reasoning appeared to shift following this instructional move. In the subsequent Task 4, students recognized that sample proportions whose corresponding intervals contained the population proportion were not merely 'near' the population proportion but must lie within the margin of error from it. This shift may be interpreted as a progression toward structuring the situation of 'repeated statistical estimation outcomes' in terms of mathematical relationships. These students determined a confidence interval length that would yield an approximate confidence level of 90% by considering the relative frequency of sample proportions within the margin of error from the population proportion. In doing so, they began to use the empirical sampling distribution as a 'model for' quantifying sampling variability and uncertainty in the context of statistical estimation. In Task 5, these students articulated in their own words how the confidence level changes as the confidence interval length increases, by considering the shape of the empirical sampling distribution.

In relation to instructional supports that likely facilitated the transition to the 'model for' stage in the LIT, we note the following. Providing students with multiple opportunities to reflect on their reasoning about the empirical sampling distribution (Chance et al., 2004) can promote reasoning about the range of sample proportions whose corresponding intervals contain the population proportion. Crucially, when students' reasoning deviated from the intended statistical idea, the teacher's timely intervention appeared to help redirect students toward more mathematically productive reasoning. In line with the principle of guided reinvention (Gravemeijer, 2004), the teacher supported students in reconstructing the intended concept for themselves by building on their informal strategies and reasoning rather than simply presenting the formal idea. Such instructional support was grounded in the teacher's noticing of pedagogically critical events (Leatham et al., 2015) to provide meaningful opportunities for students' reflection and conceptual refinement.

Emerging conjectures from students' unintended reasoning Abduction is closely tied to the nature of design-based research, which aims to develop new theory (de Beer et al., 2015). Unexpected events can indicate discrepancies between the researchers' prior understanding of students' learning processes and what actually occurs in practice. To resolve this discrepancy, researchers must reexamine their prior assumptions and further consider students' perspectives to generate new explanatory conjectures (Ibid). Accordingly, this section summarizes the unintended concepts-in-action and theorems-in-action identified from the teaching experiment and proposes emerging conjectures related to these findings.

In the 'model of' development stage, it is worth noting that a conceptual shift regarding the range of sample proportions did not occur for Group 1 students (n = 2; S1, S2). In light of our findings, continuing to rely on the 'sampling variability band image' (Pfannkuch et al., 2012) during the 'model of' development stage may hinder the subsequent transition to the 'model for.' While various factors may have contributed to the absence of a conceptual shift regarding the range of sample proportions, one possible explanation lies in the relatively short confidence interval length set by Group 1 students. Group 1 students set the confidence interval length to 0.08, which corresponded to a theoretical confidence level of approximately 56%. This relatively low success rate[3] may have encouraged students to rely on visually prominent clusters they observed across repeated simulation outcomes, leading them to focus on ranges of frequently occurring sample proportions. Students sometimes concentrate on information from technology-generated feedback that is not essential for conceptual development (Olsson, 2018). Our empirical findings suggest that, for a conceptual shift regarding the range of sample proportions, it may be essential to include a phase in which students infer and empirically justify the boundaries of sample proportions whose confidence intervals contain the population proportion. A promising approach could be to provide additional simulation opportunities in Task 2 and explicitly guide students to redetermine the range of such sample proportions. In light of the above, our first emerging conjecture is as follows.

Emerging Conjecture 1 When the confidence interval length is set too short, students may attend primarily to visually salient clusters of sample proportions across repeated simulation outcomes, rather than to the boundaries of sample proportions whose corresponding intervals contain the population proportion. Revisiting Activity 2 with explicit guidance to reexamine the simulation and redetermine the range of sample proportions whose confidence intervals contain the population proportion could facilitate these students' conceptual progression.

It is also necessary to examine why the intended concept-in-action and theorem-in-action were not identified for the majority of students in Task 3. Those students (n = 8) who had focused on the range of sample proportions whose corresponding intervals contained the population proportion in Task 1 were able to determine that range in Task 2 based on their observations. However, in Task 3, the majority of students (n = 7) attempted to determine the confidence level for a confidence interval length of 0.16 by identifying a range of sample proportions near the population proportion and calculating the relative frequencies within that range. They interpreted the range formed by eight neighboring bars (sample proportions) near the population proportion as having a width of 0.16, even though 0.02 actually represented the difference between adjacent sample proportions. The range determined by these students was asymmetric around the population proportion, suggesting that they had not yet fully coordinated their reasoning with the margin of error.

