



FORMATIVE ASSESSMENT AS A TOOL TO ENHANCE THE DEVELOPMENT OF INQUIRY SKILLS IN SCIENCE EDUCATION

Abstract. *Formative assessment (FA) is considered a powerful tool to enhance learning. However, there have been few studies addressing how the implementation of FA influences the development of inquiry skills so far. This research intends to determine the efficacy of teaching using FA in the development of students' inquiry skills. A test of inquiry skills was used as the main research instrument to identify the level of the selected inquiry skills before and after a period of teaching with or without FA in the experimental and control group respectively. In the experimental group (n=73), FA was implemented across five subjects: biology, chemistry, physics, mathematics, and informatics, while the control group (n=48) was taught without any external intervention. In both groups, mostly teacher-centred methods of teaching were applied. The research findings confirmed a low initial level of inquiry skills. Teaching using FA showed higher efficacy than teaching without FA, even without a special focus on the inquiry methods. The experimental group achieved significantly better results compared to the control group. In addition, the results confirmed higher efficacy of FA in lower performing students. Gender did not prove as a determining factor in the effective implementation of FA.*

Keywords: *formative assessment classroom techniques, inquiry skills, inquiry skills test, students' performance, science teaching, upper secondary school*

**Mária Ganajová, Ivana Sotáková,
Stanislav Lukáč, Zuzana Ješková,
Veronika Jurková, Renáta Orosová**
*Pavol Jozef Šafárik University in Košice,
Slovakia*

**Mária Ganajová,
Ivana Sotáková,
Stanislav Lukáč,
Zuzana Ješková,
Veronika Jurková,
Renáta Orosová**

Introduction

Assessment is the key process in assuring quality education (Lamanauskas & Vilkonienė, 2008). The problems of assessment – and formative assessment (FA), in particular – have been widely addressed by teachers, teachers' educators, and researchers for several decades. Formative assessment has a variety of definitions. However, different researchers have come to a consensus regarding certain characteristics pertaining to this form of assessment. Formative assessment is based on regular and interactive evaluation of students' work, it provides feedback on their learning in terms of the specified goals, and indicates what the next steps in teaching should be (OECD/CERI, 2008; OECD, 2005; Wiliam, 2011).

The efficacy of FA has been investigated by a number of researcher teams focused on education. Many of them have confirmed a positive influence on the students' learning process such as an improvement in their academic results, i.e., FA helps improve students' performance and the overall quality of education (Allal & Mottier Lopez, 2005; Bell & Cowie, 2001; Black & Wiliam, 1998a, 1998b; Brookhart, 2008; Floréz & Sammons, 2013; Fluckiger et al., 2010; Hattie & Timperley, 2007; OECD/CERI, 2008; Wiliam, 2010). Implementing FA enhances understanding of the educational contents (Schunk & Swartz, 1993). Black and Wiliam (1998a) state that lower-achieving students benefit from it the most. The efficacy of FA has also been emphasized by Brookhart (2001), Choi et al. (2001), Condie et al. (2005), Department for Education and Skills (2007), Furtak and Ruiz-Primo (2008), Kellard et al. (2008), Kingston and Nash (2011), Kirton et al. (2007), Office for Standards in Education, Children's Services and Skills (2008), Ozan and Kincal (2018), Taras (2007), Torrance and Pryor (2001), Webb and Jones (2009) and many others.

FA develops students' competences, mainly their ability to learn, i.e., students are able to learn from their mistakes, evaluate the results of their learning process, and propose how to improve their own performance (Heritage, 2007, 2010; Looney, 2011; Marshal, 2011; Myhill & Warren, 2005;

OECD, 2005). FA is also important in the process of developing social and personal competences (Starý et al., 2016).

There are studies that have not statistically confirmed any significant influence of FA on students' performance or academic achievement (e.g., Andrews, 2011; Collins, 2012; King, 2003; Tuominen, 2008; Yin et al., 2008). Benett (2011) has pointed out there is little experimental and quantitative research verifying the impact of FA on students' learning results, and the specialised literature addressing FA does not focus on the provability and interpretation of the research results. A similarly critical view of the research focused on FA efficacy can be found in Briggs et al. (2012), Dunn and Melvenon (2009), Filsecker and Kerres (2012), and McMillan et al. (2013). However, this kind of conclusion tends to occur when there are difficulties with effective implementation of FA.

Of course, implementing FA is accompanied by certain difficulties. Researchers have pointed out that the influence of FA on students' learning results is difficult to prove because it only becomes visible in the long run (e.g., Starý, 2006), this kind of research is time-consuming (e.g., Cizek, 2010; Florez & Sammons, 2013; Stiggins, 2002), the requirements for educational contents are complex and determined by the curriculum – the amount of knowledge students are supposed to learn keeps increasing (e.g., Starý, 2006; Stiggins, 2002), the number of students in the classroom is high, teachers are not trained to use FA, and there is little access to mentoring (Clark, 2010; Novotná & Krabsová, 2013).

In the Slovak educational system, summative assessment prevails. The OECD evaluation team has brought attention to the need to improve students' understanding by using formative assessment (Shewbridge et al., 2014). This situation has also been addressed by Orosová et al. (2019) and Ganajová et al. (2019). Their research has shown that Slovak teachers use summative assessment more frequently than formative assessment as some of them are not familiar with formative assessment and they do not understand its methods or aim. Their ideas of certain FA methods such as self- or peer-assessment are often incorrect.

Formative Assessment Classroom Techniques

Formative assessment is based on "purposeful, planned and often spontaneous teacher-to-student, student-to-teacher and student-to-student verbal and written interactions" (Keeley, 2008). These interactions include a variety of assessment techniques known as "Classroom Assessment Techniques" (CATs) (Angelo & Cross, 1993) or "Formative Assessment Classroom Techniques" (FACTs) (Keeley, 2008).

Using FACTs is not time consuming, yet it can provide the teacher with a good overview and feedback on the lessons and at the same time, students get feedback on their own learning (Keeley, 2008). For example, FACTs can take the form of true or false statements (Hubbard et al., 2017; Keeley, 2008; Marzano, 2010), K-W-L chart (Bailey, 2017), Frayer Model (Buehl, 2001; Frayer et al., 1969; Keeley, 2013; Wickens & Parker, 2019), self-assessment card (Keeley & Tobey, 2011), card mapping the learning process (Szarka, 2017), metacognition (Flavell, 1979; Livingston, 1997), exit card (He, 2019; Keeley, 2013;), concept map (Pendley et al., 1994; Regis et al., 1996), check list (British Columbia Institute of Technology, 2010; Ma et al., 2012), task cards (Hattie, 2009), and others. These FACTs help students engage in deeper thinking about their ideas in science and identify the progress in the development of their scientific understanding.

When assessing students' performance in science education, students' understanding of core concepts cannot be assessed separately from their abilities to use the practices of science. These two dimensions of learning should be assessed together to verify whether students can apply theoretical scientific concepts in practice – in other words, whether their level of understanding allows them to investigate the world through practices of scientific inquiry. The term "practices" indicates that engaging in scientific inquiry requires not only knowledge, but also practical skills (National Research Council, 2012, p. 30). These skills are often referred to as inquiry skills. The FACTs can be used to train these practices as suggested by Keeley (2015).

Research Problem

In the Slovak State Educational Programme for upper secondary education (ŠPÚ, 2014), specifically the "Man and nature" educational area, it has been stated that teaching science, i.e., physics, chemistry, and biology is supposed to develop scientific literacy in students who are subsequently able to apply the gained knowledge, formulate questions, and draw conclusions based on the evidence in order to understand the subject matter. Discovery and inquiry are basic approaches that allow students to acquire new knowledge and learn the basic scientific skills. These approaches also help students develop a positive attitude to the scientific approach in learn-



ing about the world. Despite this requirement, the international comparative measurement performed by OECD PISA (Programme for International Student Assessment) in cycles has repeatedly pointed out that although Slovak students learn a lot of scientific knowledge and theory, they have issues with the ability to independently reflect on scientific phenomena and see the connections, assess and formulate hypotheses, search for and propose solutions, interpret the obtained data, draw conclusions and support them with arguments (Miklovičová et al., 2017). Specialised literature uses a variety of terms to refer to the aforementioned skills. e.g., science process skills, science inquiry skills, inquiry skills, scientific inquiry skills, or scientific literacy skills. These skills can be developed using different teaching methods. In science education in particular, it is strongly advised to implement teaching and learning strategies that encourage students' active learning based on the inquiry approach (Ješková et al., 2016; Sotáková et al., 2020; Topalsan, 2020).

