Journal of Turkish Science Education, 2025, *22*(3), 584-611.
DOI no: 10.36681/tused.2025.030

Journal of Turkish Science Education
http://www.tused.org
© ISSN: 1304-6020

Effect of guided inquiry on junior high school learners' experimental design skills in solving chemistry problems

Luca Szalay[1], Zoltán Tóth[2], Réka Borbás[3], István Füzesi[4]

[1]*Eötvös Loránd University, MTA-ELTE Research Group on Inquiry-Based Chemistry Education, Research Programme for Public Education Development of the Hungarian Academy of Sciences, Hungary, Corresponding author, luca.szalay@ttk.elte.hu, ORCID ID: 0000-0003-0176-0645*
[2]*University of Debrecen, Hungary, ORCID ID: 0009-0000-4806-6840*
[3]*Szent István Secondary School, Hungary, ORCID ID: 0000-0002-9671-087X*
[4]*Eötvös Loránd University, Bolyai János Practising School, Hungary, ORCID ID: 0000-0003-4826-1819*

ABSTRACT
The main objective of this four-year empirical research project has been to develop the experimental design skills of 12-16 year-old Hungarian pupils in Grades 7–10 of junior high school through guided inquiry, using six chemistry experiment worksheets each school year. Group 1 (control group) and Group 2 follow step-by-step instructions, but after the experiment, Group 2 also answers a series of questions about the design of the experiment. Group 3 is required to design the experiments, guided by the series of questions. The impact of the intervention on 603 pupils' experimental design skills and disciplinary content knowledge were measured by structured tests at the beginning of the project and at the end of three school years. By the end of the third year the intervention only had a small positive effect on the development of experimental design skills in Group 3 (Cohen's d: 0.16), while the development of disciplinary content knowledge was slightly negatively affected for both experimental groups (Cohen's d for group 2: -0.07; Cohen's d for group 3: -0.19).

RESEARCH ARTICLE

ARTICLE INFORMATION
Received:
27.03.2025
Accepted:
18.07.2025
Available Online:
29.09.2025

KEYWORDS:
Experimental design skills, guided inquiry, disciplinary content knowledge.

To cite this article: Szalay, L., Tóth, Z., Borbás, R. & Füzesi, I. (2025). Effect of guided inquiry on junior high school learners' experimental design skills in solving chemistry problems. *Journal of Turkish Science Education, 22*(3), 584-611.

Introduction

Models in Inquiry-based Science Education

Inquiry-based teaching strategies are generally considered to be inductive in nature. They are characterised by being question- and learner-centred and use practical activities to provide data from which to develop concepts, while in the more traditional deductive instructional laboratories the role of these is to verify a concept previously presented (Bunce & Cole, 2007, p. 59). Piaget's functioning model has been used as a rationale for inquiry-based instructional strategies (Lawson et al., 1989). According to this model, if there is an incompatibility between the assimilated information and the learner's existing mental structure, either the mental structure or the perception of the information must change (Piaget, 1963). One model of inquiry education influenced by Piaget's functioning model is the learning cycle approach (Karplus & Thier, 1967), when the learner is exposed to the information, followed by activities that help to accommodate and then organise the information. Another model of inquiry that is consistent with the learning cycle approach is the Process Oriented Guided Inquiry learning (POGIL) that uses key scientific processes to develop scientific concepts (Farell et al., 1999), while pupils solve scientific problems in cooperative learning groups. Science Writing Heuristic (Burke et al., 2006) and the Argument Driven Inquiry (Walker et al., 2013) are other examples of models that emphasise asking scientific questions, designing appropriate procedures to test those questions, supporting conclusions with experimental evidence, and communicating ideas clearly (Reynders et al., 2019).

As early as 1989, there was considerable evidence that inquiry-based laboratory activities had real educational advantages over more traditional approaches (Lawson et al., 1989). However, according to a 1997 survey in the United States, fewer than 10% of college chemistry laboratories used this approach (Abraham et al., 1997). Following the recommendations of the Rocard Report (Rocard et al., 2007), the European Commission's 6th and 7th Framework Programmes supported several international projects on inquiry-based methods to stimulate interest in science taught in schools. As a result of that, and the work of many other educational researchers and school science teachers around the world, resources for inquiry-based teaching and learning are freely available online (e.g. McLoughlin et al., 2015). Training programmes for science teachers in inquiry-based education have also been developed (e.g. Bolte et al., 2012). Successful implementation of teaching/learning materials developed by science teachers and piloted in classrooms with positive feedback from pupils have been reported (e.g. Dumitrescu et al., 2014), resulting in improved achievement, even for Grade 8 pupils (Savec & Devetak, 2013).

There have been many other studies on the benefits of using inquiry-based learning. For example, the review of research on POGIL by Rodriguez et al. (2020) claimed that it is more effective in supporting learning than lecture-based approaches. More recently a successful adaptation of a guided inquiry laboratory experiment for undergraduate food chemistry students was reported by Rodríguez-Berrios and Rodríguez-Vargas (2025). Yet, according to Restucia et al. (2018) many science, technology, engineering, and mathematics (STEM) undergraduate students finish their studies with inadequate experimental design skills (EDS). This is despite the fact that curriculum based solely on recipe-like experiments does not provide learners with the opportunity to evaluate experimental techniques, design and plan experimental protocols (Domin, 1999; Fay et al., 2007).

Matthews (2018) points out that learners can gain meaningful insights into the construction of scientific knowledge through processes of inquiry, reasoning and planning only if they are properly organised and reflected upon. Given the limited prior knowledge of pupils, the gradual development of EDS at school level is even more challenging than in university courses. Negative beliefs about science and practical work, difficulties in planning such activities, persuading pupils to reflect on their experiences and outcomes, and concerns about evaluating practical inquiry also hinder the use of these pedagogical strategies (Akuma & Callaghan, 2019). In addition, during practical work, cognitive overload of the pupil's working memory can easily occur (Johnstone & Wham, 1982). Therefore, it is worth focusing on identifying independent variables, measuring dependent variables, controlling for confounding variables and recognising the relationship between them, which is the essence of experimental design (Arnold et al., 2018; Arnold et al., 2021; Cannady et al., 2019; Pedaste et al., 2015). Lazonder and Harmsen's (2016) meta-analysis of studies on guided inquiry instructions concluded that more specific instruction promotes better quality learning outcomes.

Therefore, pupils coming from a traditional learning style should be provided with adequate and appropriate scaffolding to successfully solve an inquiry-based task (Seery et al., 2019) to reduce cognitive load and student resistance, and increase their chances of successfully answering the question asked at the beginning of the guided inquiry task. One such tool is the Experimental Design Diagram developed by Julia Cothron and colleagues (2000), based on the 'fair testing' method (i.e. 'how to vary one thing at a time' or 'holding other things/variables constant'). This consists of a series of questions to help students identify and control variables, formulate a hypothesis and design the experiment. A prompt should also be chosen to arouse the learners' interest and curiosity (Cothron et al., 2000). The systematic use of this worksheet structure to do experiments is expected not only to help the learners carry out a specific experiment, but also to provide them with an opportunity to understand and generalise the process of designing experiments. In this way, puplis can acquire epistemic knowledge about scientific experimentation, which refers to how it generates knowledge and why this knowledge is justified (Lee et al., 2024).

Objectives and Research Questions (RQ) of the Present Study

The main objective of the present four-year empirical research project has been to develop the experimental design skills (EDS) of pupils in Grades 7-10 through guided inquiry. Experimental design activities are recommended in the National Curriculum of Hungary (2020) for the development of scientific literacy. However, the number of tried and tested teaching materials for the development of EDS is limited. Some of those had been prepared and used by the research team in previous studies (Szalay & Tóth, 2016; Szalay et al., 2020; Szalay et al., 2021). As with a previous longitudinal project, in the current longitudinal study, the researchers want to influence how a large sample of pupils perform experiments in schools during all four years of compulsory chemistry lessons. However, in contrast to that previous project, in the present project pupils in the experimental groups are given help in learning how to design experiments in the form of a series of questions similar to the Experimental Design Diagram developed by Julia Cothron and colleagues (2000). These questions focus on identifying independent variables, measuring dependent variables, identifying the relationship between them and controlling for confounding variables. Also of interest is whether the intervention has any impact on learners' disciplinary content knowledge (DCK). This type of instruction used had a significant positive effect on the EDS of the Grade 7 experimental group who were asked to answer these series of questions before they designed the steps of their experiments (Szalay et al., 2023), but not for the other experimental group, who performed the experiments in the same way as the control group, step-by-step following a recipe and then answered the series of questions. Although the trend changed in the second year, the intervention still resulted in a positive change in EDS for the experimental group that designed their own experiments according to the series of questions (Szalay et al., 2024). Therefore, in the third school year of the present project, the following research questions were posed in order to find answers to the long-term effects of the methods used.

RQ1: Did the intervention result in a significant change in pupils' ability to design experiments (EDS) by the end of the third year of the present project in either of the experimental groups compared to the control group?

RQ2: Did the pupils in the experimental groups score significantly differently on the disciplinary content knowledge (DCK) questions because of the intervention compared to the pupils in the control group by the end of the third year of the present project?

RQ3: Was there a difference in the development of the EDS between pupils in the two experimental groups by the end of the third year of the present project?

Methods

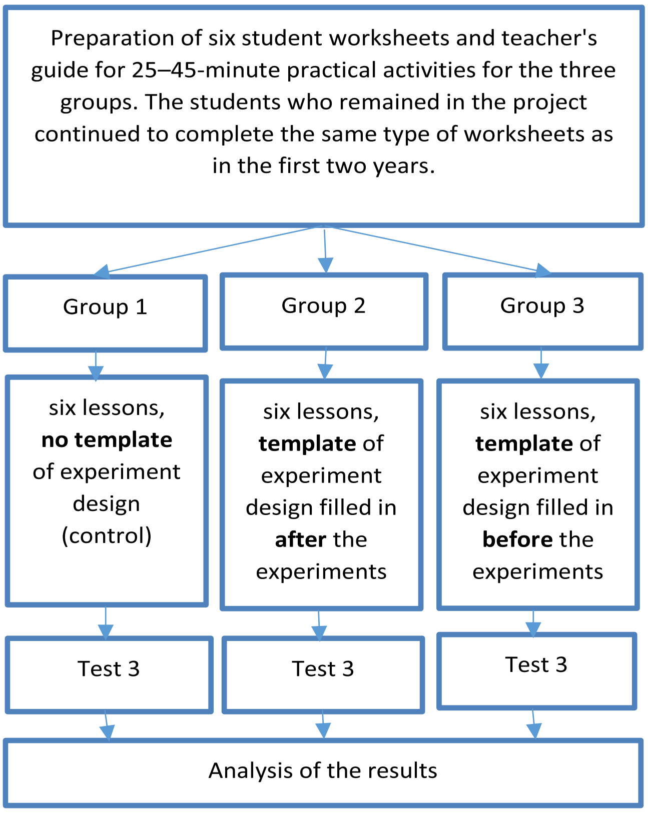
Research Design and Participants

The research model has a quasi-experimental design with a non-equivalent control group. Nine hundred thirty-one seventh-grade pupils completed a test (called Test 0, T0) in September 2021. They were enrolled in 38 classes in 25 Hungarian schools. After evaluating the results of Test 0, the classes were divided into Groups 1, 2 and 3, with no significant difference between the groups in terms of either the mean initial performance of the pupils on DCK and EDS tasks or the hypothesised parameters (school ranking, mothers' education, gender). The pupils remain in the same group they were in at the start of the project for four years. Therefore, only pupils who stay in the same school from Grade 7 to Grade 10 (from age 12–13 to age 16–17), were selected to participate in this project. Thus, the full four years of their compulsory chemistry studies can be influenced by the worksheets provided by the research group and their progress could be followed by the tests completed each year [Test 1 (T1), Test 2 (T2) and Test 3 (T3) at the end of the first, second and third school year, respectively)]. If a pupil has missed a test for any reason, their results are no longer taken into account. Therefore, 603 tests were evaluated at the end of the third school year.

Group 1 is the control group, which follows step-by-step instructions to carry out the experiments, what Banchi and Bell (2008) called a 'structured inquiry'. Groups 2 and 3 are the experimental groups. Group 2 follows the same step-by-step instructions as Group 1, but after completing the experiment they also answer a series of questions about the identification of the independent variable, the dependent variable and the constants on their worksheet. Their answers are discussed with their teachers. This teaching method can be seen as a simplified version of the approach used by Reynders et al. (2019), who assigned the relevant questions to the laboratory report during the discussion of the experimental procedure. The treatment of Group 3 can be called guided inquiry (Schoffstall & Gaddis, 2007), as the question to be investigated is given on their worksheet and learners have to design an experiment to find the answer. The set of questions (a 'template') on their worksheets can be considered as a shortened and simplified version of the Experiment Design Diagram based on the 'fair testing' method developed by Julia Cothron and colleagues (2000). Answering the series of questions concerning the identification of the independent variable, the dependent variable and the constants before the experiment helps these pupils to control variables, discuss the hypotheses and then define the steps of the experiment. The answers are discussed with Group 3 by their teachers before the experiments are carried out. The two experimental groups (Group 2 and Group 3) were treated differently to see if this set of questions really helps to develop learners' experiment design skills, which would be more effective: to use it after the step-by-step experiment (for Group 2) or to use it to plan the steps of the experiment before pupils carry out the experiment (for Group 3)? The research group prepared six worksheets for each school year, in three versions for Groups 1, 2 and 3. Teachers chose the lessons in which they used the worksheets and tests provided by the research team. The research model of the quasi-experimental design with a non-equivalent control group used in the third year of this project is summarised in Figure 1. (The set of questions to help develop experimental design skills is referred to as a 'template' for simplicity.)

Figure 1

Research procedure applied in the third school year of the current project



Thirty-four in-service chemistry teachers and five university chemistry lecturers were involved in the four-year project in September 2021. The participation of teachers was voluntary. Unfortunately, by the end of the third year, five of the teachers who worked with the participating pupils in the first year had stopped teaching or simply did not have enough time to participate in the research group anymore, so their pupils are no longer included in the sample. Three of the teachers who originally taught the participating pupils were replaced by other teachers for reasons beyond the control of the research team.

The teachers were in direct contact with the research team leader by e-mail. They asked for advice when needed, despite the very detailed instructions given in the worksheets and teacher's guides. Teachers emailed feedback to the research team leader on the piloting of the worksheets and photos of their pupils doing the experiments. The photos are available on the research team's website (MTA-ELTE Research Group, n.d.). Each school year, teachers were also asked to complete a questionnaire on the testing of the six worksheets from the previous school year (Vélemények az MTA-ELTE, n.d.). They were asked about their experiences with the piloting. Analysis of the questionnaire results showed that pupils in the experimental groups were typically able to answer the series of questions on experimental design with little or no help from their teachers.

Ethical Considerations

The research team had to develop its own ethical protocol to ensure informed consent and to protect the privacy and confidentiality of individuals (Lawrie et al., 2021), as no local procedure was available. The process was described in a paper on the results of the first year of this four-year project (Szalay et al., 2023).

Data Collection Tools

Worksheets

Each worksheet and accompanying teacher's guide describes chemistry experiments that take about 25-45 minutes. The topics of the worksheets were linked to the National Core Curriculum of Hungary (2020) and agreed with the participating teachers. Tables containing the topics and the corresponding experimental design tasks of the first two years' worksheets have already been published (Szalay et al., 2023; Szalay et al., 2024) and see Table A1.a – b in Appendix for those used in the third year. Three versions of the first 18 worksheets (for Groups 1, 2 and 3) prepared for the first three year of the project are available in English on the project website (MTA-ELTE Research Group, n.d.). The worksheets were used with pupils working in small teams.

From the second year onwards, the teacher's guides for the Group 1 worksheets include a request for the teacher to draw pupils' attention to the importance of experiments in science. In the experimental groups worksheets it is emphasised that evidence in science is collected through well-designed experiments and that understanding how to design an experiment can be useful in debunking pseudoscientific fraud.

All worksheets include a context-based task with elements of systems thinking under the heading "Let's think!", with the primary goal of maintaining interest (e.g., Chen & Xiao, 2021; del Mar López-Fernández et al., 2022; Klemeš, et al., 2021; MacDonald et al., 2022). These are identical across all three groups' worksheets. Although systems thinking and contextual pedagogy share common features, they also have important differences (York & Orgill, 2020). Systems thinking allows learners to see the system as more than the sum of its parts (Reid & Amanat Ali, 2020). Incorporating more aspects of systems thinking into chemistry education can have other clear benefits for pupil learning by developing more higher order thinking skills (Szozda et al, 2022).

The first versions of each worksheet were prepared by the participating teachers and the leader of the research group and then read by the five university lecturer members of the research group, who are experts in the development of chemistry teaching materials for school pupils. The authors then improved the worksheets based on the experts' suggestions. This second version was proof-read by one of the experts and the leader of the research group and the final changes were agreed. The worksheets used in the third school year of this project can be downloaded from the research team's website (MTA-ELTE Research Group, n.d.).

Tests

According to Vorholzer et al (2020), measuring pupils' ability to design experiments is critical in science teaching. Assessment of laboratory work should focus on developing higher-order competencies such as experimental design and understanding of epistemic practice. Since there is concern that the amount of DCK acquired during the learning process may be reduced by the time spent with the EDS tasks, the development of DCK should also be followed by tests. Therefore, both EDS, as part of the inquiry skills, and DCK had to be assessed (e.g. Cooper, 2013; Reed & Holme, 2014; Rodriguez & Towns, 2018; Underwood et al., 2018). In order to monitor the progress of the pupils' performance, each test in this project consisted of different test questions based on knowledge and skills that the learners had previously acquired by completing the worksheets provided by the reseach group. However, the questions of all four tests (Test 0, Test 1, Test 2 and Test 3) were structured in the same way and categorised according to the levels of the dimensions of cognitive processes according to the revised Bloom's taxonomy (Bloom et al., 1956; Krathwohl, 2002). Each test contained eighteen compulsory items, each worth 1 point. Nine items were used to assess EDS and the other nine for the DCK (three for recall, three for understanding and three for application).

When designing the EDS tasks, different evaluation criteria (Sirum & Humburg, 2011), evaluation tools (Chen et al., 2019; Tseng et al., 2022) and the Science Olympiad (2020) experimental design checklist were consulted for guidance. This led to the conclusion that experimental design tasks should require the identification of independent, dependent and controlled variables and the use of fair testing (Csíkos et al., 2016). The experimental design tasks should include the content knowledge needed to solve the tasks (Cannady et al., 2019). The tasks were placed in the context of everyday life, to capture the pupils' interest. The following experimental design task was used in Test 3 to compare the development of pupils' EDS in the three groups.

The 'bath bomb' is a ball made of citric acid, baking soda, starch and cocoa butter with a few drops of essential oil and colouring. It's fun to use, because it bubbles up fragrantly when added to the bath water. In water, the reaction of baking soda and citric acid produces carbon dioxide gas. Let's say you want to make lots of these bath bombs as gifts for your friends. According to a source on the Internet, 50 g of citric acid requires 100 g of baking soda. However, based on the reaction equation and the molar masses, 192 g of citric acid reacts with 252 g of baking soda. You don't want to waste the materials (which cost money), so you decide ***to experiment by reacting small amounts of citric acid and baking soda to see which mass ratio*** *produces more gas. (This can be compared, for example, to reacting the materials in bottles, with a balloon placed over the mouth of the bottle after adding water to the substances.)*

a) What would you change during the experiments?

b) Which of the products of the chemical reaction that is important for bubbling depends on the change you cause?

c) How could you test the amount of the reaction product named in (b)?

d) How would you decide which mass ratio to use?

e) Why is it always important to shake the contents of the bottles thoroughly?

f) Put a **+** *sign* ***in front of the statement in the list below if you think it is true****, and a* **-** *sign* ***in front of the statement if you think it is not true****. (You can write another sign after a clear cross-out if you change your mind.)*

☐ *The same mass of citric acid should be used in each experiment.*

☐ *The same mass of baking soda should be used in each experiment.*

☐ *The same volume of bottle should be used in each experiment.*

☐ *The bottles used for each experiment should be made of the same material.*

Note: pupils may choose either baking soda or citric acid as the independent variable in their answer to question (a), but their answers in part (f) of the task must be consistent with their choice.

Tests 0, 1, 2 and 3 were 40-minute paper-based tests each. The completed tests were coded by the teachers so that they knew the names and genders of the pupils, but the researchers only received anonymous data for statistical analysis. The pupil codes will remain the same throughout the four years of the project. Each pupil's code is entered in a row in an Excel spreadsheet, and the same row is used to show the scores they received for specific items in each test. The participating teachers corrected the tests and recorded the marks in the Excel spreadsheet as instructed, see individual instructions for Tests 0, 1, 2 and 3 on the research team's website (MTA-ELTE Research Group, n.d.).

Validity

The five university lecturers (as members of the research team) and the experienced chemistry teacher, who reviewed all the teachers' corrections and made changes where necessary, formed a panel of experts. They provided the evidence of content validity, as they were able to judge whether the items adequately sampled the domain of interest (Crocker & Algina, 2006). This measure worked against construct underrepresentation, which is a major threat to construct validity (Wren & Barbera, 2013). In order to avoid construct-irrelevant variance, the tasks in each test could be solved after the pupils had completed the tasks on the previously provided worksheets. Table 1 shows how each task in Test 3 can be matched to the corresponding content of each worksheet.

The tasks in each test should differ from previous tests for two reasons. Firstly, the measurement of transferable EDS, especially far transfer, which is needed when learners have to use these skills in a completely different context. The other reason is to avoid repetitive testing effects (Cannady et al, 2019; Schafer et al, 2023). The chances of successfully solving a task would have been higher if it had been applied a second time, because it could have caused construct-irrelevant easiness.

Table 1

Matching the content of the tasks of test 3 and the topic(s) of the worksheets for the third year

No. of task in Test 3	No. and title of worksheet and topic
1. a	Worksheet 17: Sour as vinegar (acid base reactions)
1. b	Worksheet 15: The superglue and others (rates of chemical reactions)
2. a-f	Worksheet 17: Sour as vinegar (acid base reactions)
3.	Worksheet 16: Geyser in a bottle – the Mentos-Cola story (chemical equilibrium)
4.a-b	Worksheet 18: Hydrogen peroxide as a 'miracle cure'? (redox reactions)
5.	Worksheet 14: Can you walk on water? (intermolecular forces)
6.	Worksheet 18: Hydrogen peroxide as a 'miracle cure'? (redox reactions)
7.	Worksheet 13: Exploding colours (excitation energy and flame tests)

The following item evaluation and review process was carried out for all the tasks in each test. The first version of the test and its marking instructions were prepared by the research team leader. Then, the same university lecturers of the research team who had checked the content of the worksheets checked the test tasks and the marking instructions. Modifications were made on the basis of their suggestions. Expert feedback on item content, wording, and consensus of the correct answer are all sources for evidence of expert response process validity and against construct-irrelevant variance (Wren & Barbera, 2013).
Each of the tests was piloted with two different groups of pupils who were of a similar age to the sample but who did not participate in the research. Test 3 was piloted in two classes (total N=60) with pupils aged 14-15. The chemistry teachers who organised the pilot test and corrected each pilot test suggested improvements to the wording of the tasks and the correction instructions based on their experience. The tests and their marking instructions were further improved based on the results of the pilot before they were completed by the pupils in the sample. Participating teachers had not seen any of the tests to be taken at the end of the school year before trying out all six of the worksheets. This was to ensure that the test items did not even subconsciously influence teachers' teaching behaviour, as this could have affected pupils' responses to the test questions. The test scores of Groups 2 and 3 were compared with those of Group 1 (control group) to exclude the risk of maturation (Shadish et al., 2002)

Similarly to the procedure used by Goodey and Talgar (2016), the research team tried to standardise the marking in order to ensure that the marking key is applied in the same way for all the same types of corrected test. An experienced chemistry teacher reviewed all the teachers' corrections and suggested changes to the marking instructions. After discussions within the research group, changes were made. On the basis of these, some scores given by the teachers were modified to apply a uniform evaluation process, free from individual teacher's decisions. The scoring process is consistent with the recommendation that full consensus should be reached through negotiated agreement (Watts & Finkenstaedt-Quinn, 2021).

Data Analysis

Data Collection

The number of pupils (*N*) completing all four tests (Test 0, Test 1, Test 2 and Test 3) in each group is as follows: Group 1: 163; Group 2: 224; Group 3: 216, altogether 603 (298 boys and 305 girls).
The following data were collected and analysed:

- Total scores for Test 0, Test 1, Test 2 and Test 3.
- Scores for EDS tasks for Test 0, Test 1, Test 2 and Test 3.
- Scores for DCK tasks for Test 0, Test 1, Test 2 and Test 3.
- Gender.
- School ranking. The pupil's school ranking amongst Hungarian secondary schools, according to the website (Legjobbiskola, n.d.) The ranking is based on the results of the school-leaving examinations and competency tests published by the national Education Office. The participating schools were grouped into high, medium and low-ranking categories and a categorical variable was used according to these three levels.
- Mother's education. Two categories were created according to whether or not the pupil's mother (or guardian) had a degree in higher education. This categorical variable was intended to characterise the socio-economic status of the pupil. (The mother's education was chosen because most children spend much more time with their mother than their father in the first few years of their lives. Furthermore, in the event of a divorce, children tend to live with their mother in this country.)

Statistical Methods

Chi-square tests were used to check that there were no statistically significant differences between the three groups in terms of either prior knowledge measured by Test 0 or hypothesized parameters (school ranking, mother's education, gender). Cronbach's alpha values for the four tests were acceptable: 0.740 for Test 0, 0.678 for Test 1, 0.689 for Test 2 and 0.743 for Test 3.
Statistical analysis of the data was performed using SPSS Statistics software. ANOVA and ANCOVA were performed, as ANCOVA can be used to adjust for the initial difference and reflect the effect on the dependent variable (Howell, 2012). The raw mean scores (before ANCOVA analysis) and their standard deviation (SD) of the three groups were calculated for all four tests (Test 0, Test 1, Test 2 and Test 3) in the whole test ('TOTAL'), the DCK tasks and the EDS tasks.

The effect of the intervention on the development of the experimental groups (Groups 2 and 3) was measured by the Cohen's d effect size (Cohen, 1988). These values were calculated by taking the means and standard deviations of the four differences between the four test scores (T1 - T0, T2 - T1, T3 - T2 and T3 - T0). It was assumed that, in addition to the three types of instruction methods used in the intervention for the three groups, other parameters (school ranking, mother's education, gender) and a covariate (prior knowledge, i.e., scores on the Test 0) also influenced the results. In the ANCOVA analysis, the effect sizes of these parameters and the covariate were characterized by the calculated Partial Eta Squared (PES) values. The significance value of $p<0.05$ was used to test for differences between groups. For multiple comparisons, a Bonferroni correction was applied and a significance value of $p<0.013$ was used when comparing the results of four tests.

Results

According to the chi-squared test, there is no significant difference in the composition of the groups with respect to mother's education [X^2 (2, *N* = 603) = 2.234, *p* = 0.327)] and gender [X^2 (2, *N* = 603) = 1.216, *p* = 0.545)]. However, there is a significant difference in the composition of the groups with respect to school ranking [X^2 (4, *N* = 603) =39.74, *p* = 0.000)], since the difference is significant between Group 1 and Group 2 [X^2 (2, *N* = 387) =17.45, *p* = 0.000)], Group 1 and Group 3 [X^2 (2, *N* = 466) =25.41, *p* = 0.000)] and Group 2 and Group 3 [X^2 (2, *N* = 445) =9.11, *p* = 0.011)]. The number of pupils in low, medium and high-ranking schools in each group is shown in Table A2. ANCOVA analysis was used to address this problem.

Answers to the Research Questions

Tables A3–A6 show the raw mean scores, prior to ANCOVA analysis, and their standard deviations (SD) for the three groups for the T0 – T3 tests for the whole test ('TOTAL'), the DCK tasks, the EDS tasks and the results of the ANOVA analysis. It revealed no significant difference between groups in the performance of either $T0_{TOTAL}$ or $T0_{DCK}$ or $T0_{EDS}$. Group 3's performance at the end of the first year of the project (Grade 7) exceeded that of the other two groups on the $T1_{TOTAL}$ and $T1_{EDS}$ tasks, while Group 2 performed significantly worse in $T1_{DCK}$ than the control group (Group 1) and the other experimental group (Group 3). No significant difference was found between the three groups in the sample between the T2 test ($T2_{TOTAL}$) and its subtests ($T2_{DCK}$ and $T2_{EDS}$). Group 3 again outperformed the other two groups at the end of the third year of the project (Grade 9) in the EDS tasks ($T3_{EDS}$). There was no significant difference found in $T3_{TOTAL}$ and $T3_{DCK}$ among the three groups in the end of the third school year.

For further analysis, the dependent variables were the differences between test scores. Based on the means and standard deviations of the differences between the four test scores (T1 – T0; T2 – T1, T3 – T2 and T3 – T0), Cohen's d effect size values were calculated that are presented in Figure 2 for the whole test ('TOTAL'), in Figure 3 for the DCK tasks and in Figure 4 for the EDS tasks.

These results also show that Group 3 made better progress in the EDS tasks than the control group in the first and third years of the project, while Group 2 made less progress in the EDS tasks than the control group during these years. The second year shows a reverse trend in this respect. However, taking the full three-year period into account, Group 3's Cohen's d effect size in the EDS is positive, while that of Group 2's is negative. In the DCK tasks, Group 2 performed worse than the control group in the first year and better in the second year. Group 3 made less progress in the DCK tasks than Group 1 in the third year. Over the entire three-year period, both experimental groups had a negative Cohen's d effect size in DCK compared to the control group.

Figure 2

The Cohen's d effect size values calculated from the means and standard deviations of the differences between the test scores (T1 – T0; T2 – T1, T3 – T2 and T3 – T0) for the whole test (N=603)

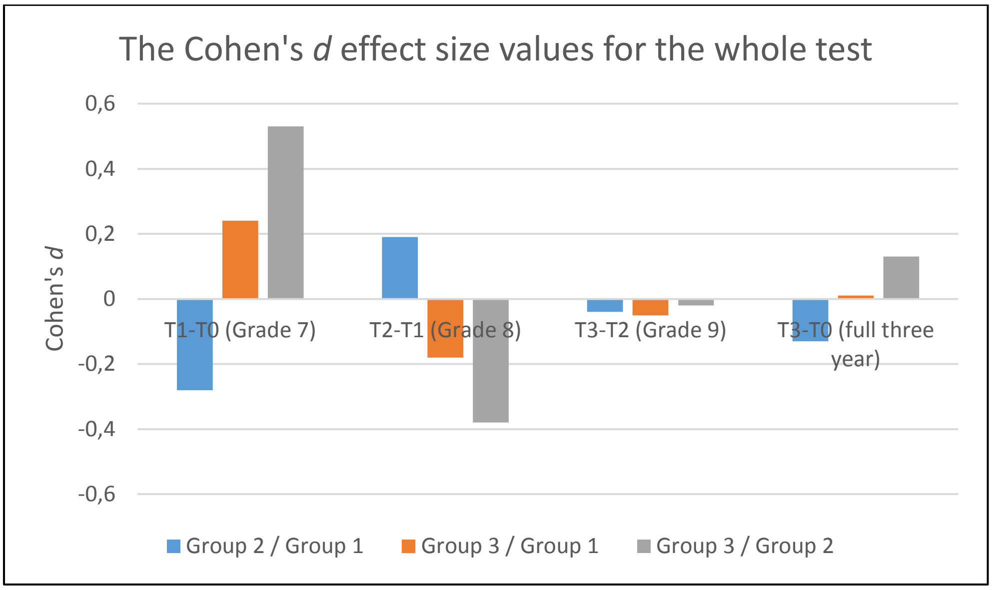
