Assessment of statistical education in Indonesia: Preliminary results and initiation to simulation-based inference

K V I Saputra[1], L Cahyadi[1] and U A Sembiring[1]

[1]Department of Mathematics, Universitas Pelita Harapan, Jl MH Thamrin Boulevard, Lippo Karawaci, Tangerang, Banten, Indonesia

E-mail: kie.saputra@uph.edu

Abstract. Start in this paper, we assess our traditional elementary statistics education and also we introduce elementary statistics with simulation-based inference. To assess our statistical class, we adapt the well-known CAOS (Comprehensive Assessment of Outcomes in Statistics) test that serves as an external measure to assess the student's basic statistical literacy. This test generally represents as an accepted measure of statistical literacy. We also introduce a new teaching method on elementary statistics class. Different from the traditional elementary statistics course, we will introduce a simulation-based inference method to conduct hypothesis testing. From the literature, it has shown that this new teaching method works very well in increasing student's understanding of statistics.

1. Introduction

In the undergraduate education in Indonesia, statistics is one of the few courses that are taught in almost every department whether the department is in engineering faculty, natural science faculty, social science faculty or in business school. Statistics course sometimes is renamed as something else such as Elementary Statistics, Introductory Statistics, or even Biostatistics in biology department but the contents of those are just the same. There have been a lot of books and resources on Elementary Statistics that first-year undergraduate students can have whether it is in Indonesian or in English but the contents of such books are also conceptually the same. The fact that many students in Indonesia are still afraid of these statistics courses means that the way of teaching statistics both conceptually and pedagogically has to be evaluated. However, there has been an effort neither from the government nor from academia to assess the statistical education in undergraduate institutions.

We believe that the Elementary Statistics course that has been taught in Indonesian Institution follows the algebra-based introductory statistics course that has been implemented in a number of popular textbooks (see for instances [1, 2, 3, 4]. These books tend to the same four parts (although some books mentioned three parts as the first two are considered to be in one part). The parts are descriptive statistics; data collection and design; probability and sampling distributions and some knowledge about basic random variables; and inferential statistics. In some literatures (see [5, 6]) they found that by the time students get to inferential statistics (confidence intervals and hypothesis testing), which is the pivotal point of this course, they end up in "survival" mode, in which they "just" memorize formulas without really understanding the underlying concepts of how to draw valid and reliable conclusions from data. With this sequencing, Cobb [7] argues that student felt comfortable learning the first two parts, however, their understanding somehow dissipated as they landed in the

third part as it conceptually difficult, and technically disconnected with real data analysis that was learned in the beginning. By the final fourth of the course, many students were not able to connect the dots in the previous part and were not able to grasp the idea of inferential statistics. This problem was also combined by the "stress" students had to experience by the end of the semester and resulted in a little understanding of what is statistics really about.

2. Methods

2.1. Recent Development of Elementary Statistics in United States

As far as the author knows, development of elementary statistics course probably started in 1997 when Moore [8] challenged all statistical educators to change their teaching pedagogy from the old model to the new model. The old model, as we also know, is more teacher-centered where students learn by absorbing information, while the new model is more student-centered, in which students learn through their own activities. Then there was relatively no result until 2005 when the American Statistics Association published the Guidelines for Assessment and Instruction in Statistical Education, also known as GAISE (see [9]). This guideline gave recommendations which at that time had not been implemented by either any instructor in their elementary statistics classes or in many statistics textbooks. Some recommendations in GAISE are: (1) to develop statistical thinking, (2) to use real data, (3) to emphasize statistical concepts rather than procedures, (4) to focus on student active learning, (5) to use technology in analyzing data, and (6) to use assessment to evaluate and to improve the learning process. However, Cobb [7] again argued that this guide along with its recommendation is not enough and urged statistical educators to reconsider the pedagogy and the content of the elementary statistics course.

Along with the above guide, in 2007 Delmas et al. [10] developed a comprehensive assessment of student understanding in the elementary statistics course, also known as the CAOS test. This assessment was the first standard, comprehensive assessment instrument for the introductory statistics course. When it first implemented before class and after class, the result was startling as the students only gaining average 9% difference, which implying most students didn't really understand about statistics. It also implied that the statistical educator must do something about this. Other assessments that have been used to evaluate elementary statistics classes are SATS (Survey Attitudes Toward Statistics) [11] and also GOALS-2 (The Goals and Outcomes Associated with Learning Statistics) [12]. Both have been successfully evaluated the impact of introductory statistics classes in U.S. institutions.

