THE EFFECT OF FORMATIVE ASSESSMENT ON THE DEVELOPMENT OF CONCEPTUAL UNDERSTANDING IN STUDENTS

[a]RENÁTA OROSOVÁ, [b]MÁRIA GANAJOVÁ, [c]IVANA SOTÁKOVÁ

Pavol Jozef Šafárik University in Košice, [a]Faculty of Arts, [b, c]Faculty of Science, Šrobárova 2, Košice, Slovak Republic
email: [a]renata.orosova@upjs.sk, [b]maria.ganajova@upjs.sk, [c]ivana.sotakova@upjs.sk

Abstract: Despite the fact that OECD education analysts recommend using formative assessment, summative assessment remains preferred in Slovak education. Formative assessment is rarely used because Slovak teachers require further training provided through projects such as Teachers and IT Akadémia. Formative assessment is an efficient tool improving the quality of the learning process, which provides targeted feedback on the quality of the education process as well. This paper presents the result of a study focused on the impact of formative assessment tools implemented in the education process, in teaching Chemistry at primary schools, on the development of conceptual understanding in students on selected levels of Bloom's taxonomy (remembering, understanding, application, analysis). The research involved Chemistry teachers from selected primary schools located in the Košice Region (N=3) and 7th grade students from parallel classes at primary schools (N=126). Quasi-experiment was used as the main research method. Research instruments in the control and experimental groups included two standardised cognitive tests (pre-test and post-test). Pre-test and post-test reliability showed $\alpha = 0.686$ and $\alpha = 0.730$ respectively, which is above the minimum value. The results of the research indicated a statistically significant effect of formative assessment on the development of students' conceptual understanding ($p<0.05$) not only in lower-order cognitive processes (remembering, understanding), but also in higher-order cognitive processes (application and analysis).

Keywords: Formative assessment, formative assessment tools, conceptual understanding, quasi-experiment, primary school teachers and students

1 Introduction

The traditional 20th century concept of school assessment preferred the summative approach focused on assessing learning outcomes. To this day, it influences teachers' prevailing tendency to simply grade students instead of evaluating the quality of their learning process. Students who do not perform well enough are simply graded, instead of receiving feedback and learning support which would actually help them improve their performance. Students are often compared to their peers, which can influence them negatively. They may feel inferior, which further decreases their academic performance. Summative assessment tools focused on evaluating learning outcomes are unable to reveal students' actual skills, capabilities, behaviour styles, actions or needs (Sándor & Nóra, 2009). They only provide students with the information about their ranking within the group. Thus, students do not know their actual status or see what they need to improve and how to approach it.

The current requirements for the assessment of learning outcomes in the Slovak Republic are defined in Act No. 245/2008 Coll. on Education and Upbringing (School Act). It allows assessing students' academic performance verbally, by grading, or to combine both these methods (School Act, 2008). Most students are graded on a five-grade numerical scale, which has not changed for decades. Grading without differentiating the assessment of students' results, process and progress, i.e. without respecting the actual functions of assessment, etc. still prevails (Kratochvílová, 2011). This kind of assessment lags behind the general world trend, which emphasizes distinguishing between the outcome – what the student knows and can do in a given period, the process of getting to that outcome – involving effort, responsibility, work habits, etc., and, very importantly – their progress (Guskey & Bailey, 2010; OECD, 2005).

The OECD Evaluation Report on Assessment in Education in the Slovak Republic of 2014 (Shewbridge, Bruggen, Nusche, & Wright, 2014), created by OECD education analysts, has shown that neither teachers, nor students and their parents understand the importance of formative assessment. The OECD evaluation team has pointed out that regular student assessment in Slovakia lacks in formative aspects and students are in strong need of feedback, which would allow them to improve their learning. Therefore, the OECD evaluation team has recommended introducing elements of formative assessment, which captures not only the "subject matter" and "learning outcomes", but also "how" the students are learning.

2 Formative assessment

The concept of formative assessment was first introduced by Michael Scriven (1967) in the late 1960s in his article "The methodology of evaluation" dealing with the evaluation of educational programmes. He proposed that summative and formative approaches should be distinguished.

Bloom, Hastings, and Madaus (1971) have built on his idea in their *Handbook on formative and summative evaluation on student learning* and introduced a concept that distinguished between the two as well. They have defined the main idea of formative assessment, i.e. that teachers should provide feedback and other information necessary for the students to improve their performance. This idea has become the foundation of the contemporary concept of formative assessment. The aforementioned publication has also explained how formative assessment can be applied in practice in teaching different subjects. The authors have proposed dividing the education process into stages and providing students with formative assessment upon completion of each stage. Teachers should use the evaluation results not only to provide students with feedback, but also to modify their style of teaching (Allal & Lopez, 2005).

In their books *Assessment and Classroom Learning* and *Inside the Black Box: Raising Standards through Classroom Assessment*, Black and William (1998a; 1998b) have elaborated the concept of formative assessment and emphasized its positive effect on student learning. They have focused on the efficiency of formative assessment and revealed a correlation between formative assessment and an improvement in student performance. Specifically, formative assessment has been shown to improve, mainly less successful, students' performance. They have also tried to refute the idea that formative assessment is considerably more time-consuming than summative assessment, therefore teachers have less time for other activities and students ultimately learn less.

According to the broadest definition (OECD, 2005), emphasis should put on student progress; formative assessment helps identify student needs and possible learning difficulties, thus allowing the teacher to modify their style of teaching accordingly.

