



## Developing University Students' Statistical Reasoning via a Research-informed Course

Binh Tri Huynh<sup>1+</sup>,  
Dung Tran<sup>2</sup>,  
An Tan Thi Nguyen<sup>3</sup>,  
Nga Thi Pham<sup>4</sup>

<sup>1</sup>Quy Nhon University, Vietnam;

<sup>2</sup>Federation University, Australia;

<sup>3</sup>University of Education, Hue University, Vietnam;

<sup>4</sup>PhD student of University of Education, Hue University, Vietnam

<sup>+</sup>Corresponding author • Email: [huynhtribinh@quu.edu.vn](mailto:huynhtribinh@quu.edu.vn)

### Article history

Received: 01 March, 2024

Accepted: 28 March, 2024

Published: 31 March, 2024

### Keywords

Statistical reasoning,  
students, quasi-experimental,  
CAOS

### ABSTRACT

Statistical reasoning (SR) is an essential skill for students in studying statistics. This study aims to develop SR for university students through an appropriately designed course in statistics. Employing a quasi-experimental design, this study evaluated students' SR changes as they participated in a research-informed statistics course, designed to integrate statistical reasoning. The results showed that the course helps students improve their SR, with their scores increased by 23% from the pre-test to the post-test. Specifically, the SR scores on measures of central tendency of the experimental class increased by 8%; while the SR scores on the measures of variability of the experimental class increased by 34%. Suggestions about curriculum and teaching are included, such as offering opportunities for statistical investigations with real data; going deeper into building and understanding formulas; and training lecturers to teach statistics using approaches that are different from traditional teaching methods, as well as evaluating students' statistical learning with reliable tools.

## 1. INTRODUCTION

Statistics is an important subject in the curriculum; however, learners encounter manifold difficulties when learning statistics, especially statistical reasoning (SR). Some research in psychology has underscored the error of "outcome-oriented" thinking, which means using probability models, rather than observing a series of events, to decide whether individual events will occur or not (Konold, 1989). Later work by Konold and colleagues documented inconsistencies in students' reasoning when they answered the same assessment questions, suggesting that the context of a problem can impact students' use (or not use) of reasoning or intuitive strategies (see Konold et al., 1997). Thus, people of all ages, even some experienced researchers, can make erroneous inferences about statistical ideas, and it is difficult to change such misconceptions.

Although SR is a necessary skill for students, there is a scarcity of research on this topic in Vietnam. A study by Nguyen and Vu (2015) shows that high school students' statistical reasoning is still limited, especially regarding the ability to reason with data representation and data distribution. The reason is that teaching and learning statistics in high schools only focuses on computational processes but disregards reasoning. In addition, students have few opportunities to practice statistical survey, so they struggle to apply statistical reasoning to solve problems arising from real life. Another study by Tran (2017) develops questions to probe statistical literacy, reasoning, and thinking of university students in medical and pharmaceutical programs in estimating confidence intervals to design better teaching and assessment activities. In 2021, Hoang also focused on training statistical literacy, reasoning and thinking

for high school students to meet the requirements of the 2018 Mathematics Education Curriculum. Based on students' level of completion, the teacher can differentiate or evaluate their student's level of statistical awareness. Recently, research by Pham and Tran (2023) on developing statistical reasoning for high school students through practical problems shows that a statistics teaching activity can simultaneously develop all three elements of statistical literacy, reasoning and thinking.

In summary, the above studies mainly propose tasks for developing students' statistical reasoning. However, they have not yet used experimental or quasi-experimental methods to evaluate the effectiveness of their programs. In addition, most studies focus on high school students; only few on university students. Therefore, we conducted this study using a quasi-experimental method on university students to evaluate the changes in their statistical reasoning.

## 2. LITERATURE REVIEW

### 2.1. Statistical reasoning

Statistical reasoning (SR) is the way people reason with statistical ideas and make sense of statistical information. Statistical reasoning may involve connecting one concept to another, or it may connect ideas about statistics and probability. Statistical reasoning also means understanding and being able to explain statistical processes and results (Garfield, 2002). Based on the definitions by Garfield and Ben-Zvi (2008) and delMas (2004), Sabbag (2017) summarizes that SR evaluates students' ability to make connections between statistical concepts, create mental representations of statistical problems, and explain relationships among statistical concepts. SR often entails more than one statistical concept and requires students to make connections between them.