Multiple publications (e.g., Angelo & Cross, 1993; Grog et al., 2017; Kelley, 2015) have stated that using FACTs in the teaching-learning process develops scientific skills, provides stimuli for discussion and scientific argumentation, encourages students to ask better questions and provide thoughtful responses. FACTs facilitate the development of scientific thinking and learning of scientific procedures. Simultaneously, students learn whether they are able to express themselves scientifically and apply the acquired knowledge in new situations.

However, most of the research has been focused on using FA in inquiry-based science education (IBSE) or project-based learning. Inquiry-based teaching and learning provides the students with an opportunity to discover and develop ideas about natural phenomena, thus arousing their natural curiosity and interest, however, students often need help with this process. Harlen (2014) has claimed that FA provides teachers with the means (i.e., FACTs), allowing students to correctly formulate their ideas and test them carefully.

The available research results have shown that using FA in IBSE develops students' inquiry skills (Barron & Darling-Hammond, 2008; Black et al., 2004; Grog et al., 2017; Harlen, 2013; Heritage, 2010; Hume & Coll, 2010).

In Slovakia and also in other countries, the teacher-centred teaching style prevails: the teacher presents concepts with logical – deductive – implications and provides examples of application while the student receives a comprehensive system of knowledge and is supposed to understand and memorise it (Balansag, 2018; Saritas, 2016). Although there are efforts to shift science education towards inquiry-based science education (IBSE), the aforementioned teaching style defined by the limited presence of active-learning elements still prevails in Slovakia. As mentioned before, most of the existing studies have focused on using FA in IBSE. Therefore, the main aim of this research was to determine the efficacy of formative assessment with regard to inquiry skill development while using the currently prevalent teacher-centred teaching style in science education. Since formative assessment can take the form of different formative assessment classroom techniques (FACTs), in this research, the implementation of FA is referred to as "teaching using FACTs".

Research Aim and Research Questions

The main research aim was to determine the efficacy of teaching using FACTs with regard to inquiry skill development while using the currently prevalent teacher-centred teaching style in science education.

The research questions were defined as follows:

1. What is the level of the selected inquiry skills in the sample of students at the beginning of the research?
2. Is teaching using FACTs effective in the development of the selected inquiry skills?
3. How do different factors (gender and overall assessment at the end of the term) affect the efficacy of teaching using FACTs in terms of developing the selected inquiry skills?

Research Hypotheses

The following hypotheses were formulated and tested:

- H₁: After FACTs are implemented into teaching, there is a significant difference in the overall level of inquiry skills in the experimental group in comparison to the control group.
- H₂: After FACTs are implemented into teaching, there is a significant difference in the development of the individual inquiry skills in the experimental group in comparison to the control group.
- H₃: In the experimental group, there is a significant difference in the inquiry skill development in students divided according to their gender.



H₄: In the experimental group, there is a significant difference in the inquiry skill development in students divided according to their overall assessment at the end of the term.

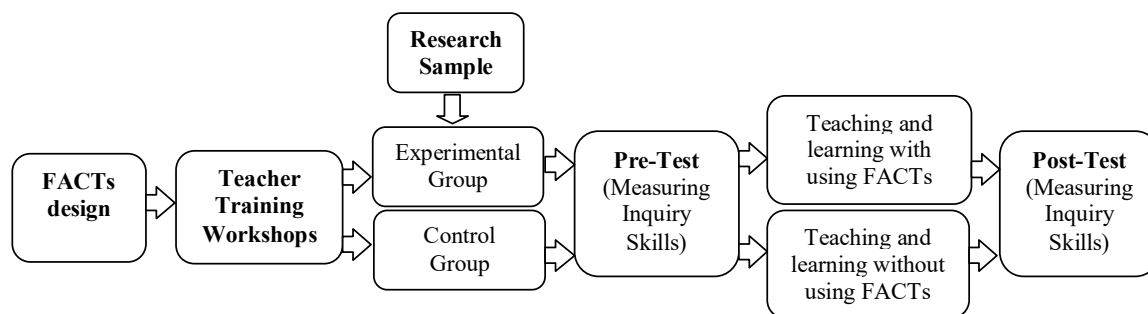
Research Methodology

Research Design

The pre-test and post-test two-group design was used in the research (Campbell & Stanley, 1963). In both experimental and control groups, science subjects (biology, chemistry, physics, mathematics, and informatics) were taught by the same teachers. Teaching in these groups did not differ in terms of content, methods, or duration. However, FACTs were used systematically to teach the experimental group. The number of lessons dedicated to each of these subjects in the students' weekly schedule, lesson contents and time available were identical in both groups. Figure 1 presents the stage of research design.

Figure 1

Stages of Research Design



Research Sample

Deliberate sampling was used to select the research sample. Three Slovak upper secondary schools were involved in the research conducted during the second term of the 2018/2019 school year. The selection of schools was deliberate and followed the following criteria. Firstly, schools inclined towards innovation whose management promoted active learning were selected. Secondly, schools whose teachers actively participated in the IT Academy – Education for the 21st Century project (<http://itakademia.sk/>) and expressed their interest to implement FA into their teaching were selected. In terms of this project, the teachers had an opportunity to get acquainted with a variety of FACTs. Based on these criteria, 11 teachers were selected for the project: 3 men and 8 women with 2 to 24 years of teaching practice.

At each school, teachers selected two 1st year forms. The form with the higher average academic achievement at the end of the term was selected as the control group, while the worse faring form was selected as the experimental group. This division did not affect the pre-test results; there was no statistically significant difference in the level of inquiry skills between the experimental and the control group. The research sample consisted of 121 students from the selected forms who took both tests (pre-test and post-test).

55 (45.5%) were males and 66 (54.5%) females. Students were aged 15–16 years. The experimental group consisted of 73 (60.3%) students, and the control group of 48 (39.7%) students. Table 1 summarises the number and percentage of students divided into groups based on their gender and overall assessment at the end of the term. Each term, the overall assessment of upper secondary school students in the Slovak Republic follows the rules specified in Guideline no. 21/2011 for the assessment and classification of upper secondary school students issued by the Ministry of Education, Science, Research and Sport of the Slovak Republic (MŠVVaŠ SR, 2011).



Table 1*Division of Students according to Gender and Overall Academic Performance at the End of the Term*

Variable		Experimental group		Control group	
		n	%	n	%
Gender	Male	34	46.6	21	43.8
	Female	39	53.4	27	56.2
Total		73	100	48	100
Overall assessment at the end of the term	Passed with distinction	35	48.0	30	62.5
	Passed very well	25	34.2	14	29.2
	Passed	13	17.8	4	8.3
Total		73	100	48	100

The students passed with distinction (PD) if their average study results in all compulsory subjects were equal to or less than 1.5.

The students passed very well (PW) if their average study results in all compulsory subjects were equal to or less than 2.0.

The students passed (P) if their average study results in all compulsory subjects were more than 2.0.

Ethical Considerations

An initial meeting was held before the research started. The research team informed the teachers about the focus of the research and their role in it as well as about the contents of the database of FACTs created, and publishing of results. Both teachers and students participated in the research voluntarily. The teachers motivated the students to accept the FACTs positively. The students were informed that the test results concerning their inquiry skills would be statistically processed for research purposes. In terms of result evaluation, each student participant was assigned a code to maintain confidentiality of their personal information.

Research Instrument

A test of inquiry skills was used as the research instrument. This test was updated from the previous version presented and described in Ješková et al. (2016); it was developed based on existing tests analyses, e.g., SclnqLiT (Wenning, 2007), TISP (Burns et al., 1985), TOSLS (Gormally et al., 2012), TGSA (Zion et al., 2005) and inquiry skills taxonomies (Fradd et al., 2001; Tamir & Lunetta, 1981; Van den Berg, 2013). The test consisted of 16 items, it was administered in the paper-and-pencil form, and took 45 minutes. The items focused on the following inquiry skills: formulating hypothesis, designing experiments, transforming data into tables and graphs, determining relationships between variables based on tables and graphs, and identifying possible sources of errors. At least two items were assigned to each skill. The context of the tasks drew from the subject matter addressed by the main science subjects (physics, chemistry, biology). Therefore, the test items related to the concepts in different subjects (Table 2). To understand the relationships described in the tasks, basic primary school knowledge was needed, but the students were required to apply their knowledge and skills in new situations. The tasks were selected deliberately – each of them required application of a single specific inquiry skill.