In October 2006, after an extensive review of the literature on formative assessment and consideration of its definitions, the FAST SCASS (Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards) Assessment Expert Panel adopted the following definition (Heritage, 2010): Formative assessment is a process used by teachers and learners during instruction that provides feedback to adapt further teaching and learning to improve the achievement of specified learning outcomes. The purpose of this feedback is to help teachers and learners make adjustments that will improve the achievement of the set objectives.

A more recent revised definition characterises formative assessment as a process that is planned, ongoing, and used by all students and teachers during learning and instruction, for the purpose and with the goal of determining how students are learning, how their understanding has improved, and where they are in the learning process (CCSSO, 2018, p. 2). At the same time, it stresses the importance of teachers' support for students to become independent individuals in the learning process.

Formative assessment is considered to be a planned process in which teachers use evidence to evaluate the ongoing instructional practices and formulate student assessment or – students modify their learning tactics in order to improve their learning (Popham, 2008).

Formative assessment provides information/feedback at a point when student performance can still be improved. The goal of feedback is to identify how students are learning, i.e. diagnose shortcomings, errors, learning difficulties and their causes in order to improve student learning. To sum up, it provides bilateral teacher <–> student feedback (Harlen, 2013). It helps both teachers and students collect information about the way students acquire and apply knowledge, learning efficiency, students' strong points and weaknesses, goals achieved, and procedures necessary to apply to improve the outcomes. Feedback allows for a comparison between the actual performance level and standards required. Feedback is usually provided by the teacher, but peer feedback is also important. It provides students with information about themselves, but it is rarely used for grading (Orna, 2010).

The question about the ideal form of school assessment does not have a clear answer. Students will always come into contact with both these forms of assessment. The relationship between summative and formative forms of assessment is influenced by the "tension" between the immediately verifiable results of summative testing and the fact that formative assessment effects are indirect and take more time to show (Slavík, 1999).

While summative assessment identifies and expresses whether students possess knowledge and understanding, formative assessment shows what exactly students know (Hattie, 2003); it aims to identify what the students are good at, how the subject matter enriched them, and what they are capable of (Laufková, 2017; Orosová, Ganajová, Szarka, & Babinčáková, 2019). Moreover, Wren and Cotton (2008) have identified a significant difference between the purpose and further use of results in summative and formative forms of assessment.

2.1 Formative assessment and the quality of education

The impact of formative assessment on the quality of education can be observed in a number of research studies indicating its efficiency.

It has been proved that formative assessment helps improve student performance as well as the overall quality of education (Allal & Lopez, 2005; Bell & Cowie, 2001; Black & Wiliam, 1998a; Black & Wiliam, 1998b; Black & Wiliam, 2005; Hattie & Timperley, 2007; OECD, 2005). It provides students with an opportunity to acquire deeper understanding of the subject matter (Schunk & Swartz, 1993). In some schools, a positive influence of formative assessment strategies has been identified also in the improved results of summative testing.

Mainly lower performing students have been proved to benefit from formative assessment, as it encourages them to start trusting themselves and become proud of their abilities (Black & Wiliam, 1998b; Flórez & Sammons, 2013). If the learning process involves formative assessment, students become more active and cooperative, but at the same time, they are turning into independent individuals (Boston, 2002; McMillan, 2007). Formative assessment also supports the process of knowledge acquisition and the ability to learn as such. Students are actively involved in the learning process (self-assessment and peer-assessment), which helps them take responsibility for their own learning (the teacher is giving way to students, which also prevents students from blaming the teacher if they fail, see e.g. Stiggins & Chappuis, 2008). This process promotes the development of meta-cognitive skills (Flórez & Sammons, 2013). Developing meta-cognitive skills is related to one of the key competences required from young Europeans in the 21st century – to be responsible for their own learning. Under the influence of formative assessment, students perceive assessment as a natural part of their life (PowerSchool, 2016) and they learn to work systematically, which improves their performance. Formative assessment develops a positive attitude to systematic work in students, which leads to improved performance (Shepard, 2005).

It also has a positive impact on the teachers who apply formative assessment in their teaching practice. It helps the teachers determine what and how they want to teach their students, which involves learning goals as well as feedback on learning. It also allows them to find out how questions revealing actual student understanding should be formulated and to provide students with constructive feedback (Flórez & Sammons, 2013). Thus, formative assessment improves teachers' quality and promotes establishing student–teacher partnership in education, which significantly improves everything that takes place in the classroom (Reddy, 2017).

However, the implementation of formative assessment involves certain difficulties as well. Teachers often complain about its time-consuming nature and point out that it may prevent them from completing the goals specified for the respective lesson due to lack of time (Reddy, 2017). Formative assessment is also much more demanding for teachers in comparison to summative assessment because the former requires teachers to guide students throughout the learning process. Teachers need to plan a broader variety of activities and measure student performance on the go to provide them with feedback, which allows students to improve and achieve the education goals and meet the standards (Akom, 2011). The effects of formative assessment are also visible only after a longer time (Starý, 2007). Moreover, some students may have difficulties with objective self-assessment. When students are asked to perform self-assessment in front of their peers, they may overestimate themselves (Brown & Harris, 2013). Students may not be able to evaluate the quality of their work in the way teachers can (Kruger & Dunning, 1999).

On the hand, there are also studies that have not proved any statistically significant impact of formative assessment on student performance (Andrews, 2011; Collins, 2012; King, 2003; Tuominen, 2008; Yin, et al., 2008). Kingston and Nash (2011), and Baird and Black (2013) have criticised the evidence showing the significantly positive impact of formative assessment.