### 2.2. Developing students' statistical reasoning

Researchers have long been interested in teaching and learning statistics in undergraduate classrooms (e.g., Zieffler, 2006). Few studies among college students have examined specific activities or interventions; others explore the use of technological tools or teaching methods (e.g., Noll, 2007). Some statisticians who teach statistics have focused on studying student learning in their classes (e.g., Chance, 2002; Lee et al., 2002; Zeleke & Wachtel, 2002; Wild et al., 1997). Most of these studies involve the researcher's own classroom, which may be a single class or a combination of classes within the same school.

#### 2.2.1. Statistical reasoning learning environment

Based on the review of relevant research, Ben-zvi et al. (2018) provide a framework including factors to consider when designing effective and active statistics learning environments that develop learners' statistical reasoning. These principles include:

- Focusing on developing understanding of statistical ideas and concepts rather than presenting statistics as tools and processes;
- Using well-designed tasks to develop statistical reasoning, promoting student learning through collaboration, interaction, discussion and interesting problem solving;
- Using real data to engage students in making and testing conjectures based on data analysis;
- Integrating appropriate technology tools into statistics teaching to allow students to test conjectures, explore and analyze data, and develop statistical reasoning;
- Promoting statistical reasoning and exchanging important statistical ideas in the classroom;
- Using assessments to monitor students' statistical learning and to evaluate instructional plans and progress.

#### 2.2.2. Using technology to develop statistical reasoning

Technology is an important tool in exploring data, performing statistical analysis, and helping learners visualize abstract concepts. Using technology in the statistics classroom can increase learners' time on statistical reasoning by reducing the need to perform calculations or draw graphics. Students can undertake statistical projects and statistical investigations. In addition to data exploration, technology is also used to explore complex statistical ideas or processes through simulation. Using computer simulations can allow learners to visualize relationships. Computer software provides opportunities for students to learn about modeling, allowing them to build their own models to describe data and create explorable simulations (Pratt et al., 2011).

Garfield and Franklin (2011) argue that teachers need to provide resources so that learners can use technology when learning statistics or provide output from computer software when learners cannot access technology, allowing students to learn how to interpret and analyze data, thereby drawing appropriate conclusions in a statistical context.

### *2.2.3. Effects of teaching and learning statistics in groups*

Keeler and Steinhorst (1995), Giraud (1997) and Magel (1998) examined different cooperative learning methods in teaching statistics and found quite positive results. Keeler and Steinhorst (1995) found that when students worked in pairs, the results were often better, and more students continued to study compared to the previous semester. Giraud (1997) found that using collaborative groups to do exercises led to higher test results than lectures. Similarly, Magel (1998) found that implementing cooperative groups in a large class also improved test results compared to classes in the previous semester that did not use group work.

Meletioui and Lee (2002) organized their curriculum according to the Project - Activity - Collaborative Learning - Exercise model, emphasizing statistical reasoning, combined with investigation of conjectures and discovery of results using data. Students are assessed on their understanding at the beginning and end of the course. Increased understanding is observed through tasks that require statistical reasoning, for example whether a data set can be randomly drawn from a particular unit of inquiry.

### *2.2.4. Developing student reasoning in a statistics course*

Zieffler (2006) focuses on the development of students' reasoning about bivariate data through an introductory statistics course. He was interested in three research questions: (1) What is the nature, or pattern of change in students' development in reasoning about bivariate data?; (2) Is the sequencing of bivariate data within a course associated with changes in the pattern of change in students' reasoning about bivariate data?; and (3) Are changes in students' reasoning about the foundational concepts of distribution associated with changes in the pattern of change in students' reasoning about bivariate data? To measure changes in students' reasoning, a scale from the Assessment Resource Tools for Improving Statistical Thinking (ARTIST) project was administered to 113 students in four sections of a course, four times during the semester. Students' reasoning about distributions was also assessed four times during the course using ten items from the Comprehensive Assessment of Outcomes in a First Statistics course (CAOS). Data were analyzed using linear mixed-effects model methodology. He found that most students had developed this type of reasoning before formally studying bivariate data. That result may be explained by the fact that the course developing general statistical reasoning helps them reason well about the distribution of bivariate data before formally studying the topic. He suggests using similar, longer-term studies to model student growth throughout training. Scheaffer (2007) also suggested the same when reporting on statistics of mathematics education research.

