Honours Project

The Education University of Hong Kong

Assessing Preservice Teachers' Misconceptions in Conditional Probability

Supervisor: Dr Leung King Shun

Student Name: Zhou Siyu

Contents

INTRODUCTION

If probability is a way of thinking in the current context, then conditional probability is a tool for reconsideration in a renewed setting. Conditional probability helps us to re-examine the likelihood of occurrence of a certain event (i.e., the conditioned event) given that another event (i.e., the conditioning event) has occurred. When these two events are interrelated instead of independent, the occurrence of the conditioning event will provide useful information to make further inferences about the probability of the conditioned event (Tarr & Lannin, 2005). Probabilities can be all conditional in the sense that everything is interrelated in some way. In the world of uncertainties, conditional probability helps us to make sound judgements and informed decisions about everyday life.

However, the notion of conditional probability can be difficult to grasp since people are susceptible to many misconceptions without even being aware (Borovcnik & Peard, 1996). As humans, we are somehow programmed to use our intuitions to interpret probabilistic problems, but some of the results related to conditional probability will turn out to be quite counterintuitive. The incentive for this study is the famous "Monty Hall Problem" creating a scenario where the player will first make a choice among three doors with only one of them representing a real car behind. Next, the host will open another door without a prize and allow the player to make a switch if desired. Even till now, I tend to believe that switching does not matter since the two unopened doors always share a fifty-fifty chance of the prize. But it has been confirmed that switching to the other door is the better strategy with a winning probability of 2/3, a surprising fact that many people refuse to accept (Sadri, 2012). This is just an illustrative example of many misconceptions documented so far.

In the context of Hong Kong, conditional probability is covered under the unit of "more about probability" in the mathematics curriculum compulsory part (Hong Kong Curriculum Development Council, 2017). It is expected that preservice mathematics teachers should possess solid subject matter knowledge as well as high resistance to the potential biases related to conditional probability. Otherwise, they will not be able to recognise and correct these biases in themselves, nor can they establish a rigorous probabilistic sense in their students. However, till now little research has been done to examine Hong Kong preservice teachers' misconceptions of neither conditional probability nor general probability. It is time that we should identify the existing problems with preservice teachers since they are the latest "fruits" as well as the new "blood" of the current education system. The results will be indicative of what has been in the past and informative of what could be improved in the future. In line with the wisdom of conditional probability: the more we know, the fewer uncertainties we have.

LITERATURE REVIEW

Current Situation

Psychologists and mathematics education researchers have conducted numerous studies regarding misconceptions in conditional probability, including studies from some parts of the world such as Australia and USA (Reaburn, 2013; Tarr & Jones, 1997). In contrast, little of the research has been found to examine the situation in Hong Kong. Disappointingly, the situation in mainland China as well as Taiwan is left nearly blank for us to draw any comparison or insights. However, when the scope of focus is enlarged to probability in a general sense, some research has been carried out in mainland China and Taiwan, with Hong Kong still out of sight. In their research, a great many misconceptions with have been identified in secondary school students with a worrying prevalence. While it would be significant to examine Hong Kong students as well to understand the current situation, this study will instead focus on Hong Kong preservice teachers' misconceptions in conditional probability in particular. Before faulting students for not doing so well, perhaps it is sensible to find out the problems in the teachers themselves.

Back to the discussion of misconceptions in conditional probability, the volume of literature keeps expanding but there is little research specified to examine preservice teachers. Among these limited resources available for review, preservice teachers' misconception in conditional probability just remains a sub-topic without detailed and sufficient account, usually under the broad analysis of all probabilistic issues (e.g., Jendraszek, 2008; Dollard, 2007). Under such circumstances, this literature review will turn to the variety of misconceptions of conditional probability that have been identified so far, along with related examples, difficulties and insights, which will be beneficial to the building of the sample questionnaire later.

1. Without Placement Situations

Fischbein and Gazit (1984) carried out an experiment where 285 students from grades 5-7 are asked to determine conditional probabilities in "with" and "without" placement situations. They found out that students generally performed better in with-placement tasks than in without-placement tasks. Regarding these two types of tasks, the percentage of correct answers dropped dramatically from 63% to 43% and from 89% to 71%, for sixth and seventh graders, respectively. It is worthy to note that the correct response rate steadily stayed at around 24% for fifth graders in both "with" and "without placement" tasks. Based on the analysis of typical incorrect solutions, the researchers identified two major misconceptions in students' thinking in conditional probability: (a) the failure to recognise the reduction in the sample space in "without placement" situations; (b) the tendency to relate the two sets of favourable outcomes before and after the first trial instead of comparing the possible outcomes against all possible outcomes. Other related literature has also addressed that conditional probability problems often get tricky in "without placement" situations (e.g., Borovcnik & Bentz, 1991; Falk, 1986).

2. Base Rate Fallacy

Bar-Hillel (1983) observed that people are inclined to ignore base rate information but favour diagnostic information at hand, which leads them to make causal attributions too soon without taking into considerations all possible cases. The author then provided an illustrative example as a teaser (as cited in Bar-Hillel, 1983, pp. 39-40): in 1973, a university accepted 44% of all male applicants but only 35% of all female applicants, given that (a) female applicants were overall more qualified and (b) there was no gender discrimination practised. While most people were trying to look for a causal explanation that only added to their confusion, this fact could be easily explained with reference to "base rate": there might just be more female applicants than male applicants this year. Despite the importance of base rates in inferential tasks, it happens when students and even professionals sometimes lack careful considerations to make full use of general information in their probabilistic reasoning.

3. Causal & Diagnostic Reasoning

Tversky and Kahneman (1980) distinguished two different commonly perceived relations between the conditioning event and the conditioned event in P(A|B). When B is perceived as a cause of A, then the relation is classified as "causal". On the other hand, if A is perceived as a possible cause of B, then the relation is classified as "diagnostic". Despite that such distinction would be irrelevant to the informativeness of data in making probabilistic judgements, people are psychologically biased to attach greater importance to causal data rather than diagnostic data of equal informativeness due to the prevalence of causal schemas in real life. Even given the hint that P(A|B) and P(B|A) could be perceived as equal in a concrete scenario, most participants still erroneously assigned greater probability to the causal relation. Moreover, there is a lack of discrimination between these two directions of conditional probability. This long-recognised confusion can be prevalent among students and professionals at all levels, usually occurring in the interpretation of medical test results (Eddy, 1982).

4. Conditioning Event

The problematic definition of the conditioning event has been demonstrated by a great many teasers. One of these is the notorious "Three Card Problem" (Bar-Hillel & Falk, 1982) originally testing undergraduates in a probability course. Three two-sided cards are in a hat, namely Card A (red on both sides), Card B (white on both sides), and Card C (red on one side while white on the other side). Imagine a card is drawn and put on the table. Given that a red side faces up, the participants were asked to give the probability that the hidden side is also red. Though the correct answer is 2/3, 66% of the students answered 1/2 in the study, claiming that "the double-white card is out" and the remaining two cards are of equal likelihood. The researchers attributed this error to misinterpreting the inferred the conditioning event: although "the double-white card is out" is a correct inference as well as a valid conclusion, it should not be the event on which the target event (i.e., "the other side is also red") should be conditioned upon. In fact, the conditioning event should be the immediate datum (i.e., "a red face shows up") as defined by the problem's sampling procedure (i.e., a random side is selected). In defining the conditioning event, an insight would be to check if it is in the sample space of possible outcomes as guaranteed by the statistical experiment.

5. Time Axis Fallacy

Falk (1989) found that many students held the belief that the conditioning event should always precede the conditioned event. Though these students naturally understood that the outcome of the first event could affect the second event but not the other way around, still they were reluctant to accept the fact that the later event could play an informative role in making inferences about the earlier event. Stereotyping conditioning events as only those "prior" incidents in a temporal order will lead to much neglect and ignorance of useful information in making probabilistic judgements. As Kelly and Zwiers (1988) noted, "the components of the inferential chain might be events in any temporal order". In other words, informativeness of the conditioning event is solely associated with its statistical dependence with the conditioned event regardless of the temporal order, a fact that many people find difficult to believe.

