Journal of Mathematics Education
June 2009, Vol. 2, No. 1, pp.15-26

Survey and Analysis of the Statistical Reasoning among High School Students in China and Dutch

Wenfang Wang

Xiaohui Wang

Guofang Chen

Northeast Normal University, China

As for the students in high school in China, it is quite common to use the intuition and misconceptions in statistical reasoning. They have difficulties in understanding "A small sample is more likely to deviate from the population" and "Independence;" they tend to make mistakes about "Equiprobability bias," "Good samples have to represent a high percentage of the population," "Groups can only be compared if they have the same size" and "Law of small numbers." Compared with the Dutch students, the third grade students in China got lower scores in the 8 correct reasoning scales, but their misconceptions' scales demonstrated a different result: in some scales, Chinese students got higher scores, while in others, they got lower ones.

***Key words*:** statistical reasoning, correct reasoning, misconceptions, comparison between the Chinese and the Dutch.

Problem Formulation

Statistical reasoning may be defined as the way people reason with statistical ideas and make sense of statistical information (Garfield, 2003). This involves making interpretations based on sets of data, graphical representations, and statistical summaries. Much of statistical reasoning combines ideas with data and chance, which leads to making inferences and interpreting statistical results. Underlying this reasoning is a conceptual understanding of important ideas, such as distribution, center, spread, association, uncertainty, randomness, and sampling. Different form traditional mathematical reasoning which emphasizes abstraction, statistical reasoning pays more attention to the context or background of the problem. It is not only an "output" of education, but also a crucial "input" in the process of learning statistics. Students enter our classes with prior reasoning skills: those corresponding to true knowledge which will ease the learning process; and those formed by misconceptions (or we can call them intuitive but faulty reasoning mechanisms). These two are crucial determinants in learning: if preconceptions are not properly addressed, newly learned correct knowledge might appear much more volatile than existing preconceptions brought into class (Tempelaar, 2004).

High school students are in a crucial period before choosing their majors; therefore, it becomes much more necessary and urgent to learn the basic knowledge and technology in probability and statistics. How to draw conclusions by reasoning according to all kinds of statistical information? How to assess and deal with the information? Which is easy for them to understand and which is decided using their intuition? In this paper, our subject is high school students, aiming at their status of using statistical reasoning. We also do some comparisons and analyses with the relating result with the Dutch.

Method

Participants

This study used survey questionnaires, observations and interviews to investigate 1360 students in 5 high schools from Changchun to Xiamen, which covers both the north and the south part of China. The sample includes 3 key high schools (where students got higher academic scores when entering the school), 2 normal high schools; 320 from the first grade, 361 from the second grade and 679 from the third grade.

Besides this, we did comparisons with the Dutch students. The score of the Dutch is from the result of Dirk (2004), who used the questionnaire of Statistical Reasoning Assessment (SRA) developed by Garfield (2003). In his sample, 1/3 was from Holland, and 2/3 was from Europe (mainly from Germany). The Dutch students received an introduction into statistics and probability in high school, including calculating statistic with the use of calculator and solving several statistical problems, with no further conceptual knowledge; while international students do not receive such an introduction about statistics, but some of them have studied a little probability. In order to make the results comparable, in our study, we chose to compare the third grade students who have learned statistics and probability with those Dutch students, so as to get rid of the context influence of no prior statistical study.