In Slauson's study (2008), the impact of hands-on experience compared to the traditional teaching method on the content of measures of variability (standard deviation, sample distribution, standard error, confidence interval and error range) was considered in an introductory statistics course at the undergraduate level. The hands-on experience method focuses on students using existing knowledge to make predictions, collecting, and analyzing data to test predictions, and then evaluating predictions based on the results. The study used the CAOS instrument to evaluate the results before and after the experiment. The results showed an improvement in statistical inference in terms of standard deviation in the experimental group, but not in the control group. However, there was no statistically significant improvement in reasoning about sampling distribution for the experimental group. Besides, there was no improvement in understanding standard errors in both groups and some students scored lower on this topic in the post-test than in the pre-test. Qualitative data analysis shows that it is important for students to understand the relationship between sampling distributions and measures of variability, as well as understand the connection between the concepts of probability and variability, to understand standard errors.

Thus, studies focusing on the teaching and learning of statistics at the undergraduate level continue to point out the difficulties that students encounter in learning and using statistics, as well as suggest some modest successes. These studies also illustrate the problems faced by university statistics instructors such as how to incorporate collaborative and active learning in a large class or how to choose an effective software tool.

## **3. MATERIALS AND METHODS**

### *3.1. Research methods and objects*

This study uses a quasi-experimental method to evaluate the effectiveness of a research-informed course in statistics on changing students' statistical reasoning. The research was conducted on 98 third-year students in two classes, 3A and 3B of the Mathematics Education major, Hue University's College of Education. At the time of the research, the students had not enrolled in any statistics courses at the university level. Their existing statistical knowledge was what they had learned in high schools. The two classes have similar ability and size (Class 3A: 48 students, Class 3B: 50 students), so Class 3B was randomly selected to become the control group, while Class 3A the experimental group.

### 3.2. Research materials

We used the CAOS 4, a product of the statistics education research group (delMas et al., 2006) from the ARTIST project, funded by the US National Science Foundation (NSF). The CAOS instrument is designed to measure students' basic reasoning about descriptive statistics, probability, bivariate data, and basic conclusions of statistics. This tool has been tested three times with 1470 students and has a Cronbach's alpha reliability coefficient of 0.82 (delMas et al., 2007). So, this tool is reliable for assessing important learning outcomes in statistics.

We only selected questions to test statistical reasoning, removing questions about statistical knowledge and keeping the questions most specific to concepts taught in the course. We selected eight questions in the CAOS 4 test and adapted to create appropriate contexts in the questions. Of those eight questions, we converted two questions into open questions that require students to explain the reasons for their choice. These are questions that ask students to explain the relationship between box plots and statistical choices (Question 2), and a given data series and their statistical choices (Question 5) (Figure 1).

**Question 5.** Giang lives about 16km from her university. There are two main possible routes to school, one through the city and one through the countryside. City roads are shorter but have more traffic lights. Rural roads are longer but have only a few traffic lights. Giang records time data (in minutes) after several days of traveling on each route.

Rural roads – 17, 15, 17, 16, 18

City roads – 18, 13, 20, 10, 16

It's important for Giang to get to class on time, but she doesn't want to arrive too early because she'll waste time waiting for class to start. Based on the data above, which road would you advise her to take? Why?

Figure 1. Question 5

Finally, the pre-test and post-test on statistical reasoning includes eight questions corresponding to two content areas:

- Questions assessing SR related to characteristics of measures of central tendency include three questions (3, 6, and 8). These questions require students to recognize the connection between statistical objects: charts and data tables, histograms, and characteristics of measures of central tendency, such as mean, median or mode, thereby making the right choice for the situations raised in the question.
- Questions assessing SR related to characteristics of measures of variability include five questions (1, 2, 4, 5 and 7). These questions require students to recognize connections between statistical objects: column chart, boxplot, data series, and characteristic measures of variability, such as variance, standard deviation, thereby making the appropriate choice for the mentioned situations in the question.

To evaluate students' SR regarding bivariate correlation relationships, we used a real-life problem as an introductory situation at the beginning of the lesson and reused it to test students' knowledge at the end of the third lesson (Figure 2).