However, a more fundamental cause may lie in the key learning goals and instructional activities of Step 3 in the conjectured LIT. Our empirical findings suggest that while Tasks 1 and 2 supported students in recognizing the range of sample proportions whose corresponding intervals contained the population proportion, they did not sufficiently support students in developing an understanding of the relationship among this range, the population proportion, and the confidence interval length. Notably, only one student (S3) generated an implicit conjecture about the location of the population proportion in Task 2 by marking a point near the center of the identified range of sample proportions (see Figure 8). S3's response opened up the possibility of formulating an emerging conjecture for refining the LIT. As a follow-up to Task 2, students could be guided to conjecture the value of the population proportion based on the range of sample proportions whose confidence intervals contain the population proportion

[3] Groups 2 and 4 set the length of the confidence interval as 0.22; Group 3 set the length as 0.16; and the theoretical confidence levels are 91.68% and 79.23%, respectively.

and to justify their conjectures. Through such an activity, students may come to recognize that individual sample proportions whose intervals contain the population proportion are located within the margin of error from the population proportion. Recognizing this relationship could provide a stronger conceptual foundation for students to use empirical sampling distributions as 'models for' quantifying sampling variability and uncertainty in later activities. In light of the above, our second emerging conjecture is proposed as follows.

Emerging Conjecture 2 To deepen students' understanding of the relationship among the population proportion, the range of sample proportions whose corresponding intervals contain the population proportion, and the confidence interval length in Activity 2, an additional task could require students to conjecture and justify the value of the population proportion.

These emerging conjectures need to be tested through further teaching experiments (de Beer et al., 2015). Such an iterative process can contribute to refining the LIT to better reflect students' ISE learning processes.

5.2. IMPLICATIONS FOR RESEARCH AND INSTRUCTION

The LIT for ISE proposed in this study can be regarded as a "potentially viable theory" (de Beer et al., 2017, p. 463). It enables other researchers and teachers to develop learning trajectories suited to their context. Beyond its role as a theoretical construct, the proposed LIT offers practical guidance for structuring instructional sequences aligned with students' actual learning processes in ISE. Building on this perspective, this section discusses implications derived from our empirical findings.

First, fostering reasoning about the range of sample statistics whose corresponding intervals contain the population parameter provides a promising pedagogical approach for enhancing middle school students' engagement in ISE. Pfannkuch et al. (2012) proposed a long-term learning trajectory in which students beginning at the middle school level developed a sampling variability band image that, at higher grade levels, evolved into an uncertainty band to strengthen their coherent understanding of confidence intervals. Taking this view further, our LIT introduces a middle-school level approach to ISE instruction. Our approach builds on students' sampling variability band image and links it to the reasoning about the range of sample statistics whose confidence intervals contain the population parameter through progressive mathematization. In particular, the ISE activities grounded in our LIT may contribute to students' development of the 'inversion argument' related to confidence intervals, which previous studies (Pfannkuch et al., 2012; Thompson & Liu, 2005) have identified as a core goal of learning statistical estimation.

Second, curriculum developers whose national curricula address statistical estimation (e.g., Australian Curriculum, Assessment and Reporting Authority, 2021; Ministry of Education, 2022) should reconsider when and how empirical sampling distributions are introduced. Statistical estimation is typically introduced only at the end of the high school curriculum, limiting students' opportunities to integrate interconnected concepts underlying confidence intervals. By contrast, a substantial body of research demonstrates that introducing empirical sampling distributions in middle school strengthens students' ISI competence and lays a stronger conceptual foundation (Nilsson & Eckert, 2024; Song et al., 2023; van Dijke-Droogers et al., 2020). These studies consistently show that simulation-based activities exploring empirical sampling distributions foster conceptual growth by enabling students to visualize variability, link theory to data, and engage in authentic inferential reasoning. Building on this body of evidence, we developed and validated a middle school LIT for ISE grounded in empirical sampling distributions. This LIT can inform the development of research-based curriculum material as well as instructional and assessment resources that bridge early conceptual development and advanced statistical reasoning in later grades.

