JAPANESE AND THAI SENIOR HIGH SCHOOL MATHEMATICS TEACHERS' KNOWLEDGE OF VARIABILITY

MASAMI ISODA
University of Tsukuba, Japan
isoda@criced.tsukuba.ac.jp

SOMCHAI CHITMUN
Tepleela School, Thailand
schitmun@hotmail.com

ORLANDO GONZALEZ
Assumption University, Thailand
ogonzalez@au.edu

ABSTRACT

In this article, the conceptions of variability held by samples of Japanese and Thai senior high school mathematics teachers were identified, based on the framework proposed by Shaughnessy (2007), using a comparative survey study. From contrasting the results of the two groups, relative tendencies of insufficient statistical knowledge for variability were found in both samples, such as a tendency of Japanese teachers to overgeneralize equiprobability, whereas Thai teachers tended to overgeneralize estimation. Based on these findings, the use of well-known tasks from the research literature for this comparative study seems useful to clarify the relative tendencies and insufficiencies in teacher knowledge and conceptions regarding variability held by both groups.

Keywords: *Statistics education research; Statistical literacy; Comparative study*

1. INTRODUCTION

Statistics is necessary to make sense of data sets, samples and distributions (Burrill & Biehler, 2011), and within the field of statistics, variability arises everywhere (Shaughnessy, 2007) and is considered as the heart of statistical thinking (Wild & Pfannkuch, 1999). So, it comes as no surprise that, in recent years, statistics started to play a more prominent role in mathematics curricula in many countries. This has been the case for Japan and Thailand. In Japan, in 2008 and 2009, the junior and senior high school mathematics Courses of Study were revised, prescribing *Analysis of Data* and *Practical Use of Data*, respectively, as one of the strands comprising the subject of Mathematics (Isoda, 2010; Ministry of Education, Culture, Sports, Science and Technology (MEXT), 2008, 2009). In Thailand, the current Core Curriculum, revised in 2009, prescribes *Data Analysis and Probability* as one of the six strands comprising the subject of Mathematics (Royal Thai Ministry of Education (MOE), 2008). Both countries emphasized the treatment of variability through the aforementioned reforms, which was practically non-existent before them. In this context, we can imagine teachers from Japan and Thailand having difficulties in teaching statistics, especially having an inappropriate understanding of variability.

During the first decade of the 21st century, research efforts on variability started to clarify how students of all levels of the educational system thought about variability in several statistical contexts (Shaughnessy, 2007; Shaughnessy & Ciancetta, 2001, 2002; Watson & Moritz, 1999). However, less is known about how teachers think about variability in such contexts. This is somewhat worrisome, because teachers who teach statistics need an appropriate conceptualization of variability to implement the 21st century statistics curriculum.

Statistics Education Research Journal, 17(2), 196–215, http://www.stat.auckland.ac.nz/serj

Under this scenario, we asked the following research questions in order to clarify the specific knowledge needed to conceptualize variability appropriately in Japan and Thailand:

1. How appropriate or inappropriate are the conceptions of variability (in the sense of Shaughnessy, 2007, pp. 984–985) held by senior high school mathematics teachers?
2. What specific knowledge base is necessary to ensure that teachers conceptualize variability appropriately in Japan and Thailand?

In order to provide answers to the research questions of the present study, we chose a comparative survey research design to enhance understanding and to assess and compare the strengths and weaknesses of Japanese and Thai teachers' conceptions of variability as well as find ways to improve their conceptions based on the literature because one can know one's status through comparison with others (e.g., Freedman & Hernández, 1998).

Focusing this study on Japan and Thailand seems appropriate, because both countries, by the time of the study, have enhanced statistics education in their revised school curricula, which were not strong before such revisions, compared to statistics education reforms worldwide. Teachers from both countries may be weak in the treatment of variability. Indeed, teachers from both countries are neither experienced nor familiar with teaching statistics (González, 2011; González, Isoda & Chitmun, 2015; Isoda, 2015; Isoda & González, 2012).

In this research, we anticipate that using Shaughnessy's (2007) framework of conceptions of variability will allow us to identify the weaknesses, strengths, and particular traits of Japanese and Thai senior high school mathematics teachers' theoretical understanding of variability while dealing with situations involving uncertainty, data and chance. This will allow us to identify how well-prepared senior high school mathematics teachers from both countries are regarding variability and to make suggestions for improving the knowledge base of teaching statistics at secondary school level in Japan and Thailand.

2. SURVEY FRAMEWORK

Statistical literacy and statistical problem solving have been emphasized as fundamental skills for the 21st century, and both require mastery of several skills, such as dealing appropriately with variability in both real-life and theoretical situations (e.g., Pfannkuch & Ben-Zvi, 2011; Shaughnessy & Pfannkuch, 2002). Here we focus on the theoretical aspects of variability and statistical knowledge. Cultural matters are outside the scope of this study.

In this section, we define variability, clarify the role of variability to make sense of statistical concepts, and describe the different conceptions of variability identified by Shaughnessy (2007), which will be serve as framework for this study.

2.1. THE NECESSITY OF VARIABILITY TO UNDERSTAND VARIATION

In order to achieve the goal of the mathematics curriculum regarding statistics education, teachers have to be able to understand variability and interpret it appropriately for understanding statistical concepts.

According to many researchers (e.g., Burrill & Biehler, 2011; Shaughnessy, 2007), variability is defined as a characteristic of an entity that is observable, which describes how much variation is present in data and how spread out the data are. Variability is more than measures of variation such as the variance, range and standard deviation. In order to understand theoretical perspectives of variability, we have to go beyond a mathematical treatment and consider uncertainty from an informal or intuitive way. On this regard, Shaughnessy (2007, pp. 984–985) found that students have mainly eight different ways in which they conceptualize variability, all influenced by the statistical context.

2.2. CONCEPTIONS OF VARIABILITY

As mentioned in the introduction of this article, most of the research efforts on conceptions of variability in several statistical contexts have focused primarily on students rather than on teachers, and therefore less is known about how teachers think about variability in such contexts. Moreover, much of what we do know about teachers' conceptions of variability is from isolated studies carried out using differing theoretical contexts (e.g., Batanero, Arteaga, Serrano, & Ruiz, 2014; Makar & Confrey, 2004; Peters, 2009).

Batanero et al. (2014) examined perceptions of randomness in a sample of 208 prospective primary school teachers from Spain by engaging these teachers in comparing random variables and deducing their mathematical properties through random experiments. These researchers identified several naïve conceptions of variability, such as focusing on individual data elements instead of on the aggregate, disregarding sampling variability in random sequences, making incorrect predictions about random experiments involving binomial distributions, and considering variability as equiprobability.

Makar and Confrey (2004) studied four American secondary mathematics teachers in order to carry out research on teachers' statistical reasoning when comparing two distributions of data. They found that participants conceptualized within-group variability as the characteristic of scores within a distribution to be liable to vary; whereas between-group variability was conceptualized as difference between two distributions regarding centers; measures of variation, or shapes.

Peters (2009) examined the conceptions of variability held by 16 American high school Advanced Placement statistics teachers. From the identified conceptions of variability, Peters classified teachers into three types: (a) teachers focusing on the design of the study and seeing variability as something that needs to be controlled; (b) teachers focusing on data analysis and seeing variability as something that needs to be explored; and (c) teachers focusing on inference, seeing variability as something that needs to be modeled and expected. Additionally, Peters found that only 5 out of the 16 teachers in the study showed connected reasoning across the three types of conceptions.

The authors of this article considered that each of these teachers' conceptions of variability can be matched to a particular student conception of variability identified by Shaughnessy (2007). That is the reason why Shaughnessy's framework was considered appropriate for the study in hand.