The SRA Instrument

Based on Dr. Garfield's SRA, our questionnaire was translated into Chinese and adopted according to the Chinese students' comprehension level and living background. We also got help from statistical doctoral students in order to make the expressions and terms more scientific and accurate. Besides this, we pretested the students from a middle vocational school in Changchun in spring, 2005. It showed that the statements and amount of the tests are suitable. The SRA is a multiple-choice test consisting 20 items. This test mainly investigates 8 correct reasoning scales (cc1-cc8) and 8 misconceptions scales (mc1-mc8) as follows: cc1-Correctly interprets probabilities (be able to explain probability by using ratio); cc2-Understands how to select an appropriate average; cc3-Correctly computes probability (knowing that probability is a ratio between 0 and 1, and can use permutational and combinatorial reasoning); cc4-Understands independence; cc5-Understands sampling variability (Knowing how samples are related to a population and what may be inferred from a sample, knowing why a well chosen sample will more accurately represent a population and why there are ways of choosing a sample that make it unrepresentative of the population; knowing to be skeptical of inferences made using small or biased samples); cc6-Dirtinguishes between correlation and causation; cc7-Correctly interprets two-way tables; cc8-Understands the importance of large samples; mc1-Misconceptions involving averages (knowing a larger, suitable sample can be more representative in representing the population); mc2-Outcome orientation (making yes or no decisions about single events rather than looking at the series of events); mc3-Good samples have to represent a high percentage of the population; mc4-Law of small numbers (Small samples are preferred over larger samples since small samples best resemble the populations); mc5-Representativeness misconception (the likelihood of a sample is estimated on the basis how closely it resembles the population); mc6-Correlation implies causation; mc7-Equiprobability bias (Events of unequal chance tend to be viewed as equally likely); mc8-Groups can only be compared if they have the same size. Each scale is calculated respectively, the score ranges between 0 and 1, the score equals to the number that the student chose for a specific concept or misconception divided by the total number of choices. Let's take cc1 score as an example: cc1 includes 2 items: #2 D and #3 D, if a student chose only one of them, then his/her score for cc1 is 1/2=0.5.

Moreover, in order to learn more about their reasoning process, we interviewed some students according to their answers, and recorded what they said at the same time.

The Content of the Questionnaire

In order to investigate the status of students' making inference and decision by using statistical reasoning under certain real life background, each item in the SRA describes a statistics or probability problem, with 4~7 choices, most of which included a statement of reasoning, explaining the rationale for the particular choice. The students need to choose what they agreed with.

Results and Analysis

Results

The results are shown in table 1.

Table 1
The SRA Results

SRA Scales	Total Chinese	Total Dutch	Chinese 3rd Grade	Dutch native students
cc1	0.61	0.69	0.65	0.71
cc2	0.63	0.71	0.65	0.77
cc3	0.36	0.41	0.35	0.43
cc4	0.41	0.60	0.51	0.62
cc5	0.26	0.25	0.26	0.30
cc6	0.73	0.70	0.71	0.78
cc7	0.74	0.77	0.71	0.81
cc8	0.66	0.72	0.68	0.73
mc1	0.34	0.37	0.33	0.38
mc2	0.25	0.23	0.23	0.23
mc3	0.31	0.15	0.29	0.15
mc4	0.34	0.28	0.34	0.24
mc5	0.25	0.15	0.19	0.15
mc6	0.13	0.23	0.11	0.19
mc7	0.49	0.57	0.54	0.56
mc8	0.32	0.26	0.28	0.27

Description and Analysis

The total statement about the Chinese high school students. As for correct reasoning, the students are able to distinguish correlation and causation; they can understand the concept of probability, two-way tables and the importance of large samples. On the other hand, they have difficulties in understanding probability calculation and independence. The high scores in “the distinguish between correlation and causation” and “two-way tables” might result from their natures of more like common reasoning skills than statistical reasoning, even though the context is something concerning probability and statistics. For the importance of large samples, since they have learned population, individual, sample and sampling method in junior middle schools, it became easier for them to “feel” and accept it. However, it is hard for them to understand that a small sample is more likely to deviate from the population: their answers are inclined to “Law of small numbers,” that is, the sample “should” be similar to the population, no matter how small it is. The reason for these phenomena is, on one hand, they are not quite clear about the relationship between frequency and probability (this is covered in junior middle school mathematics), on the other hand, it might involve the validation of judgments (Shaughnessy, 1999): When people are estimating the odds of an event that they think is easy to guess, their estimation will be influenced by their personal experience. When making conclusions based on data, Students rely more on personal experience from themselves or people around them than on the statistical result of large samples.

