Journal of Education and Learning (EduLearn)
Vol. 20, No. 4, November 2026, pp. 2723~2732
ISSN: 2089-9823, DOI: 10.11591/edulearn.v20i4.26142

Box plots-based informal inferential reasoning levels in Thai lower secondary students

Dhanachat Anuniwat[1], Supanida Silaba[2]
[1]Graduate School of Humanities and Social Sciences, Hiroshima University, Hiroshima, Japan
[2]Sriayudhya School, Bangkok, Thailand

Article Info

Article history:

Received Sep 19, 2025
Revised Mar 16, 2026
Accepted Jun 8, 2026

Keywords:

Area-represents-frequency schema
Box plot
Informal inferential reasoning
Statistical reasoning
Statistics education

ABSTRACT

This study investigated students' informal inferential reasoning (IIR) levels when interpreting and comparing box plots. This case study purposively selected 36 grade 9 students from a public secondary school in Bangkok, and 29 complete worksheets were analyzed. Using content analysis, the responses were iteratively coded using Pfannkuch's four levels of IIR (levels 0-3). Responses were distributed across all four levels, predominantly at levels 2 and 3. Level 0 responses identified single values without comparing distributions, while level 1 responses relied on everyday visual descriptions with limited statistical evidence. Level 2 responses justified conclusions using a single distributional feature, whereas level 3 responses coordinated multiple distributional features but often reflected area-represents-frequency-schema heuristics. Instructional design should emphasize integrating multiple distributional features to strengthen reasoning. These findings provide guidance for supporting students' IIR through box plot instruction.

Corresponding Author:

Dhanachat Anuniwat
Graduate School of Humanities and Social Sciences, Hiroshima University
Higashi-Hiroshima, Hiroshima, 739-8511, Japan
Email: d245626@hiroshima-u.ac.jp

1. INTRODUCTION

Inferential reasoning, the process of drawing conclusions from data to inform decisions, is a fundamental statistical skill that underpins the development of students' critical thinking [1]–[3]. This form of reasoning can be nurtured through activities that encourage students to informally explore data patterns, generate hypotheses, and justify conclusions about larger populations [4]–[6]. This emphasis is consistent with the aims of teaching and learning statistics in Thailand [7]–[9].

Inferential reasoning has traditionally been taught in higher education through formal methods, such as probability and hypothesis testing [10]–[13]. However, statistics education increasingly emphasizes informal inferential reasoning (IIR) to help students reason from data before learning formal procedures [10]. IIR is viewed as a bridge between descriptive statistics and formal inference, supporting statistical learning [3], [14]–[16]. In this study, IIR refers to the use of sample data to make and justify claims about a broader population or process by attending to distributions, variability, and context without relying on formal inferential methods [10]–[12], [17], [18].

Box plot, introduced by Tukey [19], is widely used in exploratory data analysis to summarize central tendency and variability and to compare groups efficiently [19]. However, its introduction in school education varies internationally (e.g., around age 12 in the United States, around age 14 in Japan, and around age 15 in Thailand and parts of Europe), reflecting its conceptual difficulty [17], [20]–[22]. A meaningful interpretation requires an understanding of the five distributional features (minimum, first quartile (Q1), median, third quartile (Q3), and maximum). However, research consistently shows that students find these ideas challenging [19], [20].

According to Bakker *et al*. [23], students often struggle to interpret the median as representative and may treat quartiles as indicators of data density. Lem *et al*. [24] found that students frequently rely on visual heuristics, such as assuming that wider boxes contain more data or that higher positions indicate greater frequency, rather than using statistical reasoning. These errors reflect not only limited statistical understanding but also the perceptual features of box plots. Lem *et al*. [25] further showed that many students apply the area-represents-frequency schema (ARFS), mistakenly believing that a wider box means more data, reflecting an incomplete understanding of variability and the interquartile range (IQR). More recently, Abt *et al*. [26] identified distinct cognitive profiles in students' box plot reasoning, ranging from conceptual interpretations to hybrid or persistent ARFS-based approaches. Together, these studies suggest that students' difficulties are rooted in underlying conceptual schemas rather than in surface-level misconceptions. Therefore, instruction needs to explicitly confront and restructure these schemas and not only reinforce procedural skills.

In Thailand, box plot is introduced in grade 9, and the curriculum emphasizes distribution, central tendency, and variability [9], [22]. Thus, interpreting and comparing box plots can provide a context for developing IIR. However, research shows that students often rely on superficial visual features or single values with limited distributional and variability-based reasoning [13], [17], [27].

Konold and Pollatsek [28] found that although many students can compute the mean, they struggle to interpret it as representing group tendencies when they view data as a mixture of signal and noise, making informal inference difficult. Pfannkuch [17] similarly reported that New Zealand students often rely on basic numerical comparisons, showing limited attention to distribution and variability. Difficulties with uncertainty also persist among older students. For example, Kazak *et al.* [29] showed that 17-18-year-olds analyzing large datasets often make overly certain claims and rely heavily on contextual knowledge despite underlying uncertainty. Estrella *et al.* [27] found that grade 3 students can generalize beyond observed data and coordinate context with evidence, yet they still struggle to express uncertainty and recognize variability explicitly. Together, these studies suggest that students' IIR is frequently constrained by limited distributional reasoning and weak attention to uncertainty.