**Making Prediction:**

On the People's Public Security electronic newspaper on August 13, 2015, it was written:

There are many different types of evidence that are cataloged and used during scene examination, investigation, and criminal capture. In addition to fingerprints and physical evidence left at the scene such as body fluids and bloodstains, there is other evidence such as footprints, tire tracks and soil samples that can also be used.

Footprints or shoe prints, when discovered at a crime scene, will be photographed, carefully measured and may be required to be made into a plaster mold. Foot size can help investigators determine the subject's gender, height and weight based on the footprint and its shape characteristics. ([Dấu chân, vết lốp xe và mẫu đất lên tiếng trong kỹ thuật hình sự - Báo Công an Nhân dân điện tử \(cand.com.vn\)](#))

VnExpress electronic newspaper on August 19, 2018 also recounted a story about the crime-solving skills of a footprint analyst, including the paragraph:

"Through footprint analysis, Bo Hoa determined that the two suspects who broke into the room were 1m67 and 1m7 tall, and the two suspects guarding outside were 1m72 and 1m75 tall." ([Tài phá án của chuyên gia phân tích dấu chân - VnExpress](#))

**From the data collected in Data Collection Table No. 2\* (teacher provides an Excel file), please answer the following questions:**

Q1. There is a saying: "The longer a person's feet are, the taller they are." Do you agree with this statement? What basis do you use to defend your opinion?

Q2. If three students have foot lengths of 20cm, 22cm, and 24cm respectively, what do you predict their height will be? How did you make this prediction?

\* Data Collection Table No. 2 includes 88 data cases on foot length and height of students taken actually in the classroom.

Figure 2. Situation on bivariate correlation relationships and linear regression equations

### 3.3. Research and data collection process

We conducted the research in five weeks of the first semester in the 2023-2024 school year, within two hours per week. In the first week, we introduced the project and students individually took the pre-test on SR. During the next three weeks we sequentially taught the lessons related to measures of central tendency, measures of variability, and correlation coefficients and linear regression.

We taught the control class in a traditional way, that is, stating definitions and concepts, providing calculation procedures and formulas for students, and stating the meaning of measures using Statistics and Applications textbook by Dang Hung Thang (1999).

For the experiment, the course was designed to develop students' SR, being informed by research such as Benzvi and colleagues (2018) and others (Tran & Tarr, 2018; Tran et al., 2023a, 2023b). In addition, we focused on using visual representations, relationships between two or more statistical objects, and statistical investigations to help students understand statistical concepts and their meanings in real-world situations. In addition, we regularly presented situations that use real data and require students to use software (e.g., Excel) and websites that support statistical calculations (e.g., <https://grapher.nz/>). We divided each experimental and control class into 12 groups of 4 to 5 students to improve the effectiveness of the class.

In the last week, students individually took the post-test on SR with the same eight questions as the pre-test.

Table 1. Experimental teaching plan

Week 1	Week 2	Week 3	Week 4	Week 5
Pre-test (45 minutes)	<i>Experimental teaching</i>			Post-test (45 minutes)
	Measures of central tendency	Measures of variability	Correlation coefficient and linear regression	

### 3.4. Data analysis

To analyze the data, we used quantitative and qualitative analysis methods. Quantitative data are the results of evaluating students' responses to questions. Each response was coded 0 if it was incorrect and 1 if correct. We calculated the average score for each group of questions and for the entire test and then compared the scores of two classes, 3A and 3B. In particular, we compared the pre-test results of the two classes, the pre-test and post-test results of each class, using t-test for two samples on SPSS software.

Qualitative data were collected through students' explanations for the two open questions. For example, in Question 5 asking students to choose a route for Giang, students must know how to use measures of variability to consider the problem, thereby making the correct choice, rural roads. When analyzing students' answers, we found four levels of increasing SR as follows:

- Level 0: incorrect answer or no explanation;
- Level 1: correct answer but no explanation or incorrect explanation;
- Level 2: answer correctly, explain with outliers, variability, certainty, standard deviation, ... but do not calculate the data as a proof;
- Level 3: answer correctly and calculate measures of variability (mainly variance, standard deviation) as a proof.

In addition, we analyzed the work of 12 groups in the experimental class on the tasks about correlation coefficient and linear regression to evaluate their level of development of SR on bi-variate relationships.