When concerning misconceptions, the most common mistake is “Equiprobability bias. Furthermore, “Good samples have to represent a high percentage of the population,: “Groups can only be compared if they have the same size” and “Law of small numbers” are also very common. It is relatively rare for them to make mistakes about “Outcome orientation,” “Representativeness misconception” and “Correlation implies causation.” When dealing with complicated compound events (here we take a two-step trial as an example), students are inclined to take the first one step trial as a premise of the second one, as a result, to equate two one step trials and one two-step trial. For instance, students who have an equiprobability bias think that when two dice are rolled, all the possible sums are equally likely. They do not realize that the sum of 6 for the two dice is more probable than the sum of 2. This kind of misconception is quite common when students are solving multiple-choice problems. However, when being asked to calculate the probability using permutation and combination knowledge, they can give the right answer. Apparently, when making inference and decisions, the students prefer to use the intuitive statistical reasoning for certain, rather than "making too much fuss about trifles" to calculate the respective probabilities patiently. Moreover, "Law of small numbers" is a typical misconception. Students always ignore the sample size (for example, in question 14, they think the both ratios of boys and girls born in the two hospitals should be 1:1, even though they have been informed that the average daily numbers of newborns in two hospitals are different), and they think it is suitable to get conclusions about the population from small size samples.

(2) The comparison between the Chinese students and the Dutch students

For correct concepts, the performance of Chinese students is not as good as the Dutch. But their total tendencies are the same: both of them did well in "Distinguishes between correlation and causation," "Correctly interprets two-way tables" and "Understands the importance of large samples," while, on the other hand, tended to make mistakes about computing probabilities and "Small samples are more likely to deviate from the population."

As for misconceptions, their levels are pretty even. Chinese students are good at selecting an appropriate average and they make less mistakes in "Correlation implies causation," where the second one might because the center of data is an important part in junior middle school courses, and they have learned in details how to select among median, mode and mean; And their better performance in mc6 might come from the influence of traditional exam criteria: when solving multiple-choice, they'd like to remain single than settle— if not sure about the choice, one would rather not to choose it in order to lose less points. However the Dutch students did not worry too much about this. In addition, the Chinese students made more mistakes about mc3, mc4 and mc5, which are the main components of intuitive inference and bias mentioned in amounts of papers from psychologists, mathematical and statistical education researchers (Shaughnessy, 1999). Therefore, we can say that the Chinese students make more "typical" mistakes when dealing with "typical" misconceptions. Experience shows that even for those researchers who have accepted professional training on statistical application, it is still very common to make such kind of misconceptions (Heuvel-Panhuizen & Wijers, 2005). If this is the case, then it is not surprising to see these kinds of mistakes among students. Moreover, the Chinese and Dutch students get quite even scores in mc2, mc7 and mc8, and their mc7 scores are both over 0.5, implies the commonness of equiprobability bias.

These differences may relate to some factors shown as follows:

First and foremost, it is the stress from entering universities. Chinese high school students have to prepare for the (university) entrance examinations; as a result, they pay most attention to those compulsory subjects (for the part of probability and statistics, only some calculations about probability). Because of this ignorance, they are not able to get enough training about statistical reasoning as the Dutch, who don't have stress on entering universities, therefore are freer to decide the subject and degree of statistical studying.

Equal in importance, it is the teaching philosophy. The Dutch began their reformation from traditional education to realistic mathematics education (RME) at the end of 1960s. So far, the idea and method of realistic education have been commonly accepted by government, society and the public. Consequently, their reformation is done gradually, so it is a quiet reform in the field of mathematical education. The development of what is now known as RME started around 1970. The foundations were laid by Freudenthal and his colleagues at the Institute for Development of Mathematics Education (IOWO), the oldest predecessor of the Freudenthal Institute (FI). RME consists of six principles (Heuvel-Panhuizen & Wijers, 2005): activity principle, reality principle, level principle, intertwinement principle, interaction principle, and guidance principle. These principles, especially the realistic principle, seem quite important for statistical reasoning, which is dealing with data under certain realistic context. Yet in China, although the new curriculum standard emphasizes on problem solving and the application of knowledge in real life, owing to the long-term development of traditional education stresses on exams, there are still a lot left for developing realistic thinking.