Despite these challenges, research shows that IIR can be fostered under supportive conditions. Ben-Zvi [30] demonstrated that scaffolding with growing-sample heuristics in a TinkerPlots environment can support aggregate reasoning, attention to variability, and data-based argumentation. Paparistodemou and Meletiou-Mavrotheris [31] reported that even grade 3 students can formulate and evaluate data-based inferences when supported by dynamic software. Kazak *et al.* [32] found that simulations and dialogic talk helped 11-year-olds move from intuitive judgments toward more probabilistic reasoning. Henriques and Oliveira [33] showed that tasks involving question posing, data collection, and analysis can foster grade 8 students in generating informal inferences and using probabilistic language. Collaborative settings also appear beneficial, as peer interaction can deepen IIR by drawing on shared experiences and collective reasoning [6], [34]. This development can begin from the earliest years of schooling [6], [35]‒[37].

Taken together, prior studies show persistent difficulties in using distributional features such as the median, IQR, and spread to justify informal inferences, even when students engage with rich tasks and technological tools. These difficulties are especially visible when students interpret and compare graphical representations of distributions, including box plots. However, most of this research was conducted in a Western context, and comparatively little is known about how Thai lower secondary students engage in IIR. This gap matters because box plot is introduced relatively late in Thailand and are intended to support reasoning about variability and distribution [9], [22].

Therefore, this study investigates how Thai lower secondary students interpret and compare box plots in real-world contexts. Using Pfannkuch's IIR framework [17], we examined the extent to which students demonstrate different IIR levels and how they use distributional features to justify their claims. The research question is: To what extent do Thai lower secondary school students demonstrate different levels of IIR when interpreting and comparing box plots? This study provides an empirical examination of students' IIR in Thailand. By applying an established IIR framework, it offers insights that may help refine theoretical understanding and inform educational practices in statistics education.

2. METHOD

2.1. Design and participants

This study was grounded in an interpretivist qualitative paradigm, which assumes that knowledge is constructed through human meaning-making rather than existing as an objective fact [38]. This philosophical stance aligns with the aim of the study [39] in interpreting students' reasoning processes rather than measuring quantitative outcomes [38], [40]. Accordingly, a qualitative case study design was adopted to provide an examination of how students level IIR when interpreting box plots [41], [42]. This design was appropriate because diagnosing IIR requires interpreting the structure of students' reasoning and not merely evaluating correctness. A qualitative case study also supports an in-depth analysis of written responses to identify patterns of reasoning across IIR levels.

The case study comprised 36 grade 9 students from a public secondary school in Bangkok, who were selected through purposive sampling. This school was chosen because the students had learned a statistical investigation approach since grade 7. This approach emphasizes posing statistical questions, collecting and organizing real data, representing distributions, and interpreting results to make evidence-based claims [43]. As a result, by grade 9, students were expected to have developed a classroom culture and habits of data-based reasoning, making this setting theoretically and contextually appropriate for diagnosing IIR levels and examining students' statistical reasoning.

Participants were 36 students aged 14-15 years. This sample size was appropriate for qualitative interpretive inquiry [42] because the purpose was not to generalize the findings to all Thai students but rather to capture the level of students' IIR. Through written responses and iterative analyses, the sample enabled the identification of distinct reasoning patterns and IIR levels.

2.2. Instrument

The research instrument was a worksheet titled Comparing Internet Speeds, developed by the first author and adapted from Okamoto *et al.* [44]. The task was revised to reflect the Thai educational context and align with students' cognitive levels, following the design principles of problem situations [45], [46] to strengthen its content validity, as shown in Figure 1. To further establish content validity, the instrument was reviewed by a focus group consisting of three mathematics education experts, two experienced secondary mathematics teachers, and one mathematics teacher educator. Based on their feedback, the context was refined to be more authentic and relevant to students' experiences, while the task instructions and conditions were clarified and streamlined prior to classroom implementation.

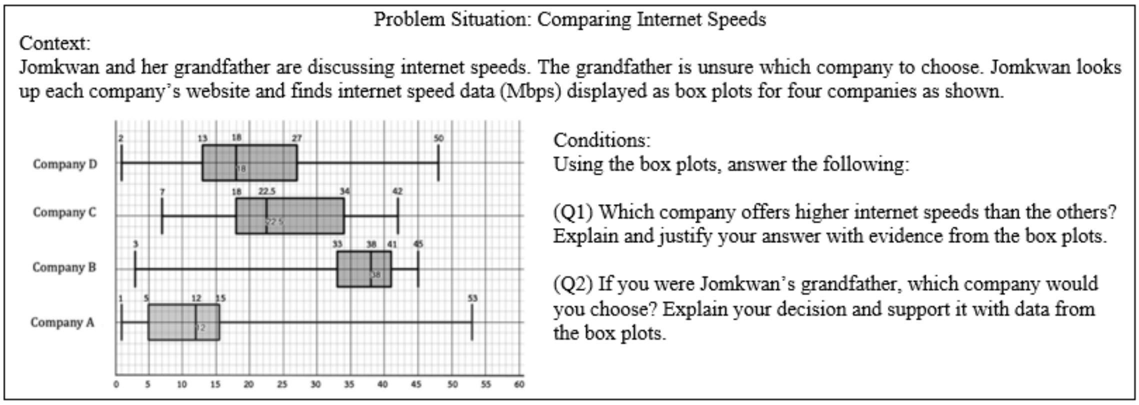

Figure 1. Worksheet, comparing internet speeds, adapted from Okamoto *et al.* [44]

2.3. Data collection

Data were collected using a cross-sectional design [47] during the first semester of the 2024 academic year. The "comparing internet speeds" worksheet was administered during a regular mathematics lesson by the class teacher (the second author) to maintain naturalistic classroom conditions and reduce anxiety. The first author attended only as a non-participating observer and had no prior teaching relationship with the students [48]. Before data collection, the researchers obtained ethics approval and informed consent from the school, teacher, and students, and anonymized all responses in accordance with educational research ethics guidelines [49], [50].

