Statistical Literacy of Undergraduate Students Before Taking the Introductory Statistics Course

Erblina Zeqiri[1], Stevo Gjorgiev[1], Irena Stojkovska[1*]

[1]*Institute of Mathematics, Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University of Skopje, N. Macedonia*

Abstract. Statistical literacy is an essential knowledge for all citizens in today's data-driven society. There are numerous definitions and descriptions for statistical literacy, statistical reasoning and statistical thinking. All of them are based on a common principal that the statistical literacy develops starting in schools and it relates mainly to the processes of evaluating, interpreting and communicating data. If the students do not meet these skills on time, the lack of statistical knowledge may interfere with the process of proper decision making further in their lives. A research for statistical literacy of undergraduate mathematics students at Faculty of Natural Sciences and Mathematics in Skopje, before taking the Statistics course is conducted. The research questionnaire is based on Gal (2002) and Watson and Callingham (2003) models for statistical literacy. Besides the statistical literacy components, student gave their attitudes towards statistics and its application in everyday life. The results show that the basic statistical literacy skills like reading graphical charts or tables are met, statistical knowledge needed for understanding sampling, randomness and sense of making statistical conclusions as well as mathematical knowledge are on satisfactory level, but contextual understanding and critical reasoning are below average. The results of the research will be used to improve teaching statistics at university level.

Keywords: Statistical literacy, Undergraduate students, Attitudes and beliefs towards statistics, Graphical representations, Research.

INTRODUCTION

The importance of statistical literacy in today's data-driven society is not questionable. That is why many curriculum frameworks and national and international education initiatives give high priority to educating future adults to become more informed citizens and employees, [1]. For instance, citizens need to understand that headlines in media are determined from a sample of population under study and the conclusions may be subject to sampling error. They should be able to discriminate between credible and incredible information and they should be able to interpret, critically evaluate and adequately react to such messages, [2].

According to Wallman, [3], *statistical literacy is the ability to understand and critically evaluate statistical results that are present in our daily lives – coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions.* Statistical literacy is more than mathematical skills that are used for statistical calculations. To be statistically literate also means to be familiar with basic statistical terms, to recognize and to be able to interpret different representations of data, appreciating the social context in which the data are set and to apply critical thinking to the statistical information that is used in everyday life and workspace.

[*] The research is supported by Ss. Cyril and Methodius University of Skopje [Grant NIP.UKIM.20-21.6].

Statistical literacy models

Gal ([1]) proposed a statistical literacy model that involves both knowledge and certain attitude or dispositional components that operate together. The knowledge component consists of five elements: (a) literacy skills, (b) statistical knowledge, (c) mathematical knowledge, (d) context knowledge and (e) critical skills, and the two dispositional elements are: (f) critical stance and (g) beliefs and attitudes. According to Gal, the components and elements in the model should not be viewed as fixed and separate entities but as a context-dependent, dynamic set of knowledge and dispositions that together produce statistically literate behaviour. The questionnaire for the current study is mainly based on the Gal's model. The items that are used to access each element, are described in the results section.

Apart of Gal's model, Watson and Callingham ([4]) proposed a hierarchical construction of the statistical literacy with six levels of understanding: (1) idiosyncratic, (2) informal, (3) inconsistent, (4) consistent non-critical, (5) critical and (6) critical mathematical. In the first two levels (1) and (2), students are only merely interacting with the language and meanings of statistical terms, at level (3) and (4), students are beginning to engage with the context and uncover the statistics embedded in the context, and in the last two levels (5) and (6), students are able to be critical and challenge claims made in statistical reports and data. The essence of both Gal's and Watson and Callingham's descriptions are very similar. Both emphasize a need for statistical knowledge and skills, the ability to communicate ideas, the centrality of context and the need to be critical. Because of this mapping between these two models, some items proposed in Watson and Callinghams' model are used in the questionnaire for the study, as it will be explained later, similar as it is done in [5].