After a thorough review of the literature on statistics education to date, Shaughnessy (2007, pp. 984-985) identified and provided examples of the following eight contexts in which students might acknowledge variability:

1. *Variability in particular values, including extremes or outliers*: People focus their attention on particular data values (such as the mean, the median or the mode) as pointers, often on very large or very small values, or very strange values in a graph or a data set.
2. *Variability as change over time:* When the data are represented graphically, with time being used as independent variable, and people consider what the overall pattern in the data is, as well as why the data may be varying, they are using this conception.
3. *Variability as whole range- the spread of all possible values:* In this conception, people explore the spread of an entire data set, moving from seeing data only as individual values that vary, towards an aggregate view of data, by recognizing that entire samples of data can also vary.
4. *Variability as the likely range of a sample*: This conception arises in repeated trials of probability experiments, particularly when people consider the likely range of a distribution of sample statistics. This conception requires the concept of relative frequency, and hence relies on proportional reasoning.
5. *Variability as distance or difference from some fixed value:* In this case, people think of variability as an actual or visual measurement from some measure of central tendency or endpoint value to one data point, or a group of points.
6. *Variability as the sum of residuals*: People with this conception associate variability with deviation-based metrics, such as the mean absolute deviation and the sum of residuals or averages of the absolute value differences from either a measure of central tendency or other fixed value.
7. *Variation as a covariation or association*: When interpreting the interaction of several variables, people with this conception think of how changes in one variable may correspond to changes in other variables.
8. *Variation as distribution:* People with this conception relate variability to characteristics of a distribution such as center, spread and skewness, as well as to theoretical probability distributions, in order to make decisions about data.

These conceptions of variability are usually discussed alongside misconceptions. Various studies on statistical misconceptions regarding variability have been carried out, as mentioned by Shaughnessy

(2007). In this study, Shaughnessy's framework was used as a filter to select the tasks for data collection and to characterize the relative differences between Japanese and Thai teachers.

3. RESEARCH METHODOLOGY

In order to elicit the conceptions of variability held by the participants in the present study, a questionnaire comprised of nine tasks was designed. The authors hypothesized that such tasks, chosen from previous research focused on students' conceptions of variability, were also appropriate to assess how teachers conceptualize variability. In this article, due to space limitations, five tasks are discussed, and the other four tasks will be referred to, as needed, to strengthen the discussion.

Each task in the questionnaire was chosen by González (2011) and Isoda and González (2012) from previous studies reported in the literature on statistics education. Tasks that could be solved theoretically were selected. The relation between the chosen tasks and the conceptions of variability that might be elicited by them is shown in Table 1.

Table 1. Conceptions of variability identified by Shaughnessy (2007) that could be applied to answer each of the survey questions

Conception of variability that could be used to answer the task appropriately	Task number
1. Variability in particular values, including extremes or outliers	Task 2, Task 5
2. Variability as change over time	Task 4
3. Variability as whole range	Task 1, Task 3
4. Variability as likely (most probable) range of a sample	Task 1
5. Variability as distance or difference from same fixed point	Task 2, Task 5
6. Variability as sum of residuals	Not surveyed[a]
7. Variation as covariation or association	Task 4
8. Variation as distribution	Task 1, Task 2, Task 3, Task 5

[a] Engagement in a hands-on statistical experiment is recommended for eliciting this conception (Shaughnessy, 2007), which would have been time-consuming for this survey.

The match between conceptions and tasks was done on the basis of the connections required to answer appropriately each task, based on statistical and mathematical knowledge. Examples of how responses to these tasks were analyzed are shown in the discussion sections for each task.

In both countries, all the government senior high schools from specific provinces were contacted. For the Japanese study, Ibaraki Prefecture was chosen because the first and third authors were working together there at that time. For the Thai study, the provinces of Prachinburi, Nakhon Nayok and Sa Kaeo, comprising the Secondary Education Services Area Office 7 (SESAO7), were chosen because the second author was working in one of these provinces at that time. The Japanese study was conducted as part of the graduate thesis of the third author. The second author replicated this study in Thailand four years later, as part of her graduate teacher training program in Japan. Both studies were conducted under the supervision of the first author. All the participating schools were following National Curriculum standards. Participant teachers from each school were asked to voluntarily and anonymously fill in the survey instrument; and the completed questionnaires were then collected.

Following this methodology, 78 mathematics teachers voluntarily participated from 56 out of the 109 public senior high schools in Ibaraki Prefecture (province), Japan, during August 2009. In the case of Thailand, during August and September 2013, 77 mathematics teachers from 41 out of the 44 public senior high schools located in SESAO7 voluntarily participated, while attending a teacher training program organized by SESAO7.

In this article, we will refer to the group of 78 Japanese high school teachers as *Japanese teachers*, and to the 77 Thai senior high school mathematics teachers as *Thai teachers*. In the case of the Japanese teachers in this study, 58 (74.4%) of them took a statistics course when in university. Only one (1.3%) of the surveyed Japanese teachers had taught statistics content every year of teaching, whereas 56 (71.8%) of them had not taught statistics content at all. In the case of Thailand, 66 Thai teachers (85.7%) took a statistics course when in university. Only 11 (14.3%) of the surveyed Thai teachers had taught statistics content every year of teaching, whereas 33 Thai teachers (42.9%) had not taught statistics content at all in the last three years.

Thailand and Japan have a centralized educational system, their own national teacher certification system under a national teacher education system regarding mathematics education, and national curriculum standards. The collected data could be useful to gain insight into the current status of the statistical knowledge of government school teachers in both countries, even though the data were collected from representative provinces.

In order to analyze the data from Japanese and Thai teachers, chi-square tests of independence, Fisher's exact tests a,nd adjusted standardized Pearson residuals (Agresti, 2013) were used.

For interpreting the results obtained in this study in a more critical and reasonable way, partial results were presented, along with our preliminary interpretation and analysis, at conferences of mathematics education societies in Thailand (Isoda, 2015) and Japan (González et al., 2015), in order to produce a critical interpretation of results through the discussion with other mathematics educators in both countries. Finally, based on the received criticism and feedback, the discussion for every task provided here was developed.

4. RESULTS, DISCUSSIONS, AND SUGGESTIONS

The results, discussion and suggestions for every task in relation to addressing the research questions are structured as follows. Firstly, the results from the collected answers were categorized and analyzed using statistical tests for clarifying which categories were significantly and non-significantly different. Secondly, in relation to our research questions, a discussion of such results with significant differences is carried out, considering both the appropriate and inappropriate use of the conceptions of variability identified by Shaughnessy (2007) and the tendencies emerging from the data collected from Japanese and Thai teachers are described. Additionally, suggestions are provided in relation to inappropriate use of the conceptions of variability. The discussion is focused on the relative differences between countries on each task.

4.1. TASK 1: THE TWO-SPINNER TASK

The two fair spinners below are part of a carnival festival. A player wins a prize only when both arrows land on black after each spinner has been spun once.

(a) If we did 10 trials of the game, how many would you expect be winners?
(b) What would you expect to happen if we played the game 10 more times?
(c) Tarō/Malee thinks he has a 50-50 chance of winning. Do you agree with Tarō/Malee that there is a 50-50 chance of winning this game? Explain why or why not.

Figure 1. "Two-spinner" task (adapted from Shaughnessy, Ciancetta, & Canada, 2004)

The purpose of this task was to examine, for each group of teachers, the meaning of mathematical equiprobability (as in the expected value and the equiprobable sample space) and statistical probability (as a likely value or likely range of values) in a problem involving random devices, such as spinners (e.g., Isoda & González, 2012; Shaughnessy, 2007; Watson, Kelly, Callingham, & Shaughnessy, 2003). Watson et al. (2003) implemented this task with the same purpose, but on Australian students in Grades 3, 5, 7, and 9.

Task 1's results For the data in Table 2, Fisher's exact test was used because four cells (40.0%) had expected counts less than 5 (Fisher's exact test = 6.41, $p > 0.05$). From the results, Thai teachers did not differ statistically from Japanese teachers.

Table 2. Frequency, percentage and percentage difference of each kind of answers given by Japanese and Thai teachers to Task 1 item (a)

Category		Japan		Thailand		Difference
		Frequency	%	Frequency	%	JP–TH (%)
C0	No response, "I don't know"	1	1.3	5	6.5	-5.2
C1	Reasonable number (either 2 or 3) with no reason	24	30.8	22	28.6	+2.2
C2	Expected value (2.5), anything can happen (theoretically correct)	25	32.0	15	19.5	+12.5
C3	*Recognition of variability (use of the words "about," "between," "probably," "will be close to 3," "2 or 3")*	*27*	*34.6*	*32*	*41.6*	*-7.0*
C4	Other number (neither 2 nor 3)	1	1.3	3	3.9	-2.6
Total		78	100.0	77	100.0	

Note. C3 is the best answer

For the data in Table 3, Fisher's exact test was used because 4 cells (40.0%) had an expected count less than 5 (Fisher's exact test = 37.675, $p < 0.05$). The adjusted standardized residuals indicated that the Thai teachers significantly differed from the Japanese for categories C1 and C2. We observe that 61.5% of the Japanese teachers answered using the expected value, whereas only 18.2% of the Thai teachers answered in the same way. In the case of Thai teachers, the category with the largest number of teachers was C3, with 37.7% of the teachers using a likely value (e.g., "about 5") or a likely range of values (e.g., "between 4 and 6") in their answers.