Research Gaps

While the above studies provide a comprehensive picture of the misconceptions of conditional probability, the subject of relative experiments have always been children or students. Beyond that, we would like to know if these misconceptions are still persistent, or at least, exist in the minds of preservice mathematics teachers. Somewhat surprisingly, some misconceptions are found to be more prevalent among prospective mathematics teachers than secondary students in a study by Fischbein and Schnarch (1997). This worrying phenomenon a decade ago, yet inconclusive of the whole picture, is ringing an alarm for us to look back at the current situation in Hong Kong. If that is the case, it would be illuminating to question how well our preservice teachers have been prepared by both secondary and tertiary probability education. Is probability education hopefully aiding them or disappointingly failing them? This unanswered question largely motivates this research as well as representing itself as one of the research questions. Hopefully, this study will act as a pause-for-reflection button for preservice teachers as well as different stakeholders to identify existing problems and make necessary improvements.

REAEARCH QUESTIONS

The primary purpose of this study is to assess Hong Kong preservice secondary mathematics teachers' misconceptions in conditional probability. Following this purpose, the detailed research questions are as follows:

(1) Do Hong Kong preservice teachers have the five misconceptions in conditional probability described in the literature review? If they do, which misconception(s) are more prevalent?

(2) Is there any effect of probability and statistics education on the five misconceptions? Does such a relationship differ for secondary and tertiary levels' education?

(3) How confident do these preservice teachers report upon completion of each question? For each question, what is the correlation between their rating of confidence and actual performance?

METHODS

Questionnaire

In the development of the online questionnaire, five misconceptions from the literature review will be taken into consideration, with each assigned two items, one item as a brain teaser and the other as a standardised one, both following a multiple-choice format. The five brain teasers are either biased or normative, requiring participants to make a "black or white" judgement. For the five standardised questions, there are more than two answers available for participants to choose from, the order of which is randomised and thus distinguished from that of the first five brain teasers. Some of the questions are validated items from previous research; some are analogues or adapted versions; some were created by the author according to the descriptions of relative misconceptions in literature. The original questionnaire items have gone through expert examination aimed to identify potential ambiguities in wording. A pilot test of the revised questionnaire was tried on some undergraduates who do not major in mathematics. For item 5, few of them doubted whether "'later' occurrences" meant "100 % certain of taking place for 'occurrences'" or "uncertainty of taking place for 'later' [future]". Therefore, we resolved this issue by replacing the old version with "the probability of a [previous] event could not be revised according to some information about [current] occurrences". We then sent the finalised version to all BEd(S)-MA undergraduates in EdUHK.

For each question, participants are required to select an option and give explanations or calculations in either English or Chinese. Their justifications will allow us to further investigate their ways of thinking when approaching probability tasks. Apart from the knowledge level, participants will be asked to give their confidence rating, on a scale of 1 to 5, about their answer and explanations upon completing each item. It is estimated that 30 minutes will be enough for average students to complete the questionnaire, but no time limit will be imposed.

As for the scoring, participants' answers will be scored as incorrect (i.e., incorrect judgement), partially correct [=1 points] (i.e., correct judgement and incorrect explanation), or completely correct [=2 points] (i.e., correct judgement and explanation). The total score ranges from 0 to 20 points.

Interview

Interviews were conducted as a supplementary qualitative component to the main quantitative design above. At the end of the questionnaire, we asked if the participants would like to join an interview afterwards. Among those who did not indicate unwillingness (i.e., "yes" or "not sure"), Joey and Alex were selected for the interview, mainly because of the representativeness of their incorrect answers to the questionnaire. Besides, since Alex was the only participant who correctly answered both questions relating to the conditioning event, his insightful probabilistic thinking would be extremely interesting when put in comparison with the biased responses. Apart from verbally expressing their ideas that cannot be communicated through written words on the questionnaire, the interviewees paraphrased those justifications they have already written down in more detail, thus allowing the researcher to further interpret their ways of thinking that could possibly lead to the potential misconceptions. During the interview, the researcher did not interrupt or give any guidance or hint to the interviewee so as to guarantee the authenticity of the collected data.

Sample

Recruitment efforts were directed to undergraduate students of all year groups enrolled in the programme Bachelor of Education in Secondary Mathematics at The Education University of Hong Kong [EdUHK]. Given the role of EdUHK in cultivating the main body of future teachers, these students are of primary interest to our study. Still, one concern is that this sampling method will exclude some other potentially preservice teachers (e.g. PGDE students in EdUHK, preservice teachers outside EdUHK, etc). Therefore, we will not claim that the future findings will be generally applicable to the whole population but will be indicative of the current situation.

The online questionnaire was sent to all eligible participants to ensure the inclusion of participants with varied levels of probability and statistics education and socio-economic background. Study materials were provided as non-monetary reward for every participant. By the end of the data collection stage, 42 responses were obtained.

RESULTS

Data Analysis

For the research question 1, we compared the mean scores of the two questions relating to each 5 misconceptions. The maximum possible score for each misconception is 4. A misconception with a relatively lower score will be identified as a prevalent misconception. For the research question 2, the participants were classified into different groups regarding their probability/statistics education at both secondary and tertiary stages. In order to see whether significant differences exist among the secondary and tertiary groups in the test performance, the ANOVA was conducted. Both the total score and the five sub-scores were examined. For the research question 3, the Pearson Correlation Coefficient was computed for each of the ten items separately to analyse the relationship between the rating of confidence and the actual test score.

General Findings

Table A

Report

	score_without placement	score_baserate	score_causal	score_conditioningevent	score_timeaxis	score_total
Mean	2.98	.57	1.10	.52	1.48	6.64
N	42	42	42	42	42	42
Std. Deviation	1.137	.770	.932	.943	1.215	2.408

All five misconceptions were identified in our preservice teachers' probabilistic thinking to varying degrees (Table A). Participants scored the relatively lowest means around 0.5 point for conditioning event and base rate, demonstrating the prevalence of misconceptions in these two areas. Followingly, the causal reasoning and the time axis fallacy come next with both mean scores over 1 point. In contrast, participants performed a lot better in without placement tasks where the mean score is nearly 3 points.

Table B

Descriptives

score_total

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
1	16	7.13	2.156	.539	5.98	8.27	4	12
2	12	6.50	2.276	.657	5.05	7.95	4	11
3	14	6.21	2.833	.757	4.58	7.85	2	10
Total	42	6.64	2.408	.371	5.89	7.39	2	12

ANOVA

score_total

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6.536	2	3.268	.551	.581
Within Groups	231.107	39	5.926		
Total	237.643	41			

Group 1: participants who have taken no probability or statistics courses in EDUHK.
Group 2: participants who have taken only MTH4155.
Group 3: participants who have taken both MTH4155 and MTH4153.

The relation between the total test score and the tertiary probability/statistics education was analysed (Table B). One direct observation from the result is that, the mean scores decrease with the probability/statistics education at tertiary stage. Despite this, it can be inferred from the AVOVA result that there is no significant difference (α=.05) among participants of different tertiary levels.

Table C

Descriptives

score_total

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
1	8	6.25	1.982	.701	4.59	7.91	3	9
2	13	6.69	2.750	.763	5.03	8.35	4	12
3	11	7.18	2.562	.772	5.46	8.90	2	10
4	10	6.30	2.312	.731	4.65	7.95	3	10
Total	42	6.64	2.408	.371	5.89	7.39	2	12

ANOVA

score_total

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5.637	3	1.879	.308	.820
Within Groups	232.006	38	6.105		
Total	237.643	41			

Group 1: participants who have taken M1 in Hong Kong secondary schools.
Group 2: participants who have taken M2 in Hong Kong secondary schools.
Group 3: participants did not take the extended mathematics modules in Hong Kong secondary schools.
Group 4: participants who have learned conditional probability in mainland secondary schools.

