

RESEARCH

Open Access



Evidence of probability misconception in engineering students—why even an inaccurate explanation is better than no explanation

Marija Kaplar^{1*} , Zorana Lužanin² and Srđan Verbić³

Abstract

Background: In the rapidly changing industrial environment and job market, engineering profession requires a vast body of skills, one of them being decision making under uncertainty. Knowing that misunderstanding of probability concepts can lead to wrong decisions, the main objective of this study is to investigate the presence of probability misconceptions among undergraduate students of electrical engineering. Five misconceptions were investigated: *insensitivity to sample size, base rate neglected, misconception of chance, illusory correlation, and biases in the evaluation of conjunctive and disjunctive events*. The study was conducted with 587 students who attended bachelor schools of electrical engineering at two universities in Serbia. The presence of misconceptions was tested using multiple-choice tasks. This study also introduces a novel perspective, which is reflected in examination of the correlation between students' explanations of given answers and their test scores.

Results: The results of this study show that electrical engineering students are, susceptible to misconceptions in probability reasoning. Although future engineers from the sample population were most successful in avoiding misconceptions of chance, only 35% of examinees were able to provide a meaningful explanation. Analysis of students' explanations, revealed that in many cases majority of students were prone to common misconceptions. Among the sample population, significant percentage of students were unable to justify their own answers even when they selected the correct option. The results also indicate that formal education in probability and statistics did not significantly influence the test score.

Conclusions: Results of the present study indicate a need for further development of students' deep understanding of probability concepts, as well as the need for the development of competencies that enable students to validate their answers. The study emphasizes the importance of answer explanations, since they allow us to discover whether students who mark the correct answer have some misconceptions or may be prone to some other kind of error. We found that the examinees who failed to explain their choices had much lower test scores than those who provided some explanation.

Keywords: Misconceptions in probability reasoning, Open-ended tasks, Decision-making under uncertainty

* Correspondence: marja.radojicic@uns.ac.rs

¹Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, Novi Sad 21000, Serbia
Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

The dynamic world in which engineers operate on a daily basis is always demanding and keeps providing fresh challenges in the form of profoundly diverse and incessant changes which require a host of technical competences, as well as non-technical skills, such as decision-making under uncertainty, critical thinking, data use skills, logical thinking, and problem solving (Foster, Wigner, Lande, & Jordan, 2018; Zilinski, Nelson, & Van Epps, 2014). In such an environment, engineers are often in a situation to make decisions in uncertain situations, as well as to argue their solutions and choices with a purpose to convince employers, colleagues, or others in the correctness of their decisions. The modern work environment requires engineers to possess competencies in the decision-making under uncertainty, where they should be able to identify, understand, and apply basic concepts from probability and statistics in both their professional and personal life (Kang & Park, 2019). Besides decision-making under uncertainty, one of the important non-technical competences of engineers is also ability to explain and justify the provided decisions or solutions. In line with this, graduates of engineering programs are increasingly expected not only to use available data to make decisions, but also to develop skills that allow them to explain their solutions, in order to be successful part of workforce (Lin, Wu, Hsu, & Williams, 2021; Stehle & Peters-Burton, 2019; Zilinski et al., 2014).

Making decisions in uncertain situations is not only important for engineers but also for the general population, which is the reason why the research of misconceptions in probability has been in focus several decades (Kang & Park, 2019; Kustos & Zelkowski, 2013; Paul & Hlanganipai, 2014; Tsakiridou & Vavyla, 2015; Tversky & Kahneman, 1974). Although many engineering curricula are updated in line with the needs of modern society, it is questionable to what extent engineering students can avoid probability misconceptions which can be crucial for making decisions under uncertainty. Another important issue is students' argumentation and explanations of their own decisions and solutions.

This study investigates the presence of five common probability misconceptions in students of electrical engineering using localized versions of tasks which were previously used in similar studies (Blanco & Chamberlin, 2019; Kahneman & Tversky, 1973; Kang & Park, 2019; Kustos & Zelkowski, 2013; Nabbout-Cheiban, 2017). We investigate the extent to which students of electrical engineering are prone to following types of misconceptions: *insensitivity to sample size*, *base rate neglected*, *misconception of chance*, *illusory correlation*, and *biases in evaluation of conjunctive and disjunctive events*. Special focus was placed on the connection between the existence of students' explanations and their test scores.

Additionally, regarding this study involved students of the first and third year, the impact of the year of study was tested. Further, the relationship between the finished secondary school (gymnasium or vocational secondary school) and the achieved test score was examined in the case of first year students, while for the third year students examined was the influence of the previously passed courses relate to probability and statistics.

Despite of existence of the vast body of studies on this topic among adults in various professions—from health professionals or bank employees (Bramwell, West, & Salmon, 2006; Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007; Kang & Park, 2019) to secondary or university students (Hirsch & O'Donnell, 2001; Khazanov & Prado, 2010; Kustos & Zelkowski, 2013; Paul & Hlanganipai, 2014)—to the extent of our knowledge, there is a lack of investigations of these misconceptions, especially among engineering students. Our aim is to contribute to narrowing the knowledge gap in this area by providing a novel perspective on the testing of this kind of misconceptions among engineers with special examination of relationship between students' explanations and their test scores, as an original contribution in this area.

Literature review

Probability misconceptions

In their early works, Tversky and Kahneman (Tversky & Kahneman, 1974) defined typical misconceptions used under uncertainty in making judgments and decisions. Within this study, we will discuss some of them:

Insensitivity to sample size is applied when people believe that the probability of the judged sample statistic is independent of the sample size. Actually, in many cases, people believe that the similarity of the sample statistic and population parameter is independent on the sample size (Kang & Park, 2019; Tversky & Kahneman, 1974). More precisely, the fact that, in a small sample, large deviations can occur from the actual value of a parameter is often overlooked.

The misconception of chance occurs when people believe that the sequence of outcomes of random experiments will have the main characteristics of the random process, even when the sequence is short (Kang & Park, 2019).

Illusory correlation denotes a tendency to overestimate the degree of covariation between two variables and takes place when people make a wrong conclusion about the relationship between categories of events (Tversky & Kahneman, 1974).

The base rate neglected (base rate bias, insensitivity to prior probability) occurs when a base set or some other important information is neglected. In fact, information about the basic set is neglected, and usually, the

conclusion is made on the basis of additional information given in the assignment, which is essentially insufficient to make the right decision. It is wrong to believe that if the description of a representative is more specific to one class, there is a greater chance that this representative belongs to that class (Kang & Park, 2019; Tversky & Kahneman, 1974).

If, on the other hand, the description is not specific to any class, respondents are usually prone to *equiprobability bias*, where they believe that all outcomes have the same probability and in such situations, the base set is also neglected (Gauvrit & Morsanyi, 2014).

The Biases in the evaluation of conjunctive and disjunctive events occur when people overestimate the probability of conjunctive events or underestimate the probability of the disjunctive events (Bar-Hillel, 1973; Bazerman & Moore, 1994; Brockner, Paruchuri, Idson, & Higgins, 2002; Kang & Park, 2019; Tversky & Kahneman, 1974). Conjunctive events refer to the intersection of compound events while disjunctive events refer to the union of compound events (Bar-Hillel, 1973). In conjunctive events, when several events must all occur in order for a certain outcome to be realized, the likelihood that all of them will happen can be easily overestimated. By contrast, probability of disjunctive events can be underestimated where the separate probability of each event might be small but the probability of the union of such events is higher.