The majority were multiple-choice items with one or two correct answers. In four items, students were offered a set of variables among which they were supposed to identify the appropriate variables and the relationships between them. Test scores were calculated based on the following criteria. In the four aforementioned tasks, only the maximum of .5 point could be achieved; multiple-choice test items (12) were assigned the values of 0 or 1, therefore students could achieve a maximum of 14 points. Most multiple-choice test items included one correct answer. If the student selected the correct answer only, they received 1 point. In all other cases, even when the correct answer was combined with another option, students got 0 points. Multiple-choice items with two correct answers were assigned 1 and .5 point if the student selected both the two, or just one of the correct answers respectively. In all other cases, 0 points were assigned (Ješková et al., 2016).



The reliability of the selected inquiry skill test was determined using the Cronbach Alpha; $\alpha = .63$. Several researchers (e. g. Cohen et al., 2007; Marx et al., 2004) have stated that lower α values are acceptable for tests with fewer test items.

Table 2 includes all test characteristics. The last column of Table 2 presents scientific practices supported by FACTs as proposed by Keeley (2015). They correspond with the tested/measured inquiry skills presented in the first column.

Table 2

Distribution of Test Items in the Tested/Measured Inquiry Skills and Subject Matter

Tested/measured inquiry skills	Number of test items	Subject	Scientific practices as proposed by Keeley (2015)
Formulate a hypothesis to test	2 (2 p)	Chemistry, physics	
Designing experiments (identify variables and their relationships)	6 (4 p)	Physics, informatics, chemistry	Planning and carrying out investigations
Transform data to standard forms (i.e., tables or graphs)	2 (2 p)	Physics, informatics, mathematics	
Determining the relationship between variables (based on tables)	2 (2 p)	Physics, informatics, mathematics	Analysing and interpreting data Using mathematics and computational thinking
Determining the relationship between variables (based on graphs)	2 (2 p)	Physics, mathematics	
Determining the accuracy (identify possible sources of errors)	2 (2 p)	Physics, biology	

Procedures

In the preparatory stage of the research (December – October 2018), a group of experts in subject didactics at the Faculty of Science, Pavol Jozef Šafárik University in Košice prepared FACTs for thematic units in the following science subjects – chemistry, biology, physics, mathematics, informatics. In accordance with the respective State Educational Programme for upper secondary education (ISCED 3), these thematic units are taught in the second term of the 1st year at upper secondary schools (ŠPÚ, 2014).

In January 2019, teacher research participants were invited to attend a specialised seminar to learn about the focus and goals of the research. These teachers were given access to the database of FACTs and became conversant with their use in teaching.

At the end of January 2019, all student participants were administered the pre-test to identify their initial level of inquiry skills. Nor teachers, neither students were informed of the pre-test results to avoid influencing the post-test. The pre-test results showed that the experimental group students and the control group students were at a statistically similar level ($p > .05$).

In both control and experimental groups, teaching took place in the same period of the school year, i.e., February to June 2019; In terms of content and teaching methods, the teaching process in both control and experimental groups was comparable. In the control group, the teachers used the teacher-centred teaching style i.e., they did not use FACTs. In the experimental group, teachers used FACTs at their own discretion in different phases of teaching/learning (Table 3). They chose the specific FACTs from the aforementioned database. Their work was coordinated to ensure that at least three subjects were taught using FACTs, and they were used at least once a week in at least one of the subjects.



Table 3*FACT Type Used in the Teaching-Learning Process*

FACT type	Average number of uses per student
True or false statements	2
K-W-L chart	.5
Fruyer Model	1
Self-assessment card	1
Card mapping the learning process	.5
Metacognition	1
Exit card	1.5
Concept map	1.5
Check list	1
Task cards	2

Every FACT addressed the content of the respective lesson.

During the 5-month implementation of FACTs, teachers and experts in subject didactics met three times to share and discuss their new knowledge and experience. At the end of the school year in June 2019, students in both the control and experimental groups were administered, the post-test and the efficacy of FACTs on the development of their inquiry skills was evaluated.

Data Analysis

The qualitative and quantitative analyses were divided into several parts according to the respective research question. The overall test score was analysed as well as specific scores for different inquiry skills. Two-sample *t*-test was used to compare the overall scores of experimental and control groups in the pre-test and post-test. A paired sample *t*-test was used to compare the differences in the level of students' inquiry skills (post-test vs. pre-test) in the experimental and control group. Regarding the effect of the given variables (gender, overall assessment at the end of the term) on the test scores, conditions of the two-way analysis of variance (ANOVA) were complied with. The Shapiro-Wilk test was used to verify the normality of test scores or differences in all parts of the data analysis. In the data analysis, the significance level was $\alpha=.05$, $p<.05$, therefore the difference was considered significant. The SPSS software (SPSS Inc., 2009) was used to analyse the data.

Research Results

The basic descriptive statistics pertaining to the pre-test and post-test results in both groups can be seen in Table 4.

Table 4*Pre-test and Post-test Results*

	Pre-test				Post-test			
	Experimental group		Control group		Experimental group		Control group	
	\bar{x}	<i>SD</i>	\bar{x}	<i>SD</i>	\bar{x}	<i>SD</i>	\bar{x}	<i>SD</i>
Total	.273	.138	.228	.102	.327	.162	.239	.126

The results in Table 4 show that the initial level of inquiry skills failed to match the relative score of .3. In the pre-test, no statistically significant difference was found between the experimental and control groups ($p=.068$, therefore $p>\alpha$).



However, the comparison of the post-test results showed a statistically significant difference between the experimental and the control group ($p=.001$, therefore $p < \alpha$). Therefore, the H_1 hypothesis was confirmed.

After the overall test results were evaluated, the results of students in the control and experimental groups respectively were grouped according to the inquiry skills measured (see Table 5).

Table 5

Results of Students in the Control and Experimental Groups according to the Inquiry Skills Measured

Inquiry skills	Pre-test				Post-test			
	Experimental group		Control group		Experimental group		Control group	
	\bar{x}	<i>SD</i>	\bar{x}	<i>SD</i>	\bar{x}	<i>SD</i>	\bar{x}	<i>SD</i>
Formulate a hypothesis	.223	.248	.154	.151	.233	.229	.181	.230
Design experiments (identify variables and their relationships)	.312	.193	.269	.172	.412	.225	.286	.222
Transform data to standard forms (i.e. tables or graphs)	.151	.297	.167	.260	.219	.289	.250	.326
Determine the relationship between variables (based on tables)	.219	.250	.250	.292	.288	.322	.146	.230
Determine the relationship between variables (based on graphs)	.370	.315	.292	.289	.449	.361	.249	.254
Determine accuracy (identify possible sources of errors)	.322	.305	.207	.260	.274	.344	.302	.338

Table 5 shows the scores obtained by the experimental and the control groups in the pre-test and post-test with regard to the measured inquiry skills. The results indicate low level of the monitored skills.

As can be seen in Table 2, in most cases, two tasks were used to measure the level of one inquiry skill. Therefore, students could gain 0, 1 or 2 points in total (or multiples of .5 point in tasks with two correct answers). The average success rate shows that students frequently scored 0 points. On the other hand, some students were able to score full 2 points. This range of results resulted in rather high *SD* values in some inquiry skills.

Testing the H_2 hypothesis, the differences in the inquiry skill development measured comparing pre-test and post-test results were analysed using the paired *t*-test. The paired *t*-test values in the experimental and control groups are listed in Table 6.

