Acta Med. Okayama, 2021
Vol. 75, No. 5, pp. 601-609

Acta Medica Okayama
http://escholarship.lib.okayama-u.ac.jp/amo/

Original Article

The Effect of Peer Instruction Lectures on Learning Attitudes in Epidemiology Education

Toshiharu Mitsuhashi*

Center for Innovative Clinical Medicine, Okayama University Hospital, Okayama 700-8558, Japan

Research suggests that the fundamental concepts of epidemiology cannot be sufficiently learned in traditional lectures, and interactive learning is necessary. However, few studies have investigated interactive epidemiology education in general, or peer instruction (PI) in particular. This study investigated the effect of PI. Study participants were fourth-year medical students. The attitude of participants in regard to PI learning was examined in a non-PI and a PI group. The Survey of Attitudes Toward Statistics (SATS) (containing six sub-categories) was conducted as a learning-attitudes index. The pre- and post-lecture scores were compared between the non-PI and PI groups using double robust (DR) estimation. The non-PI and PI groups consisted of 20 and 121 student participants, respectively. In DR estimation, affect exhibited the lowest SATS score changes, at −0.51 (95% confidence interval −0.78 to −0.24; *p*-value < 0.001), whereas effort exhibited the highest score changes of 0.01 (95% confidence interval −0.30 to 0.32; *p*-value = 0.952). The epidemiology lecture with PI did not increase the SATS scores. This might be due to issues related to the experimental design. Further research investigating the effects of interactive epidemiology education, it will be necessary to develop tools for assessing the learning of epidemiological concepts and to improve the research design.

Key words: medical students, peer instruction, epidemiological education, learning attitude, double robust estimation

Clinical research requires proper management and analysis of clinical data. Managing epidemiological and biostatistical concerns without building a detailed understanding of the given concept may damage the research validity and result in inefficient use of resources [1]. According to Ioannidis [2], most studies, from clinical trials and traditional epidemiology studies to the most modern molecular research, report inaccurate results because of study design errors and/or biases. Various measures will be needed to solve this problem, and appropriate epidemiology and biostatistics education will be one of them.

However, there are some challenges in making the necessary improvements, such as clinicians' reports of having insufficient access to learning opportunities. In fact, one study has reported that 37% of residents received no formal statistical literacy training during the residency period [3]. Nevertheless, the epidemiological study designs and statistical methods used in clinical research are becoming increasingly sophisticated and complex [4-6]. Learning to use the latest modern epidemiological study designs and statistical methods can be difficult, and developing an understanding of the fundamental concepts can help to ease such difficulty [7,8]. Therefore, learning epidemiology and biostatistics during undergraduate education is essential. However, statistics cannot be learned by only passively

Received February 13, 2021; accepted April 30, 2021.
*Corresponding author. Phone : +81-86-235-6504; Fax : +81-86-235-6505
E-mail : mitsuh-t@cc.okayama-u.ac.jp (T. Mitsuhashi)

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

listening to a lecture where the instructor talks, and the students listen [9-11]. Merely listening to a lecture, a common learning style in Japanese universities, has limited usefulness for providing learners with a detailed understanding of fundamental epidemiology and biostatistics concepts.

Physics educators and researchers have developed many interactive education methods to promote an understanding of fundamental concepts [12]. Among them, peer instruction (PI) has received significant attention [13-15]. PI is more advantageous than other interactive learning styles because, regardless of the number of students, it can be easily incorporated into conventional lectures and implemented [16].

Interactive lecture styles are often employed for teaching epidemiology [17-20]. There are some implementation problems, such as difficulties with integration into conventional lectures and limitations on the number of students. For example, in problem-based learning (PBL), students work in groups of 7 or 8 under a facilitator's supervision. PBL requires participation from facilitators as well as the instructor, and it is not possible to educate many people at the same time. In contrast, PI lectures have fewer limitations and many distinct advantages. Nonetheless, to the best of my knowledge, no reports on PI lectures for epidemiology have been published. No research has identified differences in learning attitudes between PI lectures and conventional lectures. This study aimed to investigate such differences in learning attitudes (defined by changes in attitude toward a given field of study) by comparing PI and conventional lectures.