Moreover, the difference in statistical reasoning also has to do with the different curriculum and teaching between these two countries, since we are lack of the corresponding materials about the Dutch's curriculum and teaching resources, we do not go into details here.

Conclusions, Reflection and Suggestion

Conclusions

(1) It is common to use intuitive reasoning and misconceptions. The students have difficulties with the comprehension of “small samples are more likely to deviate from the population,” and “independence,” they are not familiar with the calculation of probabilities. When using statistical reasoning, the high school students make more mistakes about “equiprobability bias,” “good samples have to represent a high percentage of the population,” “groups can only be compared if they have the same size,” and “law of small numbers.”

(2) Comparing to the Dutch students, the third grade students in China got lower scores in all the eight correct conceptions. In the aspect of misconceptions, they both have advantages and disadvantages: Chinese students made fewer mistakes concerning “misconceptions involving average”, and “correlation implies causation”, while Dutch students performed better in “good samples have to represent a high percentage of the population”, “law of small numbers” and “representativeness misconception”.

Reflection and Suggestion

About curriculum. It is not suitable to arrange all the probability and statistics subjects separately in two different semesters. In the first place, it goes against helping students get rid of the prior intuition and misconceptions. Our study and other related researches imply that, before learning probability and statistics, the students have already formed some misconceptions, and these misconceptions are quite pertinacious even after getting enough knowledge and students intend to rely more on their intuition after the studying of those subjects. Secondly, although it takes a short time to teach them the statistics conceptions and how to compute basic probabilities, it doesn’t mean the students are successful in avoiding those intuitions which are against theories completely. In fact, without the comparison between intuitive estimation and actual result, the students will depend on their intuition stiffly, and use probability and statistics knowledge unilaterally to support their decisions (for instance, using “law of small numbers” and “equiprobability” are both the one-sided uses of statistical knowledge). Therefore, the teaching of probability and statistics should be combined together. And the time for teaching these two subjects should be prolonged in order to help students “taste” uncertainty gradually.

As for the specific subjects, we need to pay more attention to the introduction of misconceptions. Tracing the problem directly to the acceptance of correct concepts and reasoning does not promise to refuse misconceptions which might be commonly used in their future study and working. Accordingly, the teaching materials should point out that in some situations, intuitive reasoning is unfavorable to us, if this is the case, be careful when using intuition. We also wish to point out that, these intuitive mistakes are not because of their wrong understanding, but just their (also including teachers' and researchers') excessive trust in the effectiveness of intuitive inference.

About teaching. According to our study, we also find that the students are a little conservative when choosing the answers: although most of the items have more than 5 choices, a great number of students only chose one or two answers, which results in "correct less but wrong less." The key factor is they are not familiar with statistical information, thus becoming suspicious and not confident when making a decision, and as a result, underestimate their abilities. Also, since the traditional science problems always have certain answers —you can get only one answer through calculation. When confronting problems about uncertainty, they continue to use traditional strategy to choose the one they are quite sure about, for the sake of getting higher scores. The students get used to computing and remembering; when they are asked to observe, judge or infer, they are at a loss where to start. Based on our study, we propose several suggestions about teaching:

(1) Providing real, abundant statistical cases. The more substantial the data and context are, the more students can realize the importance of statistical method in daily life, therefore they will devote to the reasoning effectively; the larger the amounts are, the more students are accustomed to dealing with real life problems and uncertainty. In addition, through these interesting cases, they can apply statistical method as a tool to solve realistic problems, rather than just regard them as a disorderly accumulation of unrelated formulas. Whereas it is not easy to find cases that can satisfy all students' interests, what we can do is to cover the most. There is a great amount of information about collecting real data. We may quote a simple example of inputting "Favorite Data Sets" in searching engine, there will appear scores of links, for instance: http://www88.homepage.villanova.edu/thomas.slhort/Bradstreet/ provides data about health care and medicine; http://course.ncssm.edu/math/CPTA/data/dataold.htm supplies data that can be used in education. Such kinds of data are too numerous to list, so it is not a big deal to find suitable data. Yet these real data also have their own deficiencies.