Table 6

H₂ Hypothesis Testing: Paired t-test Results

Paired samples test		Paired Differences							
		95% Confidence Interval of the Difference							
		\bar{x}	<i>SD</i>	<i>SD</i> \bar{x}	Lower	Upper	<i>t</i>	<i>df</i>	<i>p</i>
Pair E	Mean Post-test – Mean Pre-test	.054	.156	.018	.018	.090	2.952	72	.004
Pair C	Mean Post-test – Mean Pre-test	.011	.152	.022	-.033	.055	.509	47	.613

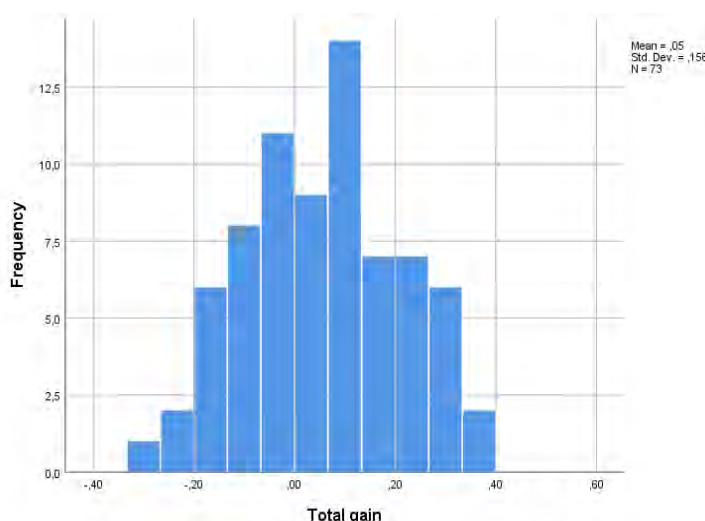


The paired t -test value in the experimental group (Pair E) was $p=.004$, therefore $p<\alpha$. In the experimental group, a statistically significant difference in the inquiry skill development, was confirmed. In the control group (Pair C), the paired t -test value was $p=.613$, therefore $p>\alpha$, so that significant difference was not identified in this case. Therefore, the H_2 hypothesis was confirmed.

An overview of the changes in the students' inquiry skills measured in the experimental group can be seen in the histogram (Figure 2). The histogram shows that the majority of students in the experimental group achieved better results in the post-test in comparison with the pre-test (most students manifested an approximately 10% improvement).

Figure 2

The Inquiry Skill Development in the Experimental Group



To compare the level of individual inquiry skills in the experimental group (pre-test and post-test), the paired t -test was used (Table 7).

Table 7

Differences in the Development of the Individual Inquiry Skills in the Experimental Group: Pre-test and Post-test

		Paired Differences							
		95% Confidence Interval of the Difference							
		\bar{x}	SD	$SD \bar{x}$	Lower	Upper	t	df	p
Pair 1	Formulate a hypothesis	.010	.310	.036	-.062	.083	.283	72	.778
Pair 2	Design experiments (identify variables and their relationships)	.100	.244	.029	.043	.157	3.507	72	.001
Pair 3	Transform data to standard forms (i.e., tables or graphs)	.068	.366	.043	-.017	.154	1.598	72	.114
Pair 4	Determine the relationship between variables (based on tables)	.079	.454	.053	-.027	.185	1.481	72	.143
Pair 5	Determine the relationship between variables (based on graphs)	.068	.366	.043	-.017	.154	1.598	72	.114
Pair 6	Determine accuracy (identify possible sources of errors)	-.048	.401	.047	-.142	.046	-1.021	72	.311



Table 7 shows how the individual inquiry skills improved in students after the experimental intervention. A statistically significant difference in the pre-test and post-test was found in the “designing experiments” skill level ($p=.001$, therefore $p < \alpha$). In other inquiry skills, no statistically significant difference was found ($p > \alpha$).

The H_3 and H_4 hypotheses were tested by comparing the scores of students in the experimental group divided according to their gender and their overall assessment at the end of the term in pre-test and post-test (see Table 8).

Table 8

Scores of Students in the Experimental Group Divided According to their Gender and their Overall Assessment at the End of the Term in Pre-Test and Post-Test

Variable		Pre-test			Post-test		
		\bar{x}	<i>MD</i>	<i>SD</i>	\bar{x}	<i>MD</i>	<i>SD</i>
Gender	M	.278	.259	.140	.363	.393	.182
	F	.267	.250	.138	.295	.306	.137
Overall assessment at the end of the term	P	.207	.214	.126	.321	.321	.152
	PD	.297	.286	.137	.317	.321	.160
	PW	.271	.268	.140	.343	.321	.175
Total		.273	.250	.138	.327	.321	.162

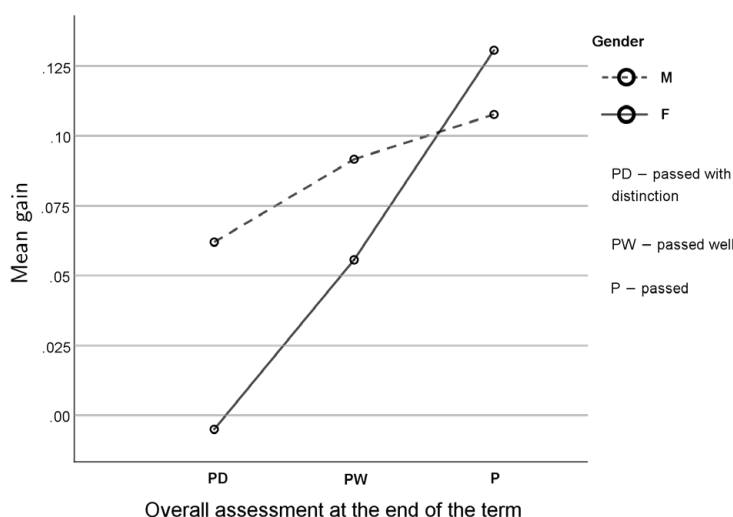
PD – Passed with distinction, PW – Passed very well, P – Passed

Table 8 shows that a bigger difference in the pre-test and post-test results was identified in males. Based on the overall assessment at the end of the term, it can be stated that the biggest difference in the pre-test and post-test results was observed in average students (P), while the least impact could be observed in their excellent counterparts.

Figure 3 shows that in males, divided according to their overall assessment at the end of the term, the development in inquiry skills was less significant than in females. In females, divided according to their overall assessment at the end of the term, the differences were significant; the biggest development could be compared in females assessed as “Passed” at the end of the term. This group of females achieved the biggest development in comparison with males in terms of their assessment at the end of the term.

Figure 3

The Inquiry Skill Development in the Experimental Group according to Students' Gender and Overall Assessment at the End of the Term



The results of the experimental group divided according to the students' gender and their overall assessment at the end of the term in pre-test and post-test were statistically evaluated using the two-factor analysis of variance (Tables 9 and 10). The test aimed to identify whether the difference in the average test scores between the groups was merely accidental or statistically significant. The interaction between the "gender" and "overall assessment at the end of the term" factors determined whether the influence of gender on the average success rate in the pre-test and post-test depended on the "overall assessment at the end of the term" factor.

Table 9*Two-factor Analysis of Variance in the Pre-test*

Source	Type III Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>p</i>
Corrected Model	.128	5	.026	1.382	.242
Intercept	3.113	1	3.113	167.481	.001
Gender	.037	1	.037	1.975	.165
Overall assessment at the end of the term	.104	2	.052	2.784	.069
Gender * overall assessment at the end of the term	.034	2	.017	.920	.404
Error	1.245	67	.019		
Total	6.795	73			
Corrected Total	1.374	72			

In the pre-test, no statistically significant difference was found between males and females ($p=.165$; $p>a$). There is no statistically significant difference between the groups of students with different overall assessment at the end of the term either ($p=.069$; $p>a$). The interaction (gender * overall assessment at the end of the term) was not confirmed: $p=.404$; $p>a$.

Table 10*Two-factor Analysis of Variance in the Post-test*

Source	Type III Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>p</i>
Corrected Model	.106	5	.021	.798	.555
Intercept	5.229	1	5.229	196.087	.001
Gender	.080	1	.080	3.001	.088
Overall assessment at the end of the term	.020	2	.010	.368	.694
Gender * overall assessment at the end of the term	.004	2	.002	.084	.920
Error	1.787	67	.027		
Total	9.678	73			
Corrected Total	1.893	72			

Similarly, the post-test did not show statistically significant difference between males and females ($p=.088$; $p>a$) or groups of students with different overall assessment at the end of the term ($p=.694$; $p>a$). Again, the interaction (gender * overall assessment at the end of the term) was not confirmed: $p=.920$; $p>a$.