Third, the concepts-in-action and theorems-in-action identified in this study offer practical and evidence-based guidance for teachers in designing and implementing ISE lessons at the middle-school level. Beyond its theoretical contribution, our LIT illustrates how teachers can respond to students' thinking in real time to identify and address conceptual gaps. For example, in Task 3, the teacher recognized a pedagogically critical event when students interpreted 0.02, which represented the difference between adjacent sample proportions in the empirical sampling distribution (see Figure 3), as the width of an individual bar segment. This misinterpretation was directly addressed through targeted feedback during whole-class discussion, prompting students to re-examine their reasoning and align it with more formal statistical concepts. Such moments illustrate how the LIT can support teachers in (a) attending to and interpreting mathematically significant moments, (b) deciding how to respond to achieve lesson objectives or connect with broader mathematical ideas, and (c) anticipating potential misconceptions at specific points in the learning trajectory. Teachers can develop refined instructional strategies for determining which students' ideas to take up as the focus of whole-class discussion by noticing mathematically significant student thinking (Leatham et al., 2015). The concepts-in-action and theorems-in-action identified in this study can serve as a reference framework for predicting students' thinking patterns, conceptions, and productive opportunities for intervention. By incorporating this framework, teachers can design ISE lessons that not only follow a planned trajectory but also adapt responsively to students' reasoning during instruction. As a result, instructional decisions can be both evidence-based and pedagogically timely.

5.3. LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH ON INFORMAL STATISTICAL ESTIMATION

This study has several limitations that may constrain the generalizability and scope of the proposed LIT for ISE. First, the teaching experiment involved a small group of relatively high-achieving, ninth-grade students with strong mathematical communication skills. Thus, the findings may not represent students with diverse achievement levels and competencies. Second, this paper focuses on analyzing the LIT's third and fourth steps, leaving the influence of prior steps beyond the scope of the present analysis. Third, the current LIT does not explicitly incorporate the sample size effect, a critical factor in students' conceptual development in statistical estimation (Pfannkuch et al., 2012).

Future research should examine the LIT's robustness with larger, more heterogeneous student groups and incorporate the sample size effect into the learning trajectory. Such expansions will strengthen the empirical validity and theoretical validity of the LIT, ultimately supporting instructional approaches that better foster students' learning of ISE.

5.4. CONCLUSION

This study developed and empirically examined a conjectured LIT for ISE grounded in the design heuristics of RME. Regarding research question 1, our analysis of ninth-grade students' emergent modeling processes showed how a descriptive 'model of' repeated statistical estimation could evolve into a 'model for' quantifying sampling variability and uncertainty. This progression was manifested in the refinement of students' concepts-in-action and theorems-in-action. It involved a shift from distinguishing between sample proportions whose corresponding intervals contain the population proportion and those whose corresponding intervals do not. Students then began to use empirical sampling distributions to determine the range of sample proportions whose confidence intervals contain the population proportion and to quantify the confidence level based on the corresponding relative frequency.

Regarding research question 2, we identified instructional supports that appeared to foster this transition and informed the refinement of the proposed LIT. Simulation-based activities seemingly supported students' objectification of repeated statistical estimation and enabled them to attend to the boundaries of sample proportions whose corresponding intervals contained the population proportion. Teacher interventions during whole-class discussion appeared to help orient students toward more mathematically productive reasoning. At the same time, unintended concepts-in-action and theorems-in-action, as well as the emerging conjectures, highlighted points where the conjectured pathway requires further refinement. Future teaching experiments with more diverse learners are needed to validate and strengthen the revised LIT, which incorporates the emerging conjectures.

Taken together, our findings illustrate how a research-informed design can make ISE instructionally workable by articulating a learning pathway and the instructional supports through which students can progressively develop the concepts and reasoning that underlie statistical estimation. Nevertheless, further attention and sustained discussion across both research and practitioner communities are needed to advance the understanding and teaching of ISE.