Table 3. Frequency, percentage and percentage difference of each kind of answers given by Japanese and Thai teachers to Task 1 item (b)

Category		Japan		Thailand		Difference
		Frequency	%	Frequency	%	JP–TH (%)
C0	No response, "I don't know"	2	2.6	6	7.8	-5.2
C1	Reasonable number (either 4, 5 or 6) with no reason	5	6.4	26	33.8	-27.4*
C2	Expected value (5), anything can happen (theoretically correct)	48	61.5	14	18.2	+43.3*
C3	*Recognition of variability (use of the words "about," "between," "probably," "will be close to 5")*	*22*	*28.2*	*29*	*37.7*	*-9.5*
C4	Other number (neither 4, 5 nor 6)	1	1.3	2	2.6	-1.3
Total		78	100.0	77	100.0	

Note. C3 is the best answer.
*$p < 0.05$, based on an analysis of the adjusted standardized residuals

For the data in Table 4, Fisher's exact test was used because four cells (50.0%) had expected counts of less than 5 (Fisher's exact test = 28.044, $p < 0.05$). The adjusted standardized residuals indicated that Thai teachers significantly differed from Japanese for all categories, with the exception of C1. We can observe that Japanese teachers were able to provide an appropriate reason to disagree with Tarō, such as the expected value or the entire sample space, which are mathematical answers; whereas the number of Thai teachers in the same category was significantly lower. Compared with Thai teachers, Japanese teachers tended to answer using the calculation of the equiprobable sample space, or the exact mathematical probability for disagreeing with the event of Tarō having a 50-50 chance of winning. On item (c), the number of Thai teachers categorized under C2 (disagree but providing either no reason or a wrong answer) was higher than those categorized as C3 (disagree and providing an appropriate reason). From this fact, it is clear that 44.2 % of Thai teachers in this study, although correctly disagreeing with a 50-50 chance of winning, were unable to provide an appropriate reason for such disagreement.

Table 4. Frequency, percentage and percentage difference of each kind of answers given by Japanese and Thai teachers to Task 1 item (c)

Category		Japan		Thailand		Difference
		Frequency	%	Frequency	%	JP–TH (%)
C0	No response, "I don't know"	0	0.0	8	10.4	-10.4*
C1	Agree with Tarō/Malee	2	2.6	7	9.1	-6.5
C2	Disagree with Tarō/Malee, but providing either no reason or a wrong answer (less than ½ chance of winner as there are 2 spinners; chance is 1/3 BB,WW, BW; anything can happen)	17	21.8	34	44.2	-22.4*
C3	*Disagree with Tarō/Malee, with an appropriate reason ("chance is ¼"; "BB,WW, BW, WB"; "½×½=¼")*	*59*	*75.6*	*28*	*36.4*	*+39.2**
Total		78	100.0	77	100.0	

Note. C3 is best answer; The name "Tarō" was used with the Japanese group, whereas "Malee" was used with the Thai group
* $p < 0.05$, based on an analysis of the adjusted standardized residuals

Task 1's discussion and suggestions The results obtained in items (a) and (b) show that more than half of the teachers from both countries were unable to provide a reasonable response acknowledging variability. This fact is worrisome, if we compare with the results reported in Watson et al.'s (2003) study, in which nearly half of the students (46.5%) from grades 3, 5, 7, and 9 were also unable to give a reasonable response that acknowledged variability to a similar task. Regarding item (c), the teachers participating in this study were outperformed by the grades 10–12 students taking advanced mathematics courses in Shaughnessy and Ciancetta's (2002) study. This study reported that 91.2% of these students answered the item correctly, with almost all of them also supplying correct reasoning for their answer.

From the research results obtained in items (a), (b) and (c), we concluded that Japanese teachers tended to answer using equiprobability, in comparison to Thai teachers. In item (b) (see Table 3), we conclude that the largest proportion of Thai teachers answered using a likely value (e.g., "about 5") or a likely range of values (e.g., "between 4 and 6"). The contrast of how the different groups of teachers answered could be explained from the perspective of Shaughnessy's (2007) conceptions of variability. Equiprobability is mathematical probability, which can be categorized as the eighth conception of variability identified by Shaughnessy, as *variation as distribution*. More Japanese teachers may have overgeneralized the use of equiprobability, which is inappropriate, or may have interpreted items (a) and (b) as asking for the expected value, given the responses to item (c), from which one could assume that Japanese teachers are more familiar with probability theory than Thai teachers. However, the nature of the questions in items (a) and (b) points to an overgeneralized use of equiprobability by Japanese teachers, because the Japanese word "expect" in "expected value" (*kitai*) implies to look forward to something you have hope for, and to wait for it, whereas the Japanese word for "expect" in items (a) and (b) (*yosō*) implies just to predict, to anticipate without any feelings, not including to wait for.

Providing answers such as "about 5" or "between 4 and 6" can be categorized as the fourth conception of variability identified by Shaughnessy (2007), as *variability as likely range of a sample*. More Thai teachers used this conception appropriately.

Based on Shaughnessy's (2007) framework, we can say that, when comparing the number of Japanese and Thai teachers using the conceptions of variability identified above, the two groups do not seem to be using the same conceptions of variability in this task. Based on this contrast, we call the Japanese teachers group "equiprobability oriented," whereas Thai teachers group is called "estimation oriented." The data collected from other survey tasks not included here, such as "The gumball machine" task (adapted from Shaughnessy et al., 2004), also supported these differences in orientation.

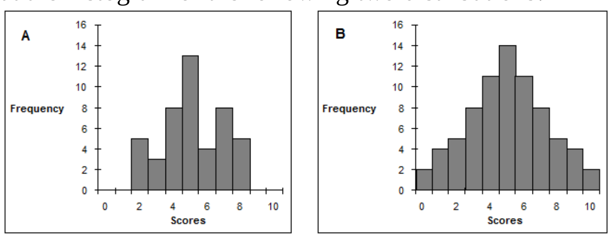
For Task 1 item (c), a large proportion of Japanese teachers provided an appropriate reason using the expected value or the equiprobable sample space. Teachers who answered appropriately and listed correctly the whole equiprobable sample space might have demonstrated the third conception of variability identified by Shaughnessy (2007) (i.e., *variability as whole range*). In comparison to Japanese teachers, fewer Thai teachers were able to do so.

For solving Task 1, we have to prepare teachers to use the conceptions *variability as likely range of a sample*, *variability as whole range* and *variation as distribution* appropriately. However, as a result of the survey, specifically, for the Japanese group of teachers, it is suggested to adapt their equiprobability ideas by using the context of statistical estimation, in order to provide statistically appropriately answers. For the Thai group of teachers, it is suggested to complement their estimations with equiprobability, in order to provide rational and formal explanations.

4.2. TASK 2: “CHOOSING DISTRIBUTION WITH MORE VARIABILITY” TASK

The purpose of Task 2 was to assess teachers’ understanding of variability in the context of comparing data distributions represented with histograms, and to examine whether teachers made inappropriate use of the normal distribution in this context. Meletiou and Lee (2003) implemented this task with a similar purpose for the university freshman students enrolled in an introductory statistics course in an American university.

Look at the histogram of the following two distributions:



Which of the two distributions do you think has more variability?

(a) Distribution A (b) Distribution B

Why do you think this?

Figure 2. “Choosing distribution with more variability” task (adapted from Meletiou & Lee, 2003)

Task 2’s results In Table 5, Fisher’s exact test was used because two cells (33.3%) had expected counts of less than 5 (Fisher’s exact test = 14.160, $p < 0.05$). An analysis of the adjusted standardized residuals indicated that Thai teachers falling into categories C2, C3, C5, and C6 significantly differed from Japanese ones. Considering surveyed teachers who chose Distribution B (which was right), we observed that 59% of the Japanese teachers and 85.7% of Thai teachers were able to make the right choice. However, only those teachers whose answers were classified as C7 (i.e., teachers whose answers evidenced simple recognition of variability by focusing just on the extremes or the ranges of each distribution) and C8 (i.e., teachers whose answers were beyond the ones in category C7 by connecting both extremes and measures of central tendency) can be regarded as answering this task appropriately: 50% of the surveyed Japanese teachers and 54.6% of surveyed Thai teachers. Analyzing the data in Table 5, we observe that, from the teachers who selected Distribution B as an answer, 8.9% of the Japanese and 31.1% of the Thai teachers were either not able to appropriately explain their choice (responses in category C5), or made an inappropriate statistical interpretation of the term *variability* (responses in category C6).