The relationship between the total test score and the secondary probability/statistics education was analysed (Table C). One direct observation from the result is that, the mean scores decrease with the probability/statistics education at secondary stage, as probability/statistics enrichment were provided to Group 1 and Group 4 participants. However, the AVOVA result indicate no significant difference (α=.05) among participants of different secondary levels.

Detailed Findings

1. Without Placement Situations

Table D

score_withoutplacement

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	16	3.00	.894	.224	2.52	3.48	2	4
2	12	3.67	.778	.225	3.17	4.16	2	4
3	14	2.36	1.336	.357	1.59	3.13	0	4
Total	42	2.98	1.137	.175	2.62	3.33	0	4

ANOVA

score_withoutplacement

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11.095	2	5.548	5.166	.010
Within Groups	41.881	39	1.074		
Total	52.976	41			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: score_withoutplacement

LSD

(I) education_tertiary	(J) education_tertiary	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-.667	.396	.100	-1.47	.13
	3	.643	.379	.098	-.12	1.41
2	1	.667	.396	.100	-.13	1.47
	3	1.310*	.408	.003	.48	2.13
3	1	-.643	.379	.098	-1.41	.12
	2	-1.310*	.408	.003	-2.13	-.48

*. The mean difference is significant at the 0.05 level.

Group 1: participants who have taken no probability or statistics courses in EDUHK.
Group 2: participants who have taken only MTH4155.
Group 3: participants who have taken both MTH4155 and MTH4153.

The two questions (Q1, Q8) in the instrument are related to without placement situations. While there is no significant difference (α=.05) in the combined test scores for these two questions among groups by secondary education, we see a significant difference (α=.05) among groups in terms of tertiary education. Therefore, the post hoc test of LSD was done (Table D). Results show a significant difference between Group 2 and Group 3, indicating that participants who have taken only MTH4155 significantly performed better in without-placement tasks than participants who have taken both MTH4155 and MTH4153. However, since we observe an opposite pattern between Group 1 and Group 2, it cannot be concluded whether preservice teachers' performance shows a particular relation to their tertiary probability/statistics education level.

For Q1, the mean score is 1.17 and the mean confidence level is 3.93. Analysis of data shows that participants' actual performance and reported level of confidence are not significantly related. 23 respondents (54.8%) gave the correct answer along with correct explanations. These participants recognised the change of the sample space due to the game rule of not putting back the already-drawn lots. When told no one has won the prize yet, they judged the prize pool to be smaller than before, making it more likely and easier to win the prize. Moreover, several participants introduced fractions to solve the problem: 1, the numerator, stands for the prize, whereas a, *the* denominator, represents the remaining lots each round. This method visualises the rise of winning probability $1/a$ when a decreases and 1 remains unchanged. 3 respondents (7.1%) gave the correct answer but incorrect explanations that were either contradictory to the correct judgment they made or neglectful of the question's given conditions. Therefore, none of these 3 responses could be classified into any sort of misconception worth discussing. 16 respondents (38.1%) gave the incorrect answer and incorrect explanations. The analysis of these explanations identified several misconceptions. The most prevalent misconception lies in the tendency to believe individuals' chance of winning should be the same since it is a fair game. Interestingly, a participant did not deny the increase of probability and considered it to make some sense to a certain degree. However, this participant eventually came around to reject the statement from the perspective of "fairness". The second type of misconception is mainly rooted in the biased interpretation of "independence". Though it seems natural to take into consideration what happened before for this without-placement task, some participants turned to the concept of independence and claimed that one's probability of winning is not related to others' results of drawing. Another type of misconception was observed in the inability to identify the problem as a conditional one but erroneously interpreted it as a compound one.

The probability was calculated to be P(A∩B) rather than P(B|A), where A represents "no one has won the prize yet" and B represents "my current chance of winning". Despite that the dependent nature between A and B was successfully identified, the failure to recognise the conditioning role of A remained a constant cause of misconception. Furthermore, interview data suggest that the first two types of illusions are not entirely exclusive. For example, our interviewee Alex combined the thinking of both "fairness" and "independence" to assist his explanations. He did notice the presence of event A but eventually judged it to be irrelevant. When guided to examine the definition of probability in the question, he decided that the probability should be based on event A and should be considered in its renewed sense when one is about to draw the lot. Misconceptions of "fairness" and "independence" were also seen in another interviewee Joey's thinking. When given another chance to change her mind, however, she hesitated but insisted in the end.

Table E

Correlations

		confidence_Q8	score_Q8
confidence_Q8	Pearson Correlation	1	.363*
	Sig. (2-tailed)		.018
	N	42	42
score_Q8	Pearson Correlation	.363*	1
	Sig. (2-tailed)	.018	
	N	42	42

*. Correlation is significant at the 0.05 level (2-tailed).

For Q8, the mean score is 1.81 and the mean confidence level is 4.07. From Table E, the Pearson correlation coefficient of 0.363 suggested that participants' actual performance and reported level of confidence are weakly correlated (α=.05). 37 respondents (88.1%) gave the correct answer along with correct explanations. Compared to Q1, more participants successfully recognised the reduction of the sample space when the first marble is red and not put back. The clear understanding of this without-placement sampling procedure allowed them to compare the numbers of red and blue marbles left to make the correct choice. For example, our interviewee Joey put it this way: "Simply because there are more blue marbles left". Some respondents even gave more detailed steps such as "{Red, Blue, Blue}, 1/3, 2/3 and 2/3 > 1/3". A respondent even used the rigorous formula for computing the two conditional probabilities: P (the second marble is red | the first marble is Red) = P (the first and the second marble are both red) / P (the first marble is red). Only 2 respondents (4.8%)

gave the correct answer but incorrect explanations. Apart from uncareful calculation mistakes, the main misconception could also be attributed to the misidentification of the conditional probability problem as a compound one. Similar to what has been discussed before, the probability was either calculated or understood to be P (the first marble is red ∩ the second marble is red) rather than P (the second marble is red | the first marble is red). Although this approach did help these two participants to arrive at the correct choice, it is in fact a misinterpretation of the conditional probabilities in our context. 3 respondents (7.1%) gave the incorrect answer and incorrect explanations. Only one out of these three respondents were aware that the first red marble was not put back in the box but somehow chose "the second marble is more likely to be red". The remaining two respondents regarded the two probabilities as equal, one out of instinct and the other with no explanation. Alex is the one who used intuition to solve this problem. In the interview, he quickly realised his neglect of the condition of not putting back and instantly changed his mind, giving the correct answer along with correct explanations as documented above.

2. Base Rate Fallacy

The two questions (Q2, Q7) in the instrument are related to the base rate fallacy. There is no significant difference (α=.05) in the combined test scores for these two questions among groups in terms of probability/statistics education at either secondary or tertiary stage.

Table F

Correlations

		score_Q2	condifence_Q2
score_Q2	Pearson Correlation	1	.348*
	Sig. (2-tailed)		.024
	N	42	42
condifence_Q2	Pearson Correlation	.348*	1
	Sig. (2-tailed)	.024	
	N	42	42

*. Correlation is significant at the 0.05 level (2-tailed).

