Statistical literacy of Slovak lower secondary school students

Mária Jurečková, Lucia Csachová
Catholic University in Ružomberok, Faculty of Education, Slovakia
maria.jureckova@gmail.com, lucia.csachova@gmail.com

Abstract. The article is focused on monitoring the level of statistical literacy of 9th grade students at lower secondary schools in the Slovak Republic. The results of several years testing of students were used for this purpose. During the reporting period, more than 50 000 pupils from all over Slovakia participated in the testing. In addition to quantitative data analysis, attention is also paid to qualitative analysis of the problems in the field of statistics that were difficult for students. The article presents a part of the results of the research focused on the analysis of critical points in school mathematics.

Keywords. Education, statistical analysis, statistical literacy.

1. Introduction
Nowadays, the characteristic features of society undoubtedly include its dynamic orientation towards the information society. Information in the form of tables, graphs and charts is common not only in news, politics and advertising, but also in many other areas of society (Frost, 2013). Given that many claims are used to misrepresent or obfuscate real data, there is a need to build people's ability to think critically about data credibility. The importance of statistics in everyday life leads to increased attention to statistical literacy.

Being statistically literate means being able to use statistical knowledge in a number of different situations in different ways. However, this requires a certain amount of mathematical and statistical knowledge and skills. This is not only about knowing the terminology, procedures and skills in performing certain operations and implementing certain procedures. Statistical literacy is the ability to creatively combine these elements in response to external demands. Statistical literacy, therefore, is not only an understanding as a purposeful use in the narrower sense, but as a key ability in establishing, formulating, interpreting and solving problems in various situations (Vrábelová, 2013).

Statistics educators, statisticians and researchers around the world have not reached a consensus on the definition of statistical literacy, and so there are several different definitions (Sharma, 2017). The statistical literacy construct from Watson (1997) consists of three levels: first, a basic understanding of the basic terminology and statistical probability; the second one is an understanding of statistical language and concepts in context, and at the end there is a level with a questioning attitude.

According to Gal (2002), a model of statistical literacy is defined as the ability to interpret, critically evaluate, and communicate about statistical information and messages. Gal proposes a model of statistical literacy which includes knowledge and a certain disposition or access
components that work together. The result is a diagram (Figure 1) describing statistical literacy.

| Knowledge elements | Dispositional elements |
|--------------------|------------------------|
| Literacy skills    | Beliefs and attitudes  |
| Statistical knowledge | Critical stance       |
| Mathematical knowledge |                 |
| Context knowledge  |                        |
| Critical questions |                        |

![Diagram](image)

**Figure 1.** The model of statistical literacy (Gal, 2002)

The components and elements in this model do not represent “fixed and separate entities but as a context-dependent, dynamic set of knowledge and disposition that together enable statistically literate behaviour” (Gal, 2004).

Statistical literacy and the teaching of statistics are relatively new research areas in the world. Garfield and Ben-Zvi (2009) say that the teaching of statistics and statistical training in comparison with other branches of education is still perceived as an emerging and challenging discipline. At schools, the emphasis is still put on the development of skills, which means that many of students are not able to think and reflect statistically, which is a challenge for a new approach to the development of statistical thinking (Moore, 1997).

Also, according to Shaughnessy (2007), statistical education in schools focuses more on procedural and computational aspects than on developing conceptual understanding. At present, the field of probability and statistics is also represented in tests of international measurements of students' knowledge (PISA).

Although statistical literacy is already a stable part of the Slovak curriculum, inadequate results of our students in PISA testing call for an increased attention to statistical literacy. The main monitored area of the PISA cycle in years 2003 and 2012 was mathematical literacy, which was tested in four areas: quantity, space and shape, change, relationships and dependence, and randomness. In PISA 2003, 18 test items of total 85 mathematical problems were from statistics. From four mentioned tested areas, success of Slovak students in the area of randomness was the worst. More than a quarter of students did not reach level 2 in this area (Kubáček, 2004).

