Journal of Science Education Vol.22 n. 2 2021

Journal of Science Education

ISSN 0124-5481

Development of experimental design skills - the final results of a longitudinal study

Luca Szalay[a*], István Füzesi[b], Réka Borbás[c], Zoltán Tóth[d]

[a] ELTE Eötvös Loránd University, Faculty of Science, Institute of Chemistry
[b] ELTE Eötvös Loránd University, Bolyai János Practicing School
[c] Szent István Secondary School
[d] University of Debrecen, Faculty of Science and Technology, Institute of Chemistry

ARTICLEINFO

Keywords:
inquiry-based science education
inquiry-based teaching
inquiry-based learning
experimental design
planning an experiment

ABSTRACT

This is a report about a longitudinal research that was originally planned for four years, but was extended to the fifth school year, because of the COVID-19 pandemic. Its goal was to investigate whether the participating students' experimental design skills could be developed effectively by the intervention methods used in the project. 920 Hungarian students were involved in the sample at the beginning (in September 2016), each 12-13 years old. They were divided randomly into three groups. Group 1 students formed the control group, who only did step-by-step student experiments according to the 6 students sheets per school year that we provided. Group 2 students did the same step-by-step experiments, but also got experimental design tasks on paper in the first school year. From the second year, the most important principles of the experimental design after the step-by-step experiments were explained to them. Group 3 students also did the same experiments, but they had to design some steps, without scaffolding in the first year. From the second year they were directly taught the relevant principles of the experimental design before they started to plan those steps. Group 2 and 3 did not achieve significantly better results than Group 1 in the paper and pencil structured test at the end of the first school year on the experimental design tasks. However, both Group 2 and 3 students scored significantly better than Group 1 at the end of the second school year, when the principles of the experimental design were taught directly to them.No significant difference was detected among three groups in terms of experimental design in the third and fourth end-of-school year tests. School effect was significant at the beginning of the project and it exceeded the effect size of the intervention on the last three end-of-school year tests.

1. Introduction

The analysis of PISA 2006 results (OECD, 2007). showed that Hungarian students performed poorly on the sub-tests measuring the competencies named "Identifying scientific issues" (483 points) and "Using scientific evidence" (497 points) compared to their average scores gained in science (504 points). Their "Knowledge about science" also proved to be unsatisfactory (492 points). However, their average score on the "Knowledge of science" scale, and especially on the sub-test called "Physical systems" (533 points) measuring content knowledge in physics and chemistry ranked 4th out of 57 participating countries. They also performed well in the sub-test "Explaining phenomena scientifically" (518 points). These results showed that in Hungary the emphasis in the teaching and learning process in science has traditionally been on the acquisition and understanding of factual knowledge. Therefore, it is seldom discussed how scientists collect and evaluate evidence in scientific research. This fact limits the applicability of the acquired knowledge and can also serve as a basis for the spreading of pseudoscience. According to the recommendation of the report evaluating the results of PISA 2006 (OECD, 2007), the ways in which the extension of scientific competencies could be achieved are needed to be investigated. By 2012, the average performance of Hungarian students in the field of science (494 points) fell below the international average (501 points), which was a cause for concern among both professionals and educational politicians (PISA, 2012). The mean score in science of the Hungarian students in PISA 2015 was even lower (477 points). Politicians and journalists, on behalf of everybody concerned, called for new methods of teaching and learning in science which many educational researchers felt obliged to investigate and explore

2. Rationable

The "Rocard Report" published in 2007 drew attention to the negative trends in science education in European countries (Rocard, 2007). The same document focused on the methods most commonly referred to as inquiry-based science education (IBSE) as one possible response to the problems. The essence of

* Corresponding author. Luca Szalay, PhD
Email: luca.szalay@ttk.elte.hu
Accepted 2 October 2021, Available online 31 December 2021

this approach is that the learning process models scientific research (Olson, Loucks-Horsley, 2000). IBSE activities involve posing questions, collecting information related to the problem, planning investigations and using tools to gather, analyse, and interpret data, proposing answers and communicating the result. In using these methods, students become actively involved both mentally and physically, which can help to develop science literacy (Minner et al., 2010, Tomperi and Aksela, 2014), as well as arouses and maintains interest under the right conditions (Hofstein, Kempa, 1985). The facilitation of differentiation between science and pseudoscience can also be expected (Finlayson et al., 2015). Since 2007, the European Union has supported a number of major projects to research and propagate IBSE methods. The corresponding author of this present paper was involved in two such projects (Mind the Gap and S-TEAM) and took part in the promotion of the dissemination of IBSE methods in Hungary by teaching pre-service and in-service chemistry teachers, by publications and conference talks.