Tintle et al. [5] as far as the author knows was the first to develop a new pedagogy in teaching statistics. They not only took a student active learning approach which also is in line with the GAISE recommendation but also completely re-ordering, re-emphasizing, adding and subtracting content of the traditional elementary statistical course. The result is quite significant as it showed that there was a strong evidence (p-value < 0.001) that the student average in the new pedagogy is higher than that in the current pedagogy. Recently, Chance et al. [13] also showed a promising evidence that the simulation-based inference (SBI) curriculum works better than the traditional curriculum. The article analyzed multi-institution data which are pre/post of student attitudes and conceptual understanding of topics in statistics course in those institutions.

2.2. Development of THE simulation-based inference (SBI) CURRICULUM

The need of changing the pedagogy also is motivated by the fact that some statistical concept was taught in high school. High school education in US has already seen much of the material in the first part of the traditional statistics course [9]. This does not only happen in US but it also happens in many countries such as Indonesia [14]. Indonesia high school education currently uses the 2013 curriculum which exposes students to learn statistics descriptive. They will learn as much as center data (mean, median), variability, histogram and other concepts of presenting data. This means that the elementary statistics material taught in first-year college or first-year university has to be reshaped.

Another motivation of changing the pedagogy is the number of exceptions given in the traditional statistics course. As an example, suppose we wanted to test of hypothesis for the population mean. We use the z distribution to perform a test if the number of samples is more than 30. However, if the number of samples is less than 30 we must use the t distribution to perform such a test. But sometimes the population distribution is known normal and the population standard deviation is given, so even if the number of samples is less than 30, we can still use the z distribution to perform this test. This exception is one of many that not only confuse students but also drive the student to just memorize the rule and the formula.

In our statistics course with simulation-based inference (SBI), we shall teach students data collection, formulation research question, hypothesis testing, and statistical inference in the first week. We believe students have to know the statistical process from the beginning until making conclusion. We also emphasize the logic of statistical inference by using simulation test, and therefore we will integrate computer technology into our curriculum. The students have to be active in our class, because we will expose them some real problems that are aligned with the topic we are discussing. We highlight the following:

1. We only review topics on descriptive statistics but do not spend explicit time discussing this topic
2. After establishing the logic of inference, students are asked to draw a connection between simulation-based and theory (asymptotic) based
3. We present confidence interval as a result of test of significance, not the other way around
4. We underline the difference between observational study and experiment
5. We utilize statistical process in a large-scale research project

We now give an example of the syllabus of the traditional elementary statistics course which one can see from Table 1. We assume that the course will be taught in 16 weeks with mid-semester exam at week 8 and final semester exam at week 16. As we have stated before, typical syllabus of traditional elementary statistics course consists of four parts, i.e. data analysis (week 1-2), descriptive statistics (week 3), probability theory and random variables (week 4-7), and inferential statistics (week 9-15). The students first learn about hypothesis testing in week 10, which is near the end of the course. Some students may be able to understand the concept of inferential but most of them will not be able to understand the first three parts and how to use them in inferential statistics.

We would like to give also an example of the syllabus of the elementary statistics with simulation-based inference. As one can see from Table 2, we already introduce statistical inference the first two weeks. We then expose the student to the logic of statistical inference every week, adding more and more statistical concepts as time goes on. The descriptive statistics will be reviewed as needed but we do not dedicate a time slot for it. We also replace the probability theory and random variables part in the traditional statistics course with simulation-based. We also will compare the simulation-based inference with the theory-based but we shall explain to the students that the theory behind the theory-based inference will be taught in the next statistics course.