## 4. RESULTS AND DISCUSSION

### 4.1. SR based on the pre-test result of the two classes

To examine students' SR of Classes 3A and 3B in the first week, a t-test for two independent samples was performed based on the results of the SR pre-test. The result shows that there is no statistically significant difference in average scores between the two classes ( $t = -0.773$ ;  $p = 0.443$ ). In other words, these two classes had equivalent SR when starting the study (see Table 2).

Table 2. Means of pre-test on SR of two classes 3A and 3B

Description	Class 3A (N=46)	Class 3B (N=50)	p-value
Mean	5.78	5.56	0.443
Standard deviation	2.010	2.012	

This result is also true when comparing the average SR scores of the two classes in each question group (see Table 3).

Table 3. Means of pre-test on SR each SR component

SR component	Class 3A (N=46)	Class 3B (N=50)	p-value
Measures of central tendency	2.59	2.36	0.133
Measures of variability	3.20	3.20	1.000

Thus, the two classes had equivalent SR at the beginning of the study.

### 4.2. Students' SR changes from the pre-test to post-test

#### Experimental class

The mean SR scores on the measures of central tendency of the experimental class increased from 2.59 to 2.8 (maximum score of 3), while its average SR score on the measures of variability increased from 3.2 to 4.3 (maximum score of 5). The SR on the measures of variability of the experimental class increased more than the SR on the measures of central tendency. Specifically, regarding the SR on measures of central tendency, Class A's figure increased by an average of 0.21 points, or 8%. Regarding the SR on the measure of variability, Class A's figure increased by an average of 1.1 points, or 34%. In sum, the overall SR of Class A increased by 1.31 points, or 23%.

Table 4. Means of SR on pre-test and post-test of Class 3A

Experimental class 3A	Mean of pre-test (N=46)	Mean of post-test (N=44)	p-value
Measures of central tendency	2.59	2.80	0.037
Measures of variability	3.20	4.30	0.000
Total	5.78	7.09	0.000

Next, to evaluate the development of students' SR in the experimental class after participating in the three experimental lessons, we used a t-test for two samples to compare the results of the test before and after the experiment. The result shows that there was a statistically significant difference in the SR scores after the study ( $t = 6.235$ ;  $p = 0.000$ ). This improvement is also clearly shown in each group of questions corresponding to each lesson (Table 4).

Thus, with the two designed lessons, the students' SR on measures of central tendency and measures of variability developed and this difference is statistically significant.

#### Control class

In the control class, 3B, the average SR score in each group of questions increased but not statistically significant. The t-test for two samples between the pre-test and post-test on the SR of the control class shows that there is no statistically significant increase ( $t = 1.580$ ;  $p = 0.121$ ) (see Table 5).

Table 5. Means of SR pre-test and post-test of Class 3B

Control class 3B	Mean of pre-test (N=50)	Mean of post-test (N=46)	p-value
Measures of central tendency	2.36	2.59	0.123
Measures of variability	3.20	3.39	0.344
Total	5.56	5.98	0.121

Therefore, it can be concluded that using traditional methods does not help students develop statistical reasoning, that is, they do not know how to connect two or more statistical concepts.

The above results show that while the control class had no change in SR, the experimental class witnessed a statistically significant development in SR. The following qualitative analysis helps illustrate the changes in SR of the experimental class while participating in the research.

#### 4.3. SR on measures of variability in open questions

We also analyzed responses to Question 5 in the SR test, requiring students to choose the correct answer and justify their choice.

The pre-test and post-test responses to Question 5 were coded as shown in Table 6 below:

Table 6. Distribution of answers according to the level of Class 3A

Answer levels	Pre-test (%)	Post-test (%)
<i>Incorrect answer or no answer</i>	14 (30.4%)	10 (22.7%)
<i>Correct answer but no explanation or incorrect explanation</i>	16 (34.8%)	2 (4.5%)
<i>Correct answer, incomplete explanation</i>	16 (34.8%)	27 (61.4%)
<i>Correct answer, correct explanation</i>	0 (0%)	5 (11.4%)
	N=46 (100%)	N=44 (100%)

In addition to the increase in correct answers (from 69.6% to 77.3%), the quality of answers was also improved. Correct answers with incorrect explanations or no explanations were significantly reduced in the post-test. The students knew how to choose the correct answer with a reasonable explanation, although it was incomplete because they did not use data as a proof. In particular, 11.4% of students provided correct answers with accurate explanations.