Teaching statistical literacy at schools and at universities

Despite the widespread emphasis on reform in the learning and teaching of statistics at schools, statistics is still viewed as an emerging and challenging discipline, when compared to other learning areas, [6, 7]. A reason for that could be that statistics at schools focuses on the procedural and computational aspects of statistics rather than developing conceptual understanding, [8]. On the other hand, at the university, statistics is usually taught in isolation without being connected with a more general framework of research methodology and experimental design that relate to real life situations, [9]. The Introductory Statistics Course for undergraduate mathematics students at Faculty of Natural Sciences and Mathematics in Skopje, since 2020, is a part of the obligatory Probability and Statistics Course during the fifth semester. Before 2020, the Statistics Course was obligatory course for the students on applied mathematics studies and an optional course for students on mathematics education and theoretical mathematics studies. The focus in the Introductory Statistics course was always on underlying mathematical aspects of statistical concepts with very low emphasize on practical aspects or computer-based teaching of these concepts. There are optional courses such as Statistical Software and Statistical Modeling with focus on the last two aspects, but these courses are either not in the lists of elective courses of all study programs or they are not often elected.

The current study

Authors' personal beliefs are that making Introductory Statistics Course an obligatory course for all mathematics students is not enough for developing a statistical literate graduate. But, before taking any new actions in teaching statistics to undergraduate mathematics students, we needed better understanding of the current level of students' statistical literacy before taking the Introductory Statistics Course. We try to answer the following research questions:

RQ1. What is the students' current level of statistical literacy, before taking the Introductory Statistics course? What are the possible reasons for the level of statistical literacy?

RQ2. What are the students' beliefs and attitudes towards statistics? Is there a relationship between the attitudes towards statistics and the level of statistical literacy?

RQ3. What actions can be taken in order to possible improve statistical literacy of our students?

The contribution of the paper is in an adoption of Gal ([1]) and Watson ([4]) statistical literacy models, as well as Cimpoeru and Roman ([5]) research, for construction of more complete and more adequate tool for assessing both students' statistical literacy and their attitudes towards statistics. As far as our knowledge, this is first attempt to assess the statistical literacy of students in Macedonia. Beyond descriptive statistics, a statistical analysis is done, in order to find possible reasons for the level of statistical literacy, a correlation analysis between the level of statistical literacy and attitudes towards statistics is made and possible directions for improvements of the level of statistical literacy are discussed.

METHODS AND METHODOLOGY

The study was conducted among university mathematics students at Faculty of Natural Sciences and Mathematics in Skopje, Macedonia, before taking the Introductory Statistics Course in the third academic year. A questionnaire for assessing both students' statistical literacy and their attitudes towards statistics was used. One day before answering the questionnarie, all mathematics students from the first and second academic years were invited to participate in the research, not knowing the subject of the research. Their participation was voluntary and anonymous.

The questionnaire design

The questionnaire was designed to answer the research questions mentioned in the introductory section of the paper. First part of the questionnaire contains general information such as gender and previous education. In the second part there are 10 questions that access students' attitudes towards statistics as well as 4 self-assesment questions towards computers, data analysis, mathematics and critical thinking. This part was designed to access the level of dispositional aspects of the statistical literacy model. A similar assessment of this component was performed by Cimpoeru and Roman ([5]). After answering the first two parts, students start with answering the third part of the questionnaire. The third part consists of 30 questions that access all five elements of students' statistical literacy knowledge component, according to Gal, [1]: literacy skills, statistical knowledge, mathematical knowledge, context knowledge and critical skills. All questions in this part refer to data collections presented with a pie chart, a bar chart and a contingency table. Literacy skills, statistical knowledge and mathematical knowledge questions are closed answer Yes / No / Don't Know questions, similar to questions in [10], while context knowledge and critical skills questions are open ended questions, with similar rating scale as in [4].

Sample description

The data set consists of total 25 questionnaires, from which 13 questionnaires, filled in by the first year mathematics students (2 boys and 11 girls), and 12 questionnaires, filled in by the second year mathematics students (all 12 girls), during April 2022, before taking the Introductory Statistics Course in the third academic year as a part of the Probability and Statistics Course. The respondents are less than the half of the targeted audience. The most of

them (21 of 25) have math profile high school and the rest (4 of 25) have social science profile high school or other.

Because of a very high majority of girl students (95%) and students with math profile high school (84%) within a small sample, we could not make comparisons of levels of statistical literacy between groups based on gender or previous education.

Statistical analysis

Numerical data collected from questionnaires in a form of scores were statistically analyzed. The Shapiro-Wilk test of normality for the statistical literacy score was applied. The Pearson's correlation significant test was used for determining a correlation between attitudes towards statistics and the level of statistical literacy. For the statistical analysis, the R software was used.

RESULTS

The questionaries' results are analysed considering the research questions RQ1-RQ3 stated in the introduction section.

Assessing students' statistical literacy knowledge component

The obtained results from processed collected data were analyzed for all elements of the statistical literacy components, regarding the scores of different questions groups included in each of them.