Earlier findings

Similar studies in various fields such as banking, investment, auditing, management, insurance, and health care have found that even well-educated adults struggle with the interpretation of information and drawing of conclusions in cases of uncertainty (Bílek, Nedoma, & Jirásek, 2018; Bramwell et al., 2006; Gigerenzer et al., 2007; Kang & Park, 2019). The study of Khazanov and Prado (2010) showed that about 80% of college students who major in the fields of accounting, business, liberal arts, and mental health were prone to some probability misconceptions such as equiprobability bias. Moreover, the study of Hirsch and O'Donnell (2001) found that about 75% of college students from various fields, mostly from psychology, were prone to probability misconceptions. In such situations, many rely on personal strategies that are, in many cases, typical and are known as heuristics (Kahneman & Frederick, 2002). Heuristics are not bad by definition (Bílek et al., 2018). On the contrary, they can represent a simplification of certain problems as well as a shortcut to reasoning. However, they can often lead to cognitive errors—misconceptions, of which the subject has no awareness of (Kahneman & Tversky, 1972; Kang & Park, 2019; Morvan & Jenkins, 2017). In terms of knowledge improvement, many studies (Batanero &

Borovcnik, 2016; Khazanov & Prado, 2010) testify that standard courses in stochastics do not have a great influence on probability misconceptions. Earlier studies (Chance, Ben-Zvi, Garfield, & Medina, 2007; Garfield & Ahlgren, 1988; Garfield & Ben-Zvi, 2007) emphasize that, although students can be very successful in calculating correct answers on probability tests, these same students frequently exhibit misunderstanding of basic ideas and concepts when making their judgments about uncertain events in real life (Smyrnaiou, Georgakopoulou, & Sotiriou, 2020). On the other hand, many studies claim that courses in probability and statistics can be very significant in solving this problem providing a different approach to teaching (Delmas, Garfield, Ooms, & Chance, 2007; Gauvrit & Morsanyi, 2014; Gigerenzer et al., 2007; Masel, Humphrey, Blackburn, & Levine, 2015; Morsanyi, Handley, & Serpell, 2013). They suggested that in the process of teaching, special attention should be paid to students' previous experiences, application of experiments, frequencies, and intuitive understanding of probability concepts. Such results indicate that some misconceptions can be avoided and stress the importance of the way of teaching. In line with this, our study will reveal to what extent engineering students in Serbia are able to avoid probability misconception, as well as whether there is a statistically significant difference in dealing with misconceptions between the students with and without previous formal education in probability and statistics. The results should indicate if there exists the need to investigate alternative approaches to teaching probability and statistics for engineering students.

Types of tasks

In the research of misconceptions, different tasks were used, based on the content and task format. The first group included research based on a test with general tasks, which means that the tasks were adjusted to the general public, and are usually used to test students in elementary and high schools, and higher education (Gürbüz, Erdem, & Firat, 2014; Kustos & Zelkowski, 2013; Triliana & Asih, 2019; Watson & Kelly, 2009). The second group of tasks was suited for a specific profession or discipline, and they were presented to doctoral students and employees (Kang & Park, 2019; Masel et al., 2015). Bearing in mind that this investigation was focused on undergraduate students, we used general tasks related to real-life examples. By putting the tasks into a realistic context, students are prone to believe that tasks are not typical mathematical problems, which reduces their anxiety.

Based on the task format, two major groups of studies can be observed. The first group consists of studies that use multiple-choice tasks (Gürbüz et al., 2014; Kang &

Park, 2019; Tsakiridou & Vavyla, 2015). Multiple-choice tasks are very common in many fields, due to their ease of grading, grading consistency, and ability to cover a vast amount of topics (Kastner & Stangla, 2011; Stankous, 2016). Despite many advantages, these tasks “push” subjects to answer questions even when they are not sure about the answer, because it allows elimination and guessing. Kastner and Stangla (2011) express the opinion that multiple-choice tasks are not suitable for testing high-level thinking and they also point out the issue of lucky guessing. Shaughnessy (1992) stated that multiple-choice tasks often give incomplete information and cannot be ideal for the investigation of students’ reasoning. Moreover, this task format enables quantitative, but not qualitative analysis. In addition, multiple-choice test-taker is asked to estimate the probability of some event, but the test does not allow one to determine how test-takers arrived at the answer (Hirsch & O’Donnell, 2001). The second groups consist of studies based on an open-ended task (Bekkink, Donders, Kooloos, De Waal, & Ruiters, 2016; Paul & Hlanganipai, 2014; Triliana & Asih, 2019). Open-ended tasks require time (Kustos & Zelkowski, 2013) and answering can be demanding, but they provide better insight into students’ reasoning. According to Chaoui: “Open-ended questions provide insights into the misconception of students and allow the teacher to evaluate the various techniques they use” (Chaoui, 2011). Use of open-ended tasks increases reliability, because guessing is minimized compared to conventional multiple-choice tasks (Kastner & Stangla, 2011). The disadvantage of using open-ended tasks is that only a few tasks can be included in the test which require an answering process. Also, in open-ended tasks, grading can be subjective and examinees with poor writing skills can be disadvantaged (Kastner & Stangla, 2011). In order to use the upsides, or at least to minimize the downsides of both types of tasks, in this study, we used combined tasks. The first part consisted of multiple-choice tasks, while the other part consisted of open-ended tasks, where students were required to explain their selection of answer. In other words, for better insight into students’ answers and their reasoning, we required open-ended explanations of their answers. Checking whether students can provide explanations for their choices reduced the noise of accidental correct answers.

Methods

Participants

The sample consisted of 587 students, all of which were attending undergraduate electrical engineering studies at two universities (409 from the University of Novi Sad, and 178 from the University of Niš). Out of 587 students, 201 of them were female, while the remaining 386 were male students. A total of 384 students were tested

in their first semester, i.e., at the first year of studies, while 203 students were tested in their sixth semester (third year of studies).

Regarding the first year students, we examined the impact of the secondary school they finished (gymnasium (general education secondary school) or the vocational (electrical engineering school)) on the test score they achieved in this study. During their secondary school education, the gymnasium students had approximately 680 math classes, of which 27 classes were dedicated to the topics of probability and statistics. As for the students with the vocational secondary school background, they had a total of 420 math classes. Although the teaching methodology and topics involved are basically identical in both types of schools, vocational schools do not provide classes related to probability and statistics. With this in mind, the goal of this investigation was to establish whether the gymnasium students have any advantage over their vocational-school counterparts when it comes to decision making in uncertain situations and critical reasoning. In this way, we will be closer to finding out whether the number of classes and present teaching methodology in gymnasiums have any effect on students’ capability to deal with the misconceptions related to concepts of probability. In our study, 269 out of a total of 384 first year students had previously finished gymnasium, while the remaining 115 went to vocational, electrical engineering secondary schools.

On the other side, the test scores of the third year students were examined in relation to the course in probability and statistics they took during university studies. Although students are supposed to have obtained formal knowledge, the question is how long to pertain such a knowledge and whether the students are able to recognize concepts of probability and statistics in the problems (tasks) which more or less deviate from the standard form and content of the university test problems. During their fifth semester, the students from our sample had an opportunity to take an elective course in probability and statistics. A total of 145 students, out of 203 third year students, attended the course, while the remaining 58 chose not to do so.

Development of the instrument

Previous studies have shown that heuristics research is feasible, by conducting such research using reconstructed judgment tasks from previous studies in the form of questionnaire tests (Kang & Park, 2019; Woods, Dekker, Cook, Johannesen, & Sarter, 2010). In a similar vein, the present study also reconstructed judgment tasks from previous experiments and used them for testing. Investigation of the presence of misconception in our study was conducted with the use of tasks that were adapted and modified from other studies that dealt with

similar topics (Blanco & Chamberlin, 2019; Kahneman & Tversky, 1972, 1973; Kang & Park, 2019; Kustos & Zelkowsky, 2013; Nabbout-Cheiban, 2017; Tversky & Kahneman, 1974). To ensure reliability and validity, this study translated the terms and conditions regarding judgment tasks used in the previous research and changed the examples of heuristics into tasks as necessary. The tasks used in this study are presented in Table 1.

As concerns the multiple-choice tasks, besides the right answer, most common misconceptions were offered as distractions, based on previous studies (Blanco & Chamberlin, 2019; Kang & Park, 2019; Kustos & Zelkowsky, 2013; Nabbout-Cheiban, 2017; Tversky & Kahneman, 1974). That is the main reason why the number of distractions varies throughout the tasks. Tasks “Lotto” and “R/W balls” (Table 1) are typical tasks that can be found in the standard curriculum of secondary

Table 1 Tasks used in this study

Tasks

Task 1—“**Maternity hospital**” (misconception type: *insensitivity to sample size*):

In one hospital on a typical day, 45 babies are born. In another hospital, on a typical day, 15 babies are born. Every year the number of born girls and boys is approximately equal. In both hospitals, the number of days was counted when 60% of the newborns were boys. Which hospital had a greater number of such days?

- a) The larger hospital (about 45 babies born per day)
- b) The smaller hospital (about 15 babies per day)**
- c) The number of such days would be approximately equal in both hospitals.

Task 2—“**Lotto**” (misconception type: *misconception of chance*):

In the game where 7 numbers are drawn out of 39, choose the combination that is most likely to win.