3 Material and Methods

General Background of Research
The *main goal of the presented research* was to identify the impact of formative assessment tools (also referred to as FA classroom techniques or FACTs) implemented in the education process, in teaching Chemistry at primary schools, on the development of conceptual understanding in students on selected levels of Bloom's taxonomy (remembering, understanding, application, analysis). In accordance with the goal, the following research questions have been formulated:

RQ_1: Does the implementation of formative assessment tools in the educational process influence the development of conceptual understanding in students?
RQ_2: Does the implementation of formative assessment tools in the educational process influence student conceptual understanding on the selected levels (remembering, understanding, application, analysis)?
RQ_3: Does students' gender influence the efficiency of formative assessment in terms of developing conceptual understanding?

The following *research tasks* have been specified:

- Determine the level of conceptual knowledge in the research groups before the teaching experiment – pre-test.
- Perform the teaching experiment (quasi-experiment) – implement the formative assessment tools in teaching.
- Determine the level of conceptual knowledge in the research groups after the teaching experiment – post-test.

In accordance with the theory presented, research goal, tasks, and questions, the following *hypotheses* were formulated:

H_1: After implementation of the formative assessment tools in teaching, there is a statistically significant difference in the level of conceptual understanding in the experimental vs. control groups of students.
H_2: After implementation of the formative assessment tools in teaching, there is a statistically significant difference in conceptual understanding in the experimental vs. control groups of students on the level of remembering.
H_3: After implementation of the formative assessment tools in teaching, there is a statistically significant difference in

the level of conceptual understanding in the experimental vs. control groups of students on the level of understanding.

H_4: After implementation of the formative assessment tools in teaching, there is a statistically significant difference in the level of conceptual understanding in the experimental vs. control groups of students on the level of application.

H_5: After implementation of the formative assessment tools in teaching, there is a statistically significant difference in the level of conceptual understanding in the experimental vs. control groups of students on the level of analysis.

H_6: After implementation of the formative assessment tools in teaching, there is a statistically significant difference in the level of conceptual understanding in boys vs. girls.

Research Sample

The research took place in three primary schools in the Košice Region in Slovakia (Primary Schol Kežmarská 28, Košice; Primary School Staničná 13, Košice, and Primary School Sídlisko II/1336, Vranov nad Topľou) in the second half of the 2019/2020 school year. Two parallel 7th grade forms at each school were involved in the research. Chemistry was taught by the same teachers in each of the two forms at respective schools. The school educational programmes at the selected schools taught the same number of Chemistry lessons (i.e. 2 lessons per week/66 lessons per year) which addressed the same content, and dedicated the same time to the respective topics within their educational plans and schedules. The selection of schools was deliberate and guided by two criteria. The first criterion was the school management's positive attitude to innovation and support for active learning. The second criterion was the teachers' participation in the IT Academy – Education for the 21st Century national project (http://itakademia.sk/) and their interest in active implementation of formative assessment in teaching. In terms of this project, teachers had an opportunity to get acquainted with a variety of formative assessment tools via inquiry-based activities.

This research involved 3 teachers and 126 7th grade students from the aforementioned primary schools. All teachers were females with 11 to 29 years of teaching practice. Students were aged 12 to 13. The experimental group involved 58 (46%) students and the control group involved 68 (54%) students. Table 1 shows the number and % of students in the experimental and control groups divided according to their gender; it provides their average academic performance in the compulsory subjects at the end of the previous term. Primary school student assessment in each term follows the Methodological Instruction No. 22/2011 on student assessment and classification at primary schools (MŠVVaŠ SR, 2011).

Table 1: Characteristics of the research sample – students, primary school

Average		**Research sample**		**Experimental group**		**Control group**	
		N	%	N	%	N	%
Gender	Boys	62	49.2	28	48.3	34	50
	Girls	64	50.8	30	51.7	34	50
Total		126	100	58	100	68	100
Average academic performance at the end of the term		1.34		1.42		1.26	

Instrument and Procedures

The research had three basic stages – preparatory, experimental, and post-experimental (Figure 1).

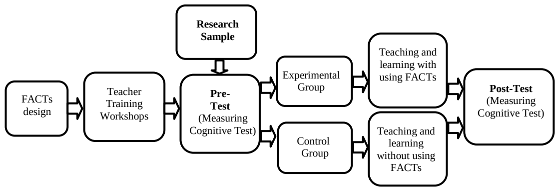

Figure 1: Research stages

The *preparatory stage* consisted of three basic activities – (a) FA tool preparation, (b) specialised seminar for teachers, and (c) pre-test administered to students.

(a) From October to December 2019, a group of experts in subject didactics from the Faculty of Science, Pavol Jozef Šafárik University in Košice (involved in the VEGA No. 1/0265/17 Formative Assessment in Teaching Science, Mathematics, and Informatics research project) was preparing the formative assessment tools for the selected thematic units in Chemistry for primary schools in accordance with the respective state educational programme (ŠPÚ, 2014). FA tools for the "Changes in Chemical Reactions" (7th grade Chemistry) were prepared.

(b) In January 2020, the teachers involved in the research completed the specialised seminar where they were informed about the focus and goal of this research. These teachers were given access to the created formative assessment tools database and instructed about their implementation in the educational process. Teachers were allowed to modify the tools as they deemed appropriate.

(c) All students involved in the research had previously completed a didactic pre-test covering the subject matter addressed during the 1st term of the school year focused on the *"Substances and Their Properties"* thematic unit. The pre-test results showed that all forms were on a statistically comparable level ($p>0.01$). Therefore, forms at the respective schools were assigned to the control and experimental groups randomly (Kireš, Ganajová, & Sotáková, 2019).