This was one of the reasons why we have focused on a more detailed analysis of the level of statistical literacy of lower secondary schools graduates in Slovakia. The purpose of this article is to present some results of this research.

### 2. Statistics in the Slovak school curriculum

Pupils get their first contacts with statistics during primary education by sorting various objects or numbers based on one or more characteristics. They learn to read a simple table, identify a row, a column and add data to it. Innovated national educational programme for lower

---

1. [http://www.oecd.org/pisa/test/](http://www.oecd.org/pisa/test/)
secondary education (hereinafter referred to as ISCED 2), which has been in Slovak Republic in force since September 2015, includes statistics mainly in the fifth, eighth and ninth grades as part of the educational process. Therefore, after completing lower secondary education, students should be able to perform within the performance standard:

- Carry out adequate statistical survey;
- Prepare and process simple project focused on statistical survey of certain event, expressing the frequency of certain phenomenon, event;
- Solve appropriate statistical tasks using the calculation of arithmetic mean;
- Process obtained values – data from own statistical survey into table;
- Interpret data in the table;
- Present values – data using several types of diagrams – graphs.

In the fifth grade, primary education skills are complemented by representing the data in graphs as well as choosing a strategy for solving tasks from everyday life. ISCED 2 defines a separate unit of probability and statistics only in the eighth grade, where procedures applicable to statistics are explicitly described. Pupils become familiar with statistical terminology and work more intensively with the interpretation of data recorded in tables or graphs. They are able to use information in tables and represent it in a column or circle diagram. They should also be able to calculate the arithmetic mean based on the data in the table. However, the school makes little use of the wide range of application of statistics in various scientific fields, it looks as if statistics belong only to mathematics.

In Slovakia, the quality of education and training is monitored and evaluated at national level by the National Institute of Certified Measurements (NICEM). The specification of competencies related to students’ statistical literacy is based on the standard of competencies that students should acquire in lower secondary education. Inspired by the distribution of students' mathematical literacy according to the OECD PISA 2003 study, NICEM lists three levels of statistical literacy competences (Vrábelová, 2013):

- **Basic level** - the reproduction level represents the ability to reproduce the learned material, perform simpler calculations, select various information from the display.
- **Intermediate level** - the level of connection allows solving tasks that are not entirely routine, but contain known elements. It is characterized by an undemanding extension to the student's familiar material and reasoning.
- **High level** - the competence that contains elements of thinking about the processes needed to solve a task. They are characterized by the need to grasp the problem, developed critical thinking, argumentation, abstraction and generalization.

To verify the level of expected skills related to the statistical literacy of lower secondary school graduates, we analyzed the results of tests provided by the company Exam testing. One of the projects of this organization is also output testing of mathematics at the end of lower secondary school. Students take part in testing on a voluntary basis, to obtain information about the level of their knowledge of mathematics.

---

2 Innovated National Educational Programme for Lower Secondary Education, Mathematics, p. 38.
3 Innovated National Educational Programme for Lower Secondary Education, Mathematics, p. 30.
4 https://www.nucem.sk/sk/merania/narodne-merania/testovanie-9
5 Exam.sk
3. Quantitative analysis of the level of statistical literacy

Participants in our research were 9th grade students who signed up for the final testing of mathematics in the years 2016 - 2018. The test exams provided us with an anonymized database of these test results. The database contains information on gender, type of school and the results of solving individual tasks of tested students. All tests created by a team of experts are tested in advance. Before using them, they go through several external reviews, which are performed by experienced teachers from practice. We used IBM Statistics 22 for statistical data analysis. Each test consisted of 20 tasks aimed at verifying skills of students in applying basic mathematical principles and processes both in mathematical and real life contexts. In each monitored year, tests included also tasks corresponding to the performance standard of lower secondary school graduate in the field of statistics (there were three statistical problems in each of the monitored years). During the monitored period, 53 551 students participated in the testing within the entire Slovakia (respondents were from public, private and church schools). The selected sample represents 46.3% of the total population of students.