- a) 1, 2, 3, 4, 5, 6, 7
- b) 12, 14, 16, 18, 20, 22, 24
- c) 2, 7, 11, 38, 24, 15, 12

d) All have equal chance

Task 3—“**Mobile phones**” (misconception type: *illusory correlation*):

Two groups participate in a survey about the use of mobile phones. 70% of participants in the first group said that they used mobile phones while driving while 40% of participants in the second group stated the same. In the first group, there were 20 women and 10 men while in the second group there were 14 women and 16 men.

- a) Men use mobile phone while driving more often than women.
- b) Women use mobile phone while driving more often than men.
- c) Based on the collected data, it is not possible to make conclusions about mobile phone usage with respect to gender.**

Task 4—“**IT engineer**” (misconception type: *equiprobability bias*):

In one company 70% of employees hold a degree in information technology while 30% of employees hold law degree. One employee was randomly selected. His name is Mark, he is young, successful, and very driven. He likes to swim and exercises regularly. What is the likelihood that Mark is an engineer in information technology?

- a) 70%**
- b) 30%
- c) 50%

Task 5—“**R/W balls**” (misconception type: *biases in the evaluation of conjunctive and disjunctive events*):

Choose the game which gives you the most chance to win.

- a) Game 1: Draw a red ball from the bowl with half white and half red balls.
- b) Game 2: Draw seven red balls from the bowl with 90% red balls and 10% white balls where after each draw the color is recorded and the ball is returned to the bowl.
- c) Game 3: Draw at least one red ball out of seven drawings from the bowl with 90% white balls and 10% red balls. After each draw, the color is recorded and the ball is returned to the bowl.**

Task 6—“**Lawyer**” (misconception type: *base rate neglected*):

Mark has been interested in theater and music since childhood. He loves art, and he is an opera fan. Which of the following claims is more likely?

- a) Mark is a member of the Belgrade’s Philharmonics and plays clarinet.
- b) Mark is a lawyer.**

Correct answers are marked bold

mathematics education (in gymnasiums) and in introductory courses of probability and statistics at universities, so one can expect that the majority of students have already encountered similar problems during their previous education. Task “Lotto” is given in the real-life context and is associated with students’ experience. In contrast, “R/W balls” has little connection to real-life experiences and represents a typical task found in probability and statistics coursebooks. Tasks 1, 3, 4, and 6 are considered to be non-standard, and it is assumed that students have not encountered similar tasks during their previous education.

The tasks were piloted during pre-test period. Pilots involved 32 students, who were offered all six tasks in a single test. This test allowed the assessment of text clarity, bearing in mind that the majority of tasks were translated from English. In addition, this pre-test gave us an opportunity to assess the time necessary for giving answers. During the pre-test, students were allowed to ask questions related to the text of the tasks. In the briefings after the pre-test, the students mentioned that providing explanations for the given answers is time consuming. This comment was most often given by students who, rather than providing mathematical explanation, tended to give verbal descriptions. Based on the obtained results, we found that our instrument is quite demanding, because our examinees took a lot of time reading and understanding explanations, so we decided to reduce the number of tasks to four per test. From the given tasks, we created two tests with four tasks each, where two tasks were common for both tests.

As our study is largely based on investigation of representativeness heuristic, similar to previous studies (Hirsch & O’Donnell, 2001; Kang & Park, 2019; Kustos & Zelkowski, 2013), each of the test versions used in this study consists of three tasks dedicated to representativeness heuristic, but also one task in each version is dedicated to availability heuristic or adjustments and anchoring heuristic. The first version of the test contained tasks 1, 2, 3, and 4 (Table 1), while in the second version were tasks 4, 5, 6, and 2 (Table 1). The first version consisted of 3 tasks related to representativeness heuristics and a single task related to heuristics availability (task “Mobile phones”). The second test version consisted of three tasks 3 related to heuristics representativeness, and a single task dedicated to heuristics adjustments and anchoring (“R/W balls”). This second version also features tasks “Lotto” and “R/W balls” in order to allow us to recognize the difference between the real-context tasks used in teaching, and tasks which are also used in teaching but are devoid of any significant relationship with the real context. It should be noted that this comparison was not the primary goal of this study, but was included to allow us additional

insight necessary for drawing valid conclusions. Test version two contains tasks “IT engineer” and “Lawyer”, where both belong to base rate neglected heuristics. Again, although not in our primary focus, the two tasks are present in the same test to allow establishment of relationship between them, thus yielding valuable information for the subsequent conclusions.

Procedure

The data were collected during the 2018–2019 academic year, as part of a voluntary activity scheduled to take place before or after regular classes. Managements of both faculties issued their approvals, setting the dates and terms for research realization. In order to allow all students to participate, the terms set coincided with the obligatory subjects in the first and third year. The testing was conducted either at the beginning or at the end of the classes, which, for that reason, were prolonged for 30 min. In this way, bias was averted in terms of selection, because all present students were allowed to participate. Eight terms were set for the research (four for the first year, and four for the third year students). Students were informed by the faculty management about the research 1 week before the term. The study was presented as a study of reasoning about uncertain situations, while the students were also informed that participation is voluntary and anonymous. All students present at the lecture were given the test. A total of 628 tests were handed out. 6.53% tests were eliminated from analysis for being returned empty (26), while the additional 15 were excluded from the sample for reasons of incorrectly given answers—either no alternatives or more than one alternative was marked. Two research assistants were entering data independently, while inconsistencies were avoided by comparison with the original data.

Data analysis

Correctly selected alternative and correctly answered tasks

As all tasks within the test had two parts, the first part was related to the selection of an alternative, while the second part required them to provide rational explanation for the choice they made. Students’ performance in particular tasks was analyzed and presented here as *correctly selected alternative* and *correctly answered tasks*. *Correctly selected alternative* implies that the student had chosen the correct answer, i.e., that he or she had selected the right alternative in the multiple-choice scenario, regardless of explanation validity. Alternatively, the *correctly answered task* implies the right answer, correctly explained.

Estimated students’ test scores and students’ answer explanation

Performance was dichotomously scored. Answers that contained correctly selected alternatives followed with

valid explanation, scored 1, while all other answers scored 0. In addition, it was analyzed how many students fell to common misconceptions.

Rather than performing the conventional summing of scores for each task, Rasch model was used to estimate students' test score. In the remaining text, syntagm "test score" shall pertain to the test score evaluated by Rasch analysis, where it should be noted that the score is normalized so that zero represents mean. The main rationale behind using the Rash model was to bind the two four-task tests with two common tasks. Tests were designed to overlap in order to enable examiners to create a common students' test score scale. Therefore, these two tests can be treated as a single test, where examinees managed to answer only four out of six tasks (Boone, 2016). That way the results of two tests became comparable. Since the test had only four tasks, test score is not very practical for the comparison of the probability-related competences of individual students. However, this test score can be used to compare big groups of students and determine which group outperformed the other.

In order to analyze misconceptions, similar to study (Kustos & Zelkowski, 2013), all answers were coded and divided into three categories: correct answer (C), main misconception (MM), and alternative incorrect answer

(AI). C is a category where students selected correct answers while also providing meaningful explanations. MM is a category containing answers and explanations which point out to the most frequent misconception according to literature (Blanco & Chamberlin, 2019; Kahneman & Tversky, 1973; Kang & Park, 2019; Kustos & Zelkowski, 2013; Nabbout-Cheiban, 2017; Shaughnessy, 1992; Tversky & Kahneman, 1974). Belonging to category AI are all other answers and explanations based on students' personal convictions or beliefs. Given in Table 2 are examples of students' answer explanations for all of these categories per task.

In order to analyze relation between students' answer explanation and estimated students' test score, all answers were coded and divided into five different categories: c—correct choice without explanation, ce—correct choice with an adequate explanation, cwe—correct choice with wrong explanation, w—wrong choice without explanation, and we—wrong choice with some explanation.