Table 1. Typical syllabus of the traditional elementary statistics course. This may vary depending on the textbook used by lecturer and also depend on the faculty this course taught

Week	Topics	Learning Outcomes
1	Introduction to Statistics	• Understand statistics and type of statistics • Understand difference between sample and population • Understand type of variables
2	Organizing and Graphing Data	• Understand how to organizing and graphing qualitative and quantitative data • Able to make cumulative frequency distributions • Able to make graphical representation (stem and leaf, dot-plots) of data

3	Descriptive Statistics	• Understand measurement of central tendency of data • Understand measurement of dispersion of data • Able to make box-plot data
4	Probability Theory	• Understand the theory of probability • Understand addition and multiplication rule to compute probability • Understand permutation and combination
5	Some Discrete Distributions	• Understand probability distribution of discrete random variables • Understand binomial, Hyper-geometric and Poisson distributions
6	Some Continuous Distributions	• Understand normal distribution and standard normal distributions • Able to determine probability of normal distribution using normal table and its inverse • Understand normal distribution approximation using binomial distribution
7	Sampling Distributions	• Understand sampling distribution • Understand sampling error and non-sampling error • Understand mean and standard deviation of $\bar{x}$ • Understand mean and standard deviation of $\hat{p}$
8	Mid Semester Exam	
9	Estimation of mean and proportion	• Understand estimation of μ when σ is known • Understand estimation of μ when σ is not known • Understand estimation of population proportion
10	Hypothesis testing about mean and proportion	• Hypothesis testing of μ when σ is known • Hypothesis testing of μ when σ is not known • Hypothesis testing of population proportion
11	Estimation and Hypothesis Testing: Two Populations	• Hypothesis testing and estimation of two population when $\sigma1$ and $\sigma2$ are known • Hypothesis testing and estimation of two population when $\sigma1$ and $\sigma2$ are unknown but equal
12	Estimation and Hypothesis Testing: Two Populations	• Hypothesis testing and estimation of two population when $\sigma1$ and $\sigma2$ are unknown and unequal • Hypothesis testing and estimation of two population for paired sample • Hypothesis testing and estimation of two population proportion
13	Chi-Square Tests	• Understand chi-square distribution • Understand a goodness of fit test
14	Analysis of Variance	• Understand F distribution • Understand one-way analysis of variance
15	Simple Linear Regression	• Able to compute simple linear regression • Understand inferences of slope • Able to compute correlation • Understand application of regression model
16	Final Semester Examination	

Table 2. Example syllabus of the elementary statistics with simulation-based inference. This is one example of syllabus that has been implemented in the Department of Mathematics,

Universitas Pelita Harapan.

Week	Topics	Learning Outcomes
1	Introduction to Statistics (P1 - P3)	• Understand six step statistics process • Understand how to explore data • Understand random process • Small Project
2	Significance: How strong is the evidence? One proportion test (1.1 - 1.5)	• Understand chance model in one proportion test • Understand measuring strength of evidence; p-value • Understand factors that affect strength of evidence • One proportion; theory-based approach • Small Project
3	Generalization: How Broadly Do the Results Apply? One mean test (2.1 - 2.3)	• Understand chance model in one mean test • Understand sampling in finite population • Understand error and significant • Small Project
4	Estimation: How Large is the Effect? Confidence Interval (3.1 - 3.3)	• Understand the concept of confidence interval • Understand 2SD method for estimation of proportion and mean populations • Small Project
5	Estimation: How Large is the Effect? Confidence Interval (3.2 - 3.5)	• Understand the theory-based method for estimation of proportion and mean populations • Understand factors that affect confidence interval • Small Project
6	Causation: Can We Say What Caused the Effect? (4.1-4.2)	• Understand definition of causation and confounding • Understand observational studies and experiments • Small Project
7	Comparing Two Proportions (5.1 - 5.3)	• Understand hypothesis testing in more general situation • Understand how to compare two proportions: simulation-based • Understand how to compare two proportions: theory-based • Small Project
8	Mid Semester Exam	
9	Comparing Two Means (6.1 - 6.3)	• Understand how to compare two means: simulation-based • Understand how to compare two means: theory-based • Small Project
10	Paired Data: One Quantitative Variable (7.1 - 7.3)	• Understand pair design vs two mean test • Understand how to do a paired test: simulation-based • Understand how to do a paired test: theory-based • Small Project
11	Generate population/chance model (Review Chapter 5,6,7)	• Understand how to generate simulation for two proportion test • Understand how to generate simulation for two mean test • Understand how to generate simulation for a paired test • Small Project
12	Comparing More Than Two Proportions	• Understanding simulation-based approach to compare multiple proportions

	(8.1-8.2)	• Understanding the theory-based approach to compare multiple proportions • Small Project
13	Comparing More Than Two Means (9.1-9.2)	• Understanding simulation-based approach to compare multiple means • Understanding the theory-based approach to compare multiple means • Small Project
14	Two Quantitative Variables (10.1-10.2)	• Able to compute correlation between two quantitative variables • Inference of correlation: simulation-based approach • Small Project
15	Two Quantitative Variables (10.3-10.5)	• Able to compute simple linear regression • Inference of slope: simulation-based approach • Inference of slope: theory-based approach • Small Project
16	Final Semester Examination	