Materials and Methods

Study subjects and settings. This cohort study targeted fourth-year medical students at Okayama University and was conducted in 2018 and 2019. A conventional lecture and a PI lecture were held as part of a 1-day epidemiology exercise course. This course was offered as part of the hygiene lecture, and until 2018, students could choose from among this 1-day epidemiology exercise course or various alternative exercise courses. All students had already attended an epidemiology lecture and thus had prior knowledge about epidemiology. All the exercises were designed for fourth-year medical students and conducted by the author. In 2018, 20 students who had chosen to attend the 1-day epidemiology exercise course participated in a conventional lecture without PI. In 2019, 137 students who attended the 1-day epidemiology exercise course participated in a PI lecture. The difference in these numbers can be attributed to the differences between the university's FY2018 and FY2019 curricula; in FY2019, all the students took the 1-day epidemiology course, as all the other exercises had been dropped.

PI was introduced in 2019; however, the lecture topics were the same as in 2018. The fundamental concepts of epidemiology were explained, and then the ConcepTest was conducted to confirm student understanding, and discussions about the ConcepTest were held among students and facilitated by the author. This PI session focused on five topics. The conventional lecture was limited to an explanation of these five fundamental topics, whereas the PI lecture covered these topics, and then followed up with PI session. The time spent on each topic was about 16 min in each lecture style (the entire lecture was about 80 min). The lecture topics were: (1) Epidemiological indicators (risk, rate and prevalence), (2) Descriptive epidemiology (spot map and epidemic curve), (3) Cohort and case-control studies (study concept and interpretation of a two-by-two table), (4) Random error (error evaluation and interpretation of confidence interval), and (5) Systematic error (selection bias, information bias, and confounding bias). The details of the lecture and ConcepTest in PI have been described in previous research [21]. After the epidemiology lecture, a statistical analysis was performed using the statistical software Epi Info 7 (www.cdc.gov/epiinfo/index.html) in the same manner and in both years.

Data collection. Data were collected from students using a paper questionnaire. Students were asked to respond before and after the epidemiology exercise. Data were not originally obtained for research purposes, but rather annually to measure the learning attitude toward lectures. An assistant entered the questionnaire responses into a computer, and the author verified they were entered correctly.

Exposure variable. Students participating in 2018 and 2019 were categorized as the non-PI group and the PI group, respectively.

Outcome variable. The Survey of Attitudes Toward Statistics (SATS) was administered [22]. The SATS is a 36-item questionnaire that uses a 7-point Likert scale to score attitudes toward statistics in 6 categories: interest, effort, affect, cognitive competence, value, and difficulty. In this study, the SATS was used as an evaluation index for epidemiology education. SATS was originally used as an index for measuring attitudes toward statistics, not epidemiology. However, since SATS was the only evaluation index available in fiscal year 2018, which was prior to the planning stage of the study, SATS was used as the evaluation index. The SATS was administered before and after the lecture, and the difference in students' SATS scores between the two time points was calculated. Group differences between the PI and non-PI groups between before and after the lectures were defined as the primary outcome, as expressed in the equation below.

$$(after\ score - before\ score)_{PI-group} - (after\ score - before\ score)_{\mathrm{non}PI-group}$$

Candidates for confounding factors. Since this was an observational study, it was subject to the influence of confounding factors when making causal inferences. Therefore, it was necessary to adjust for these factors. In this study, the following candidates for confounding factors were identified.

Gender and self-assessment questions were used as covariates and were provided in a questionnaire during the SATS. For self-assessment of proficiency in mathematics and epidemiology, responses to four additional questions were acquired using a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree). The questions addressed in the tests were as follows:

1. Self-rated mathematics skill: Were you good at mathematics in high school?
2. Expectancy of epidemiology use after graduation: In your opinion, to what extent is epidemiology used in jobs after graduation?
3. Confidence in acquiring the content of the lecture: How confident are you that you acquired the content of today's lecture?
4. Willingness to select epidemiology lectures again: If you have the opportunity to select a lecture in the future, will you choose a lecture on epidemiology?

Covariate selection. The participants' pre-lecture SATS scores, gender, and self-assessment were selected as covariates, taking into account how the course was conducted [21], gender differences in math skills [23], and factors affecting response rates [24].