However, the effectiveness and efficiency of minimally guided instruction methods were debated (Cheung, 2011, Kirschner, Sweller, and Clark, 2006, Sweller, 1988). Besides, the motivational effect of the inquiry did not seem to be universal either (Bolte, Streller and Hofstein, 2013). On the other hand, some practicing teachers have reservations on guided inquiry that are due to large class size, lack of time and the need to prepare students for public examinations, according to the summary of Cheung (2011).

However, since other researchers argued that IBSE is not just about learning content, but also develops transferable skills, like epistemic practices, self-directed learning and collaboration (Hmelo-Silver, Duncan and Chinn, 2007), it still seems to be worth doing. But teachers must manage students' cognitive load properly, e.g. by reducing the degree of freedom for the learners (Criswell, 2011). Taber (2011) emphasized that "teachers need to... plan teaching in terms of suitable, regularly reinforced, learning quanta". This could be done, for example by converting a verification experiment to a guided inquiry (Allen et al., 1986). Student should work in teams since the discussions and interactions among students have an important impact on learning (Criswell, 2011). Students also need to have previous knowledge and skills to be successful at doing the inquiry step of the experiment (Bruck and Towns, 2009).

Our research group organised a brief project in the school year 2014/2015 to measure the impact of the simplified version of the IBSE on abilities of designing experiments, disciplinary content knowledge and attitude, involving 660 students aged 15 and 16 (Szalay and Tóth, 2016). The control group did only step-by-step experiments, whereas the experimental group had to design some steps of the same experiments. According to the results, the average development of the experimental design skills of the experimental group was significantly higher than that of the control group. Consequently, the question of the effect of long-term intervention starting from an early age (12-13) arose. Therefore, an investigation planned to last for four school years was started in September 2016, which is described in the present paper. It was also enquired whether it is necessary to carry out the designed experiment, or designing experiments on paper has similar effect on the experimental design skills of the students. There was a research question concerning the influence of the intervention on the development of the students' attitude toward chemistry and chemical experiments, but those results are not reported in this paper.

3. Method

In the first year of the longitudinal research that started in the school year 2016/2017 the same approach was applied as in the brief project: students of the experimental groups were made to design experiments without any scaffolding. 24 in-service chemistry teachers and 5 university chemistry lecturers participated, and pre-service chemistry teacher students were also involved. The sample consisted of 920 students who were 12-13 years old at the beginning of the research, and came from 18 different Hungarian schools and 31 classes/groups. These students were expected to study chemistry for four years (from Grade 7 to Grade 10) in the same school. They had to do experiments on six 45 minutes lessons in each school year, following student sheets provided by our research group. The classes/groups were divided randomly into the following three groups:

- Group 1 (control): Students only carried out step-by-step experiments, given full instructions.
- Group 2: Students carried out the same step-by-step experiments as the control group, given full instructions, and attempted to design closely-related experiments.
- Group 3: Students carried out some of the same step-by-step experiments as the control group, and designed and carried out the missing experiment(s) based on those for which they had full instructions.

Six student sheets and teacher guides were written and reviewed in three versions for the three groups described above by the members of the team. Each student sheet and teacher guide were piloted by teachers working in the team and by their students. Further modifications were made upon reflections, and the revised student sheets and teacher guides were published on the website of our research group. Their English versions are also available on the same website. The research model applied in the first year is shown on Figure 1.

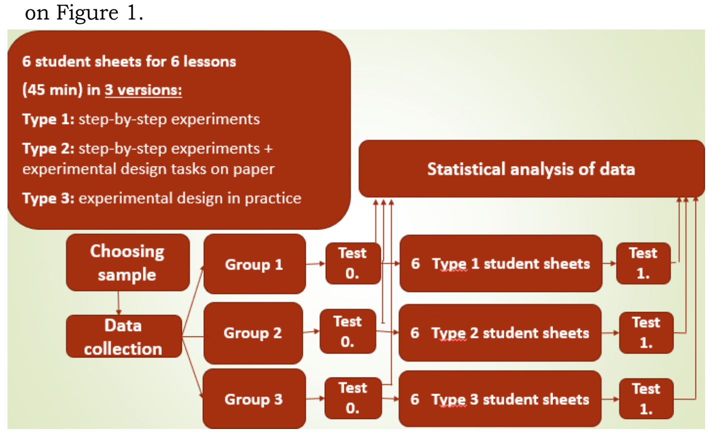

Figure 1: Research model applied in the first year of the project (school year 2016/2017)