We now would like to give an example of simulation-based inference on hypothesis testing. Suppose we want to do a hypothesis testing on population proportion. We want to check whether the null hypothesis (H0: $\pi = 0.5$) is true or the alternative hypothesis is true (Ha: $\pi > 0.5$). From the sample, we are given that there are 8 correct trials out of 10 trials. In this case, we could not use the theory-based inference, as our sample is not large enough. In case the sample is large enough, we need to compute the statistic z to determine whether we can reject H0. However, using simulation-based inference, we could easily do this. When the null hypothesis is true, we have $\pi = 0.5$, which is equivalent to a coin toss process. Therefore, we are going to toss a coin ten times and compute how many times the coin landed on head. We shall repeat this simulation many times (say 500 times) and we shall compute a number of simulations that landed on head more than 8 times. This proportion (relative to 500) is an approximation of p-value. Using this logic of inference, students will do some experiments on tossing coins. Each of them will have to toss a coin ten times and record how many times it landed on head. They will understand that if p-value is small then the sample (8 out of 10) provides a strong evidence against the null hypothesis. But if the p-value is large then the sample gives us not enough evidence to reject the null hypothesis. The student will use computer to simulate this many times. We will also introduce the theory-based inference as an alternative.

3. Results and Discussion

In this preliminary stage, we conducted a pre and post survey on students of batch 2015 from Mathematics Department at Universitas Pelita Harapan in the Even Semester 2015/2016. In this semester the traditional elementary statistics was taught. Students would complete our survey that consists of their demography, their attitude toward statistics, and their understanding of some statistical concepts, in the first week and again during the last week of the semester. We are going to compare the result of these two surveys and we want to see the difference between the two. In our next paper, we will conduct the same analysis of pre and post survey on students of batch 2016 from Mathematics Department in the Odd Semester 2016/2017. This batch, however, will use the elementary statistics using simulation-based inference (SBI).

The first analysis that we are going to do is about the student's attitude toward statistics. One can see the result of our survey in Table 3. The survey about student's attitude toward statistics was using Likert-type scale. Before the class started, student's attitude toward statistics is rather high, that is about 62%. It shows that on average students understand that statistics is important. The standard deviation is also small, meaning that there is not a lot of variation here. When we conducted the post-survey, the result is about the same. The average student's attitude toward statistics is 63% but the

variation is now increasing. We suspect that some of the students find statistics difficult and decided to have a non-statistics career.

Table 3. Comparison of student's attitude toward statistics before and after class.

	n	Average (%) student's attitude toward statistics	SD
Pre Test	32	62.7	4.8
Post Test	24	63.7	6.2

Next, we would like to analyze the student's understanding of some statistics concept. One can see the result in Table 4. In this survey, we test some concepts in statistics, such as data collection, descriptive statistics, confidence interval, scope of conclusion, significance and simulation. On average, the score of the students before the class started is 43.2% however, after the semester ended the average score of students is 40.8%. This is the thing that we expected. After the semester ended, we ought to have an increasing value but instead, we could not see the improvement of student understanding. The variation is not so much changed.

Table 4. Comparison of student's understanding to some statistical concepts before and after class.

	n	Average (%) student's understanding	SD
Pre Test	32	43.2	9.4
Post Test	24	40.8	10.4
Change	-	-2.4	1

We can also break down the concept to see which of the concept students can easily understand. One can see in Table 5. It is somehow interesting that the score after the semester ended is lower than the score before the semester started. This could be because of many reasons. The important point is that the students are not getting better in statistics.

Table 5. Comparison of student's understanding to each of statistical concepts before and after class.

	Average Percent Correct Pre	SD	Average Percent Correct Post	SD	Average Change	SD
Data collection	46.1	29.5	35.9	19.3	-10.2	-10.2
Descriptive Statistics	41.5	20.6	39.8	19.7	-1.7	-0.9
Confidence Interval	31.8	20	32.6	24.3	0.8	4.3
Scope of conclusion	54.7	34.5	41.7	28.2	-13	-6.3
Significance	51.8	16.9	51.3	12.6	-0.5	-4.3
Simulation	36.3	17.9	32.5	18.4	-3.8	0.5

Our last analysis is a pair test. In previous analysis, we assume that the data is independent, so that we can conduct a two mean test. However, it is obvious that the data is not independent. Since the number of participants is not the same, we have to only consider participants that are active in both pre

and post survey. In all cases, we also find that p-value is greater than 10%. Therefore we cannot conclude that there is a significant difference between the pre and the post.