Statistical methods. All analyses were performed using Stata 16.0 (StataCorp LP, College Station, TX, USA). $P < 0.05$ (two-sided) were considered statistically significant. Assuming a large effect of the PI lecture on SATS score improvement (*i.e.*, Cohen's $d = 0.80$) for the non-PI group of 20 and PI group of 137, the statistical power was 0.91.

All the materials needed for the analyses are reposited on GitHub (https://github.com/sankyoh/PI_epidemiology_education).

1. Descriptive statistics

Baseline characteristics were defined using the mean and standard deviation (SD) for quantitative information. Frequency and percentages were included wherever required.

2. Handling missing values

Missing values were addressed in the complete case analysis. This study may have comprised unmeasured confounders, and missing values were randomly observed. However, a complete case analysis was used in order to yield more accurate or less biased results compared to other methods, such as multiple imputation [25].

3. Crude analysis

A linear regression analysis was performed to calculate coefficients and their 95% confidence intervals (CIs) by using the lecture style and SATS scores as the explanatory and response variables, respectively.

4. Double robust analysis

A double robust (DR) analysis was performed to calculate the average treatment effects (ATEs) and their 95% CIs from the observed data. For DR calculation, weighted regression coefficients were used to compute averages of predicted outcomes of each lecture style, where the weights were estimated using propensity scores (PSs). The contrasts between these averages represent the treatment effects. Moreover, 95% CIs were estimated using a bootstrap approach [26].

The treatment probabilities were estimated as the PSs by using a logistic regression analysis; this was performed using the variables specified by the DAG and the lecture style as explanatory variables and the response variable, respectively. The weights (overlap weights) were calculated based on the PS by using the following equation [27]:

$$overlap\ \ weight = (PI\ \ group) \times (1 - propensity\ \ score) + (\mathrm{non}\ PI\ group) \times propensity\ score$$

5. Evaluating the balance and propensity score overlap

The absolute standardized mean difference (SMD) and variance ratio were used for evaluating baseline covariates before and after the weighting. Moreover, a variance ratio with a range of 0.5-2.0 is considered a balanced variance ratio [28].

To validate the PS overlap, histograms were developed for both groups.

Ethical issues. The study was approved by the Okayama University Graduate School of Medicine, Dentistry, and Pharmaceutical Sciences and Okayama University Hospital Ethics Committee (approval number K1909-037). No reward was provided to any student for participating in this study.

The research was drafted in 2019, and the information was disclosed to students on the university website in 2018. They were informed about the purpose and methods of this research and that their responses to the paper questionnaire would be used for the study. However, they were also informed that since their names were removed from the paper questionnaire, their data could not be excluded from the analysis, even if they refused to participate in the study.

For students in 2019, the purpose and methods of the study were appropriately presented in writing. After reading this material, students provided their informed consent. Students could refuse to participate for any reason.

Results

Descriptive statistics. In the PI group, 121 (88.3%) of the 137 participants answered the questionnaire, whereas in the non-PI group, all 20 (100%) participants responded. The data for 87 (63.5%) and 14 (70.0%) participants in the PI and non-PI groups, respectively, had no missing values. Table 1 shows the baseline characteristics. The numbers of female and male students were 34 and 87 in the PI group, and 10 and 10 in the non-PI group, respectively. In the self-assessment, all questions exhibited higher mean values in the non-PI group. The pre-lecture SATS scores were higher for interest, effort, affect, and value in the non-PI group, and were higher for cognitive competence and difficulty in the PI group.

Table 2 shows the differences in the SATS scores between before and after the lecture in each group. For effort, the PI group exhibited a higher increment than the non-PI group (0.529 vs 0.286). However, for all other outcomes, the non-PI group exhibited a higher increment than the PI group.