REFERENCES

Australian Curriculum, Assessment and Reporting Authority (n.d.). *Senior secondary curriculum: Mathematics* (Version 8.4). https://v8.australiancurriculum.edu.au/senior-secondary-curriculum/mathematics/

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. [Doctoral dissertation, Utrecht University]. Utrecht University Repository. https://dspace.library.uu.nl/handle/1874/893

Bakker, A., & Gravemeijer, K. P. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147–168). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-2278-6_7

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

Budgett, S., & Rose, D. (2017). Developing statistical literacy in the final school year. *Statistics Education Research Journal*, *16*(1), 139–162. https://doi.org/10.52041/serj.v16i1.221

Büscher, C., & Schnell, S. (2017). Students' emergent modelling of statistical measures: A case study. *Statistics Education Research Journal, 16*(2), 144–162. https://doi.org/10.52041/serj.v16i2.188

Büscher, C., & Prediger, S. (2019). Students' reflective concepts when reflecting on statistical measures—A design research study. *Journal Für Mathematik-Didaktik*, *40*(2), 197–225. https://doi.org/10.1007/s13138-019-00142-2

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295–324). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-2278-6

Cobb, G. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1), Article 1. https://doi.org/10.5070/T511000028

Cobb, P., & Jackson, K. (2015). Supporting teachers' use of research-based instructional sequences. *ZDM Mathematics Education*, *47*(6), 1027–1038. https://doi.org/10.1007/s11858-015-0692-5

Cobb, P., Jackson, K., & Dunlap, C. (2015). Design research: An analysis and critique. In L. D. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (pp. 481–503). Routledge. https://doi.org/10.4324/9780203448946

de Beer, H., Gravemeijer, K., & van Eijck, M. (2015). Discrete and continuous reasoning about change in primary school classrooms. *ZDM Mathematics Education*, *47*(6), 981–996. https://doi.org/10.1007/s11858-015-0684-5

de Beer, H., Gravemeijer, K., & van Eijck, M. (2017). A proposed local instruction theory for teaching instantaneous speed in grade five. *The Mathematics Enthusiast, 14*(1), 435–468. https://doi.org/10.54870/1551-3440.1406

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

Doorman, M., & Gravemeijer, K. (2009). Emergent modeling: discrete graphs to support the understanding of change and velocity. *ZDM Mathematics Education*, *41*(1), 199–211. https://doi.org/10.1007/s11858-008-0130-z

Dvir, M., & Ben-Zvi, D. (2023). Fostering students' informal quantitative estimations of uncertainty through statistical modeling. *Instructional Science*, *51*(3), 423–450. https://doi.org/10.1007/s11251-023-09622-y

Estrella, S., Méndez-Reina, M., & Vidal-Szabó, P. (2023). Exploring informal statistical inference in early statistics: A learning trajectory for third-grade students. *Statistics Education Research Journal*, *22*(2), 1–16. https://doi.org/10.52041/serj.v22i2.426

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A Pre-K–12 curriculum framework*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEPreK-12_Full.pdf

Freudenthal, H. (1991). *Revisiting mathematics education: China lectures*. Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47202-3

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer. https://doi.org/10.1007/978-1-4020-8383-9

Gravemeijer, K. (1999). How emergent models may foster the constitution of formal mathematics. *Mathematical Thinking and Learning*, *1*(2), 155–177. https://doi.org/10.1207/s15327833mtl0102_4

Gravemeijer, K. (2004). Local instruction theories as means of support for teachers in reform mathematics education. *Mathematical Thinking and Learning, 6*(2), 105–128. https://doi.org/10.1207/s15327833mtl0602_3

Gravemeijer, K., & Doorman, M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational Studies in Mathematics*, *39*(1), 111–129. https://doi.org/10.1023/A:1003749919816

Gravemeijer, K., & Prediger, S. (2019). Topic-specific design research: An introduction. In G. Kaiser & N. Presmeg (Eds.), *Compendium for early career researchers in mathematics education* (pp. 33–57). Springer. https://doi.org/10.1007/978-3-030-15636-7_2

Henriques, A. (2016). Students' difficulties in understanding of confidence intervals. In D. Ben-Zvi & K. Makar. (Eds.), *The teaching and learning of statistics: International perspectives* (pp. 129–138). Springer. https://doi.org/10.1007/978-3-319-23470-0_18

Inzunza, S. (2018). Design and evaluation of a hypothetical learning trajectory to confidence intervals based on simulation and real data. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan*. International Statistical Institute. https://icots.info/icots/10/proceedings/pdfs/ICOTS10_1B2.pdf?1531364185

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*(1), Article 1. https://doi.org/10.5070/T521000032