Of the 36 worksheets collected, 29 contained complete written responses that were retained for analysis. This sample size is appropriate for a case study, where the goal is to examine reasoning in depth within a bounded case rather than to make broad generalizations [42]. During coding, no new IIR categories appeared, indicating that data saturation had been achieved. The students were from a regular grade 9 class following the Thai national curriculum, representing a typical range of abilities and learning experiences. This number of responses, therefore, allowed for detailed coding and meaningful interpretation of students' IIR.

2.4. Data analysis

Data were analyzed using qualitative content analysis [51], which suited the interpretation of written student responses. Data were managed and coded using NVivo (version 14.23). Students' responses were segmented by question, yielding 58 analyzable items. These were categorized according to Pfannkuch's [17] IIR framework (level 0-3, as shown in Table 1), ranging from single-value reading to integrated distributional reasoning. First, the authors jointly reviewed Pfannkuch's framework to establish a shared conceptual understanding and ensure interpretive consistency. Next, each author independently read the full database to become familiar with the responses. Coding was then conducted collaboratively, with discrepancies discussed and resolved through consensus. For example, the response from item 5 (question 1) "Company A, because it has the highest maximum," was coded as level 0 because it relied on a single isolated value (the maximum) rather than a comparison of distributions.

Trustworthiness was further enhanced through iterative analysis [9]. The dataset was coded twice, one month apart (July and August 2024), to confirm the stability of the emerging categories and reinforce the reliability of the interpretations. The results are presented descriptively and reflect the systematic patterns of students' IIR.

Table 1. Level of students' IIR

Level	Description
0	Students are able to read single values from a graph (e.g., identifying the median as 200) and may perform simple calculations (e.g., IQR). However, they are not yet able to explain or compare data in terms of the overall distribution represented in the box plot.
1	Students begin to compare data by treating the box plot primarily as a picture, providing descriptions in everyday language. At this level, students lack clear connections to statistical meanings.
2	Students interpret the shape of the box plot in statistical terms, moving beyond surface-level descriptions. They use language that connects the graphical representation with concepts of data distribution.
3	Students use statistical language fluently and integrate evidence from multiple aspects of the box plot to form judgments. They are able to weigh evidence appropriately and use it in making decisions.

Source: authors' analysis (2025).

3. RESULTS AND DISCUSSION

3.1. Results

As the purpose of this study was to examine grade 9 students' IIR when interpreting and comparing box plots, 58 written responses were analyzed. Students' reasoning was distributed across all four IIR levels, with most responses concentrated at levels 2 and 3. The distribution of responses across levels is summarized in Figure 2.

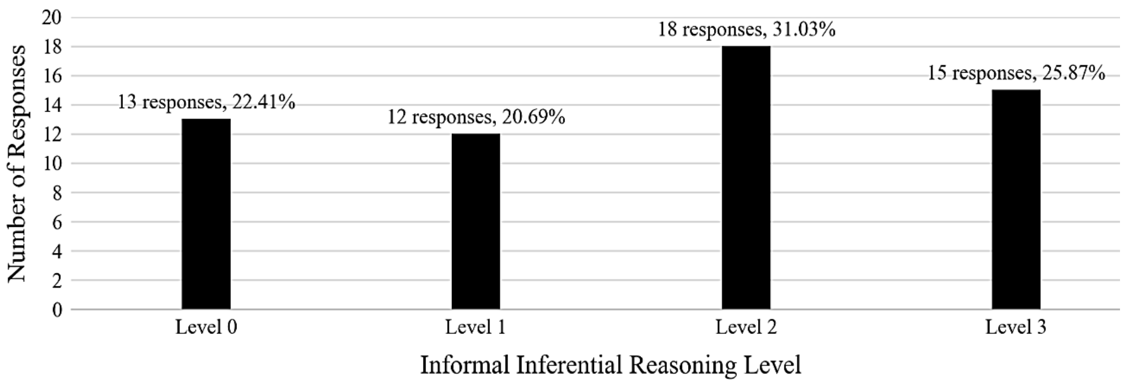

Figure 2. Summary of IIR levels

3.1.1. Level 0 of students' IIR

A closer look at level 0 responses showed that students could identify single values from box plots (e.g., maximum, minimum, quartiles, and IQR). However, when asked to explain their reasoning on the worksheet, no interpretation or comparison of the overall distribution was made. Examples of level 0 responses are presented in Table 2

Table 2. Level 0 of students' IIR

Item	Question	Students' written
5	Q1	"Company A, because it has the highest maximum."
45	Q1	"Company C because the data is the most distributed."
12	Q2	"Choose B because it is the most stable."
46	Q2	"Choose C because the company offers speed that is fast and reliable. For using the service, it gives confidence to use this company, as the service will be more efficient than other companies."