Evaluation of balance and PS overlap. Table 3 presents the SMD and variance ratio. After weighting, the mean of the SMD and variance ratio decreased to 0.00 and 1.26, respectively, both of which are considered well balanced. Table 4 shows the descriptive statis-

Table 1 Descriptive characteristics of the 141 students

		Non-PI group (*n*=20)		PI group (*n*=121)	
		Mean (SD) / *n* (%)	Missing	Mean (SD) / *n* (%)	Missing
Gender	Female	10 (50%)	0	34 (28.1%)	0
	Male	10 (50%)	0	87 (71.9%)	0
Self-assessment [a]					
	Self-rated mathematics skill	4.75 (1.84)	4	4.66 (1.64)	18
	Expected use of epidemiology after graduation	4.94 (1.18)	4	4.83 (1.11)	18
	Confidence in acquiring the content of the lecture	4.19 (1.11)	4	4.02 (1.17)	18
	Willingness to choose epidemiology lectures again	4.81 (1.05)	4	3.93 (35)	18
Pre-lecture SATS score					
	Interest	4.94 (1.07)	0	4.53 (1.09)	2
	Effort	6.05 (1.05)	0	5.32 (1.00)	0
	Affect	3.70 (0.89)	0	3.68 (0.85)	3
	Cognitive competence	3.62 (0.61)	1	3.81 (0.70)	1
	Value	5.34 (0.71)	0	5.20 (0.74)	6
	Difficulty	3.10 (0.54)	2	3.27 (0.76)	3

[a]: Additional question texts are provided in Section 2.2.3 of the main text.
Percentages refer to gender, whereas means (standard deviations) refer to other components.
SD, standard deviation; SATS, Survey of Attitudes Toward Statistics; PI, peer instruction.

tics after weighting.

Figure 1 shows the distribution of the PS for both groups. Although a difference was observed in the distribution, the overlap range was 0.45 or more.

Estimation of learning attitude. Table 5 shows the estimated values of the learning attitude of the PI group compared to the non-PI group as primary outcome. In the crude analysis, all point estimates were not statistically significant. Apart from effort, the point estimates were negative (−0.26 to −0.05) for other outcomes. In DR, ATE was significantly negative for 3 categories (affect: ATE −0.51, 95% CI −0.78 to −0.24; cognitive competence: ATE −0.49, 95% CI −0.71 to −0.28; difficulty: ATE −0.47, 95% CI −0.69 to −0.24).

Table 2 Changes in SATS scores after the lecture among 101 students (without missing data)

	non-PI group ($n=14$)	PI group ($n=87$)
Interest	0.366 (0.85)	0.158 (1.00)
Effort	0.286 (0.40)	0.529 (1.10)
Affect	0.179 (0.86)	−0.082 (0.92)
Cognitive competence	0.154 (0.54)	0.068 (0.92)
Value	0.056 (0.45)	−0.089 (0.68)
Difficulty	−0.040 (0.35)	−0.090 (0.66)

The mean and standard deviation are indicated.
SATS, Survey of Attitudes Toward Statistics.

For interest, ATE was negative (−0.11) but not significant. Other components exhibited positive ATEs; however, the absolute values were very small (0.00 and 0.01).

Discussion

The results of this study showed that the PI lecture was not associated with an increase in the SATS score compared with the non-PI lecture. That is, the introduction of the PI in an 80-min epidemiology lecture had a negative or no effect on the students.

One previous study revealed that, while PI enhanced learning effects in the short term, it decreased such effects in the long term [29]. Numerous other studies [30-32] found that PI lectures were advantageous for learning. Nevertheless, the present study did not reveal any positive effects from PI lectures. There are three possible reasons for this absence of any significant positive effects from PI in our present cohort.

First, the PI lecture utilized in this study may have been insufficient in terms of improving attitudes toward learning. Five topics were explained during the 80-min lecture, and each topic was discussed for only 16 min. Therefore, the ConcepTest included a maximum of 5 min for response and discussion, which was insufficient for facilitating in-depth discussions among the

Table 3 Absolute standardized mean differences and variance ratios of baseline characteristics among 101 students without missing values in the raw and weighted models

	Raw	Weighted
	SMD (Variance ratio)	
Gender	0.80 (0.80)	0.00 (0.94)
Self-assessment [a]		
Self-rated mathematics skill	0.13 (0.98)	0.00 (0.69)
Expected use of epidemiology after graduation	0.09 (1.15)	0.00 (1.35)
Confidence in acquiring the content of the lecture	0.36 (2.46)	0.00 (2.34)
Willingness to choose epidemiology lectures again	0.71 (2.39)	0.00 (1.78)
Pre-lecture SATS score		
Interest	0.34 (1.41)	0.00 (0.93)
Effort	0.86 (0.92)	0.00 (0.45)
Affect	0.12 (1.00)	0.00 (0.59)
Cognitive competence	0.02 (3.63)	0.00 (1.97)
Value	0.15 (0.95)	0.00 (0.83)
Difficulty	0.21 (2.32)	0.00 (1.93)
Average	0.34 (1.64)	0.00 (1.26)

[a]: Question text is provided in the main text section 2.2.3.
SMD, standardized mean differences; SATS, Survey of Attitudes Toward Statistics.