For Q2, the mean score is 0.45 and the mean confidence rating is 3.48. From Table F, the Pearson correlation coefficient of 0.348 suggested that participants' actual performance and reported level of confidence are weakly correlated (α=.05). 5 respondents (11.9%) gave the correct answer along with correct explanations. When judging if the news is "likely" to be true or false, they not only paid attention to the different base rates but also attempted to illustrate such a difference by relevant knowledge of road safety, including “Much fewer pedestrians cross the street when it is a red light, and are much more careful than when it is green light”, “Even if the probability of being killed when crossing on a green light is extremely small, the number of deaths may exceed that of ‘crossing on a red light’ due to a considerably larger base rate”. Interviewee Joey gave the latter explanation. Nevertheless, she said it was strange to see the ratio contrast at first sight and struggled for a while before figuring out what it really meant in the context, “What 3% and 5% really meant is the percentages of all pedestrians instead of the probability of being hit when it’s green or red light”. 9 respondents (21.4%) made the correct judgement but gave incorrect explanations that were ignorant of the base rate information in the background. Four of them also tried to support the statement from the perspective of road safety, such as “When crossing the street on a green light, people will be less alert than that of red lights”, “Maybe people crossing the street on a red light will be even more cautious, so they have less chance to be killed by car”. Two of them, including our interviewee Alex, found no evidence to argue against the claim: “‘Being hit when it’s red light’ and ‘being hit when it’s green light’ are independent of each other”. This explanation is irrelevant since the ratios in the problem is dealing percentages of pedestrians rather than probabilities. Alex also added that the two proportions should be unrelated and uncontradictory to each other, and that was why he judged the claim to be “likely to be true”. Another two respondents only stated that there could be some other possibility which might contribute to the contrast between 3% and 5%, without providing more explicit examples. 28 respondents (66.7%) gave the incorrect answer and incorrect explanations. Apart from few respondents with no idea, typical explanations from these respondents include “It contradicts my assumption that green light is comparatively safe to cross the street. The percentage of getting killed should be lower than that of red light, “This is a strange ratio contrast”, “I believe the chance of getting killed when crossing the street on a red light is higher than crossing the street on a green light”. As has been discussed, this common misconception was also revealed in Joey’s former thinking, where the unusual and sharp ratio contrast (3% vs 5%) kind of revokes the tendency to fault with the given information without resorting to the background information of base rate.

For Q7, the mean score is 0.12 and the mean confidence rating is 3.67. Analysis of data shows that participants’ actual performance and reported confidence level are not significantly related. 2 respondents (4.8%) chose the correct option and provided correct explanations. From the detailed calculation steps they provided [(0.8*0.15) /

(0.8*0.15+0.2*0.85) = 0.41], two main steps are identified to be useful in solving such a tricky problem: first, the identification of "the witness said it's green" as the conditioning event; second, the awareness of the existence of the two cases (taxi: green, witness is right; taxi: yellow, witness is incorrect). Only 1 respondent (2.4%) chose the correct option but showed incorrect explanations. His calculation steps, 1-0.8*0.15-0.2*0.85, was in fact the probability that "the witness said blue", which is contradictory to the given condition. The vast majority of respondents (39; 92.8%) fell into four different kinds of misconceptions and chose the incorrect options accordingly. For those 8 respondents who got a probability of 12%, the method is multiplying the percentage of green taxis (15%) by the testimony's reliability (80%). For one thing, this is a compound probability instead of a conditional one. For another, the case of "taxi: yellow, witness: incorrect" is not considered. These two errors indicate their inability to integrate the background information (i.e., base rates) and the new conditioning information (i.e., the witness said green) to calculate conditional probability. 3 respondents refused to consider the witness's testimony. They thus got a probability of 15%, equal to the ratio of green taxis in the city: "It is obvious", "Whether the witness cannot determine correctly do not affect the mentioned probability". Here, there seems to be an overestimation of the base rate information but an underestimation of the witness's testimony. An opposite kind of thinking could be observed in 16 respondents who ignored the base rate information and entirely relied on the witness's testimony whose explanations included "Since the witness correctly identified the colour of the taxi 80% of the time, the probability that a green taxi was involved in the accident is 80%", "The colour of the taxi involved in the taxi was unrelated to the proportions of yellow or green taxis. Since the accident already happened, the only uncertain factor should be the testimony of the witness". In the interview with Joey, we brought up a question "What if the proportion of the green taxis in the city is 1%? Would you change your mind?". With not the slightest hesitation, Joey said no and that she didn't think the proportion of the green taxis would make a difference, which further confirmed the overriding effects of the witness's testimony when compared with the base rate information. Interestingly, when the group of 12 respondents equally balanced these two pieces of information, they still obtained the wrong answer "0.8*0.15 + 0.2*0.85 = 0.29". Though succeeding in identifying the two cases (taxi: green, witness is right; taxi: yellow, witness is incorrect), they did not appreciate the conditioning role of the event "The witness said it's green". In the interview, Alex reported, "Though I had a feeling that 'the witness said it's green' might well have other uses, I did not know how to apply that in calculation". After all, 0.29 seems fine to me. I have no idea whether it is correct or not".

3. Causal & Diagnostic Thinking

Table G

Post Hoc Tests

Multiple Comparisons

Dependent Variable: score_causal
LSD

(I) education_secondary	(J) education_secondary	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
1	2	-.452	.401	.267	-1.26	.36
	3	-1.011*	.415	.019	-1.85	-.17
	4	-.275	.423	.520	-1.13	.58
2	1	.452	.401	.267	-.36	1.26
	3	-.559	.365	.134	-1.30	.18
	4	.177	.375	.640	-.58	.94
3	1	1.011*	.415	.019	.17	1.85
	2	.559	.365	.134	-.18	1.30
	4	.736	.390	.067	-.05	1.53
4	1	.275	.423	.520	-.58	1.13
	2	-.177	.375	.640	-.94	.58
	3	-.736	.390	.067	-1.53	.05

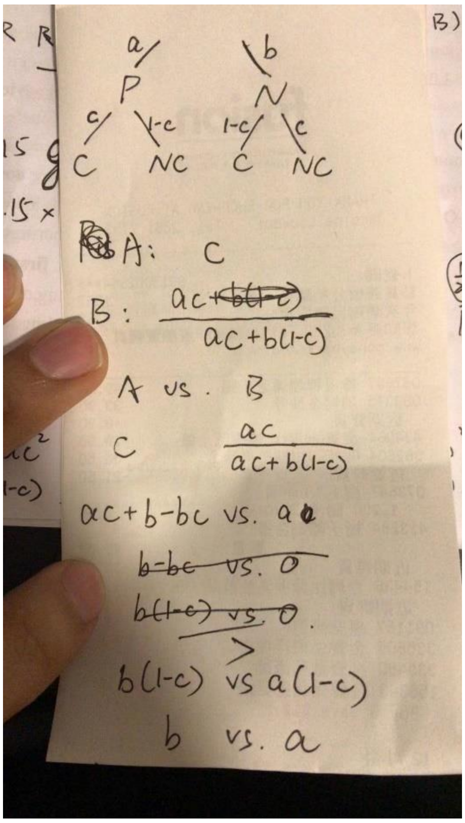
*. The mean difference is significant at the 0.05 level.

Group 1: participants who have taken M1 in Hong Kong secondary schools.
Group 2: participants who have taken M2 in Hong Kong secondary schools.
Group 3: participants did not take the extended mathematics modules in Hong Kong secondary schools.
Group 4: participants who have learned conditional probability in mainland secondary schools.

The two questions (Q3, Q10) in the instrument are related to causal and diagnostic thinking. Analysis of data shows no significant difference (α=.05) in the combined test scores for these two questions among groups in terms of probability/statistics education at either secondary or tertiary stage. Still, the contrasting performance between secondary group 1 and secondary group 3 is statistically significant when we performed the post hoc test (Table G). This result indicates that those who did not take any extended module performed significantly better than those who took M1 in secondary schools.