Larsen, S. (2018). Didactical phenomenology: The engine that drives realistic mathematics education. *For the Learning of Mathematics*, *38*(3), 25–29. https://www.jstor.org/stable/26548508

Leatham, K. R., Peterson, B. E., Stockero, S. L., & Van Zoest, L. R. (2015). Conceptualizing mathematically significant pedagogical opportunities to build on student thinking. *Journal for Research in Mathematics Education*, *46*(1), 88–124. https://doi.org/10.5951/jresematheduc.46.1.0088

Lee, J. M., You, J. S., Lee, J. A., & Lee, K. H. (2016). A study on KSAT for assessing statistical literacy by a comparative analysis with SAT. *Journal of Educational Research in Mathematics, 26*(3), 527–542. https://scholar.kyobobook.co.kr/article/detail/4010039257739

Lee, K. H., Yoo, Y. J., & Tak, B. J. (2021). Towards data-driven statistics education: An exploration of restructuring the mathematics curriculum. *School Mathematics, 23*(3), 361–386. https://doi.org/10.29275/sm.2021.09.23.3.361

Lee, S. H., Lee, K. H., & Kim, H. L. (2023). An exploration of restructuring the statistical estimation unit in high school <Probability and statistics>. *Journal of Educational Research in Mathematics, 33*(2), 291–315. https://doi.org/10.29275/jerm.2023.33.2.291

Liu, Y., & Thompson, P. (2007). Teachers' understandings of probability. *Cognition and Instruction, 25*(2–3), 113–160. https://doi.org/10.1080/07370000701301117

Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal*, *4*(1), 27–54. https://doi.org/10.52041/serj.v4i1.524

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

Makar, K., & Rubin, A. (2018). Learning about statistical inference. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 261–294). Springer. https://doi.org/10.1007/978-3-319-66195-7_8

Mayring, P. (2015). Qualitative content analysis: Theoretical background and procedures. In A. Bikner-Ahsbahs., C. Knipping., N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 365–380). Springer. https://doi.org/10.1007/978-94-017-9181-6_13

Ministry of Education. (2022). *2022 revised national curriculum for mathematics* [Notice No. 2022–33]. Government of the Republic of Korea. https://ncic.re.kr/inv/org/list.do?step1=_1&step2=_15&step3=_26

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). National Academies Press.

Moore, D. S. (1992). What is statistics? In D. C. Hoaglin & D. S. Moore (Eds.), *Perspectives on contemporary statistics* (pp. 1–17). The Mathematical Association of America.

Nickerson, S. D., & Whitacre, I. (2010). A local instruction theory for the development of number sense. *Mathematical Thinking and Learning*, *12*(3), 227–252. https://doi.org/10.1080/10986061003689618

Nilsson, P. (2023). Hypothetical learning trajectory on informal hypothesis testing in a probability context. *Statistics Education Research Journal, 22*(2), Article 11. https://doi.org/10.52041/serj.v22i2.425

Nilsson, P., & Eckert, A. (2024). Design principles for simulation-based learning of hypothesis testing in secondary school. *Mathematical Thinking and Learning*, *26*(4) 359–381. https://doi.org/10.1080/10986065.2022.2161288

Olsson, J. (2018). The contribution of reasoning to the utilization of feedback from software when solving mathematical problems. *International Journal of Science and Mathematics Education*, *16*(4), 715–735. https://doi.org/10.1007/s10763-016-9795-x

Pfannkuch, M., Wild. C. J., & Parsonage. R. (2012). A conceptual pathway to confidence intervals. *ZDM Mathematics Education, 44*(7), 899–911. https://doi.org/10.1007/s11858-012-0446-6

Prediger, S., Gravemeijer, K., & Confrey, J. (2015). Design research with a focus on learning processes: An overview on achievements and challenges. *ZDM Mathematics Education*, *47*(6), 877–891. https://doi.org/10.1007/s11858-015-0722-3

Reaburn, R. (2014). Students' understanding of confidence intervals. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* International Statistical Institute. icots.info/icots/9/proceedings/pdfs/ICOTS9_C122_REABURN.pdf

Rossman, A. J. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal, 7*(2), 5–19. https://doi.org/10.52041/serj.v7i2.467