Table 4 Descriptive characteristics of the weighted population

		non-PI group	PI group
Gender	Female	54.50%	54.50%
	Male	45.50%	45.50%
Self-assessment [a]			
	Self-rated mathematics skill	4.72 (1.75)	4.72 (1.45)
	Expected use of epidemiology after graduation	4.83 (1.02)	4.83 (1.19)
	Confidence in acquiring the content of the lecture	4.32 (0.77)	4.32 (1.18)
	Willingness to choose epidemiology lectures again	4.58 (1.20)	4.58 (0.90)
Pre-lecture SATS score			
	Interest	4.95 (0.98)	4.95 (0.94)
	Effort	6.05 (1.11)	6.05 (0.75)
	Affect	3.69 (0.95)	3.69 (0.72)
	Cognitive competence	3.75 (0.42)	3.75 (0.58)
	Value	5.28 (0.85)	5.28 (0.77)
	Difficulty	3.12 (0.56)	3.12 (0.78)

[a]: Question texts are provided in Section "3. Candidate of confounding factor" of the main text. Percentages refer to gender, whereas means (standard deviations) refer to other components. The sum of the overlap weights was 9.29 in both groups.
SATS, Survey of Attitudes Toward Statistics; PI, peer instruction.

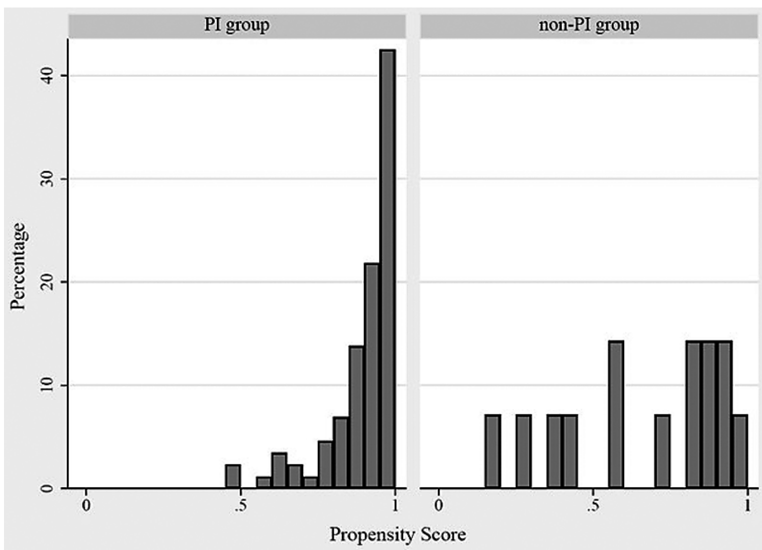

Fig. 1 Histogram of the propensity score in the PI and non-PI group.

students. In the future, it will be important to increase such discussion periods [14]. In addition, the students had little or no experience in lectures with discussions. Consequently, they may have been confused by the new format and thus felt hesitant to participate in the ConcepTest discussion. Ideally, a period of time should be set for breaking the ice and for devising approaches to promote student discussions [33,34]. In previous research, the study period and the study methods were set with an eye toward maximizing the learning effect from the PI. For example, although the discussion time in the PI was only one to a few minutes, a detailed explanation of PI was given over the first two weeks of the course [31]. In another study, the education period was as long as 4 weeks [32]. The present study also had some problems regarding poor PI efficiency [21], which is a measure of the performance on the ConcepTest, as proposed by Nitta *et al.* [35,36]. In future lectures, checking PI efficiency should be considered crucial.