The *experimental stage* lasted from February to June 2020; during this period, control and experimental groups were going through the "Changes in Chemical Reactions" thematic unit during Chemistry lessons. In the control group, teachers were not using formative assessment tools in teaching. In the experimental group, the teachers were teaching using the formative assessment tools from the database provided (and could use them as they deemed suitable). Teachers' work was coordinated to ensure that at least one formative assessment tool was used at least once per week.

7th grade Chemistry at primary schools is taught 2 lessons per week, i.e. 33 lessons per term.

The subject matter addressed within the "Changes in Chemical Reactions" thematic unit is taught during the second term of the 7th grade at primary schools. The educational standard for the "Changes in Chemical Reactions" thematic unit specified in the ISCED 2 educational programme for Chemistry in lower secondary education covers the following (ŠPÚ, 2014):

- Content standards: thermal changes in chemical reactions (exothermic and endothermic reactions), rates of chemical reactions, examples of slow and fast reactions, factors affecting the rate of chemical reactions.
- Performance standards: to list examples of exothermic and endothermic reactions known from everyday life, conduct experiments to measure thermal changes in chemical reactions, record experiment results in tables and interpret them, distinguish between slow and fast reactions, conduct and evaluate experiments on the effect of various factors on the rate of a chemical reaction.

The “Changes in Chemical Reactions” thematic unit consists of 3 basic topics and 16 subtopics. Teachers were allowed to choose from the database of formative assessment tools and modify them in terms of teaching the subtopics. The teachers reported using FACTs once a week at most (the average number of lessons = 10). An overview of the tools used can be seen in Table 2.

Table 2: The formative assessment tools implemented in teaching the “Changes in Chemical Reactions” thematic unit

Topic	FACTS used
What are chemical reactions	Self-assessment card Card mapping the learning process Before and after Frayer’s Model Mind map K-W-L method
Energy changes in chemical reactions	Self-assessment card Card mapping the learning process Before and after K-W-L method Frayer’s Model Mind map
The rate of chemical reactions and what influences it	Self-assessment card Card mapping the learning process Before and after Frayer’s Model Mind map

During the period when the FACTs were implemented in teaching, the teachers attended meetings with the experts in subject didactics (3 times over 5 months). At these meetings, the teachers shared experience with FACTs in practice and discussed the implementation issues with the experts in didactics.

The *post-experimental stage* began after the teaching experiment was completed, i.e. in June 2020. Upon completion of the teaching process, students in all groups were administered a didactic test (post-test) covering the subject matter addressed during 2nd term of their school year, i.e. the “Changes in Chemical Reactions” thematic unit in accordance with the content and performance standards defined in ISCED 2 (ŠPÚ, 2014).

Research instruments used in the control and experimental groups included two standardised cognitive tests (pre-test and post-test) (Babinčáková, Ganajová, Sotáková, & Bernard, 2020; Sotáková, Ganajová, & Babinčáková, 2020). In terms of contents, the cognitive pre-test covered the subject matter addressed during the 1st term of the school year, i.e. “Substances and Their Properties”, while the post-test covered the subject matter addressed during the 2nd term, i.e. “Changes in Chemical Reactions” in accordance with the content and performance standards defined in ISCED 2 (ŠPÚ, 2014). Each test consisted of 10 tasks focused on the four levels of the cognitive process in accordance with the revised Bloom’s taxonomy (Anderson & Krathwohl, 2001), i.e. remembering, understanding, application, analysis. In both tests, 5 tasks were single-choice questions and 5 tasks were open and structured (requiring short answer) (Table 3). All tasks have been created and verified by the National Institute for Certified Educational Measurements of the Ministry of Education, Science, Research and Sport of the SR.

Table 3: Characteristics of the tasks in the cognitive tests (pre-test and post-test)

Task number	Level	Task type
1	Remembering	Closed (single-choice)
2	Understanding	Closed (single-choice)
3	Understanding	Closed (single-choice)
4	Understanding	Closed (single-choice)
5	Understanding	Open (structured)
6	Understanding	Open (structured)
7	Application	Open (structured)
8	Application	Closed (single-choice)
9	Application	Open (structured)
10	Analysis	Open (structured)

The results obtained via the research instruments (pre-test and post-test) were statistically evaluated to identify the influence of formative assessment in teaching Chemistry on the development of conceptual understanding in students.

Data collection

The tasks pertaining to the research instruments (pre-test and post-test) were evaluated using binary scoring, i.e. correct answer = 1 point, incorrect, incomplete, or missing answer = 0 points. The data obtained were analysed using phenomenological analysis and descriptive statistics (mean, mode, standard deviation, kurtosis, skewness, range, maximum, minimum, sum, median) and inductive statistics (Kolmogorov-Smirnov test to assess the normality of data distribution and non-parametric Mann-Whitney U test to compare students' knowledge and skill levels).

Reliability of research instruments, i.e. the relationship between research instrument items and the research instrument as a whole, was ascertained through Cronbach's alpha coefficient (Cronbach, 1951). Pre-test and post-test reliability showed $\alpha = 0.686$ and $\alpha = 0.730$ respectively, which is above the minimum value.

The normality of the distribution was verified for the total score, tasks grouped into subscales (remembering, understanding, application, analysis), and also individually for each task using the Kolmogorov–Smirnov test and Shapiro-Wilk test (Table 4). The results of the tests indicated that the data obtained were not normally distributed, therefore the non-parametric Mann-Whitney U test for two independent samples was used to compare the results of the experimental and control groups.

Statistical analyses were performed using SPSS version 18 (SPSS Inc., 2009). For all statistical analyses, the p value < 0.05 was considered significant.