The effect of the intervention was measured by 40-minutes structured paper and pencil tests. 'Test 0' was taken by 883 students at the beginning of the first school year in September 2016. 'Test 1 (end of first year)' in May or June of 2017 was taken by 853 students. Disciplinary content knowledge (DCK), experiment design skills (EDS) and attitude toward chemistry and chemical experiments were assessed by the tests. Both tests were structured according to the Bloom-taxonomy. The tests were marked by the students' own teachers, and the markings were supervised and unified. Finally, the test and the revised marking scheme was uploaded on the website of the research group. Results were analysed statistically. Since the average scores of the three groups on Test 0 were significantly different, a matched pair design method was applied to form a reduced sample, in that there was no significant difference among the average scores of the three groups. Since it was found that apart from the intervention, several other factors influenced the students' scores, analysis of covariance (ANCOVA of the SPSS statistical software) was applied. The comparison of the results of Test 0 and Test 1 showed that the first year of the intervention did not produce the expected positive results. The development of experimental design skills of the Group 3 students (who designed experiments before doing them) were not detected significantly better than that of the control group's, as published by Szalay and Tóth (2016).

Since the instruction methods applied in the first year of the project did not prove to be efficient enough, changes in the research model were introduced. From the beginning of the second school year, the most important principles of the experimental design were directly taught to the experimental groups (unlike in the case of the control group, see below). The instruction methods of the three groups of the sample applied from September 2017 were as follows:

• Group 1 (control): Students only carried out 'step-by-step' experiments, given full instructions (as in the first year).

• Group 2: Students followed the same 'step-by-step' recipes as the control group and (instead of designing similar experiments on paper) they were given explanations of the experimental design of these recipes.

• Group 3: Students did the same student experiments as Group 1 and Group 2 designing one or more of those experiments before doing them, but (from the second year) with guidance about experimental design before planning and carrying out some of the experiments.

The research model applied from the second year is shown on Figure 2.

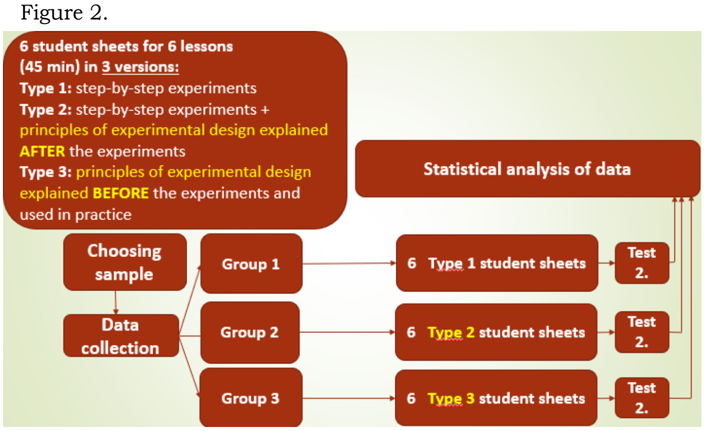

Figure 2: Research model applied from the second year of the project

According to the statistical analysis of the results of the Test 2 (completed by the students at the end of the second school year) this approach produced better results. Thereby, the work was continued in the third school year according to the modified research model. Six student sheets and teacher guides per school year were prepared in the third and the fourth school year of the project too. All were written in three versions for the three groups defined above (see an example as Appendix 1). The student sheets were implemented on the same sample, mostly by the same teachers as in the first and second year. Test 2, 3 and 4 (expected to be taken at the end of the second, third and fourth school year, respectively) were also created in a similar way, with the same structure as the previous two tests (see Test 4 as Appendix 2). On the whole, 724 students completed all the four tests (i.e. Test 0, 1, 2, 3) till the end of the third school year. The results of the analysis of the statistical data gained from the completed tests in the second and the third school year were also published (Szalay et al., 2020). The three types of the 12 student sheets and teacher guides modified according to the experiences of the pilot, the tests and their updated marking schemes based on the experiences of the corrections are all available on the website of the research group (both in English and in Hungarian).

The fourth school year of the project started in the same way as the second and the third ones. However, online teaching started in the Hungarian secondary schools from 16th March 2020 due to the COVID-19 pandemic. Students were not allowed to return to their schools before the 1st September 2020. This made the execution of the remaining student sheets impossible, since the experiments carried out by the students in teams are essential parts of the treatment of the sample. Teachers had the liberty to fit the student experiment lessons into their curriculum according to their own plans. Therefore, a different number of implemented worksheets were missing depending on the previous planning of the teachers. Consequently, the last test (Test 4) could not be taken either, since that could only be done after the completions of each student sheet. Fortunately, many teachers (all but four who taught the classes in the fourth year) agreed to finish the project in the school year 2020/2021. It was not easy to organize though, since the students involved in the sample did not learn chemistry as a compulsory subject after the 4th school year. Fortunately, several student sheets of the final year deal with organic chemistry and could be easily related to biochemistry and healthcare that are taught in biology in Grade 11. Tutorials or extra lessons could also be used for the implementation of the remaining worksheets. Hence, the experiments of the fourth-year student sheets continued from September 2020. Online teaching of the secondary school students in Hungary lasted from the middle of November 2020 till 10th May 2021 (due to the second and the third wave of the COVID-19 pandemic). Only after this could the project be finished. The statistical analysis of the data gained from the last test and the evaluation of the results were finished at the end of the fifth school year of the project.