Second, the SATS score may be inappropriate as an

Table 5 Effect estimate of the lecture with PI on SATS score difference

	Crude analysis		DR analysis	
	Coef (95%CI)	*P*-value	ATE (95%CI)	*P*-value
Interest	−0.21 (−0.77, 0.36)	0.466	−0.11 (−0.61, 0.38)	0.657
Effort	0.24 (−0.35, 0.83)	0.416	0.00 (−0.15, 0.16)	0.954
Affect	−0.26 (−0.78, 0.26)	0.321	−0.51 (−0.78, −0.24)	<0.001
Cognitive competence	−0.09 (−0.59, 0.41)	0.732	−0.49 (−0.71, −0.28)	<0.001
Value	−0.14 (−0.52, 0.23)	0.443	0.01 (−0.30, 0.32)	0.952
Difficulty	−0.05 (−0.41, 0.31)	0.785	−0.47 (−0.69, −0.24)	<0.001

SATS, Survey of Attitudes Toward Statistics; PI, peer instruction; Coef, coefficient; DR, double robust; CI, confidence Interval; ATE, average treatment effect.

effect index. At the time of this study, a valid questionnaire to measure attitudes toward epidemiology was not available, and thus the SATS was used as an alternative. However, because the SATS is a questionnaire that measures attitudes toward statistics [22], the effect of epidemiological learning may not have been adequately measured. Additionally, since the SATS was originally designed to assess knowledge before and after 15 units (90 min per unit) of learning, it is possible that the learning time in this study was insufficient and thus the SATS change was too small. In future studies, questionnaires or standard tests on epidemiological learning must be developed. Furthermore, the PI learning time should be longer.

Third, there may have been large differences in baseline characteristics between the groups. Only students who voluntarily chose epidemiology participated in the non-PI lectures; however, all students, regardless of their intentions, participated in the PI lectures. Students who actively choose an epidemiology lecture may have higher potential motivation and thus may be more likely to have higher SATS scores. Conversely, a PI group in which all students participated irrespective of their interest in epidemiology would be expected to have lower SATS scores due to relatively lower potential motivation. Therefore, the effect of PI was likely to have been underestimated in this study. The potential motivation was an unmeasured variable. Using DR, the author attempted to reduce this imbalance, but such attempts may have been inadequate.

The aforementioned factors would be expected to skew the results and underestimate the impact of PI lectures. Therefore, by resolving these factors, improvements in epidemiological learning attitudes through PI lectures could be elucidated more conclusively.

The strength of this study is that it investigated the effects of PI lectures on epidemiology. The learning attitudes toward interactive lectures have not been validated in epidemiology education. As such, this research is an early study in this field and can provide direction for subsequent research. Although the effect of confounding factors was likely to have distorted the causal relationship in this study, this effect was minimized through the use of a doubly robust method. If either of the two model assumptions (outcome or treatment model) hold, the doubly robust method yields correct estimates. This is likely to result in estimates that are closer to the true value than could be found with a conventional analysis. Moreover, in this study, the reduction of confounding effects was rigorously considered by using Directed Acyclic Graphs (DAGs). DAGs can be used for identifying variables that must be conditioned [37,38]. Figure 2 shows a DAG representing the causal relationships assumed by the author. This DAG comprises three backdoor paths between exposure (*i.e.*, the lecture style) and outcome (*i.e.*, SATS score improvement) as follows: $E \leftarrow L1 \rightarrow L2 \rightarrow Y$, $E \leftarrow L2 \rightarrow Y$, and $E \leftarrow L1 \rightarrow C \leftarrow Y$, where $E =$ lecture style, $Y =$ SATS score improvement, $L1 =$ willingness to study epidemiology, $L2 =$ SATS score pre-lecture, and $C =$ answers to the questionnaire. To block all backdoor paths, two variables, namely L1 and L2, must be conditioned. L1 was not measured in this study; however, L1 still had to be adjusted because it had a strong influence on students in 2018 who were choosing this course as well as on outcomes. Therefore, variables related to L1 were conditioned as surrogates. Thus, the participants' pre-lecture SATS scores, gender, and self-assessment were conditioned as a confounder set.