Schnell, S., & Büscher, C. (2015). Individual concepts of students comparing distributions. In K. Krainer, & N. Vondrová (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (CERME 9, pp. 754–760). European Society for Research in Mathematics Education. https://hal.science/hal-01287129/document

Shaughnessy, J. M. (2006). Research on students' understanding of some big concepts in statistics. In G. Burrill (Ed.), *Thinking and reasoning with data and chance. Sixty-eighth National Council of Teachers of Mathematics yearbook* (pp. 77–98). National Council of Teachers of Mathematics.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, *26*(2), 114–145. https://doi.org/10.5951/jresematheduc.26.2.0114

Song, C. G., Lee, S. H., & Lee, K. H. (2023). Middle school students' reasoning about ranges of sample results. *Journal of Education Research in Mathematics, 33*(4), 975–998. https://doi.org/10.29275/jerm.2023.33.4.975

Stephan, M., Bowers, J., Cobb, P., & Gravemeijer, K. P. E. (2003). Supporting students' development of measuring conceptions: Analyzing students' learning in social context. *Journal for Research in Mathematics Education Monograph No. 12*. National Council of Teachers of Mathematics. http://www.jstor.org/stable/i30037716

Tak, B. J., & Jee, Y. M. (2025). A study on the mathematics curriculum in Korea to improve teaching of representative values and dispersion measures. *School Mathematics, 27*(1), 205–227. https://doi.org/10.57090/sm.2025.03.27.1.205

Thompson, P. W., & Liu, Y. (2005). Understandings of margin of error. In G. M. Lloyd, M. Wilson, J. L. M. Wilkins, & S. L. Behm (Eds.), *Frameworks that support research and learning: Proceedings of the twenty-seventh annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Virginia Tech. https://www.pmena.org/pmenaproceedings/PMENA%2027%202005%20Proceedings.pdf van Dijke-Droogers, M., Drijvers, P., & Bakker, A. (2020). Repeated sampling with a black box to make informal statistical inference accessible. *Mathematical Thinking and Learning, 22*(2), 116–138. https://doi.org/10.1080/10986065.2019.1617025

van Dijke-Droogers, M., Drijvers, P., & Bakker, A. (2022). Introducing statistical inference: Design of a theoretically and empirically based learning trajectory. *International Journal of Science and Mathematics Education*, *20*(8), 1743–1766. https://doi.org/10.1007/s10763-021-10208-8

Vergnaud, G. (1996). The theory of conceptual fields. In L. P. Steffe, P. Nesher, P. Cobb, G. A. Goldin, & B. Greer (Eds.), *Theories of Mathematical Learning* (pp. 219–239). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203053126

Vergnaud, G. (1998). A comprehensive theory of representation for mathematics education. *Journal of Mathematical Behavior, 17*(2), 167–181. https://doi.org/10.1016/S0364-0213(99)80057-3

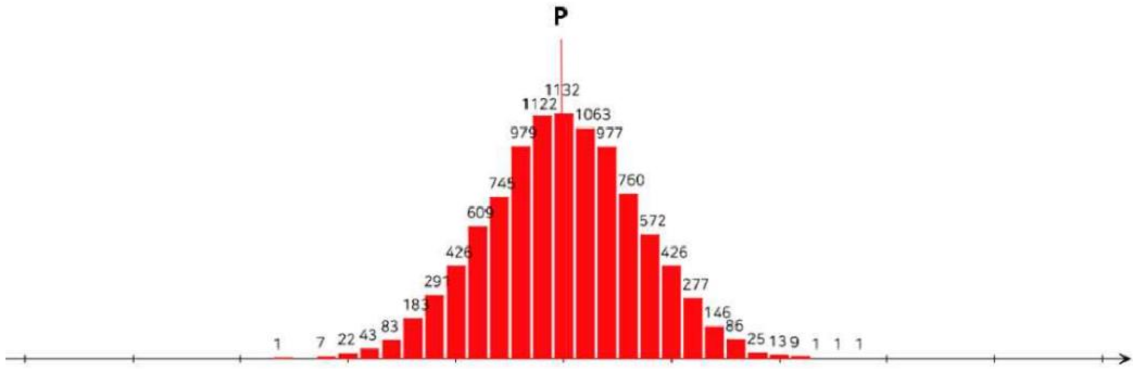
Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40–58. https://doi.org/10.52041/serj.v7i2.469