At last, all the five tests (T0-T4) were filled in by 461 (N=461) students; 130 students of Group 1, 162 students of Group 2, 169 of Group 3. There was no significant difference among the groups in terms of the total scores in T0 test (Table 1). It was assumed that apart from the three types of instruction methods used during the intervention in case of the three groups, other parameters had also influenced the results. Therefore, the statistical analysis of data was accomplished again by Analysis of covariance (ANCOVA) of the SPSS Statistics software. The dependent variables were the students' scores on the whole tests, on the disciplinary content knowledge and on the experimental design skills tasks, respectively. The independent variables and the covariate are listed below.

•Independent variables (parameters - "sources"):

–Groups (3 types of instruction methods: Group 1, Group 2, Group 3);

–School ranking (3 categories: high, medium and low ranking position of the students' school in a published ranking of the secondary schools, according to the website 'legjobbiskola.hu');

–Mother's education (2 categories: mother has/has not got a degree in higher education, HE, that can be an indicator of the socioeconomic status of the student's family);

–Gender of the student (2 categories: boys, girls);

•Covariant: result of Test 0 (continuous variable);

•Dependent variables: percent of students' scores (%) in the tests:

–on the total test (TOT);

–on DCK tasks (DCK);

–and EDS tasks (EDS)

and analysed as continuous variables;

•Bonferroni correction was applied to determine whether the results are significant at $p=0.05/5=0.01$ level;

•Partial eta-squared (PES) values were calculated and used as a measure of the effect size.

Data about the students' mark in science/chemistry; their attitude toward science/chemistry; opinion about the importance of experiments in science; the degree of how much they favour doing 'step-by-step' experiments to self-designed experiments were also collected but the results are not published in the present paper.

To broaden the feedback received from teachers, and as part of the dissemination and implementation process, chemistry teachers who are not members of our research team were asked to read, and if possible, to try the student worksheets prepared in the project. Teachers of the research group were also asked to pilot the worksheets with groups of students who are not part of the sample. Although a number of trials were obviously hindered by the pandemic, many students were involved who had not been originally part of the sample. 26 chemistry teachers (19 of them had been members of our research team and 7 were not, among whom there were four young beginners involved in the project, who got to know the project as pre-service teacher students of the ELTE University) volunteered to fill in a questionnaire, giving their opinion and experiences. The data gained from the questionnaire were statistically analysed, and the answers to the open-ended questions were summarized.

4. Results

Table 1 and 2 show the results of the ANCOVA analysis of the scores in the whole test (DCK+EDS tasks together). It seems that the intervention only had a significant positive effect on Group 2 students' development in Grade 7 (T1) who got extra theoretical experimental design tasks in that year. However, the direct teaching of the principles of the experimental design accelerated both experimental groups' (Group 2 and 3) development in Grade 8 (T2). Unfortunately, the effect was temporary, since the control group caught up with the experimental groups by Grade 10 (T4). School ranking consistently had a significant effect size (PES). The mother's education (that had been chosen to represent the social variables) mainly had a significant effect on the students' achievement on Test 0. Interestingly, it disappeared by the end of the first school year. In general, gender did not influence test results significantly. However, previous knowledge (represented by the T0TOT result) had a significant effect on the scores from Grade 8.

Table 1 The average effect sizes (*PES* values) of the assumed parameters at the beginning of the research project (T0) and at the end of each school year (T1-T4) in the whole test

	Effect size (Partial Eta Squared, PES)				
	T0	**T1**	**T2**	**T3**	**T4**
Group	0.017	0.077*	0.047*	0.019	0.004
School ranking	0.049*	0.026*	0.098*	0.157*	0.050*
Mother's education	0.077*	0.000	0.000	0.005	0.019*
Gender	0.000	0.001	0.018*	0.007	0.015*
Prior knowledge ($T0_{TOT}$)	-	0.005	0.113*	0.095*	0.062*
*p<0.01					
Estimated means					
Group 1	36.0	38.1	30.8	32.3	43.3
Group 2	40.2	45.6	41.9	38.2	40.4
Group 3	40.1	32.5	39.8	37.8	40.2
Significant difference	Group 1-2 Group 1-3	Group 1-2, Group 1-3, Group 2-3	Group 1-2, Group 1-3	Group 1-2, Group 1-3	-
Low (L) ranking school	33.8	34.0	27.1	23.8	34.0
Medium (M) ranking school	40.8	41.8	40.0	41.3	43.3
High (H) ranking school	41.7	40.3	45.4	43.3	46.6
Significant difference	L-M, L-H	L-M, L-H	L-M, L-H	L-M, L-H	L-M, L-H
Mother has no degree in HE	33.7	38.2	37.3	34.5	44.9
Mother has a degree in HE	43.8	38.3	37.7	37.7	37.6
Significant difference	yes	-	-	-	yes