The study also has many limitations. The number of

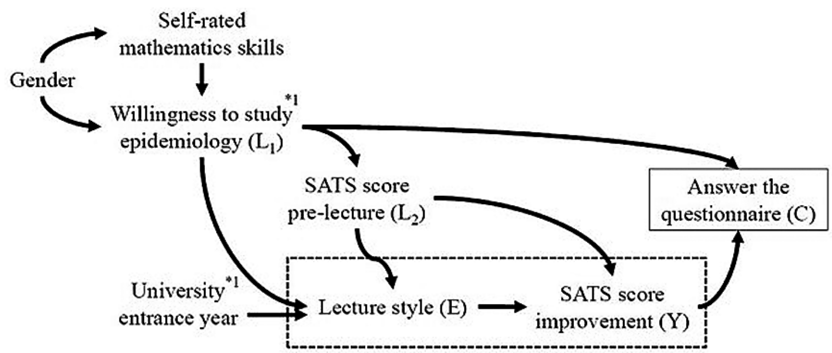

Fig. 2 Directed Acyclic Graph representing causal relationships assumed by the author.
*1: Unmeasured variable in this study.
The subjects in the study were restricted to students who answered the questionnaire. Therefore, the node of "Answer the questionnaire" is boxed. The dotted-line box is the causal relationship of interest in this study.
PI, peer instruction; SATS, Survey of Attitudes Toward Statistics.

students in the non-PI lecture group was small. Therefore, the variance in the estimated value of the effect was large, and the 95% CIs were widened. Moreover, the concerns about study design may have led to an underestimation of the effect of PI lectures.

The epidemiology lecture with an 80-minute PI did not increase the participants' SATS scores. This result may have been affected by concerns related to the study design, the lecture content, or the questionnaires for evaluating effectiveness. In particular, the differences in the unobserved variables between the groups may have been large. Furthermore, it is necessary to establish an appropriate periods of time for the lecture, the ConcepTest, and the preparation time to facilitate PI. It may be difficult for observational studies to solve such research design problems; however, a randomized control trial may be difficult to conduct. Therefore, researchers investigating educational effects need to find an environment where quasi-experimental research, such as the instrumental variable and synthetic control methods, can be performed. Furthermore, appropriate evaluations must be conducted by solving related research concerns, such as by developing questionnaires and standard tests for evaluating epidemiological learning and by improving research designs.

Acknowledgments. This study was supported by the Japan Medical Education Foundation (http://www.jmef.or.jp/). The funders had no role in the study design, data collection or analysis, the decision to publish, or the preparation of the manuscript.

I am thankful to Yoko Oka for assistance with the data management. I would like to thank Editage (www.editage.jp) and KN International (www.kninter.co.jp) for assistance in English language editing.

References

1. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, Salman R AS, Chan AW and Glasziou P: Biomedical research: Increasing value, reducing waste. Lancet (2014) 383: 101–104.
2. Ioannidis JPA: Why most published research findings are false. Plos Med (2018) 2: e124.
3. Anderson BL, Williams S and Schulkin J: Statistical literacy of obstetrics-gynecology residents. J Grad Med Educ (2013) 5: 272–275.
4. Emerson JD and Colditz GA: Use of statistical analysis. N Engl J Med (1983) 309: 709–713.
5. Horton NJ and Switzer SS: Statistical methods in the journal. N Engl J Med (2005) 353: 1977–1978.
6. Sato Y, Gosho M, Nagashima K, Takahashi S, Ware JH and Laird NM: Statistical methods in the journal—An update. N Engl J Med (2017) 376: 1086–1087.
7. Lim E: Basic statistics (the fundamental concepts). J Thorac Dis (2014) 6: 1875–1878.
8. Saucedo-Moreno EM and Fenig-Rodríguez J: Statistics in surgery, how to understand and apply basic concepts. Cir y Cir (English Ed) (2019) 87: 692–697.
9. Haller H and Stefan K: Misinterpretations of significance: A problem students share with their teachers? Methods Psychol Res Online (2002) 7: 1–20.
10. Windish DM, Huot SJ and Green ML: Medicine residents' understanding of the biostatistics and results in the medical literature. J Am Med Assoc (2007) 298: 1010–1022.
11. Goodman S and A Dirty Dozen: Twelve p-value misconceptions. Semin Hematol (2008) 45: 135–140.
12. Redish EF and Burciaga JR: Teaching physics with the physics suite. Am J Phys (2004) 72: 414.
13. Mazur E and Hilborn RC: Peer instruction: A user's manual. Phys Today (1997) 50: 68–69.
14. Crouch CH and Mazur E: Peer Instruction: Ten years of experi-