The H_3 and H_4 hypotheses were tested using the paired *t*-test (see Tables 11 and 12). The pre-test and post-test success rates were compared separately for males and females, and the groups of students with PD, PW, and P assessment at the end of the term. Each student was assigned one pair (success rate in pre-test and post-test) of mutually independent values.



Table 11*Comparison of the Inquiry Skill Level in Males and Females in the Pre-test and Post-test*

		Paired Differences							
		95% Confidence Interval of the Difference							
Paired samples test		\bar{x}	<i>SD</i>	<i>SD</i> \bar{x}	Lower	Upper	<i>t</i>	<i>df</i>	<i>p</i>
Pair 1	Post-test_M - Pre-test_M	.084	.163	.028	.022	.141	3.032	33	.005
Pair 2	Post-test_F - Pre-test_F	.027	.148	.024	-.020	.075	1.160	38	.253

In the experimental group, a statistically significant difference in the level of inquiry skills between the pre-test and post-test was identified in males ($p=.005$; $p<a$). In females, the development of inquiry skills between the post-test and pre-test was not statistically significant ($p=.253$; $p>a$). Therefore, the H_3 hypothesis was confirmed.

Table 12*Comparison of the Inquiry Skill Level in Groups of Students with Different Overall Assessment at the End of the Term in the Pre-test and Post-test*

		Paired Differences							
		95% Confidence Interval of the Difference							
Paired samples test		\bar{x}	<i>SD</i>	<i>SD</i> \bar{x}	Lower	Upper	<i>t</i>	<i>df</i>	<i>p</i>
Pair 1	P_Post-test - P_Pre-test	.114	.126	.035	.038	.190	3.258	12	.007
Pair 2	PW_Post-test - PW_Pre-test	.071	.166	.033	.003	.140	2.149	24	.042
Pair 3	PD_Post-test - PD_Pre-test	.020	.155	.026	-.034	.073	.742	34	.463

PD – Passed with distinction, PW – Passed very well, P – Passed

A statistically significant difference in the inquiry skill level between the pre-test and post-test was identified in average students (P) ($p=.007$; $p<a$) and very good students (PW) ($p=.042$; $p<a$). However, in excellent students (PD), the difference in the inquiry skill level between the pre-test and post-test was not statistically significant. Therefore, the H_4 hypothesis was confirmed.

Discussion

Considering the first research question, the research showed that the level of inquiry skills identified in the pre-test was low (see Table 4). This result may be connected with the fact that the teacher-centred teaching style dominates in many schools in Slovakia. This teaching style is based mainly on transmissive methods with small number of active-learning elements (OECD, 2009). The teacher-centred teaching style and traditional assessment has no significant impact on the development of these skills (OECD, 2018a, 2018b).

In the pre-test, no statistically significant difference was found between the control and the experimental group ($p=.068$). This fact was important for the experimental intervention.

In terms of the research, the inquiry skill levels in students were compared after teaching experimental and control groups with and without using FACTs respectively. The identified inquiry skill levels in both groups were measured using a post-test to provide data for evaluating the efficacy of teaching using FACTs referred to in the second research question.



The post-test results revealed that the experimental group scored higher in the total average than the control group, and the difference was statistically significant ($p < .05$). By comparing the results of the experimental group in the pre-test and post-test, a statistically significant difference in the students' inquiry skill level was identified ($p = .004$) while the difference in the control group was not statistically significant ($p = .613$). The research findings indicate that teaching using FACTs is more efficient than teaching without them.

The research results presented in this study supplement the results of Bell and Cowie (2001), Grog et al. (2017), and Harlen (2003) who have confirmed the positive impact of FA on the inquiry skill development in the context of IBSE. The results of this research clearly show that using FACTs – even without special intervention concerning systematic implementation of IBSE – is effective in the development of inquiry skills.

It relates to the fact that using FA in teaching helps students achieve a higher level of understanding of the subject matter (Schunk & Swartz, 1993) and develops their higher-order cognitive skills (Brookhart, 2010; Shute, 2008; Kluger & DeNisi, 1996).

In terms of the individual inquiry skill development, the experimental group achieved a statistically significant difference only in the "designing experiments" skill. Although there was some development in other inquiry skills (except the ability to determine the accuracy of data), it was not statistically significant (see Tables 5 and 7). It may relate to the fact that multiple high-frequency FACTs were used to develop the "designing experiments" skill in the learning process (see Table 3). Other skills were developed using a lesser frequency of FACTs. Based on these findings, it can be stated that it is necessary to implement multiple types of FACTs with higher frequency during the given time period to develop the respective skills.

Another research question was: "How do different factors (gender and overall assessment at the end of the term) affect the efficacy of FACTs in terms of developing the selected inquiry skill?"

For this purpose, the results of students in the experimental group were divided according to their gender and overall assessment at the end of the term, and compared (see Tables 8, 9, and 10). The pre-test and post-test showed no statistically significant difference between the inquiry skill level in males and females (see Tables 9 and 10). The research findings indicate that both genders profit from learning using FACTs almost equally. Even the Slovak students' results in scientific literacy according to the PISA 2018 national report have pointed out that scientific literacy remains the only area in which no statistically significant difference was measured across different reporting cycles in males and females in the Slovak Republic (Miklovičová & Valovič, 2019).

However, by comparing the differences between the pre-test and post-test results in the experimental group (see Table 11), a statistically significant difference in the males' inquiry skill level was identified ($p = .005$). However, there was no statistically significant difference in the females ($p = .253$). As can be seen in Figure 3, in the best performing females (PD – passed with distinction), no improvement of inquiry skills could be observed. It may have resulted from anxiety induced by too much responsibility for academic performance or the fact that their learning style focused on memorising. However, the development in inquiry skills between females and males with PW (passed very well) and P (passed) overall assessment at the end of the term did not differ significantly. The results achieved by females with PD assessment at the end of the term affected the overall improvement in inquiry skills, but it was not found significant after all. These findings are similar to the results of the research focused on motivation and engagement in male students compared to female students with regard to inquiry-based teaching (Kuo et al., 2020). However, most researchers investigating the differences between males and females state that gender is not a determining factor in effective implementation of FA (e.g., Ajobgeje, 2013; Asadifard & Afghari, 2019; Moyosore, 2015).

By comparing the pre-test and post-test results (see Table 12), a statistically significant difference in the average (P) ($p = .007$) and very good students (PW) ($p = .042$) was found in terms of their inquiry skill level. However, in excellent students (PD), the difference was not statistically significant ($p = .463$). The aforementioned indicates that using FACTs in teaching improves the inquiry skill level mainly in average-performing students. The PISA 2018 international comparative measurement of scientific literacy has identified a high number of low-performing students (31.4%) referred to as the risk group (Miklovičová & Valovič, 2019). These students are falling behind their peers due to lack of support at school or at home from parents. These students show weak internal motivation, low self-confidence, inability to endure when a complex problem needs to be solved, and also lack of confidence in their own abilities. However, using FACTs has helped them improve their confidence in their own skills and become proud of their achievements (Black & Wiliam, 1998b; Flórez & Sammons, 2013). They have also become more active in learning, more ready to cooperate and independent when provided



formative assessment (Boston, 2002; McMillan, 2007). When students become more self-confident, they learn to attribute their failure and success to the factors they can control and believe in their ability to succeed if they try hard enough. These findings are in line with the results of the study that has confirmed that the greatest influence of FA is on the lower-achieving students (Black & Wiliam, 1998a).

Research Limitations

The compared results may have been influenced by the following.

The teachers who participated in the research had access to a database of FACTs for individual subjects. They selected and used individual FACTs at their own discretion. If a teacher did not have access to pre-existing FACTs or if they had a negative attitude to FACTs, the statistical difference could be less significant.

The limitations of FA implementation in this research include a relatively small sample of upper secondary school students and teachers. However, the FACTs were applied during the same time period and in the same topics, therefore the results allow for drawing conclusions that could help implement FACTs into the teaching-learning process with regard to the development of skills in upper secondary school students.

