Assessing inquiry skills of upper secondary school students

Zuzana Ješková, Brigita Balogová, Marián Kireš
Pavol Jozef Šafárik University in Košice, Slovakia

Abstract. Inquiry-based science education (IBSE) has been in the focus of educational systems all over Europe in the last years. The main idea connected with IBSE is that students are involved in the process of learning by conducting activities to answer research questions similar way as it is done in science. This way they should gain not only scientific knowledge but also skills to do inquiry in order to understand how scientists work. In Slovakia this approach has been in force since 2008 when inquiry and the development of inquiry skills became the integral part of the science curriculum. With regard to the current science curriculum and increased interest in science and mathematics and informatics education a national project aimed at research on the efficiency of innovative teaching methods in mathematics, physics and informatics education has been running in the last few years in Slovakia. One of the project main research questions was to find out what is the current level of development of selected inquiry skills of upper secondary school students. In order to answer this research question a test assessing the level of selected inquiry skills was developed. The test was designed in cooperation with upper secondary school teachers. The test items are set in the context of physics as well as mathematics and informatics, however they test inquiry skills that should be content independent. The test was administered to 751 students aged 15-17 from the 1st and 2nd grade of upper secondary school. The test results have shown rather low level of students’ achievements with the mean score of 32.5%. It has also revealed differences between classes of different age groups, different streams and gender. Surprisingly, the 2nd grade students did not achieve better results than the 1st grade students. In accordance with our expectations, classes with focus on mathematics and informatics achieved better results than general or language classes and boys were better than girls. There were also different achievements in the test items aimed at different inquiry skills with the best results achieved in the field of designing an experiment and identifying variables and weakest results achieved in the field of argumentation. In the paper the results of the test across the sample are discussed and analysed in detail, with focus on selected inquiry skills and the context of physics.

1. Introduction

Scientific inquiry is currently one of the most used terms in science education. There are many definitions of scientific inquiry. According to National Education Standards (NRC, 1996) „scientific inquiry refers to the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work. Inquiry also refers to the activities of students in which they develop knowledge and understanding of scientific ideas, as well as an understanding of how scientists study the natural world“. Resulting from the definition, students conducting activities should gain not only scientific knowledge but also skills to do inquiry in order to understand how scientists work. This approach to science education (inquiry-based science education, IBSE) is in strong focus of European educational systems, involving Slovakia. It has been reflected in a number of projects at European and national level over the last years (e.g. 7FP: SAILS, ESTABLISH). In Slovakia the inquiry approach with focus on inquiry skills development is clearly defined in the curricular goals since 2008 not only
in science but the elements of inquiry can be found in mathematics and informatics curriculum. With regard to the curriculum that has been in force already for several years, the following research question has emerged: What is the current level of development of selected inquiry skills of upper secondary school students? Accordingly, the purpose of the presented research is to diagnose the present state of development of selected inquiry skills without any external intervention to teaching in order to find out how the curricular goals have been reflected in the students’ achievements. The presented research question represents the initial research question of the national project aimed at research on the efficacy of innovative teaching methods in mathematics, physics and informatics education (VEMIV). In the following paragraphs the answer to this research question is presented with regard to the methods of data collection and their analysis and results’ discussion in the context of physics.

2. Methods

Development of a test of inquiry skills
In order to answer the research question about the current level of development of selected inquiry skills a test assessing selected inquiry skills was designed. Firstly, different frameworks of inquiry skills based on the stages of inquiry process or level of sophistication (Wenning 2005, 2010) with regard to age group (NRC, 2000) were analysed. Based on the existing taxonomies of inquiry skills reflecting the inquiry process (Fuhrman 1978, Tamir & Lunetta 1981, Fradd et al. 2001, Van den Berg 2013) we have adapted and agreed on the framework of inquiry skills with focus on experimental activities (table 1). This taxonomy served as a framework for the test design. We have also benefitted from the several available tests of inquiry skills, e.g. Scientific Inquiry Literacy Test, ScInqLiT (Wenning 2007), Test of Integrated Process Skills (TISP, Burns et al. 1985) a Test of Scientific Literacy Skills, TOSLS, (Gormally et al. 2012).

