CLASSROOM DESIGN TO JUDGE BY STATISTICAL PROBABILITY IN A JUNIOR HIGH SCHOOL

Norihiro Nishinaka[1] and Atsushi Yoshikawa[2]
[1]Kindai University, Japan, [2]Tokyo Institute of Technology, Japan
nori1618@kindai.ac.jp

We developed teaching materials and a method for statistical education based on Cobb's research (Cobb, 1999). As a teaching material, we prepared two types of batteries in which the statistical characteristics were different. The method leads students to creating criteria for statistical judgment. We designed this method as follows:1) At first, students are given some information so that they can solve a given problem, but every group is given different information, 2) They have to solve three problems in a different context for each group, 3) They can find out other information when other groups present their answers. In this way, statistical concepts that we have not taught naturally occurred among the students.

INTRODUCTION

The purpose of this study is to develop an effective statistics lesson in a Japanese junior high school and demonstrate the students' reaction to our class. Our primary goal is that students will be good decision makers with statistical data and probability. Therefore, we believe that it is important to nurture junior high school students who can use this information. Our aim is to nurture statistical literacy, reasoning, and thinking for students.

We began to think this way because

1. in the real world, it is important to solve problems using statistics (e.g., big data society);
2. most Japanese junior high school students can calculate accurately what has been taught. However, they cannot make judgments using statistical data and probability; and
3. fostering statistical literacy, reasoning, and thinking is also a major objective of statistics education in the world (e.g., see srtl.info/).

Most mathematics teachers in Japanese junior high schools do not seek an effective teaching method in statistics classes. Statistics has not been an important unit for mathematics teachers over the past decade. However, the importance of statistical education in Japan has been gradually increasing for 10 years (MEXT, 2003). Thus, we believe that effective materials and methods for students need to be developed.

We believe that it is necessary to be a good decision maker with statistical data and probability. Statistical data and probability enable us to do the following:

- take advantage of statistical literacy, reasoning, and thinking;
- understand that the conclusion will be different if the judging criteria are different even for the same statistical data; and
- make a judgment on one's own and cooperate and make judgments.

We believe that we must train students at the junior high school stage before workable.

We organize the process using statistical data and making judgments under uncertainty. In this process, first, we determine the purpose for making judgments from the task, then select a statistical method for the purpose, create criteria, analyze data, and make judgments. This process is similar to the PPDAC cycle（Problem-Plan-Data-Analysis-Conclusion） (Wild & Pfannkuch, 1999). However, the PPDAC cycle is an investigative cycle. We consider the process of making judgments using statistical data and methods.

We note that it is very important to create criteria. If we decide what criteria we will judge first, then we can make a judgment without hesitation. This is different from the PPDAC cycle. We thought about the design of the lesson that would enable students to experience this process. Hence, we propose the following design for developing statistical thinking:

a) We prepared statistical materials, which could be interpreted in a variety of ways.
b) We provided questions to extract the students' naive concept of statistical data.
c) We presented several tasks in which students could embed statistical criteria.

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

As a result, students were able to judge different results using different statistical values, although they used the same statistical materials. They found expression in their own thinking criteria and perceived other people's thoughts. This study demonstrates that our new design can develop junior high school students' statistical thinking and ability to make judgments using probability in a state of uncertainty.

METHODS

Research on probability judgment has shown that when a judgment is made, this occurs with bias and error. From the study by Kahneman and Tversky, it is obvious that people make judgments using representative heuristics (Kahneman & Tversky 1972) and availability heuristics (Tversky & Kahneman 1973) instead of calculating complex probabilities and making decisions. Additionally, when a person makes a decision, it demonstrates a mooring and adjustment (Tversky & Kahneman 1974) that performs estimation and adjustment based on suggested reference points. It has been demonstrated that people make probability decisions for uncertain events in the presence of these cognitive biases.

We designed our method using the three steps (a,,b,,c) listed in the previous section.

(a) It is a matter to make a decision, it is a task that requires two or more comparisons. Moreover, it is a problem that any student can consider; it is a problem that the student itself can set to judge criteria. The statistical data that we used had the same mean value as far as possible, and the data distribution was different.

(b) We believe that students' naive concept of statistical data is the basis of statistical learning because it is the starting point of students' ideas.

(c) Even with the same data, we aimed to make our students experience drawing different conclusions depending on context differences and judgment criteria. For that purpose, we prepared several types of tasks and assigned each groups of students a different task (3). We refer to overall tasks and individual tasks.