The following figure provides an example of student VTL's work. This student could not provide an answer in the pre-test. However, when it came to the post-test, this student chose the correct answer, even though the standard deviation of the times taking two different paths had not been calculated.

On average, the rural road takes 16.6 minutes.  
 On average, the city road takes 15.4 minutes.  
 However, the standard deviation of the city road is higher, so I choose the rural road.

*Figure 3. VTL student's answer to Question 5*

In another example, student NQT chose the correct answer in the pre-test but did not provide an explanation. In the post-test, this student was able to provide the correct answer with a complete explanation.

Rural road: median 17, mode 17, standard deviation 1.14.  
 City road: median 16, standard deviation 3.97.  
 I would advise her to choose a rural road because the standard deviation of city road is larger than that of rural road.

*Figure 4. Student NQT's answer to Question 5*

#### **4.4. SR in correlation and regression**

The responses to Q1 at the beginning of the lessons showed that six groups disagreed with the statement (50%), one group did not answer (8.3%), and five groups agreed with the statement (41.7%). The groups who disagreed with the statement mainly used outlier examples to argue that foot length has nothing to do with height.

Q1:  
 - I do not agree with this statement  
 From the data collected from Data Collection Table No. 2 we see:  
 For two students who have 23cm feet, one is 165cm tall, the other is 167cm.

*Figure 5. Group 3's pre-lesson answer to Q1*

The groups who agreed with the statement reasoned either through observation (without supporting data) or by dividing the data about foot length into groups and averaging the data in the groups. The results show that the groups' SR related to correlation and regression was still low at the beginning of the lessons.

- I agree with the statement "The longer a person's feet are, the taller they are".  
 - Proof:  
 The average height of students has a foot length of 19 to 22: 155.05(cm)  
 The average height of students has a foot length of 22.5 to 23: 158.14(cm)  
 The average height of students has a foot length of 23.5 to 25: 164.65(cm)  
 The average height of students has a foot length over 26: 173(cm)  
 ⇒ We see that as foot length increases, average height also increases; they are proportional to each other.

*Figure 6. Group 5's pre-lesson response to Q1*



At the end of the lessons, when answering Q1 again, all 12 groups agreed with the statement and created a linear regression line with a positive coefficient  $a$  to prove their assertion.

Q1: I agree with this statement. I base it on the linear regression equation:

$$\text{Height (cm)} = 2.5684 * \text{Foot length (cm)} + 100.64$$

We see that the more  $x$  increases, the more  $y$  increases.

Figure 7. Group 3's post-lesson answer to Q1

For Q2, seven groups predicted the heights based on observations without any data. The five remaining groups did not provide an answer.

b) The prediction is respectively 155.73; 157.36; 164.65.

Figure 8. Group 5 pre-lesson answer to Q2

However, after finishing the lessons, all 12 groups relied on the linear regression equation to predict heights when knowing foot length.

Foot length	20cm	22cm	24cm
Height	152.008cm	157.114cm	162.2816cm

Because when inputting the foot length into the equation, the height being 20cm, 22cm, 24cm, then  $y$  is 152.008cm; 157.114cm; 162.2816cm respectively.

Figure 9. Group 5 post-lesson answer to Q2

The results show that with carefully designed and appropriate lesson plans, the students' SR related to bivariate relationships was developed. All groups understood the meaning of correlation coefficients and regression equations, thereby being able to provide accurate answers and predictions for specific statistical situations.

## 5. CONCLUSION

The results show that the experimental class, designed to develop SR compared to calculation procedures and formulas approach, improved their competence and this development is statistically significant. Furthermore, the performance of the experimental class was much better than that of the control class. This result corroborates that of Slauson (2008), which confirmed that there was an improvement in SR in standard deviation in the experimental group, but not in the control group. However, the current study focuses on measures of central tendency and measures of variability other than standard deviation. There is also a similar development of SR related to bivariate correlation and linear regression equations. This result further strengthens the findings of Zieffler (2006) that a course or lesson plan that is appropriately and carefully designed helps students develop their SR. Compared with previous studies in Vietnam, the results of this study move beyond proposing learning tasks to develop students' SR by including real data. In addition, this study uses a quasi-experimental method to evaluate the changes of students' SR.