For the first element of knowledge component, literacy skill, there were seven questions in the questionnaire, divided in five groups, such that, one question for reading graphical display or table, two questions for reading data from graphical display or table, also two questions for reading and comparing data from graphical display or table, one question for reading and identifying data from graphical display or table and one more question for reading and using data from graphical display or table. Non-response rate for this set of questions was near zero (0.6%). The highest rate of correct answers is for the group of questions which measure the skills of reading and using data from graphical display or table (100%). On the other hand, the lowest rate of correct answers is observed in the group of questions related to reading and identifying data from graphical display or table (92%), see Fig. 1.

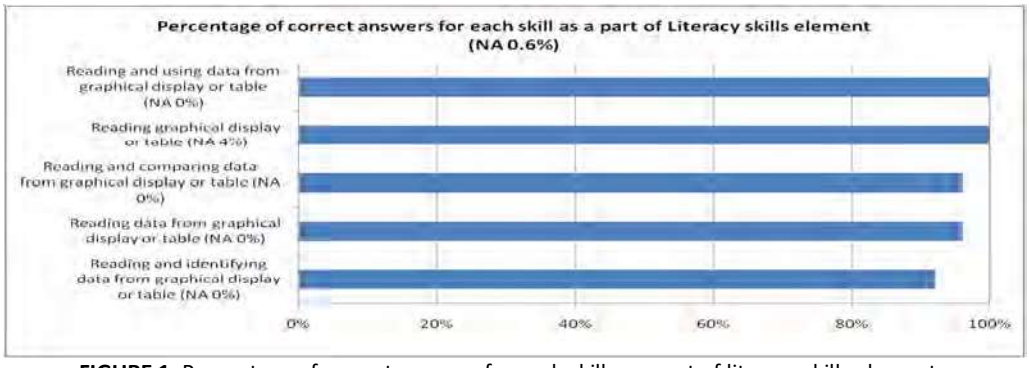

FIGURE 1. Percentage of correct answers for each skill as a part of literacy skills element

The points of the questions corresponding to these five groups, which are adjusted to the same point scale, give an overall accuracy score of 97% for the first knowledge element, literacy skill. Based on the obtained results, we can conclude that:

- Students have excellent skills for reading, using, comparing and identifying data from graphical display or table.
- Students are able to read and identify the data from pie chart, bar chart and table, and also, they know how to use the data and how to compare them.

The second element, statistical knowledge, consists of questions divided in five groups which include the concepts of: probability, sampling and inference, sampling and representativeness, mode and distribution. There was an overall accuracy score of 62% in this element and a nonresponse rate of 34.3%. The question Q5 based on the mode concept was answered correctly from 4 out of 5 students and the nonresponse rate was 80%. The lowest rate of correct answers was on Q24, with nonresponse rate of 48% and answered correctly only from one out of 13 students (8%), see Fig. 2.

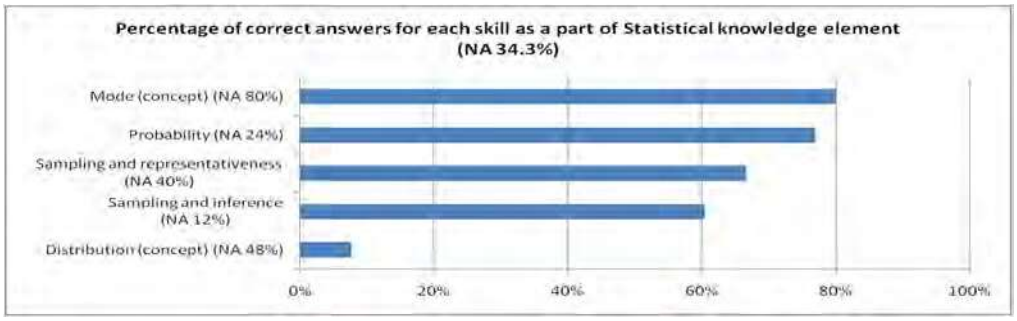

FIGURE 2. Percentage of correct answers for each skill as a part of statistical knowledge element

According to the obtained results it can be conclude that:

- Students have an average knowledge in probability.
- Also, the concept of sampling and representativeness and the concept of sampling and inference are acquired on average level.
- Students are very unfamiliar with the concepts of mode and distribution.
- High overall nonresponse rate, especially in the concept of mode.
- Generally, students have difficulty with the concepts covered with this element.