The following Table 1 shows the numbers of respondents in individual years.

| Year | Count | Total |
|------|-------|-------|
|      | 2016  | 2017  | 2018 |
| Boys | 9615  | 8937  | 8805 |
| Girls| 9303  | 8653  | 8238 |
| Total| 18918 | 17590 | 17043 |

In this article, we want to present the answers to two research questions from the realized quantitative research.

Research questions:

1. Is the success of students in statistical problems comparable to the success in other mathematical problems?
2. Is there a statistically significant difference in the success of boys and girls in statistical problems?

The following Chart 1 shows the average success rate of respondents in statistical and non-statistical problems in tests in individual years. Based on Chart 1, it is clear that every year the tested students achieved worse results in the statistical problems than in the other problems.
Chart 1. Average success rate of respondents in statistical and non-statistical problems in individual years

In order to be able to answer our first research question correctly, it is necessary to verify the statistical significance of this difference. We used the Paired-Samples T Test for a sample of respondents from the whole monitored period. The following two tables are outputs of Paired-Samples T Test realized in SPSS software package.

| Paired Samples Statistics |          |          | Std. Deviation | Std. Error Mean |
|---------------------------|----------|----------|----------------|-----------------|
| Pair 1                    | stat_success | 44,388   | 53551          | 34,086          | .147            |
|                           | nonstat_success | 58,792   | 53551          | 23,956          | .103            |

| Paired Samples Test       |          |          |                | 95% Confidence Interval of the Difference |          |          |          |
|---------------------------|----------|----------|----------------|------------------------------------------|----------|----------|----------|
|                            | Mean     | Std. Deviation | Std. Error Mean | t            | df        | Sig. (2-tailed) |
| Pair 1                    | stat_success - nonstat_success | 14,403   | 29,333         | .126         | -113,629  | 53550   | .000     |

The first part of Table 2 shows the results of basic descriptive statistics of respondents’ success in statistical and non-statistical problems. If we compare the average success achieved
in statistical and non-statistical problems from the whole monitored period, we see that the success in statistical tasks is lowered by up to 14.4%.

The second part contains the characteristics of the test itself. Based on the P-value (Sig. (2-tailed)), we reject the null hypothesis of comparable success and accept the alternative hypothesis that the difference in student success in statistical and non-statistical problems is statistically significant at each significance level. The results of tests confirmed that the level of lower secondary school graduates in the field of statistics is lagging behind the level in other fields of mathematics and statistical literacy requirements, in addition to mathematical literacy, other knowledge and dispositional elements. When looking for an answer to the second research question about the intersexual difference in success in solving problems in statistics, we start with a boxplot analysis. Chart 2 shows boxplots representing a graphical representation of monitored statistical data based on minimum, first quartile, median, third quartile, and maximum. We used the results of boys and girls from the entire monitored period. Boxplots indicate that the distribution of the monitored item is different for boys and girls. A significant difference can be seen in median values (boys 50, girls 33.3). Based on the graphs, we can identify that in the statistical problems, half of the boys had a success rate of over 50% and half of the girls did not exceed the success rate of 33.3%.

**Chart 2. Boxplots of success rates of boys and girls in statistical problems**

Table 3 provides the results of additional descriptive statistics for the distribution of monitored values that support the hypothesis that boys perform better in statistical problems than girls.

**Table 3. Descriptive statistics of the success rate of boys and girls in statistical problems**

|       | N     | Mean  | Std. Deviation | Std. Error | 95% Conf. Interval for Mean | Minimum | Maximum |
|-------|-------|-------|----------------|------------|---------------------------|---------|---------|
| Boys  | 27357 | 46.62 | 34.53          | .21        | 46.21 47.03               | 0.00    | 100.00  |
| Girls | 26194 | 42.06 | 33.46          | .21        | 41.66 42.47               | 0.00    | 100.00  |
Because boys and girls form independent samples, we used the Independent Samples Test to compare mean values. Table 4 is the output of the implemented test in SPSS. Even in this case, due to P-value close to zero (Sig. (2-tailed)), we reject the null hypothesis that the success of boys and girls in statistical problems is comparable and we can state that the difference in success of boys and girls is statistically significant.