For Q1, reasoning was often based on a single, easily identifiable value, such as the maximum (item 5) or vague references to spread (item 45). For Q2, justifications were mostly non-statistical, relying on general impressions of stability (item 12) or subjective ideas of reliability (item 46). Overall, level 0 responses showed that students relied on isolated values or personal impressions with little ability to combine the features of the distribution.

3.1.2. Level 1 of students' IIR

At level 1, students began to compare the data but relied primarily on visual impressions rather than statistical reasoning when interpreting box plots and describing what they observed in everyday language. Their links to distributional evidence were generally unclear. Level 1 responses appeared only for Q1. Some students referred to the shape or position of the box plots (item 31), but their reasoning remained largely descriptive. Others mentioned the maximum value (items 13 and 37), yet these observations were not integrated into comparative arguments across all companies. Overall, level 1 reflects a transition from reading individual data points to making descriptive comparisons, although students' reasoning remained grounded in visual impressions rather than evidence drawn from the distribution. Representative examples of level 1 responses are presented in Table 3.

Table 3. Level 1 of students' IIR

Item	Question	Students' written
13	Q1	"Company A is fastest because [we] can see from the maximum."
31	Q1	"Company B because the box is in the fast speed area."
37	Q1	"Company A because the maximum is 53 Mbps."

3.1.3. Level 2 of students' IIR

At level 2, students used statistical terms and began linking everyday language with box plot features, showing a shift toward reasoning with distributional features. In Q1, students identified features such as the median (item 7), IQR, and minimum (item 19), demonstrating an emerging ability to interpret multiple characteristics of the box plot rather than focusing on a single value. In Q2, they associated terms such as "distributed" with stability (items 32 and 52). Although these explanations were sometimes unclear, they suggest progress in connecting statistical reasoning with contextual interpretation. Representative examples of level 2 responses are presented in Table 4.

Table 4. Level 2 of students' IIR

Item	Question	Students' written
7	Q1	"Company B, because it has the highest median."
19	Q1	"Company C because the IQR is widest and the minimum speed is the highest among others."
32	Q2	"Choose Company C because it is the most distributed, meaning there is the possibility that the grandfather can use more stable internet."
52	Q2	"Choose Company C because it is the most distributed, the most stable."

3.1.4. Level 3 of students' IIR

At Level 3, students used statistical language more fluently and combined multiple features of the box plots, center, spread, and location on the scale to make decisions. Some students also performed calculations to support these claims. Examples are presented in Table 5.

These responses showed that students could coordinate evidence, weigh a high maximum speed against the position of the IQR (items 50), explicitly compare companies, and sometimes use calculations (item 36). Some misinterpreted the spread, assuming that a larger IQR indicated greater stability (items 17 and 18). Nonetheless, they demonstrated level 3 reasoning by simultaneously drawing on multiple features and balancing the competing evidence.

Table 5. Level 3 of students' IIR

Item	Question	Students' written
17	Q1	"Company C because the IQR is maximum; when the IQR is maximum the internet will have the most stable fast rate."
18	Q2	"I would choose Company C because its IQR is the largest (most stable). If not C, then D may be appropriate because its IQR is next after C."
36	Q2	"I choose Company B because it has the most stable area; the IQRs are 10.5, 8, 16, and 14 for A-D, respectively."
50	Q2	"Choose Company B: although A has the highest maximum, its overall range is largest and the IQR sits in a lower-speed region; B's maximum is lower, but its IQR is in a higher-speed region."

3.2. Discussion

3.2.1. Level of IIR of students

Regarding student levels, among the 58 responses, level 2 was the most common, followed by level 3, level 0, and level 1. Thus, level 2 was the most frequent level. At this level, students could identify and interpret the features of the box plot in statistical terms and begin to connect them to the idea of distribution. Some students reached level 3, in using statistical language more confidently, combining several features, and justifying their choices.

However, many students still find integrating multiple features and weighing evidence challenging, a difficulty that has also been reported in several studies [17], [28], [29]. One possible explanation may relate to students' cognitive development of the box plot. Box plots are initially perceived mainly as pictures. As statistical understanding develops over time, students begin to decode these pictures and eventually make judgments and arguments regarding the relevant data features [17], [24], [26]. Therefore, the links between fluency in decoding box plots, ability to evaluate evidence, and capacity to form reasoned judgments are central to IIR development.

Level 2 responses often mentioned one or two features, such as the median, IQR, or minimum, and linked them to distributional ideas. Progress to level 3 appeared to be limited by two main barriers. The first was feature isolation: students focused on one feature at a time (for instance, "the median is higher") instead of combining center, spread, and location into a coherent argument. To address this barrier, teachers can use structured comparison tasks that prompt students to integrate multiple features when justifying which distribution is preferable. The second barrier was terms such as "distributed" and "stable," which were sometimes misused and caused confusion. For example, some students described a larger IQR as "more stable," although it actually represents greater variability. The common misinterpretation among students at levels 2 and 3 regarding IQR, such as believing that a wider distribution indicates greater stability, is a misconception consistent with findings from previous research. Lem *et al.* [24], [25] showed that students often interpret the width of the box as reflecting stability rather than variability, guided by visual heuristics, whereas Abt *et al.* [26] demonstrated that such reasoning can stem from overgeneralizing an ARFS. This suggests that ARFS is not only a misconception but also a critical mechanism limiting students' progress toward evidence-based comparative judgments.