Table 6. Pair data analysis of student's understanding to each of statistical concepts before and after class.

	Average Percent Correct Pre	SD	p-value
Mean difference	-1.598	29.5	0.52
Standard deviation	10.9	20.6	

4. Conclusion

In our analysis above, we assume that the data is reliable. However, this might not be true. We have to consider the situation of students who conducted the test. When they conducted the pre-survey, they might consider this seriously, but when the semester has ended, they might just do it recklessly. We do not give them benefit for writing this survey. On top of it, the survey itself is a long survey and it consists of more than 50 questions.

Next thing we want to do is to conduct another survey on students, but in this case, we are going to use the simulation-based inference curriculum. We also want to spread this new teaching method through the website, workshop and other publications (see details on Fig. 1).
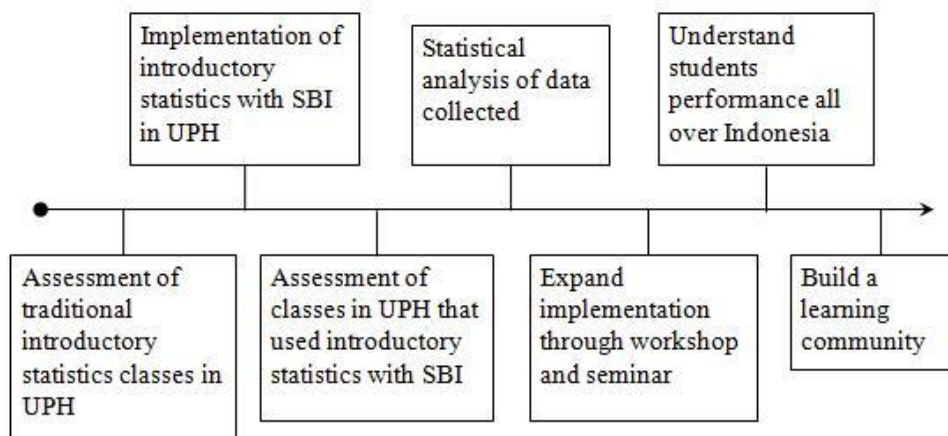

Figure 1. Timeline of our project on implementation of elementary statistics with simulation-based inference.

References

[1] Mann P S 2007 Introductory Statistics (Hoboken, NJ: John Wiley & Sons)
[2] Newbold P, Carlson W and Thorne B 2012 Statistics for business and economics (New Jersey: Pearson)
[3] Wild C J and Seber G A F 2000 Chance encounters: The first course in Data Analysis and Inference (Hoboken, NJ: John Wiley & Sons)
[4] Walpole R E et al. 1993 Probability and statistics for engineers and scientists **5** (New York: Macmillan)
[5] Tintle N, VanderStoep J, Holmes V-L, Quisenberry B and Swanson T 2011 *Journal of Statistics Education* **19** 1 pp 1-25
[6] Tintle N, Topliff K, VanderStoep J, Holmes V-L and Swanson T 2012 *Statistics Education Research Journal* **11** 1 pp 21-40
[7] Cobb G W 2007 *Technology Innovations in Statistics Education* **1** 1 pp 1-16
[8] Moore D S 1997 *International statistical review* **65** 2 pp 123-165

[9]   Aliaga M, Cuff C, Garfield J, Lock R, Utts J and Witmer J 2005 Guidelines for Assessment and Instruction in Statistics Education (GAISE): College Report. American Statistical Association.

[10] Delmas R, Garfield J, Ooms A and Chance B 2007 *Statistics Education Research Journal* **6** 2 pp 28

[11] Schau C and Emmioğlu E 2012 *Statistics Education Research Journal* **11** 2 pp 86-94.

[12] Sabbag A G and Zieffler A 2015 *Statistics Education Research Journal* **14** 2

[13] Chance B, Wong J and Tintle N 2016 *Journal of Statistics Education* **24** 3 pp 114-126

Acknowledgments