The largest proportion of teachers providing an inappropriate answer to this task was observed, in the case of the Japanese participants, for responses classified as C3, whereas in the case of Thai participants, the largest proportion of inappropriate answers was observed for responses classified as C5. In the case of category C3, 21.8% of the Japanese teachers and 3.9% of Thai teachers who selected Distribution A as the one with more variability, provided reasons such as “Distribution A isn’t symmetrical,” “Distribution A wasn’t beautiful,” and “Because Distribution A was not close to a normal distribution,” which are misconceptions related to symmetry and the normal distribution. These types of responses focus on visual cues in the graph, and not on the actual data values of the distributions, spreads of data values, or measures of central tendency and variation. In other words, a large proportion of Japanese teachers inappropriately interpreted variability as how much the given histograms fit either symmetrically or visually a normal distribution; that is, these teachers interpreted variability as an adaptation of the image of the normal distribution (i.e., bell-shaped, single-peaked, and symmetrical) to the given histograms.

Table 5. Frequency, percentage, and percentage difference of each kind of answer given by Japanese and Thai teachers to Task 2

Category		Japan		Thailand		Difference
		Num	%	Num	%	JP–TH (%)
C0	No response, "I don't know" "no idea"	2	2.6	1	1.3	+1.3
C1	Distribution A, giving no reason, just guessing, or by arguing intuitive ideas.	2	2.6	4	5.2	-2.6
C2	Distribution A, based on a mistaken calculation.	3	3.8	0	0.0	+3.8*
C3	Distribution A, based on an inappropriate interpretation related to symmetry or a poor fit to a normal distribution.	17	21.8	3	3.9	+17.9*
C4	Distribution A, based on arguments related to differences in the heights of the bars (e.g., "Distribution A is bumpier"; "the number of different heights in A is higher than in B").	8	10.3	3	3.9	+6.4
C5	Distribution B, giving no reason, just guessing, or by arguing intuitive ideas.	3	3.8	13	16.8	-13.0*
C6	Distribution B, by inappropriate interpretation (e.g., "B has a larger span in frequency than A," "B because is symmetrical," "B because is normal," "B because has more elements").	4	5.1	11	14.3	-9.2*
C7	*Distribution B, based on arguments related to simple recognition of variability (i.e., answers concerned only with extremes or the ranges of each distribution; e.g., "because it's more spread out").*	*25*	*32.1*	*31*	*40.3*	*-8.2*
C8	*Distribution B, based on arguments related to sophisticated recognition of variability, more than C7 (i.e., answers connecting both middles and extremes; e.g., "because the scores differ more from the center").*	*14*	*17.9*	*11*	*14.3*	*+3.6*
Total		78	100.0	77	100.0	

Note. Distribution B is the correct answer;
* $p < 0.05$, based on an analysis of the adjusted standardized residuals

In the case of category C5, 3.8% of the Japanese teachers and 16.8% of Thai teachers chose Distribution B as the one having more variability; however, they were unable to provide reasons to justify such a choice. On the evidence, this choice might have been made intuitively.

Task 2's discussion and suggestions The results obtained in Task 2 revealed that some Japanese (10.3%) and Thai (3.9%) teachers incorrectly regarded "bumpiness" of a histogram as "variability." This confusion was found, in a greater degree (26%), in the freshman students participating in Meletiou and Lee's (2003) study.

Based on the significant differences and findings identified on Task 2, a relative comparison between Japanese and Thai secondary school mathematics teachers is presented. On this task, we observed the overgeneralization in the use of the normal distribution by some Japanese teachers (i.e., those teachers' responses classified as C3 and many of the teachers classified as C6). This characteristic of the Japanese sample is an addition to their equiprobability orientation, which was identified from their performance in the previous tasks. So, from analyzing the results of Task 2, Japanese teachers tended to be extremely oriented towards theoretical probability distributions (in particular, the normal distribution).

We also observed that more Thai teachers were not able to provide a theory-based or appropriate answer to support their choice of Distribution B (i.e., those teachers classified as C5). Additionally, we also observed that Thai teachers' orientation towards estimation is not only used when thinking of outcomes of probabilistic experiments (as in Task 1), but also when thinking of measures of variation, such as the range and the standard deviation (as in Task 2). For example, many Thai teachers whose answers were classified as C7 compared the ranges and the standard deviations without calculating them numerically. Instead, they compared such measures qualitatively, describing the degree of spread of either the extremes or the data in relation to the central bar.

Shaughnessy's (2007) framework provides some explanations about these ways of interpreting variability. Teachers who interpreted variability as either a visual or actual measurement of the distance of each or some elements of a dataset from some measure of center (e.g., those answers categorized as C7), showed evidence of holding the fifth conception of variability identified by Shaughnessy (2007) (i.e., *variability as distance or difference from some fixed point*).

Teachers who interpreted variability as the degree of collective dispersion of the data from a center (e.g., those who answered that Distribution B has more variability because bars are distributed farther away from the center in Distribution B than in Distribution A in C8), provided a qualitative measure of the total variability of the entire distribution of data. According to Shaughnessy (2007, p. 985), thinking of variability in this way provides the foundation for statistical concepts such as standard deviation. Hence, teachers answering in this way showed evidence of holding the sixth conception of variability identified by Shaughnessy (2007) (i.e., *variability as the sum of residuals*).

About 22% of Japanese teachers overgeneralized the normal distribution and misconnected the beauty of symmetry and variability (C3), which is an inappropriate interpretation of the features and parameters of a theoretical probability distribution according to the conception *variation as distribution*.

About 17% of Thai teachers chose the right distribution without providing any reason, which is non-rational, and inappropriate for the conceptions *variability as distance or difference from some fixed point* and *variability as the sum of residuals*.

For solving Task 2, teachers need to be prepared to use the conceptions *variability as distance or difference from some fixed point, variability in particular values, including extremes or outliers,* and *variation as distribution* appropriately. Specifically, for the Japanese group of teachers, it is suggested to do more in-depth study about the meaning and appropriate use of normal distribution, and particularly its symmetry, in relation to variability. For the Thai group of teachers, it is suggested to study measures of variation in more depth, in order to provide appropriate reasons to support their choice.

Teachers who provided answers focusing on visual cues in the graph, and not on the actual data values of the distributions, spreads of data values, or measures of central tendency and variation (e.g., answers evidencing misconceptions related to symmetry and the normal distribution, such as those categorized as C3), are considered to be holding the conception we label as *variability as visual cues in the graph*; a way of thinking about variability that is not accounted for by Shaughnessy's (2007) framework (Shaughnessy, personal communication, July 19, 2013). Hence we would add this conception to his framework.

4.3. TASK 3: "TWO HOSPITALS" TASK

The purpose of this task (see Figure 3) is to assess whether teachers understand the Law of Large Numbers, as well as their reasoning about samples in order to understand sampling variability. This task leads to the use of the Law of Large Numbers, as well as the binomial distribution (similarly as in Task 1). Figure 4 illustrates the use of the binomial distribution to answer this task.

Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80% or more female births?

(a) Hospital A (with 50 births a day).
(b) Hospital B (with 10 births day).
(c) The two hospitals are equally likely to record such an event.
(d) There is no basis for predicting which hospital would have that percentage of female births.