To reduce ARFS-driven interpretations, instructional interventions should emphasize the conceptual meaning of quartiles. For instance, students can reconstruct a box plot from ordered data by marking the median and quartiles and explicitly stating that each segment represents one-quarter of the data. Pairing box plots with dot plots allows students to verify that a greater visual width reflects spread rather than stability. Additionally, comparison activities can require claims supported by multiple graphical features and explicit reasoning about why an area or width does not represent frequency. These activities shift students from perceptual cues toward statistical meaning. Together, these studies support our interpretation that students tend to prioritize perceptual cues over statistical meaning when interpreting IQR. In our context, this confusion likely stems from curriculum that emphasizes procedural calculation and visual comparison more than conceptual interpretation [8], [9], leading students to inaccurately transfer everyday meanings of words such as "stable" and "spread" into statistical contexts.

In contrast, level 3 responses demonstrated the integration of evidence. These students combined multiple features, sometimes performed calculations, and explained trade-offs; for example, they recognized that one company had a higher maximum speed, whereas another offered greater consistency. Nevertheless, for most students, a major challenge was the lack of prompts that encouraged them to weigh competing evidence or generalize beyond the sample, both essential aspects of IIR. Notably, although some students reached level 3, their reasoning reflected only the intermediate stages of the criteria; more advanced forms of IIR still require further development.

A notable finding was that a substantial proportion of students produced level 3 responses. We do not claim a causal effect. Rather, we provide two theoretically informed interpretations that warrant further empirical testing. First, this class had prior exposure to statistical investigations [43], although often in incomplete cycles. Such exposure may have provided opportunities to reason with distributions and to justify claims from data, which are foundational aspects of IIR. This interpretation is grounded in research conceptualizing IIR as emerging from engagement in investigative practices through which students learn to coordinate evidence and construct arguments that extend beyond the observed sample [12], [17], [18]. From this perspective, even partial opportunities to participate in investigation-oriented activity may support the evidential coordination characteristic of level 3. However, because we did not directly measure individual students' prior experiences or compare classes with different levels of exposure, we present this as a theoretical conjecture rather than an established explanation.

Second, task features likely shaped the observed reasoning. Although the task was contextualized for students and adapted from a problem-solving textbook [44], it did not provide explicit decision criteria or prompts to consider uncertainty or links to the population. Consequently, many students' reasoning remained at the level of descriptive reading of the graphs (level 2). This finding aligns with those of Estrella *et al.* [52], who reported that when tasks do not guide students to weigh the evidence, they tend to focus on describing graphical features rather than making inferences. Taken together, the results suggest that the task supported students' IIR but did not systematically elicit stronger evidence-weighing and generalization. Future work should examine these interpretations through interviews or classroom observations of prior instruction and by comparing classes with different levels of investigation-oriented experience.

Although this study represents a single-school case, the findings suggest transferability to similar contexts in which box plots are taught primarily through computation and visual comparison. The barriers identified feature isolation, everyday-language interference, and ARFS-related visual heuristics align with challenges reported in prior research [17], [23]–[29]. Viewed this way, the study clarifies mechanisms that may constrain students' progress from IIR level 2 to level 3, informing task design and instruction in comparable settings.

In terms of practical implications, these findings suggest several directions for educational practice and policy. For classroom practice, teachers may need to design and implement tasks that explicitly require students to compare competing claims, justify their choices using multiple features (center, spread, and shape), and articulate uncertainty. In addition, teachers may explicitly target ARFS by connecting box plots to previously learned concepts, such as dot plots, and by rebuilding the box plot tasks that foreground quartiles as equal-frequency partitions. For curriculum and policy, the results highlight the importance of embedding opportunities for statistical investigation and IIR within the intended curriculum rather than focusing predominantly on computation and graph reading. Strengthening these aspects may help students progress from descriptive reasoning (level 2) to more integrated and balanced inferential judgments (level 3).

Overall findings are also in line with those of Konold and Pollatsek [28] and Estrella *et al.* [52], which indicates that while most students can compute and understand basic statistical values, they often struggle to interpret what these representations show (in this case, box plots). Similarly, Bakker *et al.* [20] reported that students experienced difficulties in understanding the median and quartiles as representations of a distribution, although they began to use statistical language to connect certain features of the box plots.

3.2.2. Role of the task in eliciting IIR

The task employed in this study as in Figure 1 situated the IIR in a familiar context and encouraged students to discuss features such as center and spread. However, many responses focused primarily on graphical features rather than broader decision-making. Although question 2 asked students to advise an elderly family member, only a few explicitly considered the user's needs, such as tolerance for variation or preference for typical versus extreme speeds. This suggests that future tasks should encourage students not only to analyze data but also to connect statistical evidence with real-world decision-making. This outcome may stem from the underuse of a clear decision frame. Without explicit decision criteria (e.g., "consistency is more important than top speed"), students tended to describe features rather than weigh them [27]. From a pedagogical standpoint, including such decision frames can help students practice evaluating trade-offs, considering uncertainty, and linking sample-based conclusions to the population key elements of IIR.