For Q3, the mean test score is 0.83 and the mean confidence rating is 3.02. Analysis shows that participants' actual performance and reported level of confidence are not significantly related. 10 respondents (23.8%) chose the correct option and gave correct explanations. These respondents perceived the two given events as equally likely, following the presumption that the blue-eyed individuals in the two populations could be regarded as equal. This demonstrates their high resistance to the temporary order of "cause and effect". 15 respondents (35.7%) chose the correct option but gave incorrect explanations. Most of them determined that "Since the father is another contributing factor, it is impossible to make a definite judgement”. The interviewee Alex said the given conditions were insufficient for him to answer this question, but at last, he decided to gamble on “equal” and did not change his mind. It seems that the influence of genetics knowledge interferes with the retrieval of relevant information at hand (the equal proportions of the two populations). Some others indicated their choice was just a random guess. Yet still, another possibility is that they might have successfully interpreted presumption without even being aware. 17 respondents (40.5%) chose the incorrect answer and gave various incorrect explanations. Two major misconceptions were identified. First, 6 respondents attached the greatest importance to the influence of inheritance, claiming that “It is more likely that daughters will take after mothers. That is to say, the probability that a daughter has blue eyes if her mother has blue eyes will be greater”. When asked if this way of thinking contradicts the given presumption, the interviewee Joey reported, “They both seem natural and intuitive to me. But I don’t think the given presumption is relevant in my decision making”. On the contrary, another 6 respondents considered the consequence of being more informative than the antecedent when making probability comparisons. The basic idea is that: “If the daughter has blue eyes, it is very likely that both parents will have blue eyes, and so is the Mom. However, even if the mother has blue eyes, the probability that a daughter has blue eyes will also be affected by the dad, so the probability will be smaller”. The other respondents either gave no explanations or mentioned the influence of DNA without further analysis.

For Q10, the mean test score is 0.26 and the mean confidence rating is 3.12. Analysis shows that participants’ actual performance and reported level of confidence are not significantly related. Only 1 (2.4%) respondent chose the correct option and gave the correct explanation using the Bayes’ Theorem, “Let A stand for positive and B for virus. Due to the existence of false positive (i.e., those who tested positive but are not carriers), it follows that $P(B) < P(A)$. By $P(A|B) * P(B) = P(B|A) * P(A)$, I get $P(A|B) > P(B|A)$.” There were 9 (21.4%) respondents who chose the correct option but failed to support their choice with adequate evidence. Apart from 2 responses that were self-contradicting, similar explanations were provided by 7 respondents, including “by common sense”, “While there exists false positive, carriers will always test positive”, “It’s cause and effect”. 32 respondents (76.2%) chose a wrong answer. 24 respondents opted for “The events in the above two options are equally likely to take place”, whose explanations include “Testing positive and being a carrier are sufficient and necessary conditions to each other/the meaning of the two probabilities are the same” (9 respondents), “Assume there is no false positive or false negative/assume the test is 100% accurate” (2 respondents) as well as “Assume the probability of false positive to be the same as that of false negative” (2 respondents, including Alex who drafted the diagram shown below and was very confident about his answer).



Alex’s draft for Q10

However, when Alex was questioned with the reason behind his assumption, he hesitated and said, “If the probabilities of the false positive and false negative errors are different, I would probably choose ‘A person gets a positive result if this person is a carrier of the virus’; it makes more sense to me by common sense”. As for the other interviewee Joey, she said it was more “comfortable and reassuring” for her to choose the last option out of an urge to balance the first two options. This feeling could be a magic psychologic experience for some respondents though they did not always report such a feeling. The other respondents were aware of the existence of test errors but did not know how to explain. The remaining 8 respondents opted for “A person is a carrier of the virus if this person gets a positive result”. Typical explanations include “By inference: the positive result is indicative of the virus”, “A carrier of the virus may get a negative result, while most of the positive results are indicating that the person is a carrier of the virus”. Their responses demonstrate the misconception involving the ignorance of the case of “false positive” and the trust in the reliability of positive results. However, the former group of 24 respondents fell into another misconception known as “confusion over the inverse”, even though they considered the test’s accuracy.

4. Conditioning Event

The two questions (Q4, Q6) in the instrument are related to the conditioning event. There is no significant difference (α=.05) in the combined test scores for these two questions among groups regarding probability/statistics education at either secondary or tertiary stage.

For Q4, the mean test score is 0.24 and the mean confidence rating is 4.14. Analysis shows that participants’ actual performance and reported level of confidence are not significantly related. Only 3 (7.1%) respondents chose the correct option and provided correct explanations. Two of them held the idea that “Since one of the four combinations, (B, B), has been ruled out, the probability should be 1/3”. However, in the interview, Alex revealed that his 1/3 was not precisely 1/3, but actually (1/4)/(3/4). In other words, he adopted a mindset of probability. He managed to pull in each number in their place accordingly with the formula of conditional probability, which is quite intuitive by its appearance of a ratio contrast. Someone even listed out the original sample space $\Omega = \{(B, B), (B, G), (G, B), (G, G)\}$. This respondent highlighted the reduction of it upon knowing one of the kids is a girl, adding that it should be reasonable to assume each event in the sample space to be equally likely. Similarly, the other successful respondent drew a tree diagram and eliminated the impossible case. 4 (9.5%) respondents chose the correct answer but gave incorrect explanations. Three of them said, “it should be 1/4”, with only one respondent adding, “I think it is not about conditional probability; therefore, I think it is 0.5^2”. Their misconception seemed to result from the tendency to interpret a conditional probability $P(A \cap B)$ as a compound one $P(B|A)$, even with the event A given as the condition specifically. 35 respondents (83.3%) chose the incorrect option and gave incorrect explanations. These respondents all supported the claim and gave almost consistent reasons, including “The genders of the two kids are independent of each other”, “The probability will not be affected by the fact that one of Mr White’s children is girl”, “The second child can only be either male or female”. Our interviewee Joey said she was very confident about it, “Because of independence, the information ‘one of them is a girl’ does not seem relevant to me”. “This is the easiest one out of all 10 questions”. Another one respondent even confirmed this idea again with the conditional probability calculation (1/4)/(1/2). Admittedly, the idea of “independence” and “half-half chance” will work rather well for another similar problem: “Mrs White has a daughter and is expecting a second baby now. What is the probability that the second kid is also a daughter?". However, when compared to our version, this new question is totally different. For this similar question, the sample space consists of two events only, (G, B) and (G, G), as is determined by the known gender of the first kid and the random gender of the second kid to be born yet. In contrast, for our version, the sample space Ω = {(B, B), (B, G), (G, B), (G, G)} is determined by the four possible combinations, where "independence" has already played a part. Under the circumstance of "one of them is a girl", the sample space should be further restricted to {(B, G), (G, B), (G, G)}, still all equally likely. We can see that the gender of the second child is no longer independent of the condition that "at least one of the two kids is a girl". Of the three events left, only (G, G) is the desired outcome, which results in a probability of 1/3. Therefore, the main misconception here might be attributed to the confusion over different sampling procedures or simply the tendency to regard the conditioning event as independent or irrelevant.

For Q6, the mean test score is 0.29 and the mean confidence rating is 4.02. Analysis shows that participants' actual performance and reported level of confidence are not significantly related. 6 respondents (14.3%) chose the correct option and provided correct explanations. Four of them directly showed their step (1/3)/(1/2), a calculation of conditional probability. Two of them, including Alex, provided more illustrations, "There are three conditions when a side is red; A-side of double red card, B-side of double red card, and the red side of the mixed card; and there are two cases when the other side is red also, so the possibility is 2/3". 4 respondents (9.5%) chose the incorrect answer 1/4, one of them showing the correct method and calculation but somehow chose this incorrect option, one with no idea and the other showing "1/2 * 1/2 = 1/4". Once again, the confusion between conditional and compound probability was present. 32 respondents (76.2%) chose the incorrect answer 1/2 and gave the incorrect explanation. These respondents supported the claim and gave almost identical reasons that we paraphrased as "Because it must not be the all-white card. For the other two cards, one of it should be red and other should be white. Therefore, should be 1/2", which suggest the problematic interpretation of the conditioning event; Although the inference that "the all-white card is out" is correct, this should not be the event upon which the "the hidden side is also red" should be conditioned. It is simply because the possible outcomes should be one of the six faces of equal chance to show up, instead of the three cards. Here is an insight shared by the interviewee Alex when he was asked to comment on this inference: "I agree with the idea that "the all-white card is out", however, I think the other two cards are not equally likely to be the chosen one. Say, Card A has one red face, but the other Card B has two red faces, and “a red side is up” means the card B are twice likely than Card A to be the selected card. Therefore, the probability should be 2/3 instead of 1/2”. However, the respondents who obtained a probability of 1/2 overlooked this inequality between the two remaining cards but instead assigned half-half chance, as can be seen in the response: “There are two cards having at least a side is red within the three cards, so the denominator is 2, and only one card within the two cards has the other side red, so the numerator is 1”. Similarly, the interviewee Joey explained, “Because the card was selected at random, the sample space is {RR, RW, WW} with each equally likely. Now that WW has been eliminated, there are two cards left. It must be one of them, still equally likely”. In fact, “a red side is up” should be the conditioning event of the target event, “the hidden side is also red”. Out of three red sides, two sides also have the other side of red. Hence the probability is 2/3.