We developed teaching materials and methods for statistical education based on Cobb's research (Cobb, 1999; Cobb & McClain, 2005). As teaching material, we prepared two types of batteries in which the statistical characteristics were different. The method led students to create criteria for statistical judgment. Two factors underlying our choice of this material were students' familiarity with the task and mathematical thinking (see Table1).

Table 1. Overall task and individual tasks

Overall task: Investigate the life of each of the two types of batteries, and judge which one to purchase based on the data.
Individual task A: You have a toy home planetarium. This planetarium moves with two batteries, but seems to be connecting batteries in parallel. How can you use batteries from both companies to make this planetarium run for the longest time? Please think about the possibility and give the answer.
Individual task B: You have a home theater that uses two batteries. You want the battery to run as quickly as possible so that it will not cut as much as possible. How you insert the battery seems to be a serial tie. Which company's batteries should you use? Please think about the possibility and give the answer.
Individual task C: The TV remote control of your grandfather's house uses two dry batteries Since he is an old man, he does not realize that the battery will run out, and sometimes you have to check it. However, it is not easy to insert, so please select a battery so that it will last as long as possible. How to insert the battery seems to be a parallel tie. Which batteries do you want? Please think about the possibility and give the answer.

We divided the lesson into two parts. In one part, the students judged the overall task using a sample survey. In the other part, the students judged individual tasks using statistical probability. For the overall task, we asked the students to conduct simulated specimen surveys, and then asked them to make decisions.

To capture the process of students' thought, their speech was converted into data; they were urged to clarify and explain their evidence. This is because the explanations arrived at through the students' reasoning was considered to be an expression of their thought processes. We believe that the students' way of speaking is more obvious and, without fail, expresses the process of thinking clearly and explains their reasoning.

RESULTS

We describe the results of the experimental lesson divided into an overall task and individual tasks. For the individual tasks, we describe the protocols in the group announcement in the order of A, B and C for individual issues.

Table 2. Judging the overall problem

Teacher 1 This time will sample every sample for each group and conduct a sample survey and ask them to make judgments based on the data collected by them. So who is the person who wants to use Company A's batteries? Who wants to use company B's?
Student 1 In terms of stability, company A
T1 Why?
S1 Because the company data is clustered around the vicinity of 195 hours of use. (The average lifetime of the batteries.)
T1 Yes, stability is important. But, some of Company B's batteries are also very good. Each feature is different.

Judging the overall task based on the first sample survey

For each group, we conducted sample surveys on the life of the batteries of Company A and Company B and compiled the results of each worksheet. Relative power was also calculated for a comparison between the number of specimens of Company A and Company B, and a comparison among groups. We also created a histogram based on the relative frequency, which made it material for the judgment of Company A and Company B.

Table 2 shows the judgment for the overall task. As in Table 2, many students selected Company A because, as shown in Figure 1, when the distribution of Company A is compared with that of Company B, the variation is extremely small, and this result derives from its consistent distribution that concentrates around the 195 marks. Note the word "stability," which indicates the student's judgment criteria. Then, as can be seen from the fact that stability is "in a cluster around 195," it can be observed that the students tended to make their judgments by concentrating on one specific value. This observation makes it possible to express the thinking process of students. Subsequently, the overall results totaling those samples taken from each group were compiled, and whether there was any change in the judgment of both companies was checked; however, there was no change in the choice of Company B.

Scaffolding in problem solving

For individual tasks, we distributed worksheets on which the results of the survey for each group were printed to make judgments based on the results of sample surveys conducted for each group. We created eight groups with four to five people in each group. The students were initially confused about making judgments using probability. The teacher advised them that "it is easy to consider the problem by thinking about battery life as being either short or long." After this advice, the students' problem-solving behavior began to progress. This progress supports the promotion of deep learning and can be regarded as part of "scaffolding" for learners (Collins, 2005; Kazak, Wegerif & Fujita, 2015). As a result, the groups became lively, and it became possible to make a judgment using probability. Tables 4 and 5 show the students' presentations. Table 4 shows the announcement by the team who thought about individual task A. The students considered battery life as two types: short and long.

Judging second individual task A

Table 3. Protocol of the presentation by the group (individual task A)

S2 First, we decided the criteria based on the mode of the characteristic of the graph.
T1 It is based on the mode value. What is its value?
S2 Its value is 195 hours. It is longer than 195 hours. Less than 195 hours is short.
T1 So?
S2 The relative probability of AL*, AS, BL, BS from the relative frequency was obtained.
T1 What is its value?
S2 The probability of AL is calculated as 15%, the probability of AS at 85%, the probability of BL at 35%, and the probability of BS at 65%.
S2 So we chose Company B, whose batteries have a high probability of lasting for a long time. Also, from the result table of our experiments, the average value of AL was 207.8, the average value of BL was 211, and the average of Company B was also longer.