Table 1. Taxonomy of inquiry skills for experimental activities adapted on the basis of works of Fuhrman (1978), Tamir & Lunetta (1981), Fradd et al. (2001) and Van den Berg (2013). The skills selected for testing are in bold.
4. **Communication**

4.1. Share and present results in front of the class.

4.2. Discuss/defend results/form arguments.

4.3. Elaborate formal report about the gained results.

5. **Application and follow-up**

5.1. Predict on the basis of obtained results.

5.2. Formulate hypothesis for follow-up.

5.3. Apply experimental technique to a new problem.

The test was designed in cooperation with upper secondary school teachers. The individual test items are developed in order to measure the level of selected inquiry skills (table 1, in bold). The test items are set in the context of physics as well as mathematics and informatics. Firstly, a set of 12-16 test items was designed for each of these subjects. The test items were reviewed by physics, mathematics and informatics teachers who considered their difficulty and appropriateness of the test item to the corresponding inquiry skill that the test item is assigned to assess. Based on teachers’ comments and revisions, the three reviewed subject tests were implemented in four classes of upper secondary schools. As a result, there was one test with test items selected from the subject tests developed. The draft version of the test was again implemented in a selected upper secondary class. After additional smaller corrections the final version of the test was prepared. The test includes 12 test items in order to be answered within one standard 45 minutes’ lesson. Most test items are multiple-choice items with one or two correct answers and several items are open-ended. Examples of test items in the context of physics can be seen in tables 4, 5, 6.

**Research sample**

The test was implemented in the school year 2015/2016 in 11 upper secondary schools with 32 classes involving more than 800 students answering the test. From the whole research sample we gained 751 valid answers. The students included into the analysis belonged to 1\textsuperscript{st} grade classes (337 students aged 15-16) and 2\textsuperscript{nd} grade classes (414 students aged 16-17). The classes belonged to one of the three streams focusing either on foreign languages or the subjects of mathematics and informatics and the rest of the students belonged to classes with no special focus (general classes). The research sample involved 457 girls and 294 boys (table 2).

**Data analysis**

After collecting data from the test taken by the research sample, the analysis of test results was performed. The quantitative analysis involves basic descriptive statistics as well as statistical comparisons between different groups of students based on grade, gender and specialization. In order to compare achievements of different sample groups, the Shapiro-Wilk test of normality and subsequently, the Mann-Whitney test was applied. The statistical analysis was carried out in R software (http://www.R-project.org). In the qualitative analysis the students’ answers were analysed with regard to the most common misconceptions and difficulties that students experienced when answering the test.

**3. Results and discussion**

**Quantitative analysis**

The mean test score for the whole sample reached the value of 32.5% (table 2). The mean score is much lower than expected, although the test items were not considered demanding based on teachers’ evaluation of test items. This result indicates that not enough attention is paid to the activities that enable to develop these particular skills, even though it is clearly formulated in the curriculum.
Table 2. Basic statistics for the whole sample as well as for sample groups and comparison of different sample groups

|                   | Whole sample | grade | gender | specialization |
|-------------------|--------------|-------|--------|----------------|
|                   |              | 1st   | 2nd    | boys           | girls          |
| Number of students| 751          | 337   | 414    | 294            | 457            |
| Mean score (%)    | 32.5         | 31.4  | 33.4   | 36.3           | 30.0           |
| Median (%)        | 31.3         | 29.2  | 33.3   | 31.3           | 33.8           |
| St. dev. (%)      | 16.2         | 15.6  | 16.7   | 17.3           | 15.0           |
| Test of normality (p-value) | < 0.001 | < 0.001 | < 0.001 | 0.003 | < 0.001 |
| Mann-Whitney test | 0.07         | < 0.001 | < 0.001 | 0.61 | < 0.001 |

Statistical analysis shows that neither the whole sample nor the sample groups are normally distributed (p = 0.003 for the group of girls and p < 0.001 for any other group), except from the math/inf classes (p = 0.61). As a result, the comparative statistics based on Mann-Whitney test was applied showing the following results. There is no significant difference between the 1st and 2nd grade students (p = 0.07). Even though second grade students have already experienced one more year of learning compared to first grade students, it seems that this extra year spent at upper secondary school had a negligible effect on their results that are similar to their novice schoolmates. This surprising result confirms that the first year upper secondary school education does not influence the level of development of these particular skills. On the other hand, there is a significant difference between boys and girls (p < 0.001) as well as between the math/inf classes and other classes (p < 0.001).

In table 3 the individual test items score are presented with regard to the corresponding inquiry skill as well as the context in which the item is put into. The items with physical context are marked bold. The last column shows the score for the particular skill group.