Table 4. Outputs of Independent Samples T Test

| Levene's Test for Equality of Variances | t-test for Equality of Means | 95% Confidence Interval of the Difference |
|---------------------------------------|-----------------------------|----------------------------------------|
| stat_success                          |                            |                                        |
| Equal variances assumed               | 28.74,000                  | 3,978, 5,131                           |
| Equal variances not assumed           | 15.49,000                  | 3,979, 5,131                           |

We obtained similar results in the area of intersexual differences in the success of pupils in school statistics also in the research carried out in 2011-2016. In published analyses (Csachová et al., 2017, Gunčaga & Jurečková, 2017), we used the results of the nation-wide tests T9 carried out by the NICEM.

After performing the basic quantitative analysis of students' achievements in statistical problems, we focused on the qualitative analysis of problems in which students were not successful.

4. Qualitative analysis of selected problems from school statistics

In this part we would like to analyse selected problems from the mentioned Exam testing and Slovak nation-wide testing T9. We can say that both of these tests try to capture the most phenomena in mathematical literacy of lower secondary school students through their processing and selection of problems. Part of the test items related to statistics may indicate the level of statistical literacy of individual students.

In the analysis we rely on the mentioned model of statistical literacy according to Gal shown in Fig. 1. We will try to describe individual elements of the model for two selected problems. Since they are tests of students at the end of a lower secondary school, we use all the elements except Beliefs and Attitudes. We understand these elements as a part of statistical literacy in a
mature individual, which is not yet possible in the case of the 15-year-old student in terms of statistical literacy.

A common characteristic of the problems of statistics at the lower secondary school is that they are formulated as word problems, overlapping with life experience (i.e. with the context of real life). Their assignments are often a combination of text and tables or charts of different types. Therefore, literacy skills are necessary to solve these problems, as understanding is based on reading data about objects and the relationship between them with understanding.

We involved 9 students (5 girls and 4 boys) in the qualitative research. Each student received a worksheet with five statistical problems selected from Exam testing and NICEM T9 testing. For our analysis, we selected the problems that had in previous years the lowest success in the above mentioned tests. Although some problems contain four options, from which the student chooses one correct answer, we decided that it would be more advantageous to remove these options to achieve the goal. We were interested in whether they would choose other options than those listed in the original version of the problems. Students solved individual problems for 60 minutes independently. After this time, they explained and presented their procedures to other students. We analyzed each procedure together with the students. The joint analysis of individual problems allowed us to look into student thinking and also uncover mistakes that students have made. These were not only numerical, but also logical or interpretive errors.

In our article, we will focus on two of the problems that were also the most interesting from the point of view of the student's solution.

In the following Problem 1, Exam test participants achieved a 44.9% success rate.

**Problem 1**

The table shows the results of pupils in a half-year test in mathematics. What was the average test score?

| Mark | 1 | 2 | 3 | 4 | 5 |
|------|---|---|---|---|---|
| Number of pupils | 7 | 5 | 7 | 3 | 2 |

Source: Exam testing No. 17/2016

In addition to the correct procedure of solution and correct result (2 boys, 2 girls)

\[
\frac{7 \times 1 + 5 \times 2 + 7 \times 3 + 3 \times 4 + 2 \times 5}{7 + 5 + 7 + 3 + 2} = 2.5
\]

(1)

the following two incorrect procedures appeared in the solutions. In the first one, students (1 boy, 1 girl) counted the marks regardless of the numbers, e. g. \((1 + 2 + 3 + 4 + 5) / 5 = 3\). The second incorrect procedure (2 girls) was the sequence of steps \((7 + 5 + 7 + 3 + 2)/5 = 4.8\), where only the frequencies of occurrence of each mark were added. Interestingly, the students did not appear to have an average grade of 4.8 high, although the “majority” of marks was 1, 2 or 3 (lack of critical thinking). However, then most students considered that they should work with all data from the table and not only with one row. In one case it was a numerical error in the calculation of the fraction (1).