Task design, therefore, plays a crucial role. As Makar and Rubin [12] emphasized, the development of informal inference should place learners in situations where they need to use probabilistic language (e.g., "it is likely that…"), which provides a foundation for the later use of confidence intervals and hypothesis testing. Furthermore, providing explicit decision criteria [46] and encouraging attention to the population and uncertainty are likely to help students progress from simple feature identification (level 2) to a more integrated and balanced judgment (level 3). In addition, prior studies have shown that digital tools, such as TinkerPlots, the common online data analysis platform (CODAP), and excel, when combined with dialogic talk and growing-sample heuristics, can enrich students' IIR [6], [29]–[33], [53]. These approaches highlight the importance of integrating technological tools, purposeful task design, collaborative settings, and explicit attention to uncertainty, to support the advancement of students' IIR.

4. CONCLUSION

This study used Pfannkuch's four-level framework to investigate the level of Thai lower secondary school students' IIR when interpreting box plots. Level 2 was the most frequent level (18 of 58 responses), showing a core challenge: many students can identify features such as median, IQR, and minimum, but struggle to integrate center, spread, and location into coherent, evidence-based arguments. Typical obstacles included feature isolation, word confusion (e.g., everyday meanings of "distributed" and "stable"), and the misconception that a larger IQR indicates greater stability rather than greater variability. Instructional design should emphasize integrating multiple distributional features to strengthen reasoning.

This study offers three theoretical contributions. It refines the understanding of the shift from level 2 to level 3 in Pfannkuch's framework. It also theorizes ARFS as a constraining cognitive schema in IIR development. Finally, it explains how linguistic and curricular conditions influence the cross-cultural manifestation of IIR levels. Practically, the results suggest that teaching should foster integrated reasoning more deliberately by providing explicit decision criteria, prompting students to weigh multiple pieces of evidence, and situating box plot work in meaningful decision-making contexts. This study was limited by its single-school setting, case study design, and reliance on written responses. Future longitudinal and classroom-based studies using observations and interviews can further clarify students' developmental pathways in IIR and refine the application of the framework in diverse settings.

ACKNOWLEDGMENTS

The authors thank the participants for their cooperation and colleagues for their helpful comments on the manuscript.

FUNDING INFORMATION

This study was supported by JST SPRING (Grant Number JPMJSP2132).

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dhanachat Anuniwat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Supanida Silaba				✓	✓	✓	✓	✓	✓					

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [DA], upon reasonable request.

REFERENCES

[1] J. Garfield, L. Le, A. Zieffler, and D. Ben-Zvi, "Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking," *Educational Studies in Mathematics*, vol. 88, no. 3, pp. 327–342, 2015, doi: 10.1007/s10649-014-9541-7.
[2] T. Baba, "Openness of problem solving in the 21st century: mathematical or social?" in *Proceedings of the 14th International Congress on Mathematical Education*, pp. 17–32, 2024, doi: 10.1142/9789811287183_0002.
[3] J. G. Lugo-Armenta and L. R. Pino-Fan, "An approach to inferential reasoning levels on the chi-square statistic," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 20, no. 1, Art. no. em2388, 2024, doi: 10.29333/ejmste/14119.