SONGHEE LEE
Department of Mathematics Education, College of Education, Seoul National University
Bldg 10, RM 204
1 Gwanak-ro, Gwanak-gu
Seoul 08826
Republic of Korea

APPENDIX: TASKS USED IN SESSIONS 3 AND 4 OF THE TEACHING EXPERIMENT

PRELIMINARY TASK

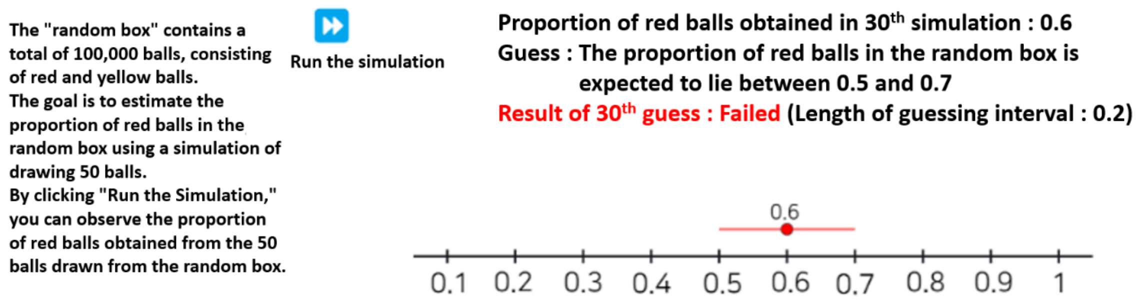
The following figure shows the outcomes of a simulation run 10,000 times, where 50 balls were drawn from a random box and the proportion of red balls was calculated. P represents the proportion of red balls in the entire random box, and the difference between adjacent bars is 0.02.



(a) When running the simulation one more time by drawing 50 balls from the random box and computing the proportion of red balls, determine a range in which the proportion of red balls will be "almost certainly" located, and mark this range on the figure.
(b) If you continue drawing 50 balls at a time and repeatedly check whether your conjecture in (a) is successful, what would the approximate success rate of your conjecture be?

TASK 1

By clicking the "Run the simulation," you can observe the outcomes of Junyoung's guessing method across steps 1–3 for each trial as shown below.



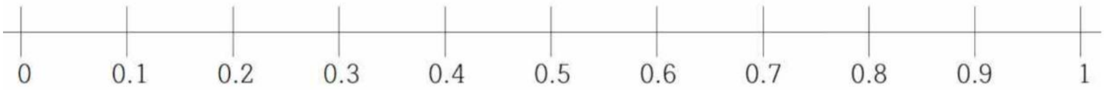
Press "Run the simulation" 200 times in total, one trial at a time, and observe the guessing outcomes. What is the approximate success rate of our team's guessing method according to the length of the interval we set? While observing the simulation outcomes, also describe any additional findings.

When the length of the guessing interval is (   ), the success rate of our team's guessing method will be approximately (   )%. Additional finding 1 :   Additional finding 2 :

TASK 2

Following the teacher's guidance, check the graph showing the outcomes of up to 200 simulations and confirm our team's success rate of the guessing method. Based on this, determine the range of sample proportions of red balls (from drawing 50 balls) whose corresponding interval contain the population proportion, and plot this range on the number line below. How did you determine the range of red ball proportions whose corresponding intervals contain the population proportion?

The range of red ball proportions whose corresponding intervals contain the population proportion will be between (   ) and (   ).

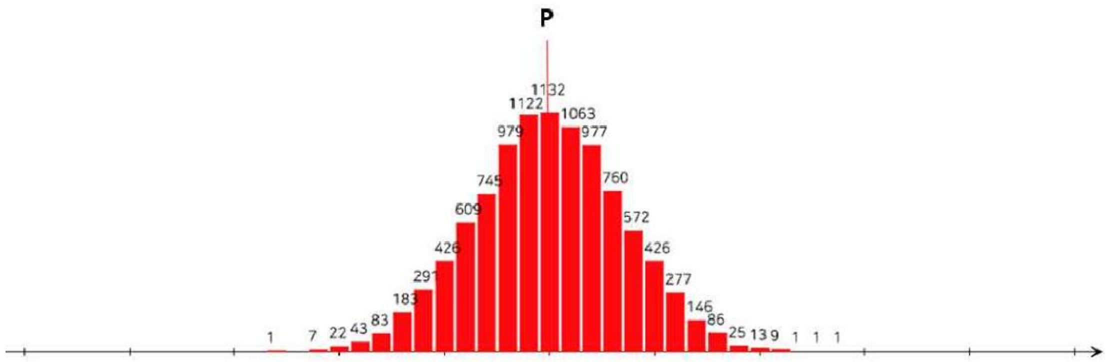


The method for determining the range of red ball proportions whose corresponding intervals contain the population proportion :