Tble 2 The average *PES* values of the experimental groups compared to that of the control group's and the average *PES* values of the high ranking schools compared to that of the medium and the low ranking schools at the beginning of the research project (T0) and at the end of each school year (T1-T4) in the whole test

	Effect size (Partial Eta Squared, PES)				
	T0	**T1**	**T2**	**T3**	**T4**
Group 2/Group 1	0.014	0.022*	0.043*	0.016*	0.003
Group 3/Group 1	0.014	0.013	0.031*	0.015*	0.004
High/low ranking school	0.039*	0.013	0.093*	0.129*	0.047*
High/medium ranking school	0.001	0.022	0.012	0.002	0.005

*p<0.01

The trends detected on the development of the disciplinary content knowledge (DCK) and on the experimental design skills (EDS) were similar to the ones that were observed on the total results of the tests. The results of the ANCOVA analysis of the scores on the DCK tasks are shown in Table 3 and 4, whereas the results of the scores on the EDS tasks can be seen in Table 5 and 6. According to the data, the positive effect of the intervention proved to be only temporary. Whereas, school ranking had a strong effect on students' scores on the tasks intending to measure EDS at Grade 8, an even greater effect at Grade 9, and it remained significant in Grade 10.

Table 3 The average effect sizes (*PES* values) of the assumed parameters at the beginning of the research project (T0) and at the end of each school year (T1-T4) on the DCK tasks

	Effect size (Partial Eta Squared, PES)				
	T0	**T1**	**T2**	**T3**	**T4**
Group	0.001	0.145*	0.050*	0.035*	0.006
School ranking	0.042*	0.023	0.091*	0.083*	0.046*
Mother's education	0.059*	0.013	0.001	0.023*	0.014
Gender	0.000	0.001	0.012	0.001	0.008
Prior knowledge ($T0_{DCK}$)	-	0.010	0.052*	0.044*	0.023*
*p<0.01					
Estimated means					
Group 1	55.0	44.5	34.7	29.4	46.7
Group 2	54.0	54.6	46.4	38.5	43.2
Group 3	55.1	36.1	46.9	37.5	42.8
Significant difference	-	Group 1-2, Group 1-3, Group 2-3	Group 1-2, Group 1-3	Group 1-2, Group 1-3	-
Low (L) ranking school	49.8	40.7	32.3	27.4	39.0
Medium (M) ranking school	56.5	46.9	45.8	41.0	49.6
High (H) ranking school	58.0	47.6	50.1	37.0	44.1
Significant difference	L-M, L-H	-	L-M, L-H	L-M, L-H	L-M
Mother has no degree in HE	50.0	47.8	42.0	31.5	47.2
Mother has a degree in HE	59.5	42.4	43.4	38.8	41.2
Significant difference	yes	-	-	yes	yes

Table 4 The average *PES* values of the experimental groups compared to that of the control group's and the average *PES* values of the high ranking schools compared to that of the medium and the low ranking schools at the beginning of the research project (T0) and at the end of each school year (T1-T4) on the DCK tasks

	Effect size (Partial Eta Squared, PES)				
	T0	**T1**	**T2**	**T3**	**T4**
Group 2/Group 1	0.001	0.040*	0.039*	0.031*	0.004
Group 3/Group 1	0.000	0.027*	0.041*	0.024*	0.005
High/low ranking school	0.031*	0.016	0.074*	0.029	0.007
High/medium ranking school	0.001	0.000	0.005	0.006	0.010

*p<0.01

Although Group 2 did significantly better than Group 3, there was no significant difference between Group 2 and Group 1 (control), nor between Group 3 and Group 1 (control) in terms of the experimental design skills in the end of the first year of the project. But both experimental groups seemed to have better results than the control group in the second year on the EDS tasks. However, no significant difference was detected among the groups in the experimental design skills in the third and the fourth year. In contrast, school ranking and prior knowledge both had a positive effect on the experimental design skills during the last three years.