Once the three coders independently evaluated tasks' answers, they have discussed criteria for the codes together. In the next stage, the three researchers coded the tasks independently. Discrepancies in coding occurred in less than 6% of cases. Those cases were subsequently re-

Table 2 Examples of students' explanations within the categories, C, MM, Als

Tasks	Categ.	Examples of students' explanations
<i>Maternity hospital</i>	C	"Smaller sample implies higher deviation."
	MM	"In the first maternity hospital more babies are born, thus there will be more boys and more days."
	AI	"I have the chance of 33.3% to get this answer right."
<i>Lotto</i>	C	"Each number can be drawn with equal probability, therefore, each combination of seven numbers has the same probability of being drawn."
	MM	"Because numbers are in no particular order."
	AI	"The lotto is rigged."
<i>Mobile phones</i>	C	"The only information we have is the ratio of men and women, and there is no information about their answers. Based only on the provided information, we cannot claim which sex uses the mobile phone more."
	MM	"In the first group the number of women exceeds the number of men and the percentage of mobile phone users is higher. In the second group the number of men exceeds the number of women, and the percentage of mobile phone users is lower."
	AI	"Because women are more cautious drivers."
<i>IT engineer</i>	C	"70% of 100% graduates have that degree."
	MM	"Regardless of the number of employees chances of Milan being an engineer and a lawyer are equal."
	AI	"intuition", "it just makes sense to me"
<i>R/W balls</i>	C	"Numerous attempts are offered, hence the chances are greater."
	MM	"It is more probable to draw one ball where the chances are 50-50, than to draw seven consecutive red or white balls." Or "Because we have 90% of red balls, and because balls are put back into the box after each draw."
	AI	"It cannot be known", "Intuition..."
<i>Lawyer</i>	C	"Percentage of lawyers is much higher than the percentage of musicians who are members of philharmonics."
	MM	"This statement is more probable because he was oriented in this direction in his childhood."
	AI	"I believe that both claims are equally probable, but I will opt for an 'a'."

analyzed and re-assessed, which resulted in unified coding.

Statistical methods

In the analysis of *correctly selected alternatives and correctly answered tasks*, both classical test analysis and the Rasch model were used. Classical analysis, which provides percentages of students who answered in a certain way, was used to enable comparison with other studies that used the same tasks, while the Rasch model was used to estimate students’ test score.

Test parameters for both analysis are given in Table 3. Discrimination indices were calculated as point-biserial correlation of task-scores with the sum of scores for all other tasks, while the task difficulty is a simple proportion of correct answers. For the Rasch model, discrimination is the same for all tasks (1.42) while difficulty coefficients are given in the table.

Thus obtained, students’ test score is further used in analysis of the relationship between students’ test score and the explanations they provided. The relationship between giving an explanation for the answer and the test score is demonstrated in two ways. Firstly, it is visually demonstrated where the mean test scores for the groups of students who responded in one of five possible ways (c, ce, cwe, w, we) for each of six tasks are shown. For testing the test-score difference between the groups (c or w, i.e., answers without explanations vs cwe or we, i.e., answers with some explanation), we used Welch two-sample *t* test. Secondly, we used broader groups of students: those who had chosen correctly the right option (c, ce, cwe) and those who gave any explanation for their choice (ce, cwe, we). Then, we have examined which component of fully correct answer (correct choice or some explanation) contributes more to the test-score. Pearson correlation coefficient between these occurrences and the test-scores are taken as a measure of that contribution.

The reliability of the tests used in this study was evaluated using the Cronbach’s alpha coefficient. The impacts of the year of study, secondary school finished, the

course in probability and statistics, and gender, on the students’ test score obtained through Rasch analysis, were tested using the Welch two-sample *t* test. *P* values of <0.05 were considered statistically significant. Data for both classical and Rasch analysis was performed in statistical software R, version 3.5.1 (using R package ltm (Rizopoulos, 2006)).

Results

Percentage of correctly selected alternatives and correctly answered tasks

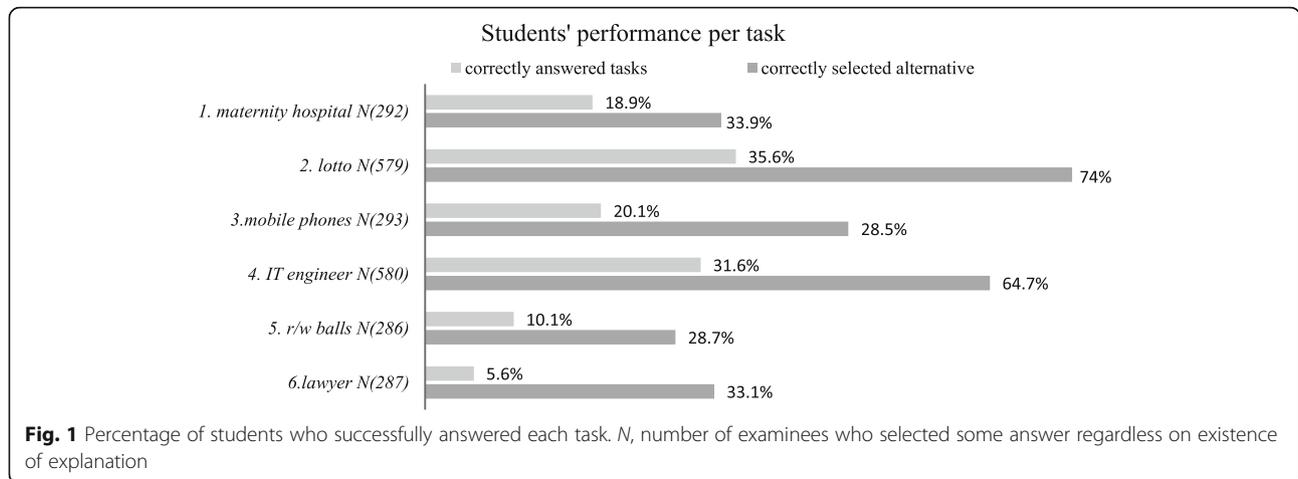
Comparison between the percentage of *correctly selected alternatives* and *correctly answered task* per task is given in Fig. 1. Percentages given in Fig. 1 pertain to individual tasks and are based on the total number of examinees who selected some answer regardless of the existence of explanation. Considering the tasks “Maternity hospital”, “Mobile phones”, “R/W balls,” and “Lawyer,” approximately 30% of students correctly selected alternatives, while this percentage was significantly higher in the case of the “Lotto” (74%), and “IT engineer” (64.7%) tasks. However, regarding the “Lotto” and “IT engineer” tasks, approximately every third student gave a *correct answer*, as opposed to the “Maternity hospital” and “Mobile phones”, where only one in five students *answered correctly*, i.e., managed to select correct alternative while giving correct justification for the answer. The “R/W balls” task yielded even lower percentage of correct answers, around 10%. Our examinees were most troubled in the case of the “Lawyer” task, where only one in twenty students (5.6% of the total number of examinees who selected alternative) managed to select correct alternative and justify their answer correctly. It can be observed that many students who selected the correct alternative in multiple-choice tasks failed to provide an adequate explanation for their selection.

Analysis of answer explanations

From the total number of students who chose an alternative, just over half of them provided some sort of explanation of the selected alternative. The least explained

Table 3 Classical analysis and Rasch model parameters for both tests

Task	Classical analysis				Rasch model
	Task difficulty		Index of discrimination		Task difficulty
	Test 1	Test 2	Test 1	Test 2	Both tests together
Maternity hospital	0.19		0.25		1.39
Lotto	0.40	0.30	0.38	0.36	0.60
Mobile phones	0.20		0.35		1.33
IT engineer	0.27	0.36	0.38	0.31	0.76
R/W balls		0.10		0.20	2.02
Lawyer		0.06		0.26	2.56



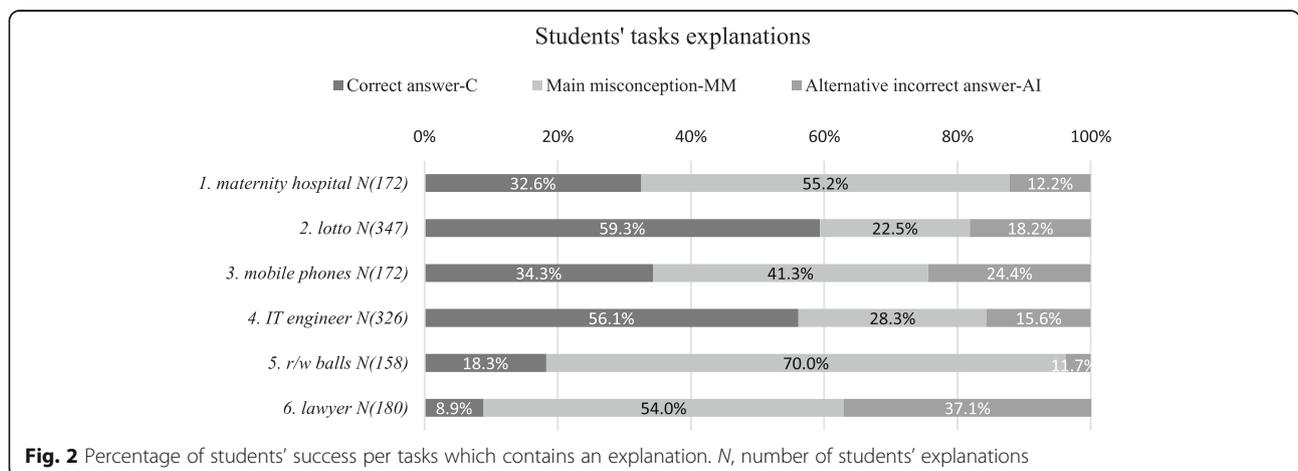
task was “R/W balls,” from the 286 students who selected an alternative, 158 of them provided some sort of explanation. On the other hand, most students’ explanations were from the “Lawyer” task, from the 287 students who selected alternative, 180 provided some sort of explanation. Figure 2 shows the proportion of students’ answers divided into categories C, MM, and AI, and presents the number of students who gave any explanation per task. Percentages shown in Fig. 2 pertain to the individual tasks and the examinees who provided any sort of answer justification.