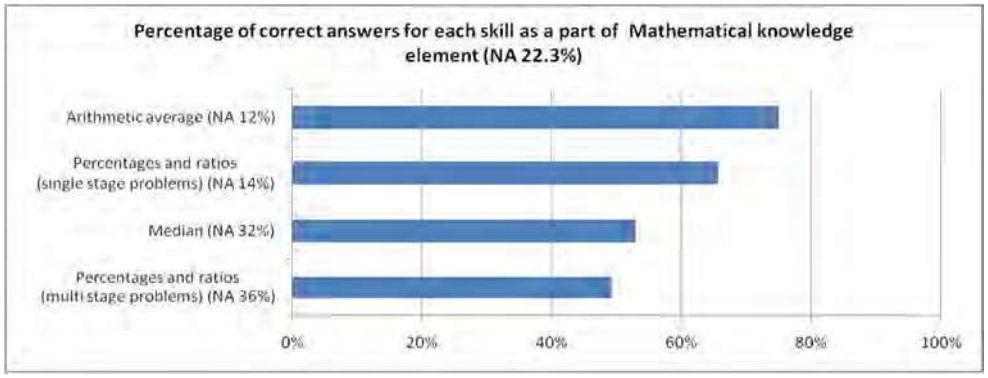

FIGURE 3. Percentage of correct answers for each skill as a part of mathematical knowledge element

The third element of the knowledge component, mathematical knowledge element was measured by the skills in the following four concepts: arithmetic average, percentages and ratios (single and multi-stage problems) and median. The correct answers rates on the questions in this element vary in the border of the interval [49%,75%]. The nonresponse rate for this set of questions is 22.3%. More precisely, the questions that develop the concept of arithmetic mean

are with a nonresponse rate of 12%, percentages and ratios (one-stage problems) at a nonresponse rate of 14%, median at a nonresponse rate of 32% and percentages and ratios (two-stage problems) at a nonresponse rate of 36 %. Our attention is especially drawn to the relatively high nonresponse rates to questions related to the concept of median and percentages and ratios (two-stage problems), as they are part of the curriculum in primary and secondary education. When asked about the median, Q26 was answered correctly by 9 out of 17 students (53%), see Fig. 3.

Some conclusions that can be derived are:

- Students have a good knowledge about arithmetic average.
- Single stage problems for percentages and ratios were with an average knowledge.
- The concepts of median and multi-stage problems for percentages and ratios were with high nonresponse rate and students were either not familiar with solving problems that include these concepts or they were not willing to solve multi stage percentages problems with increasing amounts of calculations.

The last two elements of statistical literacy component are context knowledge and critical skills. These two elements measure the ability to give an interpretation of the given data, to comment the given data and to see the student's ability for critical thinking. The overall score for these two elements is 35% and 23% respectively. Giving an interpretation of data from given graphical or table display has the highest score (60%) with nonresponsive rate of 4% and giving critical comments for given data has the lowest score (19%) with nonresponsive rate of 16%, see Fig. 4.

Based on the answers given in the questionnaires, the following conclusions for the possible reasons for the weakness in these skills, can be made:

- Students made conclusions that refer to a color that corresponds to another category of data instead of the one listed (in graphical displays two similar colors were used).
- The given sentence for interpretation was treated as a true statement, so instead of interpreting the sentence, conclusions and directions for improving things were stated, one's own views were expressed for a completely different area, different from the stated one.
- No distinction was made between two values expressed in monetary values e.g. the price of a good and the cost of raw material for production.
- Difficulties with percentages e.g. indicating a certain percentage amount for a certain category, without stating the total amount on the basis of which it was obtained.

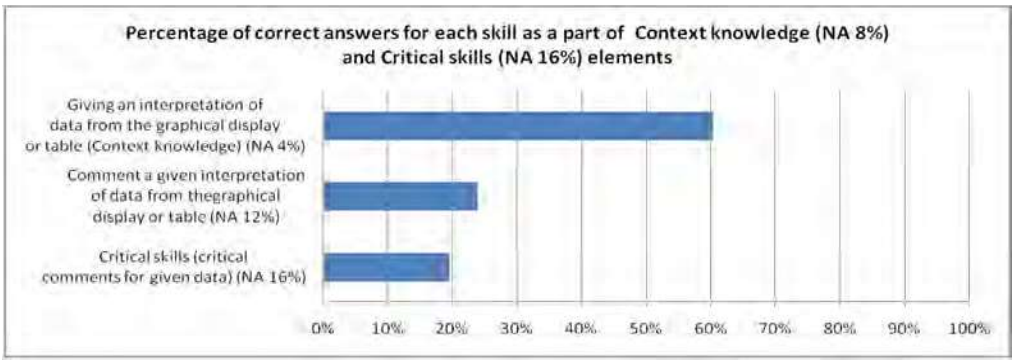

FIGURE 4. Percentages of correct answers for each skill as a part of context knowledge and critical skills elements