Figure 3. "Two hospitals" task (adapted from Garfield & Gal, 1999)

Sample size (*n*)	Probability of 80% or more female births (*X*) means…	$P(X \geq 80\%)$
10	8 ~ 10 female babies	.0546875 (≈5.5%)
50	40 ~ 50 female babies	.0000119 (≈1.2×10^{-5} %)

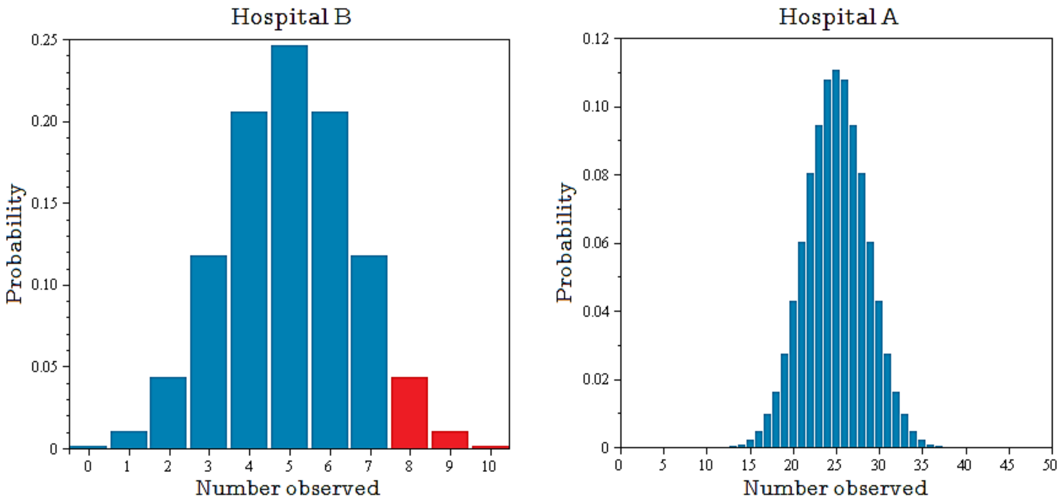


Figure 4. Cumulative probability for having 80% or more female births in hospitals A and B (red-shaded bars, which are not even visible in the case of Hospital A), using the binomial probability function

Because Task 3 sets the condition that "half of all newborns are girls and half are boys," it is assumed that, regardless of the hospital, each time a baby is delivered there is a 50% chance of it being a boy or a girl. The question for this task is the following: "Which hospital is more likely to record 80% or more female births?" Considering both the condition of having 80% or more of female births and the number of babies that are delivered daily in each hospital, and using the binomial distribution to compute such probabilities for each case, the difference between Hospital A and Hospital B can be clearly seen. From Figure 4, it is clear that the sample of 10 babies is more likely to deviate from the population trend (i.e., from the expected value) than the larger sample of 50 babies (for a sample with size $n = 10$, the probability is slightly more than 5%, whereas for a sample with size $n = 50$, the probability is around 1.2×10^{-5} %).

Task 3's results For Table 6, a chi-square test was used because no cells (0.0%) had an expected count less than 5 ($\chi^2(4, N = 155) = 65.48$, $p < 0.05$), and a significant difference between Thai and Japanese teachers was found. An analysis of the adjusted standardized residuals indicated that Thai teachers significantly differed from the Japanese ones in all categories, except in the choice of Hospital B. Teachers who chose Hospital B most likely used the Law of Large Numbers appropriately. Thai teachers (33.8%) showed a misunderstanding, or a lack of understanding, of the Law of Large Numbers. In the case of Japanese teachers, 24.4% incorrectly chose as an answer that "Two hospitals are equally likely to record such an event." Those teachers who chose option (c) could have reasoned something like this: Each time a baby is delivered, there is a 50% probability of it being a boy or a girl, and hence it makes sense that both hospitals will have a *balanced* possibility of having a particular proportion of boys and girls at a particular time.

Table 6. Frequency, percentage and percentage difference of each kind of answers given by Japanese and Thai teachers to Task 3

Category	Japan		Thailand		Difference
	Frequency	Percent	Frequency	Percent	JP–TH (%)
No response, "I don't know," "I can't say anything"	0	0.0	16	20.8	-20.8*
(a) Hospital A	0	0.0	26	33.8	-33.8*
(b) Hospital B	*39*	*50.0*	*27*	*35.1*	*+14.9*
(c) Two hospitals are equally likely to record such an event	19	24.4	2	2.6	+21.8*
(d) No basis for predicting which hospital would have that percentage of female births	20	25.6	6	7.8	+17.8*
Total	78	100.0	77	100.0	

Note. Hospital B is the correct answer.
* $p < 0.05$, based on an analysis of the adjusted standardized residuals
.

Also, we observe from Table 6 that 25.6% of the Japanese teachers chose option (d) as an answer (i.e., "No basis for predicting which hospital would have that percentage of female births"). If they reason based on the equiprobability context, this response is reasonable for them, because they cannot calculate the answer. In other words, these teachers might be unaware of the possibility of using either the binomial distribution (as in Figure 4) or the Law of Large Numbers to solve this problem.

Task 3's discussion and suggestions Teachers who chose Hospital A may have used the misconception known as *Law of Small Numbers* (Kahneman & Tversky, 1982). The Law of Small Numbers is associated with overestimating a small sample of information to draw firm conclusions that can only be correctly drawn through using the Law of Large Numbers. Teachers who chose answer (c) might have a misunderstanding about probability; they might have focused on the probability of the individual event of one baby being either a boy or a girl (which is ½, as the task sets from its initial sentence) rather than on the difference in sample sizes of babies delivered by the two hospitals. This kind of answer is closely related to the misconception known as *equiprobability bias*, which is a tendency to view several outcomes of a random experiment as equally likely, usually while focusing on the likelihood of just one event (Batanero et al., 2014; Lecoutre, Durand, & Cordier, 1990; Watson, 2005). Moreover, the results obtained in this study contrast with those reported by Watson (2000), in which 18 out of 33 Australian preservice secondary mathematics teachers correctly solved a similar version of the "Two hospitals" task.

On Task 3, more Japanese teachers used the Law of Large Numbers appropriately. The Law of Large numbers can be explained as an appropriate use of the conception *variation as distribution*. On the other hand, more Thai teachers may have used the misconception known as Law of Small Numbers by choosing Hospital A, which can be explained as an inappropriate use of the conception *variation as distribution*.

Japanese teachers who selected option (c) could hold the misconception known as *equiprobability bias*, because teachers may have focused on the fact that the probability of a baby being a boy or a girl is the same, and overgeneralized such a fact to the whole sample. They are lacking the conception *variability as whole range*, and are inappropriately using the conception *variation as distribution*.

Those teachers who chose option (d) are unaware of, or unfamiliar with, the theoretical methods to solve this problem (i.e., the binomial distribution, as in Figure 4, and the Law of Large Numbers). This provides evidence of a lack of the conceptions *variability as whole range* and *variation as distribution*.

For solving Task 3, teachers need to be prepared to use the conceptions *variability as whole range* and *variation as distribution* appropriately. Specifically, both Japanese and Thai teachers need to improve their knowledge of the use of the binomial distribution and the Law of Large Numbers.

4.4. TASK 4: "EMISSIONS" TASK

The four time plots below show annual emissions levels from 1980 through 1990 for four different pollutants. Match each plot with its corresponding verbal description.

____)a(SO_x(sulfur oxides). After a sharp 10% decline in the early 80s, later years showed more variation than trend.

____)b(Pb(lead). After a decline of more than 75% emissions held steady, with little variation, for five years.

____)c(CO(carbon monoxide). There is a linear pattern of decrease, apart from small year-to-year fluctuations.

____)d(-NO_x (nitrogen oxides).-Variation is larger, in relation to trend, than for the other graphs.

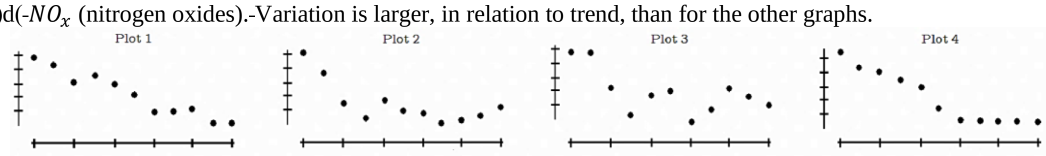


Figure 5. "Emissions" task (adapted from Cobb, 1998)

The purpose of Task 4 (adapted from Cobb, 1998) was to examine whether teachers were able to translate verbal statements on the variability of four time-dependent variables into graphical representations appropriately (Clement, 1989; Moritz, 2003, 2004). By doing so, we will be able to determine whether surveyed teachers were able to engage in transnumeration (i.e., in the dynamic process of forming and changing data representations to arrive at a better understanding; Wild and Pfannkuch 1999) appropriately. The task looks as though it is embedded in a real-life situation; however, for solving it, we only need statistical and mathematical knowledge.

Task 4's results In this task, from Plot 1 to Plot 4, 11 years of recorded emissions levels of pollutants are displayed by each graph. In the case of Thai teachers, 16 out of 77 teachers (20.8%) were able to make all the pairing matches on this task correctly. In the case of Japanese teachers, 60 out of 78 teachers (76.9%) answered this task correctly.