Based on the presented material in Fig. 2, it can be concluded that the students were most susceptible to main misconceptions in the task “R/W balls,” where about 70% of examinees provided answers which indicate some sort of misconception. “Maternity hospital” and “Lawyer” tasks were better, with somewhat over 50% of present misconceptions. Our examinees were most successful in the “Lotto” and “IT engineer” tasks, where the percentage of answer justifications which indicate misconceptions was below 30%. Finally, the “Lotto” was the task with the smallest percentage of students

who gave in to misconceptions. In the case of the base rate neglected misconception, represented in the “Lawyer” task, a large number of students (37.2%) answered led by their personal preconceptions despite the available data, while the percentage of answers containing personal preconceptions in other tasks amounted to 20%. Although the “Lawyer” task had the most answers, it featured the smallest percentage of students who gave a meaningful explanation. When it comes to correctly selected alternatives and correctly justified answers, the tasks “Lotto” and “IT engineer” yielded over 50% of examinees who answered correctly, in the case of “Maternity hospital,” and “Mobile phones” that number is about 30%, while “R/W balls” and “Lawyer” were below 20%.

Tests reliability

A test consisting of only four multiple-choice tasks cannot have great reliability. However, if we value explanations and give a full credit only to those answers where both the selection of the alternative and the explanation have to be correct, reliability becomes much greater.



When scoring multiple choice answers, Cronbach's alpha for the first version of the test (tasks 2,4,5,6) equals only 0.08 ± 0.08 . However, when the scoring rule requires explanations as well, Cronbach's alpha increases to 0.55 ± 0.04 . In a similar manner, for the second version of the test (tasks 1,2,3,4), Cronbach's alpha increases from 0.29 ± 0.06 to 0.47 ± 0.05 . Although the test is very short, scoring scheme that requires explanation enables us to treat the sum of item-scores as a test-score, thus rendering test-score comparisons between the groups of students meaningful.

The relation between students' answer explanation and estimated students' test score

All six tasks exhibited the same response characteristic: examinees who failed to explain their choice had much lower estimated students' test score compared to those who came up with some explanation, even if the choice was wrong. In Fig. 3, the mean values of the students' test scores are shown for five different groups of answers: c, ce, cwe, w, and we.

Evidently, one group of answers (c and w, i.e., answers without explanation) has consistently lower values of the estimated students' test score than the other group (cwe

and we, i.e., answers with some explanation). Statistical significance difference for each task of these groups (c, w) and (cwe, we) was determined. All differences are highly significant ($p < 1e-9$).

Also worth noting is the fact that the examinees who gave an explanation for one choice were also more likely to provide explanations for the remaining choices. About 60% of examinees provided explanation for either none (23.3%) or all four choices (36.8%).

According to earlier finding (Attali, Laitusis, & Stone, 2016), ability to explain a choice indicates willingness to earnestly think about the task, while the correct choice might be a consequence of pure guessing. The problem here was to estimate how much each of the two parts of an answer contributed to students' test scores estimation. The coefficient of correlation was therefore calculated between the students' test scores, on the one side, and selection of correct choice with or without an explanation (c, ce, cwe), and giving some explanation for the selected choice (ce, cwe, we), on the other (Table 4). It turns out that the ability to provide any explanation, even the wrong explanation for the wrong choice, contributes more to the overall students' test scores estimation than the mere selection of correct answers.

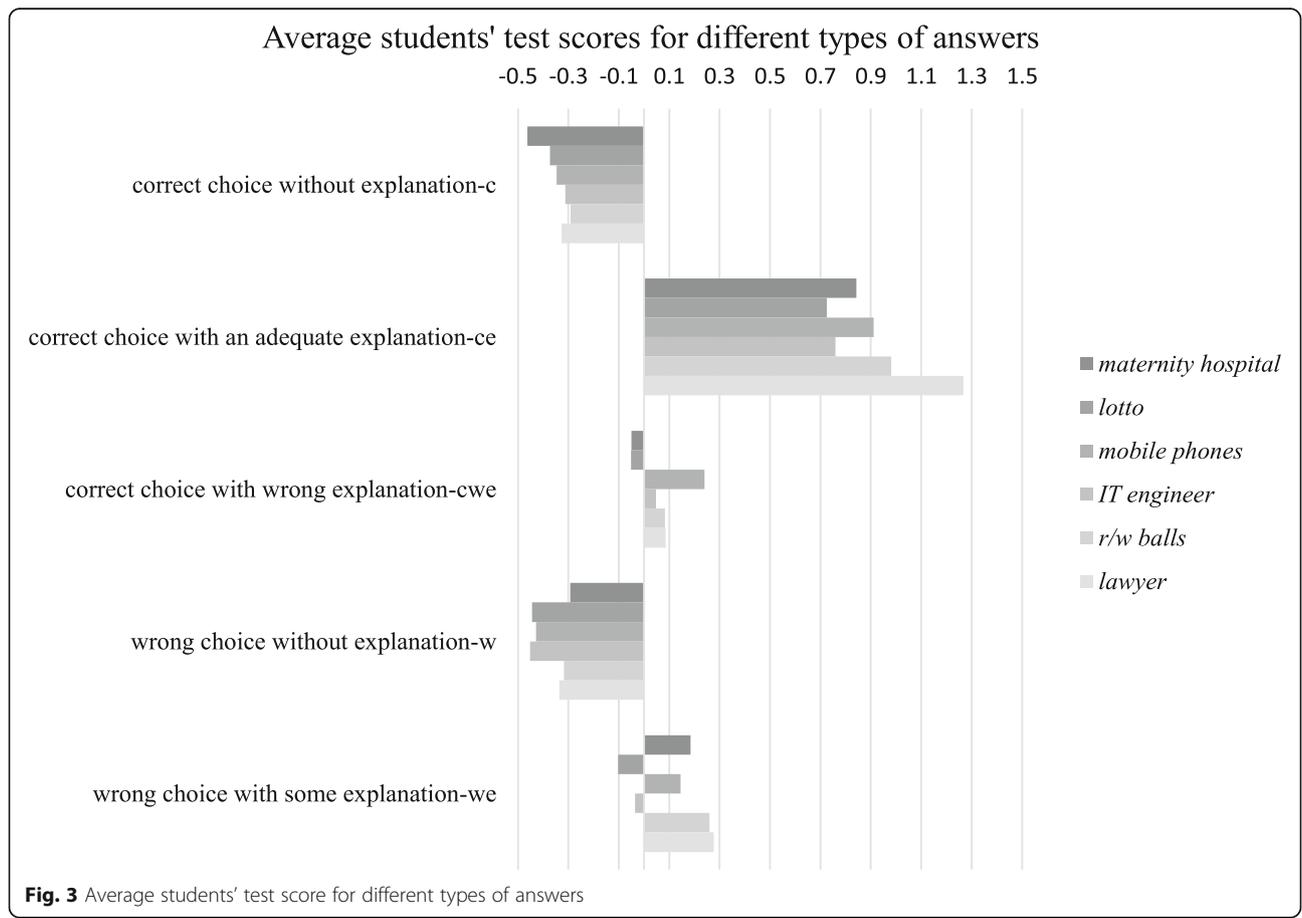


Fig. 3 Average students' test score for different types of answers

Students' test score and factors

Analysis of tests' scores between first and third year students showed a statistically significant difference in favor of first year students ($t = 2.99$, $df = 435.1$, $p < 0.01$). Further, analysis of students' scores revealed no statistically significant difference between first year students' scores based on the type of the secondary school ($t = 1.19$, $df = 229.92$, $p = 0.23$). However, among first year students' scores, there is a statistically significant difference based on gender ($t = 2.61$, $df = 266.41$, $p = 0.01$), where the boys outperformed the girls. Among the third year students, neither gender ($t = 1.72$, $df = 177.17$, $p = 0.09$) nor the completed course in probability and statistics ($t = -1.27$, $df = 115.13$, $p = 0.21$) yielded significant difference.