Table 5 The average effect sizes (*PES* values) of the assumed parameters at the beginning of the research project (T0) and at the end of each school year (T1-T4) on the EDS tasks

	Effect size (Partial Eta Squared, PES)				
	T0	**T1**	**T2**	**T3**	**T4**
Group	0.043*	0.061*	0.045*	0.011	0.008
School ranking	0.036*	0.023	0.072*	0.215*	0.103*
Mother's education	0.055*	0.001	0.003	0.000	0.001
Gender	0.000	0.000	0.010	0.032*	0.008
Prior knowledge ($T0_{EDS}$)	-	0.000	0.083*	0.052*	0.068*
*p<0.01					
Estimated means					

Group 1	19.5	34.6	21.3	27.9	41.5
Group 2	27.2	41.3	34.1	33.8	36.1
Group 3	20.1	27.2	33.4	32.6	36.5
Significant difference	Group 1-2, Group 2-3	Group 2-3	Group 1-2, Group 1-3	-	-
Low (L) ranking school	17.9	29.4	19.9	14.3	26.6
Medium (M) ranking school	25.2	37.5	34.9	38.0	37.2
High (H) ranking school	23.7	36.1	34.0	41.9	50.3
Significant difference	L-M	L-M	L-M, L-H	L-M, L-H	L-M, L-H, M-H
Mother has no degree in HE	17.5	33.7	27.9	30.9	39.1
Mother has a degree in HE	27.0	35.0	31.3	32.0	37.0
Significant difference	yes	-	-	-	-

Table 6 The average *PES* values of the experimental groups compared to that of the control group's and the average *PES* values of the high ranking schools compared to that of the medium and the low ranking schools at the beginning of the research project (T0) and at the end of each school year (T1-T4) on the EDS tasks

	Effect size (Partial Eta Squared, PES)				
	T0	**T1**	**T2**	**T3**	**T4**
Group 2/Group 1	0.031*	0.012	0.038*	0.010	0.007
Group 3/Group 1	0.000	0.014	0.034*	0.007	0.006
High/low ranking school	0.015	0.010	0.040*	0.162*	0.103*
High/medium ranking school	0.001	0.001	0.000	0.004	0.037*

*p<0.01

As for the results of the statistical analysis of the online questionnaire, 22 of the 26 chemistry teachers included in the project (out of the 26, three retired after the closure of the project) wrote that they plan to use the worksheets in the future. The worksheets that the teachers liked most were given a positive feedback for the experiments that

• could be easily carried out;
• were spectacular;
• had unambiguous outcomes;
• had a thoroughly planned experimental design;
• provided the students with useful knowledge;
• had low time-demand of preparation;
• used materials and equipment that can be found in the household;
• had everyday life context;
• taught the principles of chemical analysis;
• summarised the data in tables or otherwise in a systematic way.

The difficulty that teachers most often encountered was that filling in some of the worksheets were time-consuming (did not fit into a 45 minutes lesson). As sources of further problems, teachers mentioned the followings:

• some experimental design tasks required rather complicated calculations;
• the assumption of some previous knowledge that exceeded the students existing previous knowledge;
• occasionally ambiguous experiences of some experiments;
• some worksheets that expected the teams of students in a class to do different experiments simultaneously.

Almost half of the teachers thought that the inquiry-based/experimental design method could be made even more compatible with the present conditions of teaching chemistry in Hungary by shorter and less time-consuming worksheets that are more closely related to the curriculum. Their work would be much easier if they had more time to do experiments with their students, and if there was a lab assistant in each school to help them with the preparation of the experiments (which is sadly not the case at present in Hungary). 20 of the 26 teachers would gladly take part in a new research project, when they could pilot worksheets containing shorter and better structured experimental design tasks that also point out the connections among the various parts of the knowledge described by the requirement of the chemistry curriculum in Hungary.

5. Discussion

The results showed only a weak (although significant) effect of the intervention in the first year of the project, that was due to the achievement of Group 2. In Test 1 Group 2 achieved significantly better results on DCK tasks than the control group (Group 1). This might be due to the fact that Group 2 got extra experimental design tasks on paper as homework after doing the step-by-step experiments. They obviously discussed the solutions with their teacher at the following lessons. This means that Group 2 students probably spent more time with each worksheet than that of the other two groups. The improvement of Group 3 was not significantly different to that of Group 1 in terms of experimental design. Moreover, they scored less on DCK task than Group 1. A possible explanation is that cognitive overload might have occurred in the case of Group 3 students, since designing those six experiments without scaffolding, according to their student sheets in the first year might have been too demanding for the seventh graders.

Since no significant positive effect of the intervention in terms of experiment design skills could be detected in the case of Group 3 at the end of the first school year, it was decided that "no scaffolding" did not seem to work for younger students and for a longer period of time. Therefore, the research model was changed. Consequently, from the beginning of the second school year of the project, the principles of experimental design were taught directly to the students after doing step-by-step experiments (Group 2), or before doing the same experiments partially designed by themselves (Group 3). As a result, both experimental groups showed significantly higher development on both sub-tests, measuring the disciplinary content knowledge (DCK) and the experimental design skills (EDS) than the control group in the end of the second year of the project. However, there was no significant difference found on the EDS tasks between the students of Group 2 and Group 3. This shows that it is probably worth teaching the experimental design directly in both ways.