Table 4: Data distribution normality tests

Test	Group	Level	Kolmogorov-Smirnov test			Shapiro-Wilk test		
			Statistics	*df*	*p value*	*Statistics*	*df*	*p value*
Pre-test	Control group	remembering	0.367	65	0.000	0.632	65	0.000
		understanding	0.375	65	0.000	0.630	65	0.000
		application	0.231	65	0.000	0.863	65	0.000
		analysis	0.415	65	0.000	0.605	65	0.000
	Experimental group	remembering	0.348	58	0.000	0.636	58	0.000
		understanding	0.357	58	0.000	0.635	58	0.000
		application	0.247	58	0.000	0.831	58	0.000
		analysis	0.445	58	0.000	0.571	58	0.000
Post-test	Control group	remembering	0.359	65	0.000	0.634	65	0.000
		understanding	0.367	65	0.000	0.632	65	0.000
		application	0.233	65	0.000	0.871	65	0.000
		analysis	0.423	65	0.000	0.597	65	0.000
	Experimental group	remembering	0.462	58	0.000	0.546	58	0.000
		understanding	0.533	58	0.000	0.315	58	0.000
		application	0.393	58	0.000	0.658	58	0.000
		analysis	0.540	58	0.000	0.179	58	0.000

4 Results

Statistical significance of the relationships was verified at the significance level of 0.05. For inductive statistics, Spearman's correlation coefficient was used because the variables did not show a normal distribution ($p \leq 0.05$), which was verified using the Kolmogorov-Smirnov test. The pre-test and post-test scores in the control and experimental groups showed differences even in the basic characteristics (Table 5).

Table 5: Basic characteristics of the pre-test and post-test results in the control vs. experimental groups

		Pre-test				Post-test			
		Control group		Experimental group		Control group		Experimental group	
		$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
Task	1	0.45	0.501	0.48	0.504	0.46	0.502	0.74	0.442
	2	0.45	0.501	0.52	0.504	0.46	0.502	0.76	0.432
	3	0.31	0.465	0.31	0.467	0.29	0.458	0.66	0.479
	4	0.49	0.504	0.60	0.493	0.52	0.503	0.84	0.365
	5	0.62	0.490	0.71	0.459	0.63	0.486	0.91	0.283
	6	0.52	0.503	0.57	0.500	0.54	0.502	0.81	0.395
	7	0.46	0.502	0.52	0.504	0.45	0.501	0.74	0.442
	8	0.65	0.482	0.71	0.459	0.66	0.477	0.95	0.223
	9	0.54	0.502	0.62	0.489	0.57	0.499	0.88	0.329
	10	0.65	0.482	0.71	0.459	0.66	0.477	0.97	0.184
Level	Re	0.45	0.501	0.48	0.504	0.46	0.502	0.74	0.442
	Un	0.43	0.499	0.53	0.503	0.45	0.501	0.91	0.283
	Ap	0.55	0.272	0.61	0.357	0.56	0.283	0.86	0.226
	An	0.65	0.482	0.71	0.459	0.66	0.477	0.97	0.184
Test		0.51	0.192	0.57	0.186	0.52	0.181	0.83	0.141

Legend: Re – remembering, Un – understanding, Ap – application, An – Analysis

In the pre-test, the overall mean rate of success in the control group was 51.2% (SD = 0.192) and that of the experimental group was 57.4% (SD = 0.186). The difference in the overall mean rate of success in the experimental and control groups in the pre-test was not statistically significant (p = 0.083). In the post-test, the overall mean rate of success in the control group was 52.5% (SD = 0.181) and that of the experimental group was 82.6% (SD = 0.141). The overall average achievement of students in the control group did not change significantly compared to the pre-test. However, the overall mean achievement of students in the experimental group compared to the pre-test improved by 25.2%. In the experimental group, the improvement in the post-test can be seen on all four levels of the cognitive process, i.e. knowledge (SD = 0.442), understanding (SD = 0.283), application (SD = 0.226), and analysis (SD = 0.184).

The results of Mann-Whitney U test (pre-test and post-test, control vs. experimental groups, student gender) can be seen in Table 6 and Table 7.

Table 6: Pre-test and post-test results in the control and experimental groups – Mann-Whitney U test

		Pre-test				Post-test			
		Mann-Whitney U	Wilcoxon W	Z	p	Mann-Whitney U	Wilcoxon W	Z	p
Task	1	1816.000	3961.000	-0.405	0.686	1357.500	3502.500	-3.141	0.001
	2	1751.000	3896.000	-0.785	0.433	1325.000	3470.000	-3.346	0.001
	3	1880.000	4025.000	-0.032	0.975	1201.000	3346.000	-4.012	<0.001
	4	1675.500	3820.500	-1.231	0.218	1278.500	3423.500	-3.787	<0.001
	5	1712.500	3857.500	-1.064	0.287	1351.500	3496.500	-3.676	<0.001
	6	1798.500	3943.500	-0.508	0.611	1372.500	3517.500	-3.180	0.001
	7	1780.000	3925.000	-0.614	0.539	1328.500	3473.500	-3.304	0.001
	8	1770.500	3915.500	-0.715	0.475	1344.500	3489.500	-3.929	<0.001
	9	1730.000	3875.000	-0.918	0.359	1300.500	3445.500	-3.789	<0.001
	10	1770.500	3915.500	-0.715	0.475	1312.000	3457.000	-4.229	<0.001
Level	Re	1816.000	3961.000	-0.405	0.686	1357.500	3502.500	-3.141	0.001
	Un	1689.500	3834.500	-1.145	0.252	1003.500	3148.500	-5.470	<0.001
	Ap	1596.000	3741.000	-1.535	0.125	810.500	2955.500	-5.785	<0.001
	An	1770.500	3915.500	-0.715	0.475	1312.000	3457.000	-4.229	<0.001
Test		1547.500	3692.500	-1.732	0.083	382.500	2527.500	-7.686	<0.001