Discussion

Tasks and misconceptions

According to results of the present study, only about 19% (Fig. 1) of our sample avoided misconception *insensitivity to sample size* (task—"Maternity hospital"). The results of our study bear similarity to early studies, where 18 to 28% of examinees answered correctly in this kind of task (Kang & Park, 2019; Kustos & Zelkowski, 2013; Tversky & Kahneman, 1974). This study established that about 55% (Fig. 2) of students in the sample population who provided an explanation in this task were prone to the main misconception. This implies that they tend to relate the number of boys with a number of days, neglecting the size of the hospitals, giving explanation like this: "In the first maternity hospital more babies are born, therefore there will be more boys, namely, more days." Such answers can be justified by students' ignorance of or insufficient understanding of the Law of large numbers, which indicates the need to improve students' knowledge of the basic concepts of probability and statistics (Kahneman & Tversky, 1972). These findings indicate that even those students who were willing to provide explanations and tackle tasks are highly prone to the misconception of insensitivity to sample size.

In the analysis of the *misconception of chance* (task—"Lotto"), we found that 74% (Fig. 1) of our sample

(regardless of the existence of students' answer explanation) gave a correct answer which is in line with the results of a similar study (Kustos & Zelkowski, 2013). But the analysis of students' explanations shows that 35.6% (Fig. 1) were able to give a meaningful explanation of their answers. In both cases, in the analysis of wrong answers, the majority of students chose to answer with a disordered array of numbers which indicates the existence of the *misconception of chance*. The existence of the *misconception of chance* is additionally corroborated by students' answer explanations, which can be illustrated by the following: "Because numbers come in no particular order." This kind of task was recognized as the easiest task by our students, where about 22% (Fig. 2) of those who provided task explanations were also prone to the main misconception. According to previous findings (Tversky & Kahneman, 1974), this is due to misunderstanding of the random process concept, i.e., the students expect the sequence generated by a random process to feature main characteristics of the random process even when that sequence is short.

In the analysis of students' task explanations in the task related to *illusory correlation* (task—"Mobile phones"), analysis showed that about 41% (Fig. 2) of respondents who gave explanations were prone to illusory correlation. They made conclusions relating to the absolute number of women and the percentage of mobile users, explaining it in this way: "In the first group the number of women exceeds the number of men, while the percentage of mobile phone users is higher. In the second group the number of men exceeds the number of women, and the percentage of mobile phone users is lower." They were prone to overestimating co-occurrence of the given information, which lead them to establish unfounded relation between them. Analysis of students' answer explanation also showed that 24% (Fig. 2) of respondents who gave explanations based them on prejudices, providing explanations in the following way: "Because women are more cautious drivers" or "It is well known that women use mobile phone more than men." Such answers can be attributed to insufficient knowledge

Table 4 Correlation between the correct choice selection and the existence of explanation with the students' test-scores

Task	Correlation between the correct choice selection (c, ce, cwe)	t value	Correlation with the explanation existence (ce, cwe, we)	t value
Maternity hospital	0.25	4.46	0.54	10.9
Lotto	0.29	7.26	0.59	17.9
Mobile phones	0.30	5.49	0.60	13.0
IT engineer	0.32	8.27	0.59	17.6
R/W balls	0.16	2.84	0.53	10.7
Lawyer	0.12	2.07	0.48	9.36

All correlations are statistically significant ($p < 0.05$). Level of significance is exhibited with t-parameter of applied t-test

and ability to critically analyze and interpret data. These results indicate that many students are not quite skilled to critically read, interpret, and discuss data given in the simple form.

In the analysis of “IT engineer” task (*base rate neglected-equiprobability* bias), we found that 31.6% (Fig. 1) of our examinees were able to give the correct answer and meaningful explanation. In the analysis of students’ explanations, we found that about 28% (Fig. 2) of them believe that if there were two outcomes, chances were always equal. Despite being given explicit information about the chances, students neglected that, making their decisions based on beliefs which they expressed through comments like: “There are two options, chances are always 50-50%” or “he can be either one or the other.” One of the reasons for this phenomenon can be found in the teaching practice, where many students are more exposed to events with equally probable outcomes, such as the rolling of a dice or tossing of a coin, which lead students to observe all outcomes as equally probable (Gauvrit & Morsanyi, 2014; Smyrnaoui et al., 2020).

In the analysis of the “Lawyer” task (*base rate neglected*), many respondents chose that it is more probable that a person is a musician who plays clarinet, despite the fact that a number of such musicians is considerably lower than a number of people with law degree. Students’ results in our study are almost identical to those from the study of Kahneman and Tversky, where 5% (students of Oregon) gave the correct answer (Kahneman & Tversky, 1973), and are slightly below the results of the Kang and Park study (Kang & Park, 2019), where 14% of adults (bank employees) gave the correct answer. Analysis of students’ explanations showed that 54% (Fig. 2) of them were prone to the *base rate neglected* misconception. They made their conclusions based only on the given description of the person in the task, neglecting the ratio of musicians and lawyers in the country. Also, in this task, we found that about 32% (Fig. 2) of the students who provided explanations based them on the *equiprobability bias* or their personal opinion about the socioeconomic circumstances. Subsequently, the students tended to explain their choices with comments such as: “I believe that both claims are equally probable” or “He can’t make a living from music, he is probably a lawyer.”

Looking at the “IT engineer” and “Lawyer” tasks, which both pertain to the same misconception, out of the total of 151 students who answered both tasks, 86 correctly answered the “IT engineer”, of which only 12 managed to also correctly answer the “Lawyer” task. It turned out that even those examinees who were aware of the importance of the basic set in the “IT engineer” task—which does not feature description for one of the two employees subsets—when given the description for

a specific employees subset, were prone to neglecting the importance of the basic set. Based on these facts, one can assume that the students find it more difficult to properly apply their knowledge when there is no explicitly given rate within the basic set, as well as when they are given description of a specific subset. This assumption requires additional confirmation, bearing in mind previous studies which have reported how difficult and intricate it is to compare misconception related tasks. In addition, they have also shown the existence of massive differences in the students’ achievements, based on the type of task, even though the tasks are related to the same misconception (Gauvrit & Morsanyi, 2014; Morsanyi et al., 2013).

The task related to *biases in the evaluation of conjunctive and disjunctive events* (task—“R/W balls”) is recognized as one of the difficult tasks. According to results, about 70% (Fig. 2) of students who provided explanations were unable to correctly argument the selected choice and were prone to misconception. About half of them neglected properties of the probability of disjunctive events in favor of equal probabilities (task option a), providing arguments in this way: “Under *a* the chances are 50-50% under *b* and *c* it looks like the chances are higher but they are not.” or “Because the highest probability is 50%”. Overall, only 6 students managed to provide the appropriate calculation and mathematical explanation of this task. Despite the fact that this kind of task and its content are part of many courses of probability and statistics, students’ knowledge and deep insight in the evaluation of the probability of conjunctive and disjunctive events is still insufficient. The reason for that can be found in the students’ inability to connect the probability concept of *conjunctive and disjunctive events*, during learning, with the previous experience gained in real-life situations.

In the analysis of students’ answers and success rates, the highest correct answer rate is recorded in the “Lotto” task. The explanation can be found in the fact that the sample population is familiar with this kind of game, while both the secondary school mathematics curriculum and standard courses in probability and statistics use lotto examples in teaching. Out of the total number of 127 examinees who provided answers for both the “Lotto” and the “R/W balls” tasks, 74 gave correct answer to the “Lotto” task, while only 14 managed to also correctly answer the “R/W balls” task. In the tasks “R/W balls” students were not so successful, despite the fact that this kind of task and its content are also part of the secondary mathematics curriculum and standard university courses in probability and statistics. The reason for that could be students’ real-life experience, i.e., their familiarity with examples from real life, rather than the conventional approaches to independent events. These

results emphasize the importance of real-life experience in the teaching process.