Regarding the pairing of CO with Plot 1, the results of the corresponding statistical analysis showed a significant difference between the responses of Japanese and Thai teachers (Fisher's exact test = 78.77, $p < 0.05$; see Table 7). Furthermore, the analysis of adjusted standardized residuals indicated that the Thai teachers differed from the Japanese in all the pairings, with the exception of matching CO with Plot 3.

For Tables 7 and 8, and the pairing of SO_x with Plot 2, the corresponding chi-square test ($\chi^2(4, N = 155) = 34.84$, $p < 0.05$) indicated a significant difference between the results obtained by Japanese and Thai teachers. Furthermore, the adjusted standardized residuals' analysis indicated that Thai teachers significantly differed from the Japanese regarding to the pairing of SO_x with Plot 1, Plot 2, and Plot 3.

Table 7. Frequency and percentage of each correct matches made by Japanese and Thai teachers on Task 4, along with the corresponding statistical analysis to test for independence between the groups

Correct match	Japan		Thailand		Statistical test	
	Frequency	%	Frequency	%	Chi-square	Fisher's exact
SO_x- Plot 2	61	78.2	25	32.5	34.835*	
Pb - Plot 4	69	88.5	33	42.9		41.663*
CO - Plot 1	73	93.6	22	28.6		78.774*
NO_x - Plot 3	64	82.1	39	50.6		24.438*

*$p < 0.05$

Regarding the pairing of NO_x with Plot 3, the results of the corresponding statistical analysis (Fisher's exact test = 24.44, $p < 0.05$) showed a significant difference between the responses of Japanese and Thai teachers. From the analysis of adjusted standardized residuals, it was found that the Thai teachers also significantly differed from Japanese in matching NO_x with Plot 1.

Regarding the pairing of Pb with Plot 4, the results of the corresponding statistical analysis (Fisher's exact test = 41.66, $p < 0.05$) showed a significant difference between the responses of Japanese and

Table 8. Frequency and percentage of all the pairing matches made by Japanese and Thai teachers on Task 4

		Plot 1		Plot 2		Plot 3		Plot 4	
		Japan	Thailand	Japan	Thailand	Japan	Thailand	Japan	Thailand
Pollutant	SO_x*	2(2.6)	16(20.8)	*61(78.2)*	*25(32.5)*	11(14.1)	24(31.2)	3(3.8)	4(5.2)
	Pb*	0(0.0)	14(18.2)	8(10.3)	19(24.7)	0(0.0)	3(3.9)	*69(88.5)*	*33(42.9)*
	CO*	*73(93.5)*	*22(23.6)*	0(0.0)	16(20.8)	2(2.6)	3(3.9)	2(2.6)	28(36.4)
	NO_x*	2(2.6)	17(22.1)	8(10.3)	9(11.7)	*64(82.0)*	*39(50.6)*	3(3.8)	4(5.2)

Note. Italics indicate the correct match.
* $p < 0.05$ comparing Japanese and Thai teachers, based on either chi-square or Fisher's exact test

Thai teachers. From the analysis of adjusted standardized residuals, it was also found that the Thai teachers significantly differed from the Japanese in all pairings, with the exception of matching Pb with Plot 3.

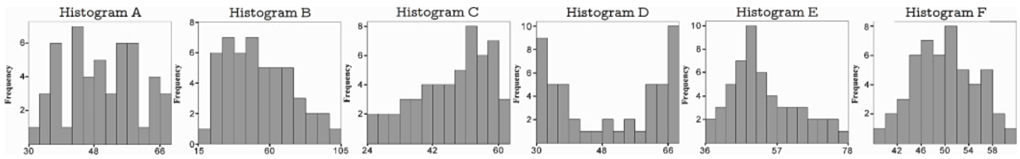
Task 4's discussion and suggestions This task can be solved with appropriate understanding of slope , percentage, proportion, ratio, linear and piecewise functions, and linear equations, without using any knowledge of the four chemical compounds involved in this task. These are regarded as fundamental aspects not only to engage appropriately in numerical graph interpretation, but also to correctly read data values from the scales of either axis of the coordinate system, not only one-by-one, but also as an aggregate, and to appropriately interpret verbal statements regarding multivariate graphs (Clement, 1989; Moritz, 2003).

On Task 4, more Japanese teachers were able to make correct matches between the verbal statements on the variability of four time-dependent variables and the given plots, which is a skill related to transnumeration. By being able to appropriately engage in transnumeration, Japanese teachers showed evidence of mastering both the second and seventh conceptions of variability identified by Shaughnessy (2007) (i.e., *variability as change over time* and *variation as covariation or association*, respectively). At the same time, more Thai teachers performed significantly lower than Japanese teachers on this task, and therefore they are lacking the knowledge regarding the conceptions *variability as change over time* and *variation as covariation or association*. In the case of Task 4, the knowledge related to these conceptions is mainly related to the ideas of proportionality and linearity.

For solving Task 4, teachers need to be prepared to use the conceptions *variability as change over time* and *variation as covariation or association* appropriately. Specifically, Thai teachers need to deepen their knowledge on concepts related to the ideas of proportionality and linearity, such as percentage, proportion, ratio, linear and piecewise functions, and linear equations.

4.5. TASK 5: "MATCHING SUMMARY STATISTICS TO HISTOGRAMS" TASK

Consider the following group of histograms and summary statistics. Each of the variables (1–6) corresponds to one of the histograms. Match each variable and its summary statistics to the correct histogram.



Variable	Mean	Median	Standard Deviation	Corresponding Histogram
1	50	50	10	
2	50	50	15	
3	53	50	10	
4	53	50	20	
5	47	50	10	
6	50	50	5	

Figure 6. "Matching summary statistics to histograms" task (adapted from Meletiou & Lee, 2003)

The purpose of Task 5 was to examine whether participants were able to make appropriate connections between real-life context-free histograms and their respective statistical summaries (Meletiou & Lee, 2003). Meletiou and Lee (2003) implemented this task with the same purpose, but with a sample comprising university students in an introductory statistics course in an American university. In this study, this is the only task covering measures of variation.

Task 5's results The poorest performance among both Japanese and Thai teachers was observed in response to this task. In the case of Japanese teachers, only 7 out of 78 (9.0%) were able to make all the matches correctly, whereas only 3 out of 77 Thai teachers (3.9%) were able to match all the histograms to the given variables correctly. No statistically significant differences between the samples were observed in this task, which means that both groups of teachers achieved poorly. Despite this fact, the best performance observed among both Japanese and Thai teachers was in matching Variable 6 with Histogram F (see Tables 9 and 10), which is a bell-shaped and almost symmetrical distribution with small standard deviation.

Table 9. Frequency and percentage of each correct match made by Japanese and Thai teachers on Task 5, along with the corresponding statistical analysis to test for significant differences between the groups

Variable	Corresponding Histogram	Japan		Thailand		Difference JP–TH (%)
		Frequency	%	Frequency	%	
1	Histogram A	29	37.2	28	36.4	+0.8
2	Histogram D	22	28.2	10	13.0	+15.2
3	Histogram E	15	19.2	27	35.1	-15.9
4	Histogram B	19	24.4	31	40.3	-15.9
5	Histogram C	30	38.5	31	40.3	-1.8
6	Histogram F	56	71.8	47	61.0	+10.8

* $p < 0.05$ comparing Japanese and Thai teachers, based on either chi-square or Fisher's exact test

Table 10. Frequency and percentage of all the pairing matches made by Japanese and Thai teachers on Task 5

	Histogram A		Histogram B		Histogram C		Histogram D		Histogram E		Histogram F	
Variable	Japan	Thailand	Japan	Thailand	Japan	Thailand	Japan	Thailand	Japan	Thailand	Japan	Thailand
1	*29(37.2)*	*28(36.4)*	17(21.8)	6(7.8)	6(7.7)	10(13.0)	0(0.0)	7(9.1)	9(11.5)	7(9.1)	14(17.9)	15(19.5)
2	28(35.9)	25(32.5)	8(10.3)	15(19.5)	5(6.4)	5(6.5)	*22(28.2)*	*10(13.0)*	11(14.1)	12(15.6)	1(1.3)	6(7.8)
3	6(7.7)	8(10.4)	17(21.8)	11(14.3)	31(39.7)	14(18.2)	5(6.4)	10(13.0)	*15(19.2)*	*27(35.1)*	1(1.3)	2(2.6)
4	7(9.0)	6(7.8)	*19(24.4)*	*31(40.3)*	3(3.8)	8(10.4)	40(51.3)	19(24.7)	6(7.7)	9(11.7)	0(0.0)	1(1.3)
5	2(2.6)	3(3.9)	13(16.7)	6(7.8)	*30(38.5)*	*31(40.3)*	6(7.7)	14(18.2)	21(26.9)	14(18.2)	3(3.8)	4(5.2)
6	3(3.8)	3(3.9)	0(0.0)	5(6.5)	2(2.6)	4(5.2)	2(2.6)	12(15.6)	12(15.4)	2(2.6)	*56(71.8)*	*47(61.0)*