It can be seen that students achieved best results in the skill to design an experiment and identify variables (49.7%). It indicates that students conduct experiments, at least, according the prescribed procedure. However, all the other results are approximately somewhere between 28-46%, except for the skill concerning discussing and forming arguments were students gain the lowest score (7.6%). This particular test item requires choosing the correct answer and also forming arguments to justify the choice. This proved to be a demanding challenge for students who either did not give relevant arguments or skip the reasoning at all. This particular skill group 4 connected to defending results and forming arguments seems to be the most problematic for students. The best results were achieved in the skill group 1 connected to Conception, planning and design of an experiment (39.5%), while in the skill group 3 aimed at skills to interpret and analyse students achieved the mean score of 32.5%, and the mean score of 18.5% was achieved in the skill group 4. These results indicates that even though quite a lot of students are able to design an experiment and identify variables, they are much worse in data analysis and their interpretation with the worst results in the field of formulating coherent arguments.
Table 3. Mean test score achieved in individual test items corresponding to certain inquiry skill and skill group (Ph-physics, I-informatics, M-mathematics). The test items with physical context are in bold. The numbering of the skill groups corresponds to the taxonomy of inquiry skills (table 1)

| Inquiry skill | Test item context | Mean score (in %) | Mean score (in %) |
|---------------|-------------------|------------------|------------------|
| 1. Conception, planning and design | | | |
| Formulate hypothesis or expectation to be tested | Ph | 30.7 | 39.5 |
| Design experiment (which variables, which relationship) | Ph | 49.7 | |
| Design experiment (which variables, which relationship) | Ph | 36.2 | |
| Design experiment (which variables, which relationship) | I | 41.3 | |
| 3. Analysis and interpretation | | | |
| Transform results into standard form (i.e. tables, graphs) | I | 28.6 | |
| Determine relationships between variables based on data in graph | M/Ph | 29.7 | |
| Determine relationships between variables based on data in table | I | 29.7 | |
| Determine relationships between variables based on data in table | I | 45.8 | 32.5 |
| Determine relationships between variables based on data in text | M | 31.7 | |
| Determine accuracy of experimental data (identify possible sources of errors) | Ph | 29.6 | |
| 4. Communication | | | |
| Discuss/defend results/form arguments | M | 7.6 | 18.5 |

**Qualitative analysis**

In the following section the specific test items are put into the context of physics are analyzed and discussed in detail with regard to most common revealed difficulties and misconceptions.

**Inquiry skill Formulate hypothesis to be tested**

In this test item (table 4, first part) students are expected to choose the correct hypothesis to test the question if the earth and earth are heated equally in the sunlight. Students achieved rather low score (30.7%). Many of them selected just one answer, mainly c. Common wrong answers were a and b. In the first case this is a question, not a statement, while the second choice is a hypothesis but not testing the question. This indicates that students do not have a proper understanding about what the hypothesis is.

**Inquiry skill Design experiment (which variable, which relationship)**

This test item (table 4, second part) with mean score of 49.7% is aimed at identifying independent and dependent variable to measure in the experiment aimed at heating different materials. The most common difficulties were related to confusion between dependent (temperature) and independent variable (time), selection of wrong variables, e.g. amount of sunlight, density, volume, mass, heat. Other problems connected with incomplete graphs, missing table or missing description of a variable were also quite frequent.
Table 4. Example of test item focusing on the skill to design an experiment (Burns et al. 1985)

Maria wondered if the earth and oceans are heated equally by sunlight. She decided to conduct an investigation. She filled a bucket with a kilogram of soil and another bucket with a kilogram of water. In sunny summer day she placed them so each bucket received the same amount of sunlight.

Which hypothesis should be tested by Maria to get the answer to her question?

a) How is water and soil heated by the sun?
b) The longer the soil and water are in the sun, the warmer they become.
c) Soil and water are heated by the sun differently.
d) Soil and water receive different amounts of sunlight at different times of the day.
e) Soil and water are heated equally.

Maria is going to conduct an experiment using buckets with water and soil. Design a table and draw the axes of a graph with physical variables to be recorded by Mary to get the answer to her question.

Name the selected physical variables.

Fig. 1. Confusion between dependent and independent variable (left) and wrong selection of variables (right)

Inquiry skill Determine relationship between variables based on graphs

This test item (table 5) was aimed at graph interpretation reaching the mean score of approximately 30%. Most students were able to identify that the intersection of the two graphs represent the moment when one runner reaches the other one, however, to decide who reaches who was not so simple (distractor c). The wrong answer d was also quite common, i.e. students assigned higher average speed to the runner whose distance vs. time graph lies higher than that of the other. This belongs to a common misconception concerning confusion between variables of distance and speed in this case.