Answer explanation and students' test score

Our analysis included students' answer explanations which had more pronounced psychometric characteristics (Table 4). In such cases, tasks appeared to be much harder, which lead to lower answer rates. However, task parameters could be better estimated, rendering the test more discriminant. Findings of the present study are in line with some previous studies which have shown that testing with only multiple-choice tasks tends to overestimate students' conceptual understanding of probability (Hirsch & O'Donnell, 2001). This study demonstrates that many students—even when they guess the answer or intuitively know the right answer—do not want to or are unable to provide adequate explanation (Fig. 1). Such results indicate either superficial knowledge and deficit of competencies that would enable students to provide correct argumentation, or just the lack of proper motivation (Douglass, Thomson, & Zhao, 2012).

Analysis of students' test scores indicates that answers with explanations give higher measurement precision (Table 4). Also, it is found that students who provided any explanation for given answers had higher test scores than the students who just opted for the right alternative (Fig. 3). Such effects are explained as the property of open-ended tasks which encourages more mindful engagement (Attali et al., 2016). The difference between the two scoring schemas (with and without explanations) could mean that our test also measured examinees' effort to explain the choice. Future research should be focused on the probability and statistics knowledge and the readiness to explain chosen answers. In order to make this possible, numerous multiple-choice tasks must be developed with better psychometric characteristics, which would decrease the guessing noise. We suggest a combination of multiple-choice tasks and open-ended explanations as a solid model for data literacy assessment in general.

Students' test score and factors

The results of this study imply that, regardless of their advantage in the number of math classes and completed topics in probability and statistics, the first year students of engineering faculties with gymnasium background did not have the edge over their counterparts from the vocational secondary schools. In the case of the third year students, the analysis also yielded no significant difference between the test scores in students based on the previous completion of a course in probability and statistics. Moreover, when looking at the test score, our results showed that first year students were more successful than third year students. One of the reasons

for that may be in the teaching practice (Gauvrit & Morsanyi, 2014), where, for example, during learning, students are more exposed to tasks with equally probable outcomes and route thinking which later lead them to wrong conclusions, but for more reliable assertion, further investigations are needed. In general, our results imply that the completion of course in probability and statistics did not significantly influence engineering students' ability to cope with the common misconceptions. Our results are in compliance with the previous studies which imply that additional classes or courses in probability and statistics will not contribute to overcoming the misconceptions related to probability concepts (Chance et al., 2007; Garfield & Ben-Zvi, 2007), unless the very approach to teaching changes (Gauvrit & Morsanyi, 2014; Gigerenzer et al., 2007; Masel et al., 2015; Morsanyi et al., 2013).

In terms of the gender, the first year students exhibited a statistically significant difference in test scores, where the males had higher test scores than females. Our findings are in line with the study of Paul and Hlanganipai (2014) who investigated probability misconceptions in secondary school students and reported that the males outperformed females. Some previous studies have shown that among the STEM students, males do better in mathematics than females (Delaney & Devereux, 2019; Ro & Loya, 2015), differently from our finding regarding third year students. Our finding is in line with the study of Hyde and Mertz (2009), who, through analysis of massive amounts of data, showed that females have reached parity with males in mathematics performance. Bearing in mind the complexity of the gender issue, further discussion requires a wider set of data, i.e., additional investigation.

Conclusion

Implication for instruction

This study shows to what extent electrical engineering students are prone to certain types of probability misconceptions, while at the same time introduces a new line of research that focuses on the cross-sectional interaction between students' test scores and the explanations of selected alternatives in multiple-choice probability tasks. The results of this study indicate that electrical engineering students are susceptible to misconceptions in probability reasoning. Findings of the present study showed that students were most successful in avoiding misconceptions of chance, but there were only 35% of examinees who were able to select the correct option and give meaningful explanation—even if the example was known from both the real life and formal education. In analysis of students' explanations, we found that in many cases the majority of students were prone to the common misconception. Students were misguided by

the context and neglected given data, which lead them to form conclusions biased by previously formed beliefs and prejudices.

The results of this study also indicate that there is no significant difference between the first year students' test scores based on their secondary school background. Similarly, their third year counterparts exhibited no difference in test scores based on the completion of the course in probability and statistics. Such results urge us to try and change the basic approach to teaching probability and statistics in both the secondary schools and at the university level. That would require special focus on the students' previous experience, as well as a broader application of experiments and real-context tasks in teaching, which would eventually lead to a more efficient adoption and understanding of fundamental concepts in probability and statistics.

Presented study also emphasizes the importance of students' answer explanations. Focusing on multiple-choice tasks with explanations, we investigated the presence of probability misconceptions in the sampled groups of electrical engineering students. Examination of their explanations allowed us to determine the quality of their reasoning, as well as to assess the presence of most frequent types of misconceptions. The results confirmed earlier findings that using only multiple-choice tasks tends to over-estimate students' ability to avoid probability misconceptions. Accordingly, the findings of the present study showed that a high percentage of students from our sample population could not justify their own answers even when they selected the correct option (Fig. 1). Such results indicate a need for the development of deeper understanding of probability concepts as well as the need for the development of competencies that enable students to argue their answers. The results also revealed an interesting relationship between students' ability to provide explanations to answers, and their estimated test score. Students being motivated enough to provide any explanation for given answers, scored higher achievements than the students who just opted for the right alternative (Fig. 3).

In line with the given results, it is necessary to enhance students' knowledge in probability reasoning, as well as their competencies which will enable them to provide the right argumentation for their answers. Having in mind the importance of making decisions under uncertainty, ability to argument own solutions, and types of tasks that are used in estimations of students' test scores, we believe that our study brings valuable contributions both for educators dealing with the education of engineers and more broadly, for educational policies. Moreover, we believe that this kind of study per specific field, in our case engineers, can be purposeful for the improvement of teaching strategies in different educational areas.

Abbreviations

C: Correct answer; MM: Main misconception; AI: Alternative incorrect answer; c: Correct choice without explanation; ce: Correct choice with an adequate explanation; cwe: Correct choice with wrong explanation; w: Wrong choice without explanation; we: Wrong choice with some explanation

Acknowledgements

This research was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Grant No. 451-03-68/2020-14/ 200125 and III 47003) and Provincial Secretariat for Higher Education and Scientific Research of Province of Vojvodina, Serbia grand No. 142-451-2705/2017-01.

Authors' contributions

Marija Kaplar and Zorana Lužanin were involved in all steps of this research: creating the instrument, conducting testing, entering data, analyzing data, and writing the paper. Specifically, Marija Kaplar analyzed and interpreted data regarding probability misconceptions. Srđan Verbić analyzed and interpreted data and also wrote sections regarding students' explanations and test scores. All authors read and approved the final manuscript.

Funding

No funding.

Availability of data and materials

The datasets generated and analyzed during the current study are available in the GitHub repository, <https://github.com/marijakaplar/probabilitymisconceptions-stem>

Competing interests

The authors declare no competing interests.

Author details

¹Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, Novi Sad 21000, Serbia. ²Faculty of Sciences, University of Novi Sad, Trg Dositeja Obradovića 3, Novi Sad 21000, Serbia. ³FEFA, Metropolitan University, Bulevar Zorana Đinđića 44, Beograd 11000, Serbia.

Received: 28 August 2020 Accepted: 16 February 2021

Published online: 17 March 2021

References

- Attali, Y., Laitusis, C., & Stone, E. (2016). Differences in reaction to immediate feedback and opportunity to revise answers for multiple-choice and open-ended questions. *Educational and Psychological Measurement*, 76(5), 787–802. <https://doi.org/10.1177/0013164415612548>.
- Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance*, 9(3), 396–406. [https://doi.org/10.1016/0030-5073\(73\)90061-5](https://doi.org/10.1016/0030-5073(73)90061-5).
- Batanero, C., & Borovcnik, M. (2016). *Statistics and probability in high school*. Rotterdam: Sense Publisher. <https://doi.org/10.1007/978-94-6300-624-8>.
- Bazerman, M. H., & Moore, D. A. (1994). *Judgment in managerial decision making*. New York: Wiley.
- Bekkink, M. O., Donders, A. R. T. R., Kooloos, J. G., De Waal, R. M. W., & Ruiter, D. J. (2016). Uncovering students' misconceptions by assessment of their written questions. *BMC Medical Education*, 16(1), 1–7. <https://doi.org/10.1186/s12909-016-0739-5>.
- Bilek, J., Nedoma, J., & Jirásek, M. (2018). Representativeness heuristics: A literature review of its impacts on the quality of decision-making. *Scientific papers of the University of Pardubice. Series D, Faculty of Economics & Administration*, 25(43), 29–38 Retrieved from <https://www.scimagojr.com/journalsearch.php?q=21100390414&tip=sid&clean=0>.
- Blanco, T. G., & Chamberlin, S. A. (2019). Pre-service teacher statistical misconceptions during teacher preparation program. *Mathematics Enthusiast*, 16(1–3), 461–484.
- Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education*, 15, rm4. <https://doi.org/10.1187/cbe.16-04-0148>.
- Bramwell, R., West, H., & Salmon, P. (2006). Health professionals' and service users' interpretation of screening test results: Experimental study. *British Medical Journal*, 333(7562), 284–286. <https://doi.org/10.1136/bmj.38884.663102.AE>.