5. Time Axis Fallacy

The two questions (Q5, Q9) in the instrument are related to the conditioning event. There is no significant difference (α=.05) in the combined test scores for these two questions among groups by probability/statistics education at either secondary or tertiary stage,

For Q5, the mean test score is 0.81 and the mean confidence rating is 3.14. Analysis shows that participants’ actual performance and reported level of confidence are not significantly related. 6 respondents (14.3%) chose the correct option and provided correct explanations, who said this statement could be true under some circumstances, “The probability of the earlier event could be revised if the two events are related”, “Although that will happen later, as long as it affects the outcome, I think we should take it into account when we calculate the probability”, “Suppose event A happened first, and then event B happened”. The fact that A is not affected by B does not exclude the case when event A affects event B. Therefore, we can deduce the probability of event A from the occurrence of B". Some pointed out the weakness of the statement, “Revising the previous event after the later occurrence does not imply conflicts”. Some others even explained more specifically on “how” the revision can be done, “The occurrence of the latter event may eliminate some possibilities of the early event”. There was also an insightful example demonstrating how “P(the earlier event is a mistake)” can be revised due to the later occurrences, whose idea is quite easy to follow, “If the first event was a mistake that happened without being noticed, the later correction event could change people's acknowledgement so that they will realise the first event is a mistake”. Indeed, the probability of a particular event is not fixed but might vary depending on how it is perceived when new information comes into sight. There were 21 responses (50%) with the correct option but incorrect explanations. Most of these responses restated the latter part of the statement without giving actual reasons; some said they had learned this fact but did not know how to explain, and some directly said they had no idea. Nevertheless, some responses were found to be rather interesting and made sense to a certain degree despite that the language they used could not be considered mathematically correct. These included "When you assume a coin is fair, then getting a head should be 1/2. But when you keep tossing it, you found that it gets head every time. Then the information will [change] (not exactly 'change', but the 'allow people to infer/revise the probability of') the previous event", "Because the things that took place later may [change your mind] (does it have anything to do with the probability of the previous event?)", "The current occurrences can be used to [avoid] (The meaning of avoid is ambiguous because it could possibly mean 'going back in time') something in the past". Though these responses could be interpreted as the partially correct mode of thinking if we lower the criteria for a bit, the inappropriate words they used might suggest a deficiency in their mathematical register, which may be further attributed to their insufficient knowledge construct. For instance, the interviewee Joey borrowed the example of "the possibility of taking an umbrella with me when going out" and "the weather forecast". While the possibility will be revised according to the weather forecast that day, "taking an umbrella" takes place after the weather forecast, therefore contradicting the order prescribed in the question. It is difficult to tell whether the problem is simply a matter of comprehension or failure to exemplify the scenario properly. When the point of temporal order was made to Joey, she admitted: "I'm not so sure if I understood the question so well. I'll just be frank that I might switch my answer to 'TRUE', if you ask me a second time. The statement seems natural and intuitive to me". The remaining 15 respondents (35.7%) made the incorrect judgement and gave incorrect explanations, which could be sorted into two main types of misconception. The first misconception, as described in the literature, lies in the thinking that the past happened already and cannot be changed. A respondent even mentioned how and when the probability should be determined: "The probability has been determined in the specific situation at the time of happening, and subsequent changes will not affect the original probability". Moreover, it is found that the second misconception lies in the subjective wishful thinking that the two events are independent. However, the recognition of the existence of dependent situations is what is needed to adopt the correct thinking mode. When we further questioned the interviewee Alex on his assumption of independence, he held back and, upon the realisation of dependents events, changed his mind.

For Q9, the mean test score is 0.67 and the mean confidence rating is 3.83. The 14 respondents (33.3%) who chose the correct option also gave correct explanations that could be categorised as three approaches. The first approach is the most direct but a little counterintuitive in terms of the time axis, "If the second marble is blue, then there are three marbles left for the first pick (red, blue, blue), so the probability that the first marble is red is 1/3, the only red one out of all three marbles". Compared to Q5, more respondents succeeded in resisting the temptation of the time axis, probably because they were presented with a concrete situation instead of being required to develop a suitable example to back up their choice. The second approach is the calculation of the conditional probability by formula, "P(the second marble is red) = (1/2)(1/3)+(1/2)(2/3) = (1/2), P(the first marble and the second marble are red) = (1/2)(1/3) = (1/6), P(the first marble is red | the second marble is red) = (1/6)/(1/2) = 1/3". Not only did these respondents recognise the second event's conditioning function on the first event, but they also recalled the relevant knowledge and formula needed to solve the problem. The third approach is a more indirect one. Instead of calculating P(R1|R2) directly, a respondent adopted the Bayes' Theorem: P(R1|R2) = P(R2|R1)*P(R1)/P(R2) = (1/3)*(1/2)/{[(2/4)*(1/3)]+[(2/4)*(2/3)]} = 1/6. Although more complex than the first two approaches, this method cleverly avoids the "face-to-face" encounter with the time axis fallacy where the trust in formula outweighs the biased intuitions in the human mind. Among the wrong answers, 21 respondents (50%) chose "1/2" and gave similar explanations, including "When drawing the first marble, there are four marbles in the box including two red and two blue, so the probability of drawing a red for first marble is 2/4 = 1/2" and "The colour of second marble does not affect the probability of the colour of the first marble". This type of misconception is caused by the tendency to determine the probability of some certain event at the exact time of happening, once and for all, regardless of what subsequent results might be. In other words, the concept of conditional probability is judged to be conditional only on the immediate background rather than events at a later stage. Our interviewee Joey confirmed the above reasoning and added, "At the beginning, we didn't know if a red marble would be chosen at the second pick. Hence, that marble should also be taken into account, making up the denominator of 4 in 2/4 = 1/2". When asked if she disagreed with the opinion that "The red marble at the second pick will not show up at the first pick", she said, "I agree with it, but this opinion is irrelevant to this question. I will stick to my answer". 4 respondents (9.5%) multiplied 2/4 by 1/3 and thus obtained "1/6", which again implies the tendency to interpret conditional problem P(A|B) as compound one $P(A \cap B)$. 3 respondents (7.1%) chose “1/4”, claiming that “Knowing that the second marble is red will affect the probability of the first pick”. Despite this correct reasoning, another misconception is uncovered by our guessing of the meaning of 1/4: 1 stands for the only red marble left, and 4 stands for the total number of marbles before the first pick. While the former part is exactly what would be expected in a normative response, the latter part is still under the time axis fallacy effect. This reasoning approach suggests the co-existence of two contradictory reasoning in these three respondents. Furthermore, as our interviewee Alex reported, “It's pure instinct”, we may infer that these three respondents might not be even aware of such a problem at the time of decision-making.