Other limitations of the FACTs implementation include the fact that their preparation is time-consuming and can be difficult, teachers lack training in FA, and their effect only becomes visible in the long run. These particular limitations were eliminated in this research as the FACTs were prepared in advance to make it easier for the teachers; material equipment and methodological guidance were also provided.

Conclusions and Implications

The results of this research confirmed the efficacy of teaching using FACTs in terms of the inquiry skill development while using the currently prevalent teacher-centred teaching style in science education. This finding adds to the available evidence supporting the importance of FA and its role in education.

A detailed analysis showed that using FACTs in teaching and learning developed the selected inquiry skills (design experiments, transform data to tables or graphs, determine relationship between variables) in students. The statistical analysis revealed a statistically significant difference only in the "designing experiments" skill, in which FACTs were used with the highest frequency. The efficacy with which the selected skills were developed correlates with the frequency of using FACTs.

Moreover, it was found that teaching using FACTs stimulated learning mainly in the lower performing students, in this case, the highest efficacy in terms of inquiry skill development could be observed. The research indicates that the problems faced by the low-performing students (*whose number is very high in Slovakia according to OECD PISA 2018*) can be tackled by using FACTs in teaching.

Gender was not a determining factor in effective implementation of FA. The fact that inquiry skills improved more significantly in males than in females after the experimental intervention may result from the female's anxiety about maintaining outstanding academic performance reflected in their overall assessment at the end of the term as well as their learning style. This finding needs to be verified by future research.

Based on this research, it is recommended to include FA in (future) teacher training. If FA is to be successfully introduced in Slovak schools, the deeply rooted beliefs of Slovak teachers regarding assessment need to change first. It is also necessary to provide them with proper support in implementing FA into their day-to-day practice on the level of educational policy (investing into teachers' professional development, mentoring and peer-support, FA research). In the initial stages of the process, ready-made FACTs databases for individual subjects could also be provided.

However, these requirements are not addressed systematically. They have been partially addressed by the authors of this paper who are involved in VEGA (Scientific Grant Agency, Ministry of Education & Slovak Academy of Sciences) and KEGA (Cultural and Educational Grant Agency, Ministry of Education) research projects focused on building the support system for implementing FA by creating teacher materials addressing the goals and ways to implement FA into education; digital collections of FACTs; methodological materials addressing the implementation of FA into the educational process based on research and verification; systems of contact and online teacher education focused on new assessment forms.

Last but not least, methodological and professional guidance must be continually provided to (future)



teachers regarding their selection and adaptation of the learning content, and creation of the school educational programmes with integrated FA. In this way, we can be certain that students are exposed to the long-term and complete benefits of the influence of FA.

Currently, the methodology centres in Slovakia are starting to provide this kind of training focused on further teacher education under the "Teachers" (professional development) project supported by the European Social Fund. The knowledge and experience gained during this research will be applied in the creation of the "Formative assessment for promoting learning in students" educational project and its implementation.

Acknowledgements

This research was supported by the VEGA No. 1/0265/17 "Formative Assessment in Natural Sciences, Mathematics and Informatics" and KEGA No. 004UPJŠ-4/2020 "Creation, Implementation, and Verification of the Effectiveness of Digital Library with the Formative Assessment Tools for the Natural Sciences, Mathematics and Informatics at the Elementary School" grants.

References

- Ajogbeje, O. J. (2013). Effect of formative testing with feedback on students' achievement in junior secondary school mathematics in Ondo State Nigeria. *International Journal of Education Research*, 1(2), 8-20.
- Allal, L., & Mottier Lopez, M. L. (2005). Formative assessment of learning: A review of publications in French. In J. Looney (Ed.), *Formative assessment: Improving learning in secondary classrooms* (pp. 241-264). OECD.
- Andrews, T. L. (2011). *The use of goal setting and progress self-monitoring with formative assessment in community college to increase academic achievement and self-efficacy* [Doctoral dissertation, Temple University]. ProQuest. <https://search.proquest.com/docview/853329849>
- Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). Jossey-Bass.
- Asadifard, A., & Afghari, A. (2019). The effect of systematic implementation of formative assessment on male and female EFL learners' academic achievement. *Research in English Language Pedagogy*, 7(1), 71-90. <https://doi.org/10.30486/RELP.2019.663423>
- Bailey, L. A. (2017). Adaptation of know, want to know, and learned chart for problem-based learning. *Journal of Nursing Education*, 56(8), 506-508. <https://doi.org/10.3928/01484834-20170712-11>
- Balansag, S. (2018). *Improvement of the teaching style: From traditional teacher-centered to student-centered teaching style*. GRIN Verlag. <https://www.grin.com/document/496335>
- Barron, B., & Darling-Hammond, L. (2008). Teaching for meaningful learning: A review of research on inquiry-based and cooperative learning. In L. Darling-Hammond, B. Barron, P. D. Pearson, A. H. Schoenfeld, E. K. Stage, T. D. Zimmerman, G. N. Cervetti, & J. L. Tilson (Eds.), *Powerful learning: What we know about teaching for understanding* (pp. 11-70). Jossey-Bass/John Wiley & Sons.
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85(5), 536-553. <https://doi.org/10.1002/sce.1022>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25. <https://doi.org/10.1080/0969594X.2010.513678>
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). *Working inside the black box: Assessment for learning in the classroom*. GL Assessment.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles Policy & Practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 118-119. <https://doi.org/10.1002/hrm>
- Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation*, 8(9), 1-6.
- Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, 31(4), 13-17. <http://dx.doi.org/10.1111/j.1745-3992.2012.00251.x>
- British Columbia Institute of Technology. (2010). *Developing checklists and rating scales*. <https://open.bcit.ca/oer/islandora/object/oer%3A33>
- Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice*, 8(2), 153-169. <https://doi.org/10.1080/09695940123775>
- Brookhart, S. M. (2008). *How to give effective feedback to your students*. Association for Supervision and Curriculum Development.
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.
- Buehl, D. (2001). *Classroom strategies for interactive learning* (2nd ed.). International Reading Association.
- 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.



- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Rand McNally & Company.
- Choi, K., Nam, J. H., & Lee, H. (2001). The effects of formative assessment with detailed feedback on students' science learning achievement and attitudes regarding formative assessment. *Science Educational International*, 12(2), 28-34.
- Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. Andrade & G. Cizek (Eds.), *Handbook of formative assessment* (pp. 3-17). Routledge.
- Clark, I. (2010). Formative assessment: There is nothing so practical as a good theory. *Australian Journal of Education*, 54(3), 341-352. <https://doi.org/10.1177/000494411005400308>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. Routledge.
- Collins, N. M. (2012). *The impact of assessment for learning: Benefits and barriers to student achievement* [Doctoral dissertation, Cardinal Stritch University]. ProQuest. <https://search.proquest.com/docview/1033338293>
- Condie, R., Livingston, K., & Seagraves, L. (2005). *Evaluation of the assessment is for learning programme: Final report*. The Quality in Education Centre, University of Strathclyde, Faculty of Education. <https://pureportal.strath.ac.uk/files-asset/493484/strathprints027726.pdf>
- Council of Chief State School Officers. (2018). Revising the definition of formative assessment. CCSSO. <https://ccsso.org/sites/default/files/2018-06/Revising%20the%20Definition%20of%20Formative%20Assessment.pdf>
- Department for Education and Skills. (2007). Assessment for learning: Eight schools project report. *DfES*. <https://dera.ioe.ac.uk/7600/1/1f1ab286369a7ee24df53c863a72da97-1.pdf>
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education. *Practical Assessment, Research and Evaluation*, 14(7), 1-11.
- Filsecker, M., & Kerres, M. (2012). Repositioning formative assessment from an educational assessment perspective: A response to Dunn & Mulvenon (2009). *Practical Assessment, Research & Evaluation*, 17(16), 1-9. <https://doi.org/10.7275/xrkr-b675>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906-911.
- Flórez, M. T., & Sammons, P. (2013). *Assessment for learning: Effects and impact*. CfBT Education Trust. <https://www.educationdevelopmenttrust.com/EducationDevelopmentTrust/files/5a/5a6d6203-ec49-4d33-9c5d-42c188184807.pdf>
- Fluckiger, J., Vigil, Y. T., Pasco, R., & Danielson, K. (2010). Formative feedback: Involving students as partners in assessment to enhance learning. *College Teaching*, 58(4), 136-140. <https://doi.org/10.1080/87567555.2010.484031>
- 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.
- Fraye, D., Frederick, W. C., & Klausmeier, H. J. (1969). *A schema for testing the level of cognitive mastery*. Wisconsin Center for Education Research.
- Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education*, 92(5), 799-824. <https://doi.org/10.1002/sce.20270>
- Ganajová, M., Lukáč, S., Kimáková, K., Jurková, V., Ješková, Z., Sotáková, I., Szarka, K., Šnajder, L., & Guniš, J. (2019). Formatívne hodnotenie v prírodných vedách, matematike a informatike na Slovensku [Formative assessment in natural sciences, mathematics and informatics in Slovakia]. *Biológia, ekológia, chémia*, 23(2), 7-14.
- Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a Test of Scientific Literacy Skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *Cell Biology Education – Life Sciences Education*, 11(4), 364-377. <https://doi.org/10.1187/cbe.12-03-0026>
- Grob, R., Holmeier, M., & Labudde, P. (2017). Formative assessment to support students' competences in inquiry-based science education. *Interdisciplinary Journal of Problem-Based Learning*, 11(2), 6. <https://doi.org/10.7771/1541-5015.1673>
- Harlen, W. (1999). Purposes and procedures for assessing science process skills. *Assessment in Education: Principles, Policy & Practice*, 6(1), 129-144. <https://doi.org/10.1080/09695949993044>
- Harlen, W. (2003). *Enhancing inquiry through formative assessment*. Exploratorium.
- Harlen, W. (2013). *Assessment & inquiry-based science education: Issues in policy and practice*. Global Network of Science Academies (IAP) Science Education Programme (SEP).
- Harlen, W. (2014). Helping children's development of inquiry skills. *Inquiry in Primary Science Education*, 1(1), 5-19.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hattie, J. A. C., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- He, Y. (2019). Traffic light cards: A cross and modification between the minute paper and muddiest point. *College Teaching*, 67(1), 70-72. <https://doi.org/10.1080/87567555.2018.1522612>
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140-145. <https://doi.org/10.1177/003172170708900210>
- Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Corwin Press.
- Hubbard, J. K., Potts, M. A., & Couch, B. A. (2017). How question types reveal student thinking: An experimental comparison of multiple-true-false and free-response formats. *CBE Life Sciences Education*, 16(2), 1-13. <https://doi.org/10.1187/cbe.16-12-0339>
- Hume, A., & Coll, R. K. (2010). Authentic student inquiry: The mismatch between the intended curriculum and the student-experienced curriculum. *Research in Science & Technological Education*, 28(1), 43-62. <https://doi.org/10.1080/02635140903513565>



- Ješková, Z., Lukáč, S., Hančová, M., Šnajder, L., Guniš, J., Balogová, B., & Kireš, M. (2016). Efficacy of inquiry-based learning in mathematics, physics and informatics in relation to the development of students' inquiry skills. *Journal of Baltic Science Education*, 15(5), 559-574.
- Keeley, P. (2008). *Science formative assessment: 75 practical strategies for linking assessment, instruction, and learning*. Corwin Press.
- Keeley, P. (2013). Is it a rock? Continuous formative assessment. *Science and Children*, 50(8), 34-38.
- Keeley, P. (2015). *Science formative assessment, Volume 1: 75 practical strategies for linking assessment, instruction, and learning* (2nd ed.). Corwin Press.
- Keeley, P., & Tobey, C. R. (2011). *Mathematics formative assessment: 75 practical strategies for linking assessment, instruction, and learning*. Corwin Press.
- Kellard, K., Costello, M., Godfrey, D., Griffiths, E., & Rees, C. (2008). *Evaluation of the developing thinking and assessment for learning development programme*. Welsh Assembly Government.
- King, M. D. (2003). *The effects of formative assessment on student self-regulation, motivational beliefs, and achievement in elementary science* [Doctoral dissertation, George Mason University]. ProQuest. <https://search.proquest.com/docview/305239912>
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28-37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Kirton, A., Hallam, S., Peffers, J., Robertson, P., & Stobart, G. (2007). Revolution, evolution or a Trojan horse? Piloting assessment for learning in some Scottish primary schools. *British Educational Research Journal*, 33(4), 605-627. <https://doi.org/10.1080/01411920701434136>
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Kuo, Y. R., Tuan, H. L., & Chin, C. C. (2020). The influence of inquiry-based teaching on male and female students' motivation and engagement. *Research in Science Education*, 50(2), 549-572. <https://doi.org/10.1007/s11165-018-9701-3>
- Lamauskas, V., & Vilkonienė, M. (2008). *European dimension in integrated science education (Training material for students)*. Palacky University Press.
- Livingston, J. A. (1997). *Metacognition: An overview*. https://people.ucsc.edu/~gwells/Files/Courses_Folder/documents/LivingstonMetacognition.pdf
- Looney, J. W. (2011). Integrating formative and summative assessment: Progress toward a seamless system? *OECD Education Working Papers* (No. 58). <http://dx.doi.org/10.1787/5kghx3kbl734-en>
- Ma, I. W. Y., Zalunardo, N., Pachev, G., Beran, T., Brown, M., Hatala, R., & McLaughlin, K. (2012). Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. *Advances in Health Sciences Education*, 17(4), 457-470. <https://doi.org/10.1007/s10459-011-9322-3>
- Marshall, B. (2011). *Testing English: Summative and formative assessment in English*. Continuum.
- Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tal, R. T. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching*, 41(10), 1063-1080. <https://doi.org/10.1002/tea.20039>
- Marzano, R. J. (2010). *Formative assessment & standards-based grading*. Marzano Research Laboratory.
- McMillan, J. H. (2007). *Formative classroom assessment: Theory into practice*. Teachers College Press.
- McMillan, J. H., Venable, J. C., & Varier, D. (2013). Studies of the effect of formative assessment on student achievement: So much more is needed. *Practical Assessment, Research & Evaluation*, 18(2), 1-15. <https://doi.org/10.7275/tmwm-7792>
- Miklovičová, J., Galabová, A., Valovič, J., & Gondžúrová, K. (2017). PISA 2015: Národná správa Slovensko [PISA 2015: Slovakia national report]. *Národný ústav certifikovaných meraní vzdelávania / National Institute for Certified Educational Measurements*. https://www.nucem.sk/dl/3482/NS_PISA_2015.pdf
- Miklovičová, J., & Valovič, J. (2019). PISA 2018: Národná správa Slovensko [PISA 2018: Slovakia national report]. *Národný ústav certifikovaných meraní vzdelávania / National Institute for Certified Educational Measurements*. https://www.nucem.sk/dl/4636/Narodna_sprava_PISA_2018.pdf
- Moyosore, O. A. (2015). The effect of formative assessment on students' performance in secondary school mathematics. *International Journal of Education and Research*, 3(10), 481-490.
- MŠVVaŠ SR. (2011). Metodický pokyn č. 21/2011 na hodnotenie a klasifikáciu žiakov stredných škôl [Guideline no. 21/2011 for the assessment and classification of upper secondary school students]. *Ministerstvo školstva, vedy, výzkumu a športu Slovenskej republiky / Ministry of Education, Science, Research and Sport of Slovak Republic*. <https://www.minedu.sk/metodicky-pokyn-c-212011-na-hodnotenie-a-klasifikaciju-ziakov-strednych-skol/>
- Myhill, D., & Warren, P. (2005). Scaffolds or straightjackets? Critical moments in classroom discourse. *Educational Review*, 57(1), 55-69. <https://doi.org/10.1080/0013191042000274187>
- National project IT Academy – education for the 21st century. <http://itakademia.sk/>
- National Research Council. (2012). *A Framework for K-12 Science Education: Practices, crosscutting concepts, and core ideas*. National Academy Press. <https://doi.org/10.17226/13165>
- Novotná, K., & Krabsová, V. (2013). Formativní hodnocení: případová studie [Formative assessment: A case study]. *Pedagogika / Pedagogy*, 63(3), 355-371.
- OECD/CERI. (2008). Assessment for learning: Formative assessment. *OECD/CERI Publishing*. <http://www.oecd.org/site/educeri21st/40600533.pdf>
- OECD. (2005). Formative assessment: Improving learning in secondary classrooms. *OECD Publishing*. <http://www.oecd.org/education/ceri/35661078.pdf>