*AL and BL indicate the long-life batteries of Company A and Company B, respectively, and AS and BS indicate the short life batteries of Company A and Company B, respectively.

Individual task A involved a parallel connection of batteries. In a parallel connection, even if a battery with a short lifetime runs out, if the other battery has a long life, then it can be used. Therefore, the students considered the judgment criteria as “a combination that can last a long time.” Another group suggested not only Company A or Company B but also Company A and Company B simultaneously to select how to combine. They arrived at an average for reducing risk; that is, Company A’s battery life is shorter than Company B’s; hence, Company A’s batteries do not have a long lifetime, but neither are they short-lived. Using Company A’s batteries performance is average, and if you use the long-life battery of Company B, an even longer battery lifetime is achieved. By contrast, even when shortening the lifespan of Company B’s battery, Company A’s battery has a higher probability of lasting a long time; hence, it can be used up to the life of Company A’s battery. Thus, the risk is reduced.

Judging second individual task B

Table 4 shows the announcement by the team who considered individual task B. They considered battery life as three types: short, normal, and long. In the presentation, they mentioned only two: short and long. Considering each combination, they calculated the probability and made a judgment using the result. Because individual task B involved a series, if there was a battery with a short lifetime, it stopped. Hence, it was necessary to compare the probabilities of long lifetimes. Therefore, comparisons between AL – AL and BL – BL became necessary.

Table 4. Protocol of the presentation by the group (individual task B)

S3 From our graph, we set the class value of 195 or less for a short time (S).
T1 Then?
S3 We have achieved 205 hours of use or a longer time (L). We extracted the relative frequency from our table and found the probability of S and L.
T1 How did you determine it?
S3 We determined the probability of AL – AL, AL – AS. AS – AS, BL – BL, BL – BS, BS – BS.
T1 You considered the combination and got the probability. How?
S3 We calculated that the probability of AL – AL is 0.07, the probability of AL – AS is 0.19, and the probability of AS – AS is 0.54. Also, the probability of BL – BL was 0.29, the probability of BL – BS was 0.25, and the probability of BS – BS was 0.21.
T1 You got the probability. What’s wrong?
S3 It is better that the probability of L appearing in problem B is better because we used Company B’s batteries rather than using company A’s battery.
T1 Which probability did you compare to choose Company B?
S3 AL – AL and BL – BL.

Judging second individual task C

Because individual task C involved a parallel connection, it was better to use a battery with a long lifetime. Students who worked on task C ranked a battery life of 205 hours as the criteria for assigning the lifetime of a battery as either short or long. Therefore, the probability of a long-life expectancy was 7/46 for Company A and 35/78 for Company B.

The challenge was that the students had to change batteries because their grandfathers could not and became confused when the batteries ran out. The questions that arose were: Is it possible to match the timing of the expected battery life if it can be predicted? Is it impossible to predict battery life? When the battery life is stable, is it easy to predict. Conversely, if the difference between a long and short lifetime is large, then it is difficult to predict. Therefore, answering "Company A" would be an equally valid response. Unfortunately, in this class, there were no teams that provided that answer.

Table 5. Protocol of the presentation by the group (individual task C)

S4	Since it is a parallel connection, it is only necessary for one of the batteries to be resilient. .
T1	A battery with a long life is better. How did it go?
S4	We thought about how to calculate the probability of batteries with a longer life.
T1	How did you set the criteria for batteries with shorter lives??
S4	We set a criteria of 205 hours or more for long, and under 205 hours for short. Then the probability that Company A is 7/46 and Company B is 35/78 was calculated. Since Company B is the one that mostly scores 205 or more, we chose Company B.

DISCUSSION

In this research, we configured the class design so that a judgment could be made using statistical probability. Initially, students alone could use the statistical probability to make a judgment from the table of battery life experiments. The specimen survey, conducted to consider relative frequency as a probability, was well received by the students. Converting battery life to clip color was also a good idea. It was good to take out the clip from a well-stirred bag. Taking out what is contained in a bag is often used in probability problems, and it is also a mathematical subject familiar to students. Additionally, by conducting the act leading to the sample survey, it was possible to connect the probability and sample survey.