For instance, they are too complicated and not easy to analyze. Consequently, we need to trade it off by doing some adjustments, on condition that the data should be based on real problems. It is also very efficient and effective to use data around students.

(2) For those misconceptions, we need to pay more attention. For those contents, not only do we need to explain the corresponding correct concepts, reasoning, and instruct them how to ponder, but we also need to begin with "the reverse cases," and execute "error correction," for the purpose of letting students know what is not correct, and where the problems lie. By doing so, the misconceptions will impress them; "The reverse cases" can also train students' critical thinking. In addition, since students always use response tendency under certain background, for example, they tend to pay attention to minor details, deduce the examiner's purpose according to some words or statements, and thereby ignore the key part. In order to get rid of this, the teacher should emphasize on the common mistakes at the beginning of teaching, not only analyze the connotation and thinking, but also execute "case teaching." Those counterexamples can increase their sensitivity about misconceptions and eliminate inertia. In addition, asking students to compare their estimation with the actual result can also help them to realize intuitive mistakes and recognize correct reasoning.

(3) Carrying out "intimacy teaching." A cordial/amiable/friendly relationship between teachers and students can influence a lot on the effect of teaching statistical reasoning. A good relationship includes: solicitude, dedication, communication and attraction (Bradstreet, 1996). Since there is no "best" statistical method, but only "better" ones, it is more suitable to use intimacy teaching. It means to put the teacher in the same status as students, constructing an equal, cooperative relationship among them. When solving a new problem, the teacher can use the way of "thinking aloud" in the class, that is, providing the total reasoning process, and encourage students to question what "the authority" said. It is not serious for the teacher to make a mistake, moreover, sometimes the teacher should make some mistakes on purpose, so as to show that everybody is not genius, yet everyone can find better ways. Similarly, teachers and students can analyze a set of new data together: no one knows the right in advance. We suggest so because we can always learn a lot from failure and mistakes, not only from ourselves, but also from others around us. Thus the teacher should dare to make mistakes in front of students, by doing so, the teacher can shorten the distance with the students, and increase intimacy.

(4) Making most of the cases from the internet. The internet provides recent designs and cases, which can be used in developing students' statistical reasoning. A case in point is through the activity of combining oral description with data scatter plots, we can develop students' reasoning about data and distribution (Rossman & Chance, 2001); by encouraging students to think the factor affecting standard deviation, we help them to develop the reasoning about variation; leading the students to simulate the picking of random samples, changing the sample size and population parameters can help them to understand sampling distribution. However, we need to mention that the internet resources are double-edged swords, therefore, the teacher needs to design teaching methods according to students' status and teaching conditions.

Acknowledgements

During this study, we got a lot of help from Dr. Dirk Tempelaar in Lawrence University, who shared his experience, information and also gave us quite constructive suggestions. We also thank for Dr. Joan B Garfield, who developed the SRA instrument.

References

Bradstreet, T. (1996). Teaching introductory statistics courses so that nonstatisticians experience statistical reasoning. *The American Statistician*, *50*(1), 69.

Garfield, J. (2003). Assessing statistical reasoning. *Statistical Education Research Journal*, *2*(1), 22-38.

Heuvel-Panhuizen, M., & Wijers, M. (2005). Mathematics standards and curricula in the Netherlands. *ZDM, 37*(4), 287-293.

Rossman, A., & Chance, B. (2001). *Workshop statistics: Discovery with data.* New York: Springer Verlag.

Shaughnessy, M. (1999). Research on statistics learning and reasoning. In Grouws D.A. (Eds.), *Handbook of research on mathematics teaching and learning*. Shanghai, China: Shanghai Education Publishing House.

Tempelaar, D. (2004, Month). Statistical reasoning assessment: An analysis of the SRA instrument. Presentation to Assessment in Statistics, Lawrence University. Appleton, Wisconsin.

Authors:
Wenfang Wang
Northeast Normal University, China
Email: wwflzq6@yahoo.com.cn

Xiaohui Wang
Northeast Normal University, China
Email: wangxh950@nenu.edu.cn

Guofang Chen
Northeast Normal University, China
Email: wangxh950@nenu.edu.cn