TASKS 3 & 4

The following figure shows the outcomes of a guessing simulation run 10,000 times from a random box. P represents the proportion of red balls in the entire random box, and the difference between adjacent bars is 0.02. Answer the following questions.

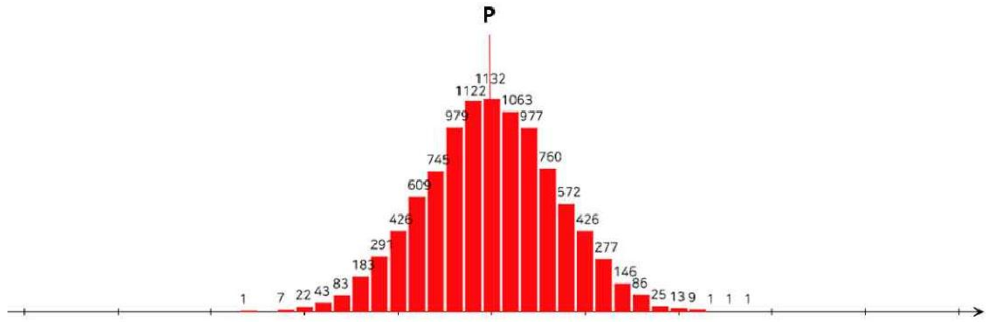
TASK 3



(a) When the guessing interval length is 0.16, indicate the range of red ball proportions whose corresponding intervals contain the population proportion on the graph.

(b) At this time, the success rate of the guessing method will be (   )%.
Solution :

TASK 4

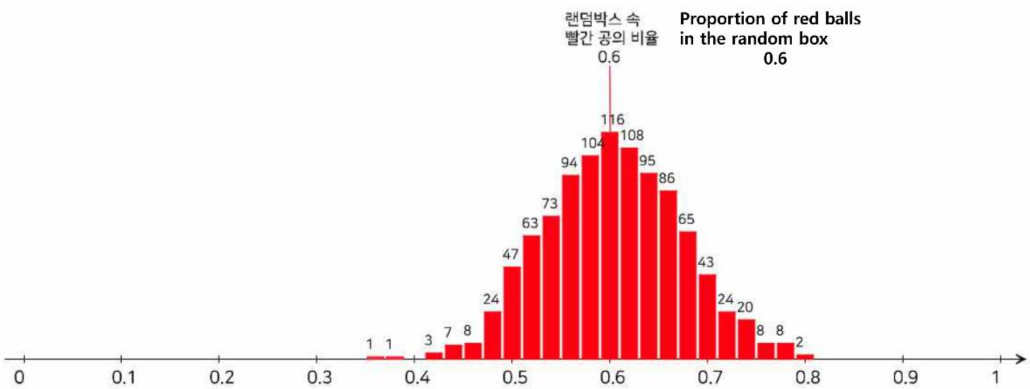


(a) When the length of the guessing interval is (   ), the success rate of the guessing method is expected to be approximately 90%.

(b) Indicate the range of red ball proportions whose corresponding interval contain the population proportion on the graph.

TASK 5

The following graph shows the outcomes of running the guessing simulation 1,000 times from a random box in which the proportion of red balls is 0.6. Answer the following questions.
(a) Estimate the success rate of the guessing method for the given guessing interval length of *x*, and explain how you obtained it.
*Guessing interval length *x* by group: *x* = 0.1 (Group 1, 2); *x* = 0.2 (Group 3, 4); *x* = 0.3 (Group 5)
(b) Explain the relationship between guessing interval length and the success rate of the guessing method.



(a) When the guessing interval length is (   ), the success rate of our guessing method will be (   )%.

Solution :

(b) The relationship between the guessing interval length and the success rate of the guessing method :

Reason :