In the third year (Test 3), the experimental groups only had a somewhat better achievement on the DCK sub-test than the control group. However, no significant difference was found among the three groups on the EDS sub-test (intended to measure the development of the experimental design skills). This was a very surprising outcome after the achievement of the second year. It might be due to the fact that by this age most students must have achieved Piaget's formal operational stage (Cole and Cole, 2006). Therefore, many of the ones involved in the project (the students of the control group as well) might have done enough experiments to be able to work out how to design an experiment in a correct way when they were asked to do so.

However, it is also possible that the motivation of the students to complete Test 3 was hugely influenced by the fact that chemistry taught according to the curriculum in grade 9 is much more abstract than the one they had learnt during the previous years, and many students might have found that really hard. In addition, most students would know by this age whether they would need chemistry for their higher education or not. If they do not have to pass a final exam in chemistry at the end of their secondary school studies (that is only needed to attend certain university courses), they might be reluctant to put a lot of energy in completing a test that does not even count in their mark in chemistry at the end of the school year.

Because of the COVID-19 disruptions, the implementations of the fourth-year student sheets and the sitting in of T4 test were only finished in the fifth year (by June 2021), under very unusual circumstances and only by 491 students, among whom 461 filled in all the 5 tests. Some conclusions might be drawn based on the results of the last test, but only with very careful consideration of those unusual circumstances. Since no significant difference was found among the achievement of the groups (neither in the aspect of DCK tasks, nor in the EDS sub-test), the other effects explained above might have overcome the effect of the intervention.

One must keep in mind that the final outcome of a longitudinal project of this size is influenced by many factors. Examination of the effect of other parameters, like mother's education and school ranking proved to be important, because they help to put the results into perspective. The effect size of school ranking did exceed the effect of the intervention in the last three years of the project. The students of the 'higher ranking' schools did better on the experimental design tasks than the students of the 'lower ranking' schools. This could be explained by the fact that the Hungarian school system is very selective, which is consequently shown by the PISA results (OECD 2007, OECD 2019). 'Higher ranking' schools had a chance to select their students from many more applicants than the 'lower ranking' schools. The influence of mother's education could also be explained by a mother's effect on the development of her young child and/or by the assumption that mothers with a degree in higher education probably encourage their children to work harder at school. One can also reasonably assume that parents with a degree in higher education provide a more favourable environment to the development of their children' cognitive skills.

6.Conclusions

Looking at the general trends in the students' progress, it can be assumed that 12-13 years old students could not generalize the experiences of the concrete experiment design tasks of the six worksheets provided during the first year of the project. Cognitive overload also might have occurred in the case of Group 3 students who were expected to design experiments without any help during that year. Although Group 2 showed significantly better results, that might have been at least partially due to the extra time they spent with the teacher-guided discussion of the experiments designed on paper. From the test results of Group 2, it may be concluded and suggested to teachers, that discussing paper-based experimental design tasks with students after they have done a step-by-step experiment is a worthwhile activity, although it is obviously time-consuming. However, asking students to design experiments, even in a scaffolded way (as with Group 3) needs careful planning to avoid cognitive overload.

As often read in literature (OECD 2005, Snook et al., 2009), our students' results had been positively influenced by their socioeconomical status, represented by the mother's education - at least at the beginning of the project. Later its influence decreased, as the school effect became stronger. It is also well-known that schools (and especially teachers) had a very strong influence on the students' achievement (Gray et al., 1986, Harker, 1996, Hattie, 2003). A feasible explanation of this effect, exceeding any other effect in our research project in the long run, is that the entrance exam required to get into the schools of the students in the sample is hard. Therefore, only the best students of their own cohort can attend these institutions in which chemistry is taught for four years in the same group (which was a required criteria of sampling in our study). The school gradually takes over, and the school variables (principally the teachers, and also the motivating atmosphere of the peers) influence the students' development most.

Our research group intends to start another research project from 1st September 2021. Just like in the present project, classes/groups of students who learn chemistry for four years in the same school is aimed to be involved. Learning from the results and the conclusions of this project described above, the cognitive load of the students at the time of designing experiments is intended to be further reduced. Since the lack of motivation also seemed to be an issue in the present project, an approach called systems thinking (of that a good summary how to use it in chemical education was published by Orgill, et al. in 2019) will be used. There is hope that students would understand the importance of learning chemistry a bit better, if it is shown them how the subjects of the worksheets fit into the 'big picture'. It is also obvious from the teachers' answers given to our questionnaire last autumn that the student sheets have to be both shorter and less time consuming. Keeping all these in mind, in the next project we try to get a bit closer again to our main goal that we set for ourselves in the longitudinal research described in this paper.