[4] A. Bargagliotti *et al.,* "Pre-K–12 guidelines for assessment and instruction in statistics education II (GAISE II): a framework for statistics and data science education," *American Statistical Association*, 2020.
[5] T. Balkaya and G. Kurt, "Middle school students' statistical reasoning about distribution in their statistical modeling processes," *Statistics Education Research Journal*, vol. 22, no. 1, p. 7, 2023, doi: 10.52041/serj.v22i1.312.
[6] S. Saplamidou and C. Sakonidis, "Cognitive and sociocultural aspects of second-grade students' informal inferential reasoning," *Statistics Education Research Journal*, vol. 24, no. 1, p. 6, 2025, doi: 10.52041/serj.v24i1.750.
[7] M. Inprasitha, "A new model of mathematics curriculum and instruction system in Thailand," in *Mathematics Education – An Asian Perspective*, pp. 51–77, 2019, doi: 10.1007/978-981-13-6312-2_4.
[8] D. Anuniwat and Y. D. W. K. Ningtyas, "Intended statistics curriculum at the elementary level: Thailand vs. Indonesia," *Journal of Instructional Mathematics*, vol. 5, no. 2, pp. 108–120, 2024, doi: 10.37640/jim.v5i2.2131.
[9] D. Anuniwat, "Comparing the statistics curricula of Thailand and New Zealand: Structure and concepts," *International Education Studies*, vol. 18, no. 5, pp. 150–161, 2025, doi: 10.5539/ies.v18n5p150.
[10] A. Rubin, J. Hammerman, and C. Konold, "Exploring informal inference with interactive visualization software," in *Proceedings 7th International Conference on Teaching Statistics (ICOTS)*, 2006.
[11] A. Bakker, P. Kent, J. Derry, R. Noss, and C. Hoyles, "Statistical inference at work: Statistical process control as an example," *Statistics Education Research Journal*, vol. 7, no. 2, pp. 130–145, 2008, doi: 10.52041/serj.v7i2.473.
[12] K. Makar and A. Rubin, "A framework for thinking about informal statistical inference," *Statistics Education Research Journal*, vol. 8, no. 1, pp. 82–105, 2009, doi: 10.52041/serj.v8i1.457.
[13] M. G. Tobías-Lara and A. L. Gómez-Blancarte, "Assessment of informal and formal inferential reasoning: a critical research review," *Statistics Education Research Journal*, vol. 18, no. 1, pp. 8–25, 2019, doi: 10.52041/serj.v18i1.147.
[14] A. Weinberg, E. Wiesner, and T. J. Pfaff, "Using informal inferential reasoning to develop formal concepts: analyzing an activity," *Journal of Statistics Education*, vol. 18, no. 2, pp. 1–24, 2010, doi: 10.1080/10691898.2010.11889494.
[15] A. L. Gómez-Blancarte and M. G. Tobías-Lara, "The integration of undergraduate students' informal and formal inferential reasoning," *Educational Studies in Mathematics*, vol. 113, no. 2, pp. 251–269, 2023, doi: 10.1007/s10649-022-10205-w.
[16] Y. Uegatani, H. Otani, and T. Fujita, "Decentralising mathematics: mutual development of spontaneous and mathematical concepts via informal reasoning," *Educational Studies in Mathematics*, vol. 118, no. 2, pp. 229–248, 2025, doi: 10.1007/s10649-024-10366-w.
[17] M. Pfannkuch, "Year 11 students' informal inferential reasoning: a case study about the interpretation of box plots," *International Electronic Journal of Mathematics Education*, vol. 2, no. 3, pp. 149–167, 2007.
[18] M. Pfannkuch, "The role of context in developing informal statistical inferential reasoning: a classroom study," *Mathematical Thinking and Learning*, vol. 13, no. 1–2, pp. 27–46, 2011, doi: 10.1080/10986065.2011.538302.
[19] J. W. Tukey, Exploratory Data Analysis. Reading, MA, USA: Addison-Wesley, 1977.
[20] A. Bakker, B. Chance, L. Jun, and J. Watson, "Working group report on curriculum and research in statistics education," in *Curricular development in statistics education*, pp. 279-282, 2005.
[21] H. Otani, "A study on the development of the statistics curriculum in school mathematics," (in Japanese), Ph.D. dissertation, Hiroshima University, Hiroshima, Japan, 2018.
[22] D. Anuniwat and S. Silaba, "Context in statistical tasks: Insights from lower secondary mathematics textbooks in Thailand," in *Proceedings of the 9th ICMI-East Asia Regional Conference on Mathematics Education (EARCOME9)*, At: Seoul National University, Siheung Campus, Korea, 2025.
[23] A. Bakker, R. Biehler, and C. Konold, "Should young students learn about box plots?" in *Curricular Development in Statistics Education*, Sweden, 2004.
[24] S. Lem, P. Onghena, L. Verschaffel, and W. Van Dooren, "On the misinterpretation of histograms and box plots," *Educational Psychology*, vol. 33, no. 2, pp. 155–174, 2013, doi: 10.1080/01443410.2012.674006.
[25] S. Lem, P. Onghena, L. Verschaffel, and W. Van Dooren, "The heuristic interpretation of box plots," *Learning and Instruction*, vol. 26, pp. 22–35, 2013, doi: 10.1016/j.learninstruc.2013.01.001.
[26] M. Abt, K. Loibl, T. Leuders, W. Van Dooren, and F. Reinhold, "Understanding student errors in comparing data sets with boxplots," *Educational Studies in Mathematics*, vol. 120, no. 1, pp. 169–193, 2025, doi: 10.1007/s10649-025-10387-z.
[27] S. Estrella, M. Méndez-Reina, and P. Vidal-Szabó, "Exploring informal statistical inference in early statistics: a learning trajectory for third-grade students," *Statistics Education Research Journal*, vol. 22, no. 2, pp. 1–16, 2023, doi: 10.52041/serj.v22i2.426.
[28] C. Konold and A. Pollatsek, "Data analysis as the search for signals in noisy processes," *Journal for Research in Mathematics Education*, vol. 33, no. 4, pp. 259–289, 2002, doi: 10.2307/749741.
[29] S. Kazak, T. Fujita, and M. P. Turmo, "Students' informal statistical inferences through data modeling with a large multivariate dataset," *Mathematical Thinking and Learning*, vol. 25, no. 1, pp. 23–43, 2021, doi: 10.1080/10986065.2021.1922857.
[30] D. Ben-Zvi, "Scaffolding students' informal inference and argumentation," in *Working Cooperatively in Statistics Education: Proc. 7th Int. Conf. Teach. Stat. (ICOTS 7)*, 2006.
[31] E. Paparistodemou and M. Meletiou-Mavrotheris, "Developing young students' informal inference skills in data analysis," *Statistics Education Research Journal*, vol. 7, no. 2, pp. 83–106, 2008, doi: 10.52041/serj.v7i2.471.
[32] S. Kazak, T. Fujita, and R. Wegerif, "Students' informal inference about the binomial distribution of bunny hops: a dialogic perspective," *Statistics Education Research Journal*, vol. 15, no. 2, pp. 46–61, 2016, doi: 10.52041/serj.v15i2.240.
[33] A. Henriques and H. Oliveira, "Students' informal inference when exploring a statistical investigation," in *CERME 9 - Ninth Congress of the European Society for Research in Mathematics Education*, Charles University in Prague, Faculty of Education; ERME, Prague, Czech Republic. pp.685-691, 2015.
[34] S. C. Radke, R. Krishnamoorthy, J. Y. Ma, and M. L. Kelton, "Your truth isn't the truth: data activities and informal inferential reasoning," *The Journal of Mathematical Behavior*, vol. 69, Art. no. 101053, 2023, doi: 10.1016/j.jmathb.2023.101053.
[35] M. Meletiou-Mavrotheris and E. Paparistodemou, "Developing students' reasoning about samples and sampling in the context of informal inferences," *Educational Studies in Mathematics*, vol. 88, no. 3, pp. 385–404, 2015, doi: 10.1007/s10649-014-9551-5.
[36] A. Leavy, M. Meletiou-Mavrotheris, and E. Paparistodemou, "Statistics in early childhood and primary education," *Early Mathematics Learning and Development (EMLD),* Singapore: Springer, 2018, doi: 10.1007/978-981-13-1044-7.
[37] S. Estrella, M. Méndez-Reina, R. Olfos, and J. Aguilera, "Early statistics in kindergarten: analysis of an educator's pedagogical content knowledge in lessons promoting informal inferential reasoning," *International Journal for Lesson and Learning Studies*, vol. 11, no. 1, pp. 1–13, 2022, doi: 10.1108/IJLLS-07-2021-0061.
[38] L. Cohen, L. Manion, and K. Morrison, *Research methods in education*. 8th ed, London, UK: Routledge, 2017, doi: 10.4324/9781315456539.
[39] T. S. Kuhn, *The structure of scientific revolutions*. Chicago, IL, USA: University of Chicago Press, 1962.