Legend: Re – remembering, Un – understanding, Ap – application, An – Analysis

Table 7: Pre-test and post-test results in the control and experimental groups according to student gender – Mann-Whitney U test

		Pre-test				Post-test			
		Mann-Whitney U	Wilcoxon W	Z	p	Mann-Whitney U	Wilcoxon W	Z	p
Level	Re	1840.500	3670.500	-0.290	0.772	1743.000	3759.000	-0.874	0.382
	Un	1630.500	3646.500	-1.517	0.129	1398.000	3414.000	-3.049	0.002
	Ap	1824.500	3840.500	-0.347	0.728	1768.500	3784.500	-0.653	0.514
	An	1675.500	3505.500	-1.338	0.181	1626.000	3456.000	-1.946	0.052
Test		1752.500	3768.500	-0.705	0.481	1609.500	3625.500	-1.433	0.152

Legend: Re – remembering, Un – understanding, Ap – application, An – Analysis

The statistical verification of the post-test and pre-test results in the control and experimental groups showed statistically significant changes in the experimental group. The Mann-Whitney U test was used to test the research hypotheses; the criterion for accepting a hypothesis is $p < 0.05$ and for rejecting a hypothesis is $p > 0.05$. The results of hypothesis testing are listed in Table 8.

Table 8: Hypothesis testing results – Mann-Whitney U test

Hypothesis	Z	p value	Conclusion
H_1 (Post-test)	-7.686	0.000	Accepted.
H_2 (Remembering)	-3.141	0.001	Accepted.
H_3 (Understanding)	-5.470	0.000	Accepted.
H_4 (Application)	-5.785	0.000	Accepted.
H_5 (Analysis)	-4.229	0.000	Accepted.
H_6 (Gender)	-1.433	0.152	Rejected.

5 Discussion and conclusion

The presented research aimed to identify the effect of formative assessment implemented into teaching Chemistry at primary schools on the development of conceptual understanding in students. The results indicated that the implementation of selected formative assessment tools in teaching was more efficient in terms of developing conceptual understanding as can be seen in the fact that Hypothesis 1 was accepted. The post-test results indicated overall better results in the experimental group in comparison to the control group and the difference was statistically significant ($p < 0.05$). This result is consistent with other research results investigating the effect of formative assessment on the improvement of conceptual understanding in students (Herman & Choi, 2008; Kennedy, Brown, Draney, & Wilson, 2005; Ozan & Kıncal, 2018; Shute, 2008), student performance, and the overall quality of education (Allal & Lopez, 2005; Bell & Cowie, 2001; Black & Wiliam, 1998a; Black & Wiliam, 1998b; Black & Wiliam, 2005; Brookhart, 2008; Flórez & Sammons, 2013; Fluckiger, Vigil, Pasco, & Danielson, 2010; Hattie & Timperley, 2007; OECD, 2005; Wiliam, 2010).

A detailed analysis of the test tasks showed the positive effect of formative assessment not only on the lower-order cognitive processes (remembering and understanding), but also the higher-order ones (application and analysis). The statistical analysis confirmed significant differences between the experimental and control groups on the levels of knowledge, understanding, application, and analysis. The results may result from the fact that the implementation of formative assessment tools in the teaching process promotes the development of higher-order cognitive processes by asking questions and training critical thinking (Brookhart, 2010; Butakor, 2016; Ghani, Ibrahim, Yahaya, & Surif, 2017; Kluger & DeNisi, 1996; Sadler, 2010). FACTs have been designed with the aim to encourage students not only to search for the answers to their questions, but also to explain their ideas and propose solutions, discuss them with their peers, and provide arguments to justify their conclusions. FACTs focused on feedback collection via questions also promote the development of critical thinking, which provides students with an opportunity to acquire a deeper understanding of the subject matter (Marshall, 2007; OECD, 2005). In turn, lower performing students are improving (Boston, 2002; McMillan, 2007).

However, no statistically significant difference was identified in the level of conceptual understanding in boys vs. girls. Therefore, Hypothesis 6 was rejected. The results presented are consistent with the results of the PISA science literacy testing of Slovak students in 2018: "Science literacy is the only area in which, across all cycles, there is no statistically significant difference between the performance of SR boys and SR girls" (Miklovičová & Valovič, 2019).

As the presented results suggest, in terms of developing conceptual understanding, the implementation of formative assessment tools in teaching Chemistry is more than justified. However, the pedagogical practice faces a number of obstacles that need to be gradually eliminated. The structured interviews performed by the experts in subject didactics with the teachers involved in the research after the completion of the experimental stage showed the following: (1) these teachers are open to new challenges related to the implementation of formative assessment tools, (2) they are willing to accept assistance in the form of instruction seminars where they can exchange knowledge and experience, (3) they have access to a FACTs database from which they can choose tools based on the specificities of their classroom groups, students, and subject matter, (4) using FACTs prevents stereotype in the education process, (5) teachers prefer freedom in decision-making, planning, and use of FACTs in their teaching. The teachers' ability to choose FACTs on their own has been previously shown to be beneficial . For example, Babinčáková, Ganajová, Sotáková, and Bernard (2020) have investigated what would happen if they chose FACTs for primary school Chemistry teachers in advance without letting them modify the tools. The teachers involved in their study were teaching the "Mixtures" thematic unit in the 7th grade. Both liberal and strict approaches to the choice of specific FACTs for teaching and the teachers' ability to modify them have been shown to improve conceptual understanding, knowledge, and skills in students. However, based on the available comparison, it can be stated that if a teacher can choose FACTs based on the composition of students and classroom climate, it helps them avoid slipping into their mechanical use.