The overall statistical literacy knowledge component scores are summarized in Table 1. We can see that the first element - literacy skills has the highest score with 97% of overall performance rate. The lowest score is attained in the critical skills, with just 23% of overall performance rate.

TABLE 1. Overall students' scores in statistical literacy knowledge component

Elements	Average score	Overall performance rate (%)	Overall median performance rate (%)	Nonresponse rate (%)
Statistical Literacy Knowledge Component	**36.30 points/ out of 60**	**61%**	**61%**	**17%**
1-Literacy Skills	13.52 points / out of 14 points	**97%**	100%	1%
2-Statistical knowledge	8.61 points / out of 14 points	**62%**	67%	34%
3-Mathematical knowledge	8.88 points / out of 14 points	**63%**	67%	22%
4-Context knowledge	3.19 points / out of 9 points	**35%**	28%	8%
5-Critical skills	2.10 points / out of 9 points	**23%**	22%	16%

The statistical literacy knowledge component scores are normally distributed as it is shown by the Shapiro-Wilk's test for normality (W = 0.97931, p-value = 0.8711 > 0.05).

The histogram below gives a better view about the distribution of students' statistical literacy scores. From the histogram we can see that approximately 76% of the students have a score higher than 30 points which is 50% of the maximal number of points in the statistical literacy knowledge component test.

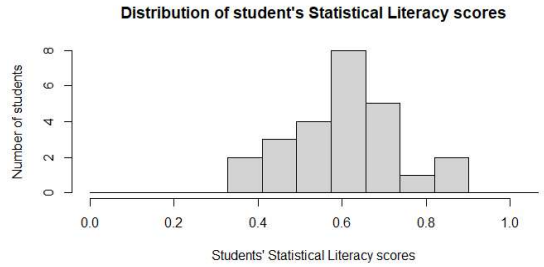

FIGURE 5. Distribution of students' statistical literacy knowledge component scores

Attitudes and beliefs towards statistics

The attitudes and beliefs part in the questionnaire is formed from 8 questions on a Likert scale, with possible answers from 1 ("strongly disagree") to 5 ("strongly agree"). Before averaging the scores, for some questions the responses were reversed if the statement is negatively stated. Higher scores mean a more positive attitude (the maximum being 5), while lower scores mean more negative attitude or belief for statistics (the minimum is 1).

The following statements are main conclusions for students' attitudes and beliefs towards statistics:

- Students on average have a positive attitude towards statistics.

- Students have the highest score (4.56 out of 5) in the question which says “I can learn statistics” so they have a positive thinking for learning statistics.
- Also, they have a very good score (4.44 out of 5, after reversing the points) in the question “Statistics is not useful for my future profession-mathematician”, that means they strongly agree that statistics is useful.

The attitudes and beliefs scores appear to be normally distributed, according to Shapiro-Wilk’s test for normality (W = 0.94446, p-value = 0.1875 > 0.05), see Fig. 6.

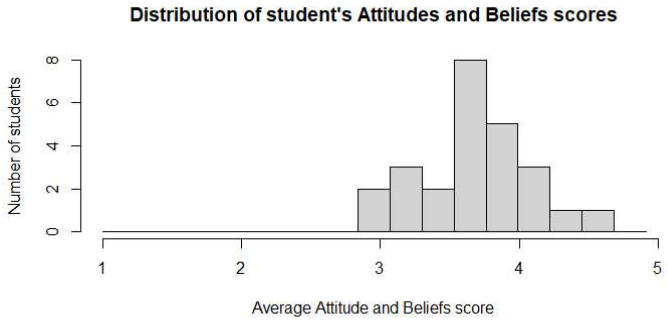

FIGURE 6. Distribution of student’s attitudes and Beliefs scores

Correlation analysis

Because of the homogeneity of the sample we couldn’t make correlation analysis between groups according to gender or previous education. The correlation between the four elements of self-assessment i.e. self-assessment in computers, data analysis, mathematics and critical thinking and the five elements of statistical literacy knowledge component is made (Fig. 7). All of the components resulted following a normal distribution tested with Shapiro Wilk’s test for normality, as previously shown, which allowed us to use Pearson’s correlation test using a correlation matrix (Fig. 7).