Acknowledgments

This study was funded by the Content Pedagogy Research Program of the Hungarian Academy of Sciences. (Project No.: 471026) Many thanks for all the colleagues' and students' work.

References

Allen, J. B., Barker, L. N. & Ramsden, J. H. (1986). Guided Inquiry Laboratory, *J. Chem. Educ.*, 63, 533-534.

Bruck, L. B. & Towns, M. H. (2009). Preparing Students to Benefit from Inquiry-Based Activities in the Chemistry Laboratory: *Guidelines and Suggestions, J. Chem. Educ.*, 56, 820-822.

Cheung, D. (2011). Teacher Beliefs about Implementing Guided-Inquiry Laboratory Experiments for Secondary School Chemistry, *J. Chem. Educ.*, 88, 1462–1468.

Cole, M. & Cole S. R. (2006). Fejlődéslélektan, Osiris Kiadó, Budapest, pp. 481-505; pp. 642-656., Hungarian translation of Cole M. and Cole S. R., (2001), *The Development of Children*, 4th edition, New York, Worth Publishers.

Criswell, B. (2012). Framing Inquiry in High School Chemistry: Helping Students See the Bigger Picture, *J. Chem. Educ.*, 89, 199–205.

Finlayson, O., Maciejowska, I. & Čtrnáctová, H. (2015). Inquiry Based Chemistry Instruction, In: Maciejowska, I. & Byers, B. (eds.) *A Guidebook of Good Practice for the Pre-Service Training of Chemistry Teachers* (p. 119), Faculty of Chemistry, Jagiellonian University, Krakow, http://www.ec2e2n.info/news/2015/1604_201510

Gray, J., Jesson, D. & Jones, B. (1986). Towards a framework for interpreting examination results, in: R. Rodgers (ed.), *Education and social class* (pp. 51-57). London: Falmer Press.

Harker, R. (1996). On 'First year university performance as a function of type of secondary school attended and gender.' *New Zealand Journal of Educational Studies*, 32, 197198.

Hattie, J. (2003). Teachers make a difference: What is the research evidence? Paper presented to the Australian Council for Educational Research annual conference: Building teacher quality, retrieved 13 February 2009 from http://www.leadspace.govt.nz/leadership/articles/teachers-make-a-difference.php

Hofstien, A. & Kempa, R. F. (1985). Motivating strategies in science education: attempt of an analysis. *Eur. J. Sci. Edu.*, 3, 221-229.

Minner, D. D. et al. (2010) Inquiry_based Science Instruction – What Is It and Does It Matter? Results from a Research Synthesis Years 1984 to 2002, *J. Res. Sci. Teach.*, 47, 474-496.

OECD (2005). Teachers matter: Attracting, developing and retaining effective teachers. Overview. *Paris: OECD*, p.2., http://www.oecd.org/dataoecd/39/47/34990905.pdf

OECD (2007) PISA 2006: *Science Competences for Tomorrow's World*, Volume 1: Analysis, 63.

Schleicher, A. (2019). PISA 2018, Insights and Interpretations, OECD, https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf

OECD (2019), PISA 2018 Results (Volume I): What Students Know and Can Do, *PISA, OECD Publishing*, Paris, https://doi.org/10.1787/5f07c754-en

Orgill, M. K., York, S. & MacKellar, J. (2019). Introduction to Systems Thinking for the Chemistry Education Community, *Journal of Chemical Education*, 96, 2720-2729.

Rocard, M. (2007). Science Education NOW: A Renewed Pedagogy for the Future of Europe. Brussels: European Commision. *Directorate-General for Research*.

Snook I., O'Neil J., Clark J., O'Neil A. & Opneshaw R. (2009). Invisible Learnings: A commentary on John Hattie's book Visible learning: A synthesis of over 800 metaanalyses relating to achievement, *New Zealand Journal of Educational Studies* 44(1):93-106.

Szalay, L. & Tóth, Z. (2016). An inquiry-based approach of traditional 'step-by-step' experiments, *Chemistry Education Research and Practice*, 17, 923-961.

Szalay, L., Tóth, Z. & Kiss, E. (2020). Introducing students to experimental design skills, *Chemistry Education Research and Practice*, 21, 331-356.

Szalay, L., Tóth, Z. & Borbás, R. (2021). Teaching of experimental design skills: results from a longitudinal study, *Chemistry Education Research and Practice*, 2021, DOI: 10.1039/D0RP00338G

Taber, K.S. (2011). Inquiry teaching, constructivist instruction and effective pedagogy. *Teacher Development*, 15(2), pp.257–264.

Tomperi, P. & Aksela, M. (2014). In-service Teacher Training Project On Inquiry Based Practical Chemistry. *LUMAT*, 2(2), 2015-2