In conclusion, the structured interviews indicated that the teachers involved in this research developed an internal conviction about the importance of formative assessment. They expressed their wish to continue using the FACTs database in their teaching and even create their own FACTs for more topics.

Research limitations

The results presented could have been influenced by the following factors.

The teachers involved in this research had access to a database of formative assessment tools. These teachers chose and used these FACTs as they deemed suitable. However, if a teacher does not have access to ready-made FACTs or has a negative attitude to using formative assessment tools in general, the ultimate results may not be that significant.

In this case, the research sample consisted exclusively of active teachers open to new methods; they had attended instruction seminars and learned how to use FACTs in practice.

Other limits include the relatively small sample of teachers and primary school students. On the other hand, the formative assessment tools were implemented during the same period and in teaching of the same topics, therefore it can be stated that using FACTs in teaching indeed does improve conceptual understanding on all levels of the Bloom's taxonomy.

The limitations of teaching using formative assessment as such include its time-consuming and more demanding nature, the necessity of training, and the fact that its results only become visible after a longer period of time, which may demotivate the teachers. In this research, the aforementioned limitations were eliminated by preparing formative assessment tools in advance to minimise the preparation time and also by providing expert consulting and methodological support to the teachers involved.

Acknowledgements

This research was supported by the 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" grant.

References

1. Akom, G. V. (2011). *Using formative assessment despite the constraints of high stakes testing and limited resources: A case study of chemistry teachers in Anglophone Cameroon* [Doctoral dissertation]. Western Michigan University. https://scholarworks.wmich.edu/dissertations/331/
2. Allal, L., & Lopez, L. M. (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 Publishing. http://www.oecd.org/education/ceri/35337948.pdf
3. Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman. https://www.uky.edu/~rsand1/china2018/texts/Anderson-Krathwohl%20-%20A%20taxonomy%20for%20learning%20teaching%20and%20assessing.pdf
4. 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]. ProOuest. https://search.proquest.com/docview/853329849
5. Babinčáková, M., Ganajová, M., Sotáková, I., & Bernard, P. (2020). Influence of formative assessment classroom techniques (FACTs) on student's outcomes in chemistry at secondary school. *Journal of Baltic Science Education, 19*(1), 36-49. https://doi.org/10.33225/jbse/20.19.36
6. Baird, J.-A., & Black, P. (2013). Test theories, educational priorities and reliability of public examinations in England. *Research Papers in Education, 28*(1), 5-21. https://doi.org/10.1080/02671522.2012.754224
7. 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
8. 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
9. Black, P., & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment.* King's College London.
10. Black, P., & Wiliam, D. (2005). Changing teaching through formative assessment: Research and practice. In J. Looney (Ed.), *Formative assessment: Improving learning in secondary classrooms* (pp. 223-240). OECD Publishing. http://www.oecd.org /education/ceri/34260938.pdf
11. Bloom, B. S., Hastings, T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning.* McGraw-Hill.
12. Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation, 8*(9), 1-6.
13. Brookhart, S. M. (2008). How to give effective feedback to your students. Association for Supervision and Curriculum Development.
14. Brookhart, S. M. (2010). How to assess higher-order thinking skills in your classroom. ASCD.