DISCUSSION

This study analysed preservice mathematics teachers' performance in an undergraduate programme from the same university on several conditional probability reasoning tasks. The results demonstrated distinctive patterns of relationship between probability/statistics education and the five misconceptions. For groups by secondary probability/statistics education, for causal and diagnostic thinking questions, preservice teachers who took M1 (calculus and statistics) performed significantly worse than those who did not take any mathematics extended modules. This finding suggests the possibility that this type of misconception may increase with probability/statistics education in the context of Hong Kong. For groups by tertiary probability/statistics education, preservice teachers who have taken both MTH4155 and MTH4153 performed significantly worse than those who have only taken MTH4155 for without placement tasks. It is somewhat disappointing that we did not observe a trend that preservice teachers' performance increases with probability/statistics education. As for the other three misconceptions, it is somewhat surprising that no significant differences were detected amongst groups by secondary or tertiary probability/statistics education in the total test score they obtained. This indicates that our preservice teachers' education in probability/statistics did not affect their general performance when solving conditional probability tasks for the items used in this study.

Concerning the literature review, the five types of misconceptions in the literal sense of their definitions were present in our preservice teachers. The first step to avoid any misunderstanding is simply a matter of comprehension. We recommend that preservice teachers read the given question cautiously, appreciate each key word's reference in the specific context, and avoid any kind of insufficient or biased interpretation of the conditions in the problem. Relating to comprehension, for example, a misconception occurs when the given clear-defined percentages are misinterpreted as probabilities, which is not just a matter of carelessness but also a negative effect of the base rate fallacy (Q2).

Apart from the five misconceptions that are distinctive in the contexts where they will occur, in the analysis of biased responses, we discovered another subset of misunderstandings rooted in the minds of preservice teachers. Relating to the conditional probability P(A|B), misconceptions seem to be consistent with: (1) the conditioning event B; (2) the target event A; (3) the relationship "|" between. A major one, "Ignorance of the conditioning event", is the disregard of the conditioning event B's presence. Despite resorting to the new information provided by B, some respondents claimed that P(A) should be fixed rather than changeable, justifying that A and B should be independent even in the case of dependence (Q1, Q5, Q7). A related misconception, "Probability all by itself", lies in the thinking that P(A) could be determined in the absence of the conditioning event B, and the reluctance to accept the dynamic nature of P(A) when more conditioning events becomes available (Q1, Q5, Q7). Since this type of misconception often comes parallel to "Ignorance of the conditioning event" and is thus difficult to be differentiated, we are not certain whether either or both of them are present in the biased probabilistic reasoning process. Interestingly, the converse misconception "Ignorance of the background information" also arises when overwhelming importance was attached to B, whereas A became completely irrelevant (Q7). The most prevalent misconception "Conjunction or Conditional?" lies in the misinterpretation of conditional probability P(A|B) as compound probability P(A∩B) when faced with the type of question "If/Given/When B, what is the probability of A" without any ambiguity in wording (Q1, 4, 8, 9). Another common misconception about general probability was described as the "Equiprobability Bias" by Lecoutre (1985) as the tendency to consider two random events as equiprobable even in case of different probabilities, with no explanations or out of instinct. This is related to the misidentification of the conditioning event B since only the primary outcomes should be equiprobable but not the secondary ones upon inference (Q6). Hence, it would be beneficial if preservice teachers could acknowledge the common committed misconceptions and make conscious efforts to resist their biased modes of thinking. What we could think of now is to share the results of this research will be shared with pre-service teachers at EdUHK with the hope that they may find it interesting and illuminating.

However, due to a small sample size, our findings might not be transferrable to either all preservice teachers in EdUHK or the community beyond. It is also important to note that some possible biases may exist in the responses to the questionnaire regarding the five brain teasers. It is likely that some respondents may choose an unusual (seemingly correct) response instead of their favoured (seemingly wrong) one since it might be considered the best strategy for handling questions in the form of brain teasers. A shred of evidence would be the inconsistency between the choice and the explanation in some responses we collected. Sometimes "no explanations" responses were also observed, probably due to the indecisiveness over the choice of "true or false" or the use of intuitive thinking without further attempts at reasoning. Hence, the quantitative data and results need to be interpreted with caution. The results for the qualitative part, on the other hand, are comparatively more reliable to a certain extent since the data were obtained directly from the questionnaire responses as well as two interviews that followed. One issue emerging from the interviews is that one of the interviewees, Alex, changed his mind and opted for the correct option when given a second time. In the interview, we noticed that when Alex became more elaborate in his explanations, he gradually discovered his biased reasoning weaknesses and made a smart switch. An implication of this for the questionnaire design would be to integrate a compulsory one-minute-pause session following each question's completion, allowing respondents to examine their explanations on the grounds of the given conditions. In this way, it is more likely to obtain valid responses upon careful considerations to enhance qualitative and quantitative data's authenticity. To sum up, it is advisable that further research should be done with a larger sample size and that efforts should be made to the building of a more reliable and effective instrument.

CONCLUSION

This research aims to assess Hong Kong pre-service mathematics teachers' misconceptions in conditional probability. A comprehensive questionnaire based on the five misconceptions in the literature review. The presentation of brain teasers as questionnaire items is educationally significant, which poses the urge for pre-service teachers to think carefully and independently for themselves to distinguish the seemingly true lies from the seemingly false facts. It is a disappointment when we found the five misconceptions were all present and sometimes persistent in the minds of our pre-service teachers. Apart from the misconception relating to without placement situations, we witness a worrying prevalence of the other four misconceptions. For each misconception, different kinds of biased reasonings were identified and analysed. It is rather surprising when pre-service teachers' performance was found to be unrelated to their probability or statistics education at secondary and tertiary stages. Only for two misconceptions, significant differences were observed, which indicates the possibility that the respective misconceptions may increase with probability or statistics education to a certain degree. We also investigated the relationship between the actual performance and the confidence rating. Generally, there was no relationship between performance and confidence, though a weak positive correlation was observed for two questionnaire items. In the follow-up interview session, it was frequent when the two pre-service teachers abandoned the incorrect option they had made but switched to the correct one when given a second chance. This suggests the coexistence of two contradictory ways of probabilistic thinking, the tendency to lean towards the incorrect one at first sight and the self-correcting switch to the correct one upon careful considerations. Therefore, it will be interesting for further research to investigate: (1) What are the condition(s) when such "a wise switch" can be made? (2) Is "the wise switch" related to being questioned by others, or it is just a self-correction upon awareness with caution? (3) Is there any effect of one's cognitive ability on making "the wise switch"? (4) Is such a switch always "wise"? These are only examples of questions to be answered. Apart from pre-service teachers, further research could be expanded to other groups such as secondary students and regions beyond the scope of Hong Kong that allows comparison studies to be conducted. Looking back at where we started, this study was undertaken in response to the insufficient understanding and documentation of pre-service teachers' misconceptions in conditional probability in the context of Hong Kong. The unsatisfying results are causing concerns that pre-service secondary mathematics teachers have not been so well prepared by probability education. It is time that pre-service teachers should examine and correct their probabilistic thinking as soon as possible before they become in-service teachers who will be held accountable for future generations. While biased intuitions get in the way, the efforts to resist such intuitions will not be in vain.

REFERENCES

Bar-Hillel, M., & Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition*, *11*(2), 109-122.

Bar-Hillel, M. (1983). The Base Rate Fallacy Controversy. In R. W. Scholz (Ed.), *Decision Making Under Uncertainty* (pp. 39-61). North Holland: Elsevier Science Publishers.

Borovcnik, M. G., & Bentz, H. J. (1991). Empirical research in understanding probability. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education* (pp. 73-105). Dordrecht, The Netherlands: Kluwer.

Borovcnik, M., & Peard, R. (1996). Probability. *International Handbook of Mathematics Education*, 289-325.

Dollard, C. (2007). Preservice elementary teachers' thinking about situations involving probability.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic and A. Tversky (Eds.) *Judgment under uncertainty: Heuristics and biases* (pp. 249-267). New York: Cambridge University Press.

Hong Kong Curriculum Development Council (2017). *Supplement to Mathematics Education - Key Learning Area Curriculum Guide Learning Content of Senior Secondary Mathematics (secondary 4-6)*. Retrieved from https://www.edb.gov.hk/attachment/en/curriculum-development/kla/ma/curr/ssmc2017_e.pdf

Jendraszek, P. A. (2008). Misconceptions of probability among future teachers of mathematics.