In the classes for individual tasks, students did not react much at first. Students knew that relative frequency was judged as a statistical probability; however, they were not accustomed to considering many probabilities. The teacher noticed this fact and proposed that the students divided the life of the battery into two categories: long and short. This is a proposal similar to the aforementioned "scaffolding." Thus, it became easier for the students to calculate the probability, and problem solving improved. This indicates that the students were not used to considering a large number of probabilities and that teachers needs to work more in this area.

As a result, in individual tasks, the students were able to make judgments based on their own experimental results by considering the relative frequency as a probability and calculating the probability. The group that solved individual task B calculated the probability using the law of multiplication of the probability assuming that it was independent, considering the combination of two companies batteries. The law of multiplication of probabilities and law of sum are not typically taught in classes for junior high school second graders. However, because the students had already learned the two laws, they were using them. To make a judgment using probability, learning about probability calculation methods was necessary.

Additionally, students S2 and S4 arrived at the standards of 195 and 205; that is, the students themselves set the standards. We believe that this is an indication that standards are not absolute but considered to be relative. Groups with different criteria may have opposing results. The teacher's advice clearly showed numerical values but did not indicate standards. Therefore, we believe that the aims of the research were achieved. It was easy for the teacher to show clear criteria, which made it easier for the students to elucidate the criteria. However, when the teachers specifically set criteria, it was unlikely that the students would be able to devise their own criteria and make presentations as they did in the research. We believe that the way the students were able to share the importance of setting criteria by themselves was a significant educational achievement.

The impressions of the students after the class largely reflected the fact that they found the lesson difficult, but that using probability in this manner was convenient. One student cited the benefits of using probability saying: "You can see which is better at a glance." This student had realized the convenience of being able to make judgments with ease by comparing the size of the numerical values; that is, it can be inferred that the student had understood the efficacy of using probability in making judgments. Another student stated that it had been able to learn about the merits of probability and its use through the lesson. The student further stated that it had realized the wider potential of using probability in its daily life and would try to do so in the future.

CONCLUSION

In this research, we designed a teaching method to show students' thinking processes and share them and described the practice. As a result, students were able to judge different results using different statistical values, although they used the same statistical materials. They found expression in their own thinking criteria, and perceived other people's thoughts. We also found: (1) The students demonstrated their own judgment, which is important, then showed the data on which the judgment was based and could show an interpretation of the data. We believe that the students were able to express the process of thought toward the uncertain event that we were aiming for. (2) The students were initially confused about making judgments using probability. The teacher advised them that it was easy to think by considering battery life in terms of short and long. As a result, the students' problem-solving behavior was stimulated. (3) Based on the results of the sample survey, the students felt that they could make probability judgments considering this combination. Similar to the task considered in this research, learning to think about probability and make judgments based on statistical data has not been used in conventional learning; however, by considering it using the approach proposed in this research, we believe that students were able to understand the situation and how to use probability naturally. We believe that the development and practice of teaching materials and lessons that not only emphasize the calculation of probability but also make use of it and allow judgment based on probability will be necessary in the future.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number JP16K00979.

REFERENCES

Ben-Zvi, D. & Garfield, J. (2004). Statistical Literacy, Reasoning, and Thinking: Goals, Definitions, and Challenges. In D. Ben-Zvi & J. Garfield (Eds.) *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* , 3-15. Springer.

Cobb, P. (1999): Individual and Collective Mathematical Development: The Case of Statistical Data Analysis. *Mathematical Thinking and Learning*, *1*(1), 5-43.

Cobb, P. & McClain, K. (2005). Guiding inquiry-based math learning. In Keith Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences*, 171-185, Cambridge University Press.

Collins, A. (2005). Cognitive Apprenticeship. In Keith Sawyer (Eds.) *The Cambridge Handbook of the Learning Sciences*,47- 60, Cambridge University Press.

Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness *Cognitive Psychology, 3*(3), 430–454.

Kazak, S., Wegerif, R., & Fujita, T. (2015). Combining scaffolding for content and scaffolding for dialogue to support conceptual breakthroughs in understanding probability, *ZDM*, *47*(7), 1269–1283.

MEXT. (2008). Course of Study for junior high schools, MEXT.

Nishinaka, N. & Yoshikawa, A. (2011). Class practice for developing statistical thinking in junior high school education. *Journal of Science Education in Japan, 35*(2), 153-166.

Tversky, A. & Kahneman, D. (1974). Judgement under Uncertainty: Heuristics and biases. *Science, 185*(4157), 1124 – 1131.

Tversky, A. & Kahneman, D. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237 – 251.

Wild, C. J. & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review*, *67*(3), 223-248.