The current research has several limitations. The number of assessment questions is only eight, and other important statistical concepts and relationships between two variables have not been tested. In the future, researchers could expand the study by including more questions to evaluate student performance in other statistical topics and observe the process of changing in other areas more comprehensively.

Based on the findings of this study, we would recommend teaching statistics focusing on developing SR. In other words, courses need to be designed appropriately with careful activities and conducted in groups. Instructors could restrain from teaching statistics by providing calculation procedures and formulas, but instead selecting activities that promote students' curiosity and interest by using more practical scenarios, especially situations that require connecting two or more statistical concepts. Through these situations, students can learn the meaning of concepts. In addition, instructors need to apply statistical software to reduce the burden of calculating and processing data for students and use the software to provide visual images and charts to help students understand statistical concepts. To accomplish these goals, there needs to be synchronous changes. First, there are needed changes in curriculum and

textbooks (cf., Tran et al., 2016; Tran & Tarr, 2018) that offer opportunities for statistical investigations with real data. Second, it is necessary to reduce formulas memorization but focus on building and understanding the formulas. Next, it is necessary to train lecturers to teach statistics with methods that are different from traditional teaching ones. These rely on teacher professional knowledge to teach in a different way (Nguyen & Tran, 2023; Tran & O'Connor, 2023). The above changes will contribute to help students develop their SR, thereby preparing them for today's information-rich society.

**Conflict of Interest:** No potential conflict of interest relevant to this article was reported.

## REFERENCE

- Ben-Zvi, D., Gravemeijer, K., & Ainley, J. (2018). Design of statistics learning environments. *International Handbook of Research in Statistics Education*, 473-502. [https://doi.org/10.1007/978-3-319-66195-7\\_16](https://doi.org/10.1007/978-3-319-66195-7_16)
- Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, 10(3). <https://doi.org/10.1080/10691898.2002.11910677>
- delMas, R. C. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79-96). Kluwer Academic Publishers. [https://doi.org/10.1007/1-4020-2278-6\\_4](https://doi.org/10.1007/1-4020-2278-6_4)
- delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6 (2), 28-58. <https://doi.org/10.52041/serj.v6i2.483>
- Garfield, J. (2002, November). The challenge of developing statistical reasoning. *Journal of Statistics Education*, 10(3). <https://doi.org/10.1080/10691898.2002.11910676>
- Garfield, J., & Ben-Zvi, D. (2008). Preparing school teachers to develop students' statistical reasoning. In C. Batanero, G. Burrill, C. Reading, and A. Rossman (Eds.), *Proceedings of the Joint ICMI/IASE Study on Statistics Education in School Mathematics: Challenges for Teaching and Teacher Education*, Monterrey, Mexico: ITESM. <https://doi.org/10.13140/2.1.3452.5765>
- Garfield, J., & Franklin, C. (2011). Assessment of learning, for learning, and as learning in statistics education. *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study: The 18th ICMI Study*, 133-145. <https://doi.org/10.1007/978-94-007-1131-0>
- Giraud, G. (1997). Cooperative learning and statistics instruction. *Journal of Statistics Education*, 5(3). <https://doi.org/10.1080/10691898.2008.11889557>
- Hoang, L. M. (2021). Rèn luyện hiểu biết thống kê, suy luận thống kê và tư duy thống kê cho học sinh trung học phổ thông [Training statistical literacy, statistical reasoning and statistical thinking for high school students]. *Vietnam Journal of Education Science*, 47, 24-28.
- Keeler, C. M., & Steinhurst, R. K. (1995, July). Using small groups to promote active learning in the introductory statistics course: A report from the field. *Journal of Statistics Education*, 3(2). <https://doi.org/10.1080/10691898.1995.11910485>
- Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6, 59-98. [http://dx.doi.org/10.1207/s1532690xci0601\\_3](http://dx.doi.org/10.1207/s1532690xci0601_3)
- Konold, C., Pollatsek, A., Well, A. D., & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 151-168). International Statistical Institute.
- Lee, C., Zeleke, A., & Wachtel, H. (2002). Where do students get lost? The concept of variation. In B. Phillips (Ed.), *Developing a statistically literate society: Proceedings of the sixth international conference on the teaching of statistics (ICOTS-6)*, Cape Town, South Africa, [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.
- Magel, R. C. (1998, November). Using cooperative learning in a large introductory statistics class. *Journal of Statistics Education*, 6(3).