The main difficulty with solving of Problem 1 was the misinterpretation of the table. After explaining to the students that the table can be broken down: 7, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3,
4, 4, 4, 5, 5, the problem was easy for them. Our sample of students thought similarly to the participants of the Exam testing, as almost all their results they had the highest representation in the official testing.

In the previous problem, there is a “typical” table that students encounter also in math textbooks. Next, we show a statistical word problem that reminds the usual situation from the social environment of children (e.g. a poll and its subsequent evaluation), but it is characterized by high difficulty. The Problem 2 assignment describes the estimation of the volume of the trash bin in the class by students at the physics lesson. The estimated volume values are recorded on the board and the numbers of students (who said the individual numbers) are listed next to them. The numbers of students are not given in the figures, but they are represented by commas. The success rate of students in this task in national testing was only 36% and up to 11.5% of students didn’t solve it at all.

**Problem 2**

At the physics lesson, students estimated the volume of the trash bin in the classroom. On the board there is a record of 20 students' answers. The real volume of this trash was 12 litters.

How many litters does the mean of student's estimate differ from this value?

![Image](image.png)

Source: NICEM T9 No. 08/2018

The aim of the problem is to test reading and interpreting data from the blackboard, calculating the arithmetic mean and comparing it to the real value. The correct solution is based on calculating

\[
V = \frac{5 \times 3 + 6 \times 3 + 6 \times 8 + 9 \times 1 + 10 \times 6 + 15 \times 1}{3 + 3 + 6 + 1 + 6 + 1} = 8.25
\]

(2)

and comparing this value with the value 12 from the assignment. The all required skills are often trained, but the problem belongs to difficult ones.

In this problem only 1 girl and 2 boys had the right solution (3,75). The most common incorrect result for calculating the arithmetic mean of the basket volume was 2.65 (2 boys and 1 girl). It is determined as the ratio of numbers 53 and 20, where 53 is the sum of the estimated volume values, no matter how many times the given estimates of volume occur, and the number 20 is the number of students. Even two boys reported this value as a result of the task. The students showed a lack of understanding of the text, because their task was to find a deviation of the mean estimate of the real volume of the trash. In three cases, students had a problem with the subtraction two numbers, one of which was a decimal and the result was also a decimal.
Thus, we can conclude that the difficulty of this problem lies not only in statistical and mathematical knowledge but also in context knowledge and critical questions. The value 2.65 was too low to be an average of 5–15. However, it is also possible that students' success in this problem would be higher if the data in the table were of the same type as in the previous problem.

Both problems can be included in the intermediate level of statistical literacy competence in terms of the classification provided by NICEM. We think these two problem's assignments are simple and clear, their grasping does not require any special literacy skills, the difficulty was not even in mathematical knowledge. The reason of incorrect solutions may be in a formal acquisition of the statistical knowledge. Students know that calculating the arithmetic mean means “something needs to be added and something divided”, but they do not understand it exactly. In addition, when the problem is already calculated, students do not ask critical questions to check the result.

5. Conclusion
The presented results represent part of the research focused on the analysis of critical points in school mathematics, which we have been dealing with since 2015 (Jurečková, 2018, Csachová & Jurečková, 2017). We were motivated by the results of international measurements PISA and TIMSS (Kubáček, 2003), which indicate that the mathematical literacy of graduates of Slovak schools is below the average level of OECD states over a long period. The aim of our research is to identify critical points in school mathematics based on results of professional testing and analyse in detail the causes of students' failure. In this article, we focused on one of the important competences of lower secondary graduates – statistical literacy.

Analyses carried out within the research confirm that it is not enough to increase mathematical skills and statistical knowledge for increasing the level of statistical literacy, but that it is also very important to develop their analytical and critical thinking. We perceive statistical literacy as a gradual and long-term developing skill, which develops to different levels and at different rates in individuals. Achieving a certain level depends not only on computational skills, which is dominant in school mathematics, but is also significantly affected by the level of context knowledge and the ability to interpret and evaluate statistical information.