- OECD (2009). Creating effective teaching and learning environments: First results from TALIS. *OECD Publishing*. <http://dx.doi.org/10.1787/9789264068780-en>
- OECD. (2018a). PISA for development assessment and analytical framework: Reading, mathematics and science. *OECD Publishing*. <https://doi.org/10.1787/9789264305274-en>
- OECD. (2018b). Education 2030: The future of education and skills [Position paper]. *OECD Publishing*. [http://www.oecd.org/education/2030/E2030%20Position%20Paper%20\(05.04.2018\).pdf](http://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf)
- Office for Standards in Education, Children's Services and Skills. (2008). Assessment for learning: The impact of National Strategy support. *Ofsted*. <https://dera.ioe.ac.uk/9309/1/Assessment%20for%20learning%20-%20the%20impact%20of%20National%20Strategy%20support.pdf>
- Orosová, R., Ganajová, M., Szarka, K., & Babinčáková, M. (2019). Hodnotenie v prírodovedných predmetoch v podmienkach slovenského školstva [Evaluation in natural science subjects in the current context of Slovak education]. *Scientia in Education*, 10(1), 17-32. <https://doi.org/10.14712/18047106.1320>
- Ozan, C., & Kincal, R. Y. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences: Theory and Practice*, 18(1), 77-92. <https://doi.org/10.12738/estp.2018.1.0216>
- Pendley, B., Bretz, R., & Novak, J. (1994). Concept maps as a tool to assess learning in chemistry. *Journal of Chemical Education*, 71(1), 9-15. <https://doi.org/10.1021/ed071p9>
- Regis, A., Giorgio Albertazzi, P., & Roletto, E. (1996). Concept maps in chemistry education. *Journal of Chemical Education*, 73(11), 1084-1088. <https://doi.org/10.1021/ed073p1084>
- Saritas, E. (2016). Relationship between philosophical preferences of classroom teachers and their teaching styles. *Educational Research and Reviews*, 11(16), pp. 1533-1541. <https://doi.org/10.5897/ERR2016.2787>
- Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, 18(3), 337-354. <https://doi.org/10.1006/ceps.1993.1024>
- Shewbridge, C., Van Bruggen, J., Nusche, D., & Wright, P. (2014). *OECD reviews of evaluation and assessment in education: Slovak Republic 2014*. OECD Publishing. <http://dx.doi.org/10.1787/9789264117044-en>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189. <https://doi.org/10.3102/0034654307313795>
- Sotáková, I., Ganajová, M., & Babinčáková, M. (2020). Inquiry-based science education as a revision strategy. *Journal of Baltic Science Education*, 19(3), 499-513. <https://doi.org/10.33225/jbse/20.19.499>
- SPSS Inc. (2009). *PASW statistics for Windows, Version 18.0*. SPSS Inc.
- Starý, K. (2006, November 23). *Sumativní a formativní hodnocení* [Summative and formative assessment]. <http://clanky.rvp.cz/clanek/c/G/992/SUMATIVNI-A-FORMATIVNI-HODNOCENI.html/>
- Starý, K., Laufková, V., Stará, J., Novotná, K., Šťastný, V., & Svobodová, V. (2016). *Formativní hodnocení ve výuce* [Formative assessment in teaching]. Portál.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765. <https://doi.org/10.1177/003172170208301010>
- Szarka, K. (2017). *Súčasný trendy školského hodnotenia: Konceptia rozvíjajúceho hodnotenia* [Current trends in evaluation: The conception of assessment for learning]. Kompres.
- ŠPÚ. (2014). Štátny vzdelávací program pre vyššie sekundárne vzdelávanie [The State Education Programme for Upper Secondary Education]. *Štátny pedagogický ústav / National Institute for Education*. <https://www.statpedu.sk/sk/svp/inovovany-statny-vzdelavaci-program/inovovany-svp-gymnazia-so-stvorrocnyim-patrocnym-vzdelavacim-programom/>
- Tamir, P., & Lunetta, V. N. (1981). Inquiry-related tasks in high school science laboratory. *Science Education*, 65(5), 477-484.
- Taras, M. (2007). Assessment for learning: Understanding theory to improve practice. *Journal of Further and Higher Education*, 31(4), 363-371. <https://doi.org/10.1080/03098770701625746>
- Topalsan, A. K. (2020). Development of scientific inquiry skills of science teaching through argument-focused virtual laboratory applications. *Journal of Baltic Science Education*, 19(4), 628-646. <https://doi.org/10.33225/jbse/20.19.628>
- Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615-631. <http://dx.doi.org/10.1080/01411920120095780>
- Tuominen, K. R. (2008). *Formative assessment and collaborative teaming with support involving middle school mathematics teachers* [Doctoral dissertation, University of California, Irvine]. ProQuest. <https://search.proquest.com/docview/304830113>
- Van den Berg, E. (2013). The PCK of laboratory teaching: Turning manipulation of equipment into manipulation of ideas. *Scientia in Education*, 4(2), 74-92.
- Webb, M., & Jones, J. (2009). Exploring tensions in developing assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 16(2), 165-184. <https://doi.org/10.1080/09695940903075925>
- Wenning, C. J. (2007). Assessing inquiry skills as a component of scientific literacy. *Journal of Physics Teacher Education Online*, 4(2), 21-24. <http://www.phy.ilstu.edu/jpteo>
- Wickens, C. M., & Parker, J. (2019). Supporting vocabulary acquisition in physical education settings. *Journal of Physical Education, Recreation and Dance*, 90(5), 16-22. <https://doi.org/10.1080/07303084.2019.1580635>
- Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18-40). Routledge.
- Wiliam, D. (2011). What is assessment for learning. *Studies in Educational Evaluation*, 37(1), 3-14. <https://doi.org/10.1016/j.stueduc.2011.03.001>



- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., Tomity, M. K., & Young, D. B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, 21(4), 355-359. <http://dx.doi.org/10.1080/08957340802347845>
- Zion, M., Michalski, T., & Mevarech, Z., M. (2005). The effects of metacognitive instruction embedded within an asynchronous learning network on scientific inquiry skills. *International Journal of Science Education*, 27(8), 957-983. <https://doi.org/10.1080/09500690500068626>

Received: September 28, 2020

Accepted: March 05, 2021

Cite as: Ganajova, M., Sotakova, I., Lukač, S., Ješkova, Z., Jurkova, V., & Orosova, R. (2021). Formative assessment as a tool to enhance the development of inquiry skills in science education. *Journal of Baltic Science Education*, 20(2), 204-222. <https://doi.org/10.33225/jbse/21.20.204>

Mária Ganajová
(Corresponding author)

PhD, Associate Professor, Pavol Jozef Šafárik University in Košice, Faculty of Science, Department of Didactics of Chemistry, Šrobárová 2, Košice, Slovakia.
E-mail: maria.ganajova@upjs.sk

Ivana Sotáková

PhD, Lecturer, Pavol Jozef Šafárik University in Košice, Faculty of Science, Department of Didactics of Chemistry, Šrobárová 2, Košice, Slovakia.
E-mail: ivana.sotakova@upjs.sk
ORCID: <https://orcid.org/0000-0002-8910-5786>

Stanislav Lukáč

PhD, Associate Professor, Pavol Jozef Šafárik University in Košice, Faculty of Science, Department of Didactics of Mathematics, Jesenná 5, Košice, Slovakia.
E-mail: stanislav.lukac@upjs.sk

Zuzana Ješková

PhD, Associate Professor, Pavol Jozef Šafárik University in Košice, Faculty of Science, Department of Didactics of Physics, Park Angelinum 9, Košice, Slovakia.
E-mail: zuzana.jeskova@upjs.sk

Veronika Jurková

PhD, Lecturer, Pavol Jozef Šafárik University in Košice, Lifelong Learning Centre and Projects Support, Šrobárová 2, Košice, Slovakia.
E-mail: veronika.jurkova@upjs.sk

Renáta Orosová

PhD, Lecturer, Pavol Jozef Šafárik University in Košice, Faculty of Arts, Department of Education, Moyzesova 9, Košice, Slovakia.
E-mail: renata.orosova@upjs.sk