- Brockner, J., Paruchuri, S., Idson, L. C., & Higgins, E. T. (2002). Regulatory focus and the probability estimates of conjunctive and disjunctive events. *Organizational Behavior and Human Decision Processes*, 87(1), 5–24. <https://doi.org/10.1006/obhd.2000.2938>.
- Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education Journal*, 1(1). Retrieved from <http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art2/>.
- Chaoui, N. A. (2011). *Finding relationships between multiple-choice math tests and their stem-equivalent constructed responses*, CGU Theses & Dissertations, 21. https://scholarship.claremont.edu/cgu_etd/21 (). <https://doi.org/10.5642/cguetd/21>.
- Delaney, J. M., & Devereux, P. J. (2019). Understanding gender differences in STEM: Evidence from college applications. *Economics of Education Review*, 72(June), 219–238. <https://doi.org/10.1016/j.econedurev.2019.06.002>.
- Delmas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28–58.
- Douglass, J. A., Thomson, G., & Zhao, C. M. (2012). The learning outcomes race: The value of self-reported gains in large research universities. *Higher Education*, 64(3), 317–335. <https://doi.org/10.1007/s10734-011-9496-x>.
- Foster, C., Wigner, A., Lande, M., & Jordan, S. S. (2018). Learning from the parallel pathways of Makers to broaden pathways to engineering. *International Journal of STEM education*, 5(1), 1–16. <https://doi.org/10.1186/s40594-017-0098-8>.
- Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education*, 19(1), 44–63. <https://doi.org/10.2307/749110>.
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372–396. <https://doi.org/10.1111/j.1751-5823.2007.00029.x>.
- Gauvrit, N., & Morsanyi, K. (2014). The equiprobability bias from a mathematical and psychological perspective. *Advances in Cognitive Psychology*, 10(4), 119–130. <https://doi.org/10.5709/acp-0163-9>.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, Supplement*, 8(2), 53–96. <https://doi.org/10.1111/j.1539-6053.2008.00033.x>.
- Gürbüz, R., Erdem, E., & Firat, S. (2014). The effect of activity-based teaching on remedying the probability-related misconceptions: A cross-age comparison. *Creative Education*, 05(01), 18–30. <https://doi.org/10.4236/ce.2014.51006>.
- Hirsch, L. S., & O'Donnell, A. M. (2001). Representativeness in statistical reasoning: Identifying and assessing misconceptions. *Journal of Statistics Education*, 9(2). <https://doi.org/10.1080/10691898.2001.11910655>.
- Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences*, 106(22), 8801–8807.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristic and biases: The psychology of intuitive judgment*, (pp. 49–81). New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3), 430–454. [https://doi.org/10.1016/0010-0285\(72\)90016-3](https://doi.org/10.1016/0010-0285(72)90016-3).
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237–251. <https://doi.org/10.1037/h0034747>.
- Kang, M., & Park, M. J. (2019). Employees' judgment and decision making in the banking industry: The perspective of heuristics and biases. *International Journal of Bank Marketing*, 37(1), 382–400. <https://doi.org/10.1108/IJBM-04-2018-0111>.
- Kastner, M., & Stangla, B. (2011). Multiple choice and constructed response tests: Do test format and scoring matter? *Procedia - Social and Behavioral Sciences*, 12, 263–273. <https://doi.org/10.1016/j.sbspro.2011.02.035>.
- Khazanov, L., & Prado, L. (2010). Correcting students' misconceptions about probability in an introductory college statistics course. *ALM International Journal*, 5(1), 23–35.
- Kustos, P., & Zelkowski, J. (2013). Grade-continuum trajectories of four known probabilistic misconceptions: What are students' perceptions of self-efficacy in completing probability tasks? *Journal of Mathematical Behavior*, 32(3), 508–526. <https://doi.org/10.1016/j.jmathb.2013.06.003>.
- Lin, K. Y., Wu, Y. T., Hsu, Y. T., & Williams, P. J. (2021). Effects of infusing the engineering design process into STEM project-based learning to develop preservice technology teachers' engineering design thinking. *International Journal of STEM Education*, 8(1), 1–15. <https://doi.org/10.1186/s40594-020-00258-9>.
- Masel, J., Humphrey, P. T., Blackburn, B., & Levine, J. A. (2015). Evidence-based medicine as a tool for undergraduate probability and statistics education. *CBE Life Sciences Education*, 14(4), 1–10. <https://doi.org/10.1187/cbe.15-04-0079>.
- Morsanyi, K., Handley, S. J., & Serpell, S. (2013). Making heads or tails of probability: An experiment with random generators. *British Journal of Educational Psychology*, 83(3), 379–395. <https://doi.org/10.1111/j.2044-8279.2012.02067.x>.
- Morvan, C., & Jenkins, W. J. (2017). *An analysis of Amos Tversky and Daniel Kahneman's judgment under uncertainty: Heuristics and biases*. CRC Press. <https://doi.org/10.4324/9781912282562>.
- Nabbout-Cheiban, M. (2017). Intuitive thinking and misconceptions of independent events: A case study of US and French pre-service teachers. *International Journal of Research in Undergraduate Mathematics Education*, 3(2), 255–282. <https://doi.org/10.1007/s40753-016-0038-x>.
- Paul, M., & Hlanganipai, N. (2014). The nature of misconceptions and cognitive obstacles faced by secondary school mathematics students in understanding probability: A case study of selected Polokwane secondary schools. *Mediterranean Journal of Social Sciences*, 5(8), 446–455. <https://doi.org/10.5901/mjss.2014.v5n8p446>.
- Rizopoulos, D. (2006). ltm : An R package for latent variable modeling. *Journal Of Statistical Software*, 17(5), 1–25 Retrieved from <http://rsirt.googlecode.com/files/ltmPackage.pdf>.
- Ro, H. K., & Loya, K. I. (2015). The effect of gender and race intersectionality on student learning outcomes in engineering. *Review of Higher Education*, 38(3), 359–396. <https://doi.org/10.1353/rhe.2015.0014>.
- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*, (pp. 465–494). New York: Macmillan.
- Smyrniou, Z., Georgakopoulou, E., & Sotiriou, S. (2020). Promoting a mixed-design model of scientific creativity through digital storytelling—the CCO model for creativity. *International Journal of STEM Education*, 7(1), 1–22. <https://doi.org/10.1186/s40594-020-00223-6>.
- Stankous, N. V. (2016). Constructive response vs. multiple-choice tests in math: American experience and discussion (Review). *European Scientific Journal*, 12(10). <https://doi.org/10.19044/esj.2016.v12n10p%25p>.
- Stehle, S. M., & Peters-Burton, E. E. (2019). Developing student 21st Century skills in selected exemplary inclusive STEM high schools. *International Journal of STEM Education*, 6(1), 39. <https://doi.org/10.1186/s40594-019-0192-1>.
- Triliana, T., & Asih, E. C. M. (2019). Analysis of students' errors in solving probability based on Newman's error analysis. *Journal of Physics: Conference Series*, 1211(1). <https://doi.org/10.1088/1742-6596/1211/1/012061>.
- Tsakiridou, H., & Vavyla, E. (2015). Probability concepts in primary school. *American Journal of Educational Research*, 3(4), 535–540. <https://doi.org/10.12691/education-3-4-21>.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>.
- Watson, J. M., & Kelly, B. A. (2009). Development of student understanding of outcomes involving two or more dice. *International Journal of Science and Mathematics Education*, 7(1), 25–54. <https://doi.org/10.1007/s10763-007-9071-1>.
- Woods, D. D., Dekker, S., Cook, R., Johannesen, L., & Sarter, N. (2010). *Behind human error*, (2nd ed.,). Routledge. <https://doi.org/10.1201/9781315568935>.
- Zilinski, L. D., Nelson, M. S., & Van Epps, A. S. (2014). Developing professional skills in stem students: Data information literacy. *Issues in Science and Technology Librarianship*, 77. <https://doi.org/10.5062/F42V2D2Z>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.