Note. Correct matches are in italics

* $p < 0.05$, comparing Japanese and Thai teachers, based on either chi-square or Fisher's exact test

Task 5's discussion and suggestions The poor performance of both groups of teachers in this task, and the fact that the best performance observed was when participants matched a bell-shaped and almost symmetrical distribution to its statistical summary (i.e., Variable 6 with Histogram F), seem to indicate that participant teachers were somehow able to establish a connection between statistical summaries and graphical representations of a dataset only when the shape of the graph is symmetrical. Meletiou and Lee (2003), in the intervention stage of their study, considered fundamental the emphasis that statistics teaching must put on helping students improve their ability to read and understand graphical representations and to appropriately establish connections between statistical summaries and features of a distribution such as shape and spread.

The poor performance of both groups on this task showed a lack of knowledge of the relationship among the mean and the median and the histogram skewness, and a lack of knowledge of the tail weight (i.e., the width of the sides of a distribution; see Figure 7). From the best performance observed among both groups of teachers on this task (i.e., matching Variable 6 with Histogram F, which is a bell-shaped and almost symmetrical distribution), it seems that some teachers of both groups have a limited

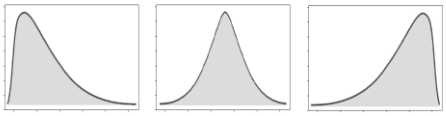
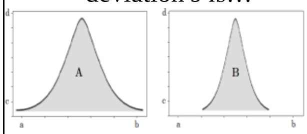
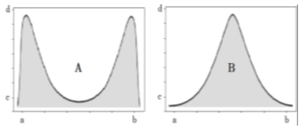
(a) Relation between mean, median, and skewness	(b) Relation between range and standard deviation	(c) Relation between tails and standard deviation
If the distribution skewness γ_1 is known, then the relation between mean and median is…  $\gamma_1 > 0$ Mean > median; $\gamma_1 \cong 0$ Mean ≅ median; $\gamma_1 < 0$ Mean < median	If two distributions have similar shape, but different ranges, then the standard deviation s is…  $s_A > s_B$	If two distributions have similar ranges, but one has heavier tails, then the standard deviation s is…  $s_A > s_B$

Figure 7. How to determine mean, median and standard deviation based on the distribution features

understanding of the normal distribution. Theoretically, probability distribution theory for continuous functions supports the rules shown in Figure 7. In order to solve Task 5 appropriately, these rules need to be extended to discrete data distributions. Thus, we can say that knowledge of ideas depicted in Figure 7 is necessary to solve Task 5 correctly. Most of the teachers from both groups may not have knowledge of these ideas. It is suggested that they need to increase their knowledge of these ideas regarding distribution features.

Both Japanese and Thai teachers need to improve their knowledge of the connection between the theoretical properties of histograms and their respective statistical summaries.

Inability to make an appropriate connection between histograms and their respective statistical summaries (mean, median and standard deviation) can be explained by a lack of the appropriate conception of *variation as distribution*, especially regarding measures of variation (such as standard deviation) and central tendency (such as mean and median).

Furthermore, Japanese mathematics teachers are expected to be familiar with the idea of standard deviation, because at senior high school level, grading on the normal curve is typically used to evaluate student achievement. This grading practice is not common in Thailand.

5. CONCLUSIONS AND LIMITATIONS

In relation to the research questions, conclusions for each task were provided in the results, discussion and suggestions, particularly about the specific knowledge base needed to conceptualize variability appropriately for Japanese and Thai teachers. This comparative survey study used tasks from previous research to elicit teachers' conceptions of variability, in the sense of Shaughnessy's (2007) framework. The results of this survey clarified Japanese and Thai teachers' conceptions of variability from the tasks, and they also provided empirical evidence on the key elements of the knowledge base required to solve those tasks that are needed for these teachers. For solving these tasks appropriately, teachers must possess appropriate conceptions of variability, and the gap between such conceptions and the current ones held by the teachers justifies the need to develop such a knowledge base among Japanese and Thai senior high school mathematics teachers.

The results obtained by this survey showed that teachers from the two countries have different understandings of variability. The Japanese teachers, who had a tendency to use equiprobability (Task 1), were labelled *mathematical equiprobability-oriented*, whereas Thai teachers, who relatively tended to engage in estimation (Task 1 and Task 2), were designated as *estimation-oriented*. An equiprobability orientation tends to consider inappropriate answers as appropriate, such as considering "2.5 times" correct in Task 1. This orientation was also confirmed by other tasks in the survey, such as the "The gumball machine" task (not included for discussion in this article).

Thai teachers showed a limited understanding of the normal distribution in relation to measures of variation and representative values, such as in the images shown in Figure 7 (Task 5). Japanese teachers showed a tendency to overgeneralize the use of the normal distribution (Task 2), but also had a limited understanding of normal distributions in relation to measures of variation and representative values (Task 5). In order to overcome these issues, a deeper knowledge of the relations between particular features and parameters of theoretical probability distributions, shown in Figure 7, is necessary for teachers in both countries. However, overgeneralized images of normal distribution held by Japanese teachers are based on an ill-conceived theory, and should be reconsidered. In the case of Japan, this problem was reported by Isoda and González (2012), and related tasks for addressing this issue, accompanied with an appropriate understanding of the relation between representative values and distribution features (in the sense of Figure 7), appeared in the 2016 junior high school mathematics textbook. All this was not included in the 2012 edition with both editions (2012 and 2016) being under the same curriculum standards (MEXT, 2008).

In relation to the Law of Large Numbers, Japanese teachers seemed to think of probability from the perspective of the Law of Large Numbers (Task 3) as well as equiprobability. Thai teachers seemed to think of probability from the ratio perspective, without using the Law of Large Numbers (Task 3). Content about the Law of Large Numbers in Thailand is missing from the national mathematics textbooks, which define probability by ratio, without using the Law of Large Numbers, and do not define equiprobability. So, Task 3 implies the need for Thai teachers to study the Law of Large Numbers and equiprobability.

This survey identified the difference in theoretical knowledge regarding variability between Japanese and Thai teachers, and it showed the need for their knowledge base to include an appropriate conceptualization of variability. The survey contained well-known tasks used in previous research studies, and they were useful to clarify these differences.

At the methodological level, the authors acknowledge the lack of data triangulation as the main limitation of this study, whose findings were not confirmed through the use of data collection methods different to teachers' responses to the survey instrument (e.g., interviews or observations). Another methodological limitation was the limited scope of the tasks used in the study, which provided preset data and did not require teachers to either engage in the collection and analysis of genuine data or use technological data exploration tools.

In relation to the theoretical knowledge observed from Japanese and Thai teachers, limitations of Shaughnessy's (2007) framework were found. His framework gave us some perspectives for searching the background theoretical view but did not describe every aspect of the theoretical knowledge of variability itself. Teachers' knowledge was not always found to be appropriate on this survey. For example, there were many misconceptions by teachers shown on Task 2 such as the *variability as visual cues in the graph*, which we would add to Shaughnessy's framework. The theoretical knowledge of teachers was discussed in relation to the tasks, for example, for solving Task 5, we needed the theory in Figure 7 to explain the results. Shaughnessy's framework, however, only provided the perspective of *variation as distribution*. Our findings are based on the chosen tasks.

Statistical literacy is usually defined as the combination not only of statistical and mathematical knowledge, but also of context knowledge (Gal, 2004). Therefore, addressing simultaneously the theoretical view and the real-world view in statistics is necessary. However, this article focused on the theoretical view of variability, because of the selection of five out of nine tasks for discussion in this paper.