Kelly, I. W. & Zwiers, F. W. (1988). Mutually Exclusive and Independence: Unravelling Basic Misconceptions in Probability Theory. In: R. Davidson and J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*, Victoria BC.: University of Victoria.

Lecoutre, M. P. (1985). Effect d'informations de nature combinatoire et de nature fréquentielle sur les judgements probabilistes. *Recherches en Didactique des Mathématiques*, *6*, 193–213.

Li, J. (2000). Chinese student's understanding of probability. (Doctoral thesis, National Institute of Education, Nanyang Technological University, Singapore). Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/dissertations/00.Li.Dissertation.pdf

Liu, Tzu-Chien, Lin, Yi-Chun, & Tsai, Chin-Chung. (2009). IDENTIFYING SENIOR HIGH SCHOOL STUDENTS' MISCONCEPTIONS ABOUT STATISTICAL CORRELATION, AND THEIR POSSIBLE CAUSES: AN EXPLORATORY STUDY USING CONCEPT MAPPING WITH INTERVIEWS. *International Journal of Science and Mathematics Education, 7*(4), 791-820

Reaburn, R. L. (2013). Students' understanding of conditional probability on entering university. In *36th annual conference of the Mathematics Education Research Group of Australasia (MERGA 36)* (pp. 570-577).

Sadri, P. (2012). Clarifying the Language of Chance Using Basic Conditional Probability Reasoning: The Monty Hall Problem. *Open Journal of Discrete Mathematics*, *2*(4), 164-168.

Tarr, J. E., & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence, *Mathematics Education Research Journal*, *9*(1), 39-59.

Tarr, J. E., & Lannin, J. K. (2005). How Can Teachers Build Notions of Conditional Probability and Independence? In G. A. Jones (Ed.), *Exploring probability in school* (pp. 215-238). Boston, MA: Springer.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgment under uncertainty. In M. Fishbein (Ed.), Progress in social psychology. Hillsdale, N.J.: Lawrence Erlbaum.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 84-98). Cambridge: Cambridge University Press.

Falk, R. (1986). Conditional probabilities. Insights and difficulties. In R. Davidson & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*. Victoria, BC: University of Victoria.

Falk, R. (1989). Inference under uncertainty via conditional probabilities. In R. Morris (Ed.), *Studies in mathematics education: The Teaching of statistics* (pp. 175-184). Paris: UNESCO.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics, 15*(1), 1-24.

Fischbein, E., & Schnarch, D. (1997). The Evolution with Age of Probabilistic, Intuitively Based Misconceptions. *Journal for Research in Mathematics Education, 28*(1), 96-105.

APPENDIX: Questionnaire

- The official online version can be assessed at https://forms.gle/FscHHuDRMYhr1jiNA
- Correct judgements are given in **bold**.

Item 1

a. "At a party, everyone will take turns to draw lots (without replacement) to win a prize. Theoretically, everyone has an equal chance to win the prize. Interestingly though, your chance of winning will increase when it is your turn now but no one has won the prize yet." Is this statement **true** or false?
b. Explain why you think it's true or false (in either English or Chinese).
c. Rate how confident you feel about the answers you gave above (1=very unconfident, 2=unconfident, 3=neutral, 4=confident, 5=very confident).

Item 2

a. (adapted from Bar-Hillel, 1983). "A local newspaper in Hong Kong claimed that 5% of all pedestrians got killed by car when crossing the street on a green light. In contrast, only 3% of all pedestrians got killed by car when crossing the street on a red light." Is this claim likely to be **true** or false?
b. Explain why you think it's true or false (in either English or Chinese).
c. Rate how confident you feel about the answers you gave above (1=very unconfident, 2=unconfident, 3=neutral, 4=confident, 5=very confident).

Item 3

a. (Tversky & Kahneman, 1982). "Compared with the probability that a daughter has blue eyes if her mother has blue eyes, the probability that a mother has blue eyes if her daughter has blue eyes is greater". (Note that you may consider the proportions of blue-eyed individuals in the two generations as equal). Is this statement true or false?
b. Explain why you think it's true or **false** (in either English or Chinese).
c. Rate how confident you feel about the answers you gave above (1=very unconfident, 2=unconfident, 3=neutral, 4=confident, 5=very confident).

Item 4

a. "Mr. White has two children. One of them is a girl. The probability that Mr. White's other child is also a girl is 1/2." Is this statement true or **false**?
b. Explain why you think it's true or false (in either English or Chinese).
c. Rate how confident you feel about the answers you gave above (1=very unconfident, 2=unconfident, 3=neutral, 4=confident, 5=very confident).

Item 5

a. “The occurrence of an event could not be affected by some events that took place later. For the same reason, the probability of a previous event could not be revised according to some information about current occurrences.” Is this statement true or **false**?
b. Explain why you think it's true or false (in either English or Chinese).
c. Rate how confident you feel about the answers you gave above (1=very unconfident, 2=unconfident, 3=neutral, 4=confident, 5=very confident).

Item 6

a. (Bar-Hillel & Falk, 1982) Three cards are in a hat. One is red on both sides, one is white on both sides, and one is red on one side and white on the other. One card is drawn from the hat and put on the table with a red side facing up. What is the probability that the hidden side is also red?

   A) 1/2
   B) 1/4
   C) 2/3
   D) 2/5

b. Explain why you think it's true or false (in either English or Chinese).
c. Rate how confident you feel about the answers you gave above (1=very unconfident, 2=unconfident, 3=neutral, 4=confident, 5=very confident).

Item 7

a. (Adapted from Tversky & Kahneman, 1982). A taxi was involved in a traffic accident and disappeared into the midnight. It is known that 85% of the taxis are yellow and 15% are green. A witness reported the taxi as green. The court tested the reliability of this testimony and concluded that the witness correctly identified the colour of the taxi 80% of the time. Given this information, what is the probability that a green taxi was involved in the accident?

   A) 12%
   B) 15%
   C) 29%
   D) 41%
   E) 80%

b. Explain why you think it's true or false (in either English or Chinese).
c. Rate how confident you feel about the answers you gave above (1=very unconfident, 2=unconfident, 3=neutral, 4=confident, 5=very confident).

Item 8

a. (Falk, 1986). Amy received a box containing two red marbles and two blue marbles. Amy picks up two random marbles from the box, one by one, without placement, and puts them aside without looking. Given that the first marble is red, which answer is true?

   A) The second marble is more likely to be red.
   B) The second marble is more likely to be blue.
   C) The probability for the second marble to be red and the probability for the second marble to be blue are the same.

b. Explain why you think it's true or false (in either English or Chinese).
c. Rate how confident you feel about the answers you gave above (1=very unconfident, 2=unconfident, 3=neutral, 4=confident, 5=very confident).

Item 9

a. Amy received a box containing two red marbles and two blue marbles. She picks up two random marbles from the box, one by one, without placement, and puts them aside without looking. If the second marble is red, what is the probability that the first marble is red?

   A) 1/6
   B) 1/4
   C) 1/3
   D) 1/2

b. Explain why you think it's true or false (in either English or Chinese).
c. Rate how confident you feel about the answers you gave above (1=very unconfident, 2=unconfident, 3=neutral, 4=confident, 5=very confident).

Item 10

a. (Analogue to Eddy, 1982). A COVID-19 test is administered to all residents in a city. A positive result is indicative of the virus and a negative result is not. Which of the following statements is more likely?

   A) A person is a carrier of the virus if this person gets a positive result.
   B) A person gets a positive result if this person is a carrier of the virus.
   C) The events in the above two options are equally likely to take place.

b. Explain why you think it's true or false (in either English or Chinese).
c. Rate how confident you feel about the answers you gave above (1=very unconfident, 2=unconfident, 3=neutral, 4=confident, 5=very confident).