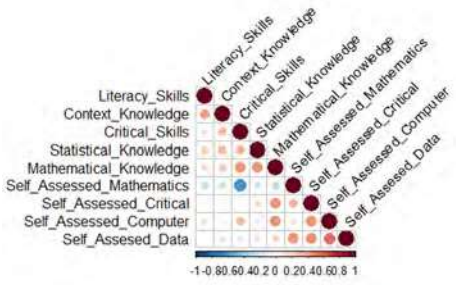

FIGURE 7. Correlation matrix for self-assessment and statistical literacy elements

From the results obtained from the first correlation tests, we can see that there is a statistically significant correlation between most of the four self-assessment elements, the most corelated pair is self-assessment in computers and self-assessment in data analysis (r=0.54, p-value=0.0057 < 0.05). The rest of self-assessment significant correlations have positive Pearson r coefficient less then, but near to 0.5. A statistically significant negative correlation is attained between self-assessment mathematics and critical stance (r=-0.59, p-value=0.0021 < 0.05). One maybe not unexpected finding is that students with a lower score about self-assessment tend to have higher scores on overall statistical literacy skills. Another result is that there is a

statistically significant, but not high correlation between mathematical knowledge element and critical stance element (r=0.43, p-value=0.033 < 0.05), also between mathematical knowledge element and statistical knowledge element (r=0.43, p-value=0.034 < 0.05).

CONCLUSION AND DISSCUSION

Ability to understand and critically evaluate statistical results that are present in our daily lives is an essential knowledge for all citizens in today's data-driven society. Statistical literacy is a complex construct that should be considered with all its elements. Teaching statistical literacy at schools and at universities should not be replaced with teaching how calculations are done, without interpretation and critical discussions.

A research among the undergraduate university students before taking the introductory statistics course is conducted, and the following main conclusions have been made:

- The level of statistical literacy score of the undergraduate students before taking the statistics course is 61%.
- Students have great literacy skills (97%), above medium statistical knowledge (62%) and mathematical knowledge (63%) and low context knowledge (35%) and critical skills (23%).
- There is a statistically significant, but not high correlation between mathematical knowledge and statistical knowledge, such as between mathematical knowledge and critical skills, which means that mathematical knowledge is important, but for overall statistical literacy skill, statistical knowledge and critical skills are essential.
- One way to improve statistical literacy among university students in the domain of context knowledge and critical skills is working on projects using real data.

REFERENCES

1. Gal I., Adults' Statistical Literacy: Meanings, Components, Responsibilities, *International Statistical Review,* **70**, (2002), pp. 1-52.
2. Sharma S., Definitions and models of statistical literacy: a literature review, *Open Review of Educational Research,* **4** (1), (2017), pp. 118-133.
3. Wallman K. K., Enhancing statistical literacy: Enriching our society, *Jornal of teh American Statistical Association,* **88** (421, (1993), pp. 1-8.
4. Watson J., Callingham R., Statistical Literacy: A Complex Hierarchical Construct, *Statistics Education Research Journal*, **2** (2), (2003), pp. 3-46.
5. Cimpoeru S., Roman M., Statistical Literacy and Attitudes Towards Statistics of Romanian Undergraduate Students, *Journal of Social and Economic Statistics* , **7** (1), (31 August 2018), pp. 1-18.
6. Garfiels J., Ben-Zvi D., How Students Learn Statistics Revisited: Acurrent Review of Research on Teaching and Learning Statsitics, *International Statistical Review,* **75** (3), (2007), pp. 372-396.
7. Garfield J., Ben-Zvi D., Helping students develop statistical reasoning: Implementing a statsitical reasoning learning environment, *Teaching Statistics,* **31** (30), (2009), pp. 72-77.
8. Shaughessy J. M., Research on statistics learning and reasoning, In Lester Jr. F. K. (Ed.), *Second handbook of research on mathematics teaching and learning*, (2007), pp. 957-1009.
9. Nikiforidou Z., Lekka A., Pange J., Statistical literacy at university level: the current trends, *Procedia Social and Behavioral Sciences*, **9**, (2010), pp. 795-799.
10. Shield M., Statistical literacy survey analysis: Reading graphs, tables of rates and percentages, *ICOTS-7 2006,* (2006), pp. 1-6.