15. Brown, G. T., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *SAGE Handbook of research on classroom assessment* (pp. 367-393). SAGE.
16. Butakor, P. K. (2016). The role of formative feedback in promoting higher order thinking skills in classrooms: A theoretical model. *African Research Review, 10*(5), 147-160. http://dx.doi.org/10.4314/afrrev.v10i5.11
17. Collins, N. M. (2012). *The impact of assessment for learning: Benefits and barriers to student achievement* [Doctoral dissertation, Cardinal Stritch University]. ProOuest. https://search.proquest.com/docview/1033338293
18. 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 %20F ormative%20Assessment.pdf
19. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. https://doi.org/10.1007/BF02310555
20. 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-42c18818480 7.pdf
21. 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
22. Ghani, I., Ibrahim, N., Yahaya, N., & Surif, J. (2017). Enhancing students' HOTS in laboratory educational activity by using concept map as an alternative assessment tool. *Chemistry Education Research and Practice, 18*(4), 849-874. https://doi.org/10.1039/c7rp00120g
23. Guskey, T. R., & Bailey, J. M. (2010). *Developing standards-based report cards.* Corwin.
24. Harlen, W. (2013). *Assessment & inquiry-based science education: Issues in policy and practice.* Global Network of Science Academies (IAP) Science Education Programme (SEP).
25. Hattie, J. (2003, October). Teachers make a difference: What is the research evidence? *Building Teacher Quality: What does the research tell us? ACER Research Conference* (pp. 1-17), Melbourne, Australia. https://research.acer.edu.au/cgi/viewcontent.cgi ?article=1003&context=research_conference_2003
26. 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
27. Heritage, M. (2010). Formative assessment: Making it happen in the classroom. Corwin Press.
28. Herman, J. L., & Choi, K. (2008). *Formative assessment and the improvement of middle school science learning: The role of teacher accuracy*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.510.7481&rep=rep1&type=pdf
29. Kennedy, C. A., Brown, N. J., Draney, K., & Wilson, M. (2005). *Using progress variables and embedded assessment to improve teaching and learning.* Berkeley: Berkeley Evaluation & Assessment Research (BEAR) Center.
30. 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]. ProOuest. https://search.proquest.com/docview/305239912
31. 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
32. Kireš, M., Ganajová, M., & Sotáková, I. (2019). The influence of formative assessment on the development of students´ scientific skills in natural science teaching at lower secondary schools. In L. Gómez Chova, A. López Martínez, & I. Candel Torres (Eds.), *ICERI 2019 Proceedings: 12th International Conference of Education, Research and Innovation* (pp. 6932-6940). IATED Academy.
33. Kratochvílová, J. (2012). Aktivní spoluúčast žáka při hodnocení – zdroj inspirace rozvoje osobnosti žáka a pokládání základů zodpovědnosti za kvalitu svého života. In Z. Helus, J. Kratochvílová, H. Lukášová, K. Rýdl, V. Spilková, & T. Zdražil (Eds.), *Proměny pojetí vzdělávání školního hodnocení* (s. 151-180). Asociace waldorfských škol ČR. http://www.iwaldorf.cz/obr/c1309002/Obsah_Promeny%20pojeti%20vzdelani.pdf
34. 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
35. Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. https://doi.org/10.1037/0022-3514.77.6.1121
36. Laufková, V. (2017). Formativní hodnocení v praxi základní školy. *Pedagogika, 67*(2), 126-146. https://doi.org/10.14712/23362189.2017.424
37. Marshall, J. C. (2007, August 15). *4E x 2 Instructional Mode.* http://media.clemson.edu/education/iim/documents/4e-x-2-overview-graphic.pdf
38. McMillan, J. H. (2007). *Formative classroom assessment: Theory into practice.* Teachers College Press.
39. Miklovičová, J., & Valovič, J. (2019). *PISA 2018: Národná správa Slovensko.* Národný ústav certifikovaných meraní vzdelávania. https://www.nucem.sk/dl/4636/Narodna_sprava_PIS A_2018.pdf
40. MŠVVaŠ SR. (2011). Metodický pokyn č. 22/2011 na hodnotenie žiakov základných škôl [Methodological Instruction No. 22/2011 on student assessment and classification at primary schools]. *Ministerstvo školstva, vedy, výzkumu a športu Slovenskej republiky/ Ministry of Education, Science, Research and Sport of the Slovak Republic*. https://www.minedu.sk/vnutorne-rezortne-predpisy-vydane-v-roku-2011/
41. OECD. (2005). *Formative assessment: Improving learning in secondary classrooms.* OECD Publishing. https://doi.org/10.1787/9789264007413-en
42. Orna, M. V. (2010). *SourceBook and 21st Century Chemistry Education. A SourceBook Module.* http://dwb4.unl.edu/ChemSource/ChemSource.html
43. Orosová, R., Ganajová, M., Szarka, K., & Babinčáková, M. (2019). Hodnotenie v prírodovedných predmetoch v podmienkach slovenského školstva. *Scientia in Educatione, 10*(1), 17-32. https://doi.org/10.14712/18047106.1320
44. Ozan, C., & Kıncal, 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
45. Popham, J. W. (2008). *Transformative assessment.* ASCD.
46. PowerSchool. (2016, May 17). *9 Benefits of using formative assessment to increase student growth.* https://www.powerschool.com/resources/blog/9-benefits-of-using-formative-assessment-to-increase-student-growth/
47. Reddy, K. (2017). *Formative evaluation: Importance, advantages & disadvantages.* https://content.wisestep.com/formative-evaluation-importance-advantages-disadvantages/
48. Sadler, R. D. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535-550. https://doi.org/10.1080/02602930903541015
49. Sándor, L., & Nóra, R. (2009). *Fejlesztő értékelés.* Gondolat Kiádo.
50. School Act. (2008). Act No. 245/2008 Coll. on Education and Upbringing [Zákon č. 245/2008 Z. z. o výchove a vzdelávaní (školský zákon) a o zmene a doplnení niektorých zákonov v znení neskorších predpisov]. https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2008/245/
51. 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
52. Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of Curriculum Evaluation* (pp. 39-83). Rand McNally.
53. Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership, 63*(3), 66-70.

54. 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
55. Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. https://doi.org/10.3102 /0034654307313795
56. Slavík, J. (1999). *Hodnocení v současné škole: Východiska a nové metody pro praxi.* Portál.
57. Sotáková, I., Ganajová, M., & Babinčaková, 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
58. SPSS Inc. (2009). *PASW statistics for Windows, Version 18.0.* SPSS Inc.
59. Starý, K. (2007). Formativní hodnocení ve školní výuce. In D. Greger, & V. Ježková (Eds.), *Školní vzdělávání: Zahraniční trendy a inspirace* (s. 221-242). Karolinum.
60. Stiggins, R., & Chappuis, J. (2008). Enhancing student learning. *District Administration, 44*(1), 42-44.
61. ŠPÚ. (2014). Štátny vzdelávací program pre 2. stupeň základnej školy v Slovenskej republike (ISCED 2). [The State Education Programme for Lower Secondrary Education]. *Štátny pedagogický ústav/National Institute for Education.* https://www.statpedu.sk/sk/svp/inovovany-statny-vzdelavaci-program/inovovany-svp-2.stupen-zs/
62. Tuominen, K. R. (2008). *Formative assessment and collaborative teaming with support involving middle school mathematics teachers* [Doctoral dissertation, University of California, Irvine]. ProOuest. https://search.proquest.com/docv iew/304830113
63. 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.
64. Wren, D. G., & Cotton, J. A. (2008). Using formative assessment to increase learning. Research Brief: Report From *The Department of Research, Evaluation, and Assessment, 1*(1), 1-8.
65. Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., Tomita, 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), 335-359. https://doi.org/10 .1080/08957340802347845

Primary Paper Section: A

Secondary Paper Section: AM