ence and results. Am J Phys (2001) 69: 970–977.
15. Lasry N, Mazur Ec and Watkins J: Peer instruction: From Harvard to the two-year college. Am J Phys (2008) 76: 1066–1069.
16. Crouch CH: Peer instruction: An interactive approach for large lecture classes. Opt Photonics News (1998) 9: 37.
17. Baker PRA, Francis DP and Cathcart A: A mock randomized controlled trial with audience response technology for teaching and learning epidemiology. Asia-Pacific J Public Heal (2017) 29: 229–240.
18. Goldmann E, Stark JH, Kapadia F and McQueen MB: Teaching epidemiology at the undergraduate level: Considerations and approaches. Am J Epidemiol (2018) 187: 1143–1148.
19. Dankner R, Gabbay U, Leibovici L, Sadeh M and Sadetzki S: Implementation of a competency-based medical education approach in public health and epidemiology training of medical students. Isr J Health Policy Res (2018) 7: 13.
20. Sohn S, Lee Y-M, Jung J, Cha ES and Chun BC: The flipped classroom model for an undergraduate epidemiology course. Korean J Med Educ (2019) 31: 103–113.
21. Mitsuhashi T: Evaluation of epidemiological lectures using peer instruction: Focusing on the importance of ConcepTests. PeerJ (2020) 8: e9640.
22. Schau C, Stevens J, Dauphinee TL and Vecchio AD: The development and validation of the survey of attitudes toward statistics. Educ Psychol Meas (1995) 55: 868–875.
23. Murakami Y and Borgonovi F: Japan needs gender equality. Science (2018) 362: 133–133.
24. Heberlein TA and Baumgartner R: Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. Am Sociol Rev (1978) 43: 447–462.
25. Choi J, Dekkers OM and Le Cessie S: A comparison of different methods to handle missing data in the context of propensity score analysis. Eur J Epidemiol (2019) 34: 23–36.
26. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA and Davidian M: Doubly robust estimation of causal effects. Am J Epidemiol (2011) 173: 761–767.
27. Li F, Thomas LE and Li F: ddressing extreme propensity scores via the overlap weights. Am J Epidemiol (2018) 188: 250–257.
28. Zhang Z, Kim HJ, Lonjon G and Zhu Y; written on behalf of AME Big-Data Clinical Trial Collaborative Group: Balance diagnostics after propensity score matching. Ann Transl Med (2019) 7: 16–16.
29. Alcalde P and Nagel J: Why does peer instruction improve student satisfaction more than student performance? A randomized experiment. Int Rev Econ Educ (2019) 30: 100149.
30. Cortright RN, Collins HL and DiCarlo SE: Peer instruction enhanced meaningful learning: ability to solve novel problems. Adv Physiol Educ (2005) 29: 107–111.
31. Vázquez-García M: Collaborative-group testing improves learning and knowledge retention of human physiology topics in second-year medical students. Adv Physiol Educ (2018) 42: 232–239.
32. Compton NJ, Cary JA, Wenz JR, Lutter JD, Mitchell CF and Godfrey J: Evaluation of peer teaching and deliberate practice to teach veterinary surgery. Vet Surg (2019) 48: 199–208.
33. Martin F and Bolliger DU: Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. Online Learn J (2018) 22: 205–222.
34. Ilias B: Icebreakers for business school students; in Learning and Teaching in Higher Education, Daniels K, Elliott C, Simon F and Chapman C eds, Edward Elgar Publishing, Cheltenham (2019) pp 26–35.
35. Nitta H: Mathematical theory of peer-instruction dynamics. Phys Educ Res (2010) 62: 020105-1–020105-4.
36. Nitta H, Matsuura S and Kudo T: Implementation and analysis of peer-instruction in introductory physics lectures. J Sci Educ Japan (2014) 38: 12–19.
37. Greenland S, Pearl J and Robins JM: Causal diagrams for epidemiologic research. Epidemiology (1999) 10: 37–48.
38. Hernan M and Robins J: Graphical representation of causal effects; in Causal Inference: What If. Chapman & Hall/CRC, Boca Raton (2020) pp 69–82.