Gal (2004) pointed out that literacy cannot be viewed only as a skill or ability, but also as a set of cultural practices that people engage in, and that is why cultural factors play an important part in the development of the components of statistical literacy, such as developing a statistical knowledge base. The cultural factor –a country's teaching and learning culture–is regarded as relevant to this study as it affects the way that people teach and learn statistics (Shaughnessy, 2007, p. 964). With all this in mind, another limitation of this study was that cultural factors for each country were left out of the scope of the research, although they could serve as a source of interpretation of the findings.

The conclusions presented here are also limited by the participants included in the study. Even though we called them *Japanese teachers* and *Thai teachers*, they were a sampled group of teachers. Both groups were volunteers: Japanese participants were teachers who responded to a questionnaire which was sent to all schools in a prefecture and Thai teachers were the participants of training program in a district. Also, every interpretation was provided by the differences between the data of the two groups of teachers for the questionnaire and was not based on the standardized tasks.

Despite the acknowledged limitations of the present study, the authors feel that the findings may have relevant implications for the professional practice of Japanese and Thai mathematics teachers, specifically regarding their theoretical knowledge on variability and the way they acknowledge variability.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Jane Marie Watson for her insightful suggestions and helpful comments in reviewing an earlier version of this article. The authors would also like to thank Dr. Mike Shaughnessy for personal communications that clarified the authors' understanding of various matters discussed in this paper.

We are also grateful to *SERJ* Associate Editor and the two reviewers for their helpful comments and suggestions which allowed us to greatly improve the presentation of our paper.

REFERENCES

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). New York: John Wiley & Sons.

Batanero, C., Arteaga, P., Serrano, L., & Ruiz, B. (2014). Prospective primary school teachers' perception of randomness. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 345–366). Dordrecht, The Netherlands: Springer.

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, and C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education*. (pp. 57–70). New York: Springer.

Clement, J. (1989). The concept of variation and misconceptions in Cartesian graphing. *Focus on Learning Problems in Mathematics*, *11*(2), 77–87.

Cobb, G. W. (1998, April). *The objective-format question in statistics: Dead horse, old bath water, or overlooked baby?* Invited paper presented at the annual meeting of the American Educational Research Association (AERA), San Diego, CA.
[Online: https://apps3.cehd.umn.edu/artist/articles/Cobb.pdf]

Freedman, K., & Hernández, F. (Eds.) (1998). *Curriculum, culture, and art education: Comparative perspectives*. New York: State University of New York Press.

Gal, I. (2004). Statistical literacy: Meaning, components, responsibilities. In J. Garfield & D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht, The Netherlands: Kluwer.

Garfield, J., & Gal, I. (1999). Teaching and assessing statistical reasoning. In L. Stiff (Ed.), *Developing mathematical reasoning in grades K–12* (pp. 207–219). Reston, VA: NCTM.

González, O. (2011). *Survey on Japanese elementary, middle and high school teachers' statistical literacy: Focusing on variability* (Unpublished master's thesis). University of Tsukuba, Tsukuba, Japan.

González, O., Isoda, M., & Chitmun, S. (2015, February). *Senior high school mathematics teachers' difficulty to teach statistics: The cases of Japan and Thailand.* Keynote lecture presented at the APEC-Tsukuba International Conference IX: Innovation of Mathematics Teaching and Learning through Lesson Study, Tokyo, Japan.
[Online: http://www.criced.tsukuba.ac.jp/math/apec/apec2015/1100 Orlando Gonzalez APEC IX FINAL.pptx]

Isoda, M. (Ed.) (2010). *Junior high school teaching guide for the Japanese course of study: Mathematics.* Center for Research on International Cooperation in Educational Development (CRICED), University of Tsukuba.
[Online: www.criced.tsukuba.ac.jp/math/apec/ICME12/Lesson_Study_set/JuniorHighSchoolTeaching Guide-Mathmatics-JP-EN.pdf]

Isoda, M. (2015, January). *Survey on elementary, junior and high school teachers' statistical literacy: How difficult to teach statistics for us*. Keynote lecture presented at the first national conference of the Thailand Society of Mathematics Education (TSMEd), Khon Kaen, Thailand.

Isoda, M., & González, O. (2012). Survey on elementary, junior and high school teachers' statistical literacy: The need for teacher training in variability (in Japanese). *Journal of Science Education in Japan*, *36*(1), 61–76.

Kahneman, D., & Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 3–20). Cambridge, UK: Cambridge University Press.

Lecoutre, M. P., Durand, J. L., & Cordier, J. (1990). A study of two biases in probabilistic judgements: Representativeness and equiprobability. In J. P. Caverni, J. M. Fabre, & M. Gonzales (Eds.), *Cognitive biases* (pp. 563–575). Amsterdam, The Netherlands: North-Holland.

Makar, K. M., & Confrey, J. (2004). Secondary teachers' reasoning about comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenges of developing statistical literacy, reasoning, and thinking* (pp. 327–352). Dordrecht, The Netherlands: Kluwer.

Meletiou, M., & Lee, C. (2003). Studying the evolution of students' conceptions of variation using the transformative and conjecture-driven research design. In C. Lee (Ed.), *Reasoning about variability: A collection of current research studies*. Mt. Pleasant, MI: Central Michigan University.

Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) (2008). *Junior high school teaching guide for the Japanese course of study: Mathematics (in Japanese)*. Tokyo: Kyoiku Shuppan Publishers.

Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) (2009). *Senior high school teaching guide for the Japanese course of study: Mathematics (in Japanese)*. Tokyo: Jikkyo Shuppan Publishers.

Moritz, J. B. (2003). Interpreting a scattergraph displaying counterintuitive covariation. In L. Bragg, C. Campbell, G. Herbert, & J. Mousley (Eds.), *Mathematics education research: Innovation, networking, opportunity. Proceedings of the 26th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 523–530). Sydney, Australia: MERGA.

Moritz, J. B. (2004). Constructing coordinate graphs: Representing corresponding ordered values with variation in two-dimensional space. *Mathematics Education Research Journal*, *15*(3), 226–251.

Peters, S. A. (2009). *Developing an understanding of dispersion: AP statistics teachers' perceptions and recollections of critical moments* (Unpublished doctoral dissertation). Pennsylvania State University, University Park, PA.

Pfannkuch & Ben-Zvi, (2011). Developing teachers' statistical thinking. In C. Batanero, G. Burrill, and C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study* (pp.323–333). Dordrecht: Springer.

Royal Thai Ministry of Education (MOE) (2008). *The Basic Education Core Curriculum B.E. 2551 (A.D. 2008)*. Bangkok, Thailand: MOE.
[Online: http://www.act.ac.th/document/1741.pdf]

Shaughnessy, J. M. (2007). Research on statistics' learning and reasoning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp. 957–1009). Charlotte, NC: Information Age Publishers.

Shaughnessy, J. M., & Ciancetta, M. (2001, August). Conflict between students' personal theories and actual data: The spectre of variation. Paper presented at the *Second Statistical Reasoning, Thinking, and Literacy Forum (SRTL-2)*. Armidale, New South Wales, Australia.

Shaughnessy, J. M., & Ciancetta, M. (2002). Students' understanding of variability in a probability environment. In B. Philips (Ed.), *Developing a statistically literate society: Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS-6)* Cape Town, South Africa. Voorburg, The Netherlands: International Statistical Institute.
[Online: https://iase-web.org/documents/papers/icots6/6a6_shau.pdf]

Shaughnessy, J. M., & Pfannkuch, M. (2002). How faithful is Old Faithful? Statistical thinking: A story of variation and prediction. *Mathematics Teacher*, 95(4), 252–259.

Shaughnessy, J. M., Ciancetta, M., & Canada, D. (2004). Types of student reasoning on sampling tasks. In M. Høines & A. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education. Vol. 4* (pp. 177–184). Bergen, Norway: Bergen University College.

Watson, J. M. (2000). Preservice mathematics teachers' understanding of sampling: Intuition or mathematics. *Mathematics Teacher Education and Development*, *2*, 121–135.

Watson, J. M. (2005). The probabilistic reasoning of middle school students. In G. A. Jones (Ed.), *Exploring probability in schools: Challenges for teaching and learning* (pp. 145–170). New York: Springer.

Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, 37, 145–168.

Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, *34*, 1–29.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–265.

MASAMI ISODA
Faculty of Human Sciences
University of Tsukuba
1-1-1 Tennodai, Tsukuba-shi
Ibaraki 305-8572
JAPAN