Results of the research indicate that increased attention needs to be paid to this field of education by all teachers, rather than only by teachers of mathematics. Evaluation and interpretation of data of any kind, presented in the form of tables or graphs, can naturally be included in the content of any teaching subject. For this reason, it is desirable to train the statistical skills of students not only in mathematics but also in other subjects taught in primary and lower secondary schools. We aim to apply the knowledge from the conducted research to training of future mathematics teachers, but also to courses within the lifelong professional education of primary and secondary school teachers.

Acknowledgments
The research was carried out with the support of VEGA 1/0079/19 Analysis of critical areas in school mathematics and identification of factors affecting the attitude of pupils to mathematics and KEGA 010KU-4/2019 Innovative forms of multimedia support for statistical methods in social and education research. We would also like to thank the company Exam testing for the provided data.
References

[1] L. CSACHOVÁ, J. GUNČAGA, M. JUREČKOVÁ: The Educational Research of Mathematical Competence, New York: Nova Science Publisher, K. Patterson, *Focus on Mathematics Education Research* (pp. 31-62). ISBN 978-1-53611-826-1 (2017).

[2] L. CSACHOVÁ, M. JUREČKOVÁ: Notes to the problems with figures at school mathematics. Proceedings of the International Conference on Applied Mathematics. Bratislava, Slovakia (pp. 270-279). ISBN 978-80-227-4765-3 (2018).

[3] J. L. DELORE: Probability and Statistics for Engineering and Sciences, ISBN-13: 978-0-538-73352-6 (2010). Retrieved from https://fac.ksu.edu.sa/sites/default/files/probability_and_statistics_for_engineering_and_the_sciences.pdf

[4] J. FROST: Why statistics is important. The World of Statistics (2013). Retrieved from http://www.worldofstatistics.org/2013/03/04/why-statistics-is-important/

[5] I. GAL: Adults’ Statistical Literacy: Meanings, Components, Responsibilities, *International Statistical Review*, vol. 70 (pp.1-51), (2002).

[6] I. GAL: Statistical literacy: Meanings, components, responsibilities, In J. B. Garfield & D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht: Kluwer (2004).

[7] J. GARFIELD, D. BEN-ZVI: Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics*, 31(30), 72–77, (2009)

[8] J. GUNČAGA, M. JUREČKOVÁ: Statistical analysis of the level of mathematics at Slovak lower secondary schools. Proceedings of the International Conference on Applied Mathematics. Bratislava, Slovakia (pp. 596-605). ISBN 978-80-227-4650-2, (2017)

[9] M. JUREČKOVÁ: Identification of factors determining the mathematical competencies of pupils, Proceedings of the International Conference Research in Didactics of Science PLUS, Prague, Czech Republic (pp. 153-159). ISBN 978-80-7444-065-6 (2018).

[10] Z. KUBÁČEK et al: Statistical Literacy. (Report). (2004). Retrieved from https://www.iuventa.sk › files › spravy › davm016 › pisa_mat_gram_sprava

[11] D. S. MOORE: New pedagogy and new content: The case of statistics. *International Statistical Review*, 65, 123–137 (1997).

[12] S. Sharma: Definitions and models of statistical literacy: a literature review, *Open Review of Educational Research*, vol. 4 (pp. 118-133) (2017).

[13] J. M. SHAUGHNESSY: Research on statistics learning and reasoning. In F. K. Lester, Jr (Ed.), Second handbook of research on mathematics teaching and learning (pp. 957–1009). Reston: The National Council of Teachers of Mathematics. (2007).

[14] M. VRABELOVÁ et al: Collection of statistical literacy tasks, ISBN 978 80 89638 079 in Slovak (2013) Retrieved from https://www.nucem.sk/dl/3560/Statisticka_gramotnost_20130908_web.pdf

[15] J. M. WATSON: Assessing statistical thinking using the media, In I. Gal and J. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 107-121). Amsterdam, The Netherlands: IOS Press. (1997).