Inquiry skill Determine accuracy of experimental data (identify possible sources of errors)

This test item (table 6) with mean score of 29.6% was aimed at identifying possible sources of errors in data taken by videomeasurement, in this case. Nevertheless, the most commonly selected answer a) was not connected with reasoning the errors but it interpreted the presented graph, even in a wrong way. It seems that students, having a quick glance at the graph decided immediately about decelerated motion (declining graph) without deeper thinking and reading the text with understanding.
Table 5. Example of test item focusing on the skill to determine a relationship based on data in the graph

In the picture you can see a distance (from the start of the race) vs. time graph for two runners.

Decide about the correct statement:

a) The first kilometer was taken sooner by B.
b) A took more than 4 km in the first 10 minutes.
c) At a distance of 4 km from the start runner A reached runner B.
d) During the first 16 minutes A was running at higher average speed than B.
e) B was running faster than A from the end of 10th min to the end of 11th min.

Table 6. Example of test item focusing on the skill to identify possible errors

In the graph the results of videomeasurement of a motion of a sailboat moving uniformly towards the lakeside is presented. Student has measured its distance from the lakeside clicking on the selected point on each videotaping frame while videoanalysing software record time and corresponding distance. He has found the best fit line. What is the most likely reason of the light scatter in the graphed data?

a) Sailboat does not move uniformly, it slows down.
b) **Student did not mark the selected point accurately on each frame.**
c) Video was not recorded precisely.
d) Sailboat does not move uniformly, it speeds up.
e) Software does not record time accurately.

4. Conclusion

To sum up the results of inquiry skills assessment it can be concluded that the level of development of inquiry skills is not satisfactory not reaching the average value of 40%. Surprisingly, no differences between older and younger students were identified, however, boys achieved better results than girls and the students of special mathematics/informatics classes achieved the best results. The best results were achieved in the skill group of designing experiment and identifying variables and relationships while there were problems identified concerning skills to analyse and interpret data. The worst results were identified in the skills of formulating hypothesis, identifying possible sources of errors and argumentation skills.

The level of development of inquiry skills of upper secondary school students diagnosed by assessment test indicate that students are likely not involved in inquiry activities that is an inevitable assumption of inquiry skills development. This is a signal for teachers to change their attitude towards
the way how they teach and how their students learn in order to implement intentionally more student-centred and inquiry-based strategies.

**Acknowledgment**

This work was supported by the Slovak Research and Development Agency under the contract no. APVV-0715-12 Research on the efficiency of innovative teaching methods in mathematics, physics and informatics education.

**References**

Burns J.C., Okey J.R. & Wise K.C. (1985) Development of an integrated process skill test: TIPS II. *Journal of Research in Science Teaching* **22**(2), 169-177.

ESTABLISH project. Available on http://establish-fp7.eu/.

Fradd S.H., Lee O., Sutman F.X. & Saxton M.K. (2001) Promoting Science Literacy with English Language Learners Through Instructional Materials Development: A Case Study. *Bilingual Research Journal* **25**(4), 417-439.

Fuhrman M. (1978) Development of a laboratory structure and task analysis inventory and an analysis of selected chemistry curricula, Unpublished master’s thesis. University of Iowa.

Gormally C., Brickman P. & Lutz M. (2012) Developing a Test of Scientific Literacy Skills (TOSLS): Measuring Undergraduates’ Evaluation of Scientific Information and Arguments. *CBE – Life Sciences Education* **11**(4), 364-377.

NRC, National Research Council (1996) National Science Education Standards. Washington, DC: National Academies Press.

NRC, National Research Council (2000) Inquiry and the National Science Education Standards, A Guide for Teaching and Learning, Washington, DC: National Academies Press.

SAILS project. Available on http://sails-project.eu.

Slovak national VEMIV project. Available on http://ufv.science.upjs.sk/_projekty/vemiv/projekt.php.

Tamir P. & Lunetta V.N. (1981) Inquiry-Related Tasks in High School Science Laboratory. *Science Education* **65**(5), 477-484.

Van den Berg E. (2013) The PCK of Laboratory Teaching: Turning Manipulation of Equipment into Manipulation of Ideas. *Scientia in educatione* **4**(2), 74-92.

Wenning C.J. (2005) Levels of inquiry: Hierarchies of pedagogical practices and inquiry processes. *Journal of Physics Teacher Education Online* **5**(2), 3-16.

Wenning C.J. (2007) Assessing inquiry skills as a component of scientific literacy. *Journal of Physics Teacher Education Online* **4**(2), 21-24.

Wenning C.J. (2010) Levels of Inquiry. Using inquiry spectrum learning sequences to teach science, *Journal of Physics Teacher Education Online* **5**(4), 11-19.