[40] D. Yadav, "Criteria for good qualitative research: a comprehensive review," *The Asia-Pacific Education Researcher*, vol. 31, no. 6, pp. 679–689, 2022, doi: 10.1007/s40299-021-00619-0.
[41] J. W. Creswell and J. D. Creswell, *Research design: qualitative, quantitative, and mixed methods approaches*. 5th ed, Thousand Oaks, CA, USA: Sage Publications, 2017.
[42] R. K. Yin, *Case study research and applications: design and methods*. 6th ed, Thousand Oaks, CA, USA: Sage Publications, 2017.
[43] C. J. Wild and M. Pfannkuch, "Statistical thinking in empirical enquiry," *International Statistical Review / Revue Internationale de Statistique*, vol. 67, no. 3, p. 223, Dec. 1999, doi: 10.2307/1403699.
[44] K. Okamoto *et al.*, *Gateway to the future mathematics 2*. (in Japanese). Tokyo, Japan: Keirinkan, 2021.
[45] M. Inprasitha, "Lesson study and open approach development in Thailand: a longitudinal study," *International Journal for Lesson and Learning Studies*, vol. 11, no. 5, pp. 1–15, 2022, doi: 10.1108/IJLLS-04-2021-0029.
[46] M. Inprasitha, "Blended learning classroom model: a new extended teaching approach for the new normal," *International Journal for Lesson and Learning Studies*, vol. 12, no. 4, pp. 288–300, 2023, doi: 10.1108/IJLLS-01-2023-0011.
[47] J. R. Turner, "Cross-sectional study," in *Encyclopedia of Behavioral Medicine*, Springer New York, pp. 522–522, 2013, doi: 10.1007/978-1-4419-1005-9_1010.
[48] G. Guest, E. E. Namey, and M. L. Mitchell, *Collecting qualitative data: a field manual for applied research*. Sage Publications, 2013, doi: 10.4135/9781506374680.
[49] British Educational Research Association (BERA), *Ethical guidelines for educational research*. 4th ed, London, U.K.: British Educational Research Association, 2018.
[50] O. F. von Feigenblatt, "Research ethics in education," in *Ethics in Social Science Research*. Springer Nature Singapore, pp. 97–105, 2024, doi: 10.1007/978-981-97-9881-0_7.
[51] P. Mayring, *Qualitative content analysis: a step-by-step guide*. London, UK: Sage Publications, 2022.
[52] S. Estrella, M. Méndez-Reina, R. Salinas, and T. Rojas, "The mystery of the black box: an experience of informal inferential reasoning," in *Advances in Mathematics Education*, Springer International Publishing, pp. 191–210, 2023, doi: 10.1007/978-3-031-29459-4_16.
[53] G. Kurt, "Young children's probabilistic and statistical reasoning in the context of informal statistical inference," *Statistics Education Research Journal*, vol. 22, no. 2, p. 4, 2023, doi: 10.52041/serj.v22i2.434.

BIOGRAPHIES OF AUTHORS

Dhanachat Anuniwat is a doctoral student in Mathematics Education in the Division of Educational Sciences, Graduate School of Humanities and Social Sciences, Hiroshima University, Japan. He earned a Bachelor of Education in Mathematics Education from Khon Kaen University, Thailand, and a Master of Education in Professional Development of Teachers from the University of Fukui, Japan. His current research focuses on mathematics education and statistics education. He can be contacted at email: d245626@hiroshima-u.ac.jp.

Supanida Silaba is a secondary school mathematics teacher in Bangkok, Thailand. She holds both a Bachelor of Education and a Master of Education in Mathematics Education from Khon Kaen University, Thailand. Her current research focuses on the teaching and learning of statistics and probability in school mathematics. She can be contacted at email: supanida.s@sriayudhya.ac.th.