- Meletiou, M., & Lee, C. (2002, December). Teaching students the stochastic nature of statistical concepts in an introductory statistics course. *Statistics Education Research Journal*, 1(2), 22-37. <http://dx.doi.org/10.52041/serj.v1i2.563>
- Nguyen, D. N., Vu, T. N (2015). Năng lực suy luận thống kê của học sinh trung học phổ thông [Statistical reasoning ability of high school students]. *Vietnam Journal of Education, Special Issue of May*, 162-165.
- Nguyen, T. D., & Tran, D. (2023). High school mathematics teachers' changes in belief and knowledge during lesson study. *Journal of Mathematics Teacher Education*, 26, 809-834. <https://doi.org/10.1007/s10857-022-09547-2>
- Noll, J. (2007). *Statistics teaching assistants' statistical content knowledge and pedagogical content knowledge*. Unpublished doctoral dissertation, Portland State University, Portland, OR.
- Pham, T. Q., Tran, T. (2023). Phát triển năng lực suy luận thống kê cho học sinh trung học phổ thông qua các bài toán có nội dung thực tiễn [Developing statistical reasoning ability for high school students through problems with practical content]. *Vietnam Journal of Education*, 23(6), 1-6.
- Pratt, D., Davies, N., & Connor, D. (2011). The role of technology in teaching and learning statistics. *Teaching statistics in school mathematics-challenges for teaching and teacher education: A Joint ICMI/IASE Study: The 18th ICMI Study*, 97-107. [http://dx.doi.org/10.1007/978-94-007-1131-0\\_13](http://dx.doi.org/10.1007/978-94-007-1131-0_13)
- Sabbag, A., Garfield, J. and Zieffler, A. (2017). Assessing statistical literacy and statistical reasoning: the REALI instrument. *Statistics Education Research Journal*, 17(2), 141-160. <http://dx.doi.org/10.52041/serj.v17i2.163>
- Scheaffer, R. L. (2007). *Using statistics effectively in mathematics education research: A report from a series of workshops organized by the American Statistical Association with funding from the National Science Foundation*. The American Statistical Association.
- Slauson, L. (2008). *Students' conceptual understanding of variability*. Unpublished Doctoral Dissertation, The Ohio State University.
- Tran, D., & O'Connor, B. (2023, online first). Teacher curriculum competence: What is it? *Journal of Curriculum Studies*. <https://doi.org/10.1080/00220272.2023.2271541>
- Tran, D., & Tarr, E. J. (2018). Examination of bivariate data tasks in US high school textbooks through the statistical investigation and cognitive demands frameworks. *International Journal of Science and Mathematics Education*, 16, 1581-1603. <https://doi.org/10.1007/s10763-017-9851-1>
- Tran, D., Nguyen, A., Nguyen, D., Ta, P., Huynh, B., Phan, P., Pham, N., & Nguyen, N. (2023a). Changes in lesson plans as teachers participate in a professional development on statistical literacy. *International Journal of Learning, Teaching and Educational Research*, 22(5), 281-301. <https://doi.org/10.26803/ijlter.22.5.14>
- Tran, D., Nguyen, A., Nguyen, D., Ta, P., Huynh, B., Phan, P., Pham, N., & Nguyen, N. (2023b). Assessing lesson plans for developing students' statistical literacy: a research-informed framework. *Vietnam Journal of Education*, 7(2), 74-81. <https://doi.org/10.52296/vje.2023.227>
- Tran, D., Reys, B., Teuscher, D., Dingman, S., & Kasmer, L. (2016). Curriculum Standards Analysis: An important research area. *Journal for Research in Mathematics Education*, 47(2), 118-133. <http://dx.doi.org/10.5951/jresmetheduc.47.2.0118>
- Tran, T. H. (2017). Phát triển hiểu biết, suy luận, tư duy thống kê của sinh viên y dược trong ước lượng khoảng tin cậy [Developing statistical literacy, statistical reasoning, and statistical thinking of medical and pharmaceutical students in estimating confidence intervals]. *Journal of Science, Hue University*, 126(6), 45-56.
- Wild, C. J., Triggs, C., & Pfannkuch, M. (1997). Assessment on a budget: Using traditional methods imaginatively. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education*. IOS Press.
- Zieffler, S. A. (2006). *A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course*. Unpublished Ph.D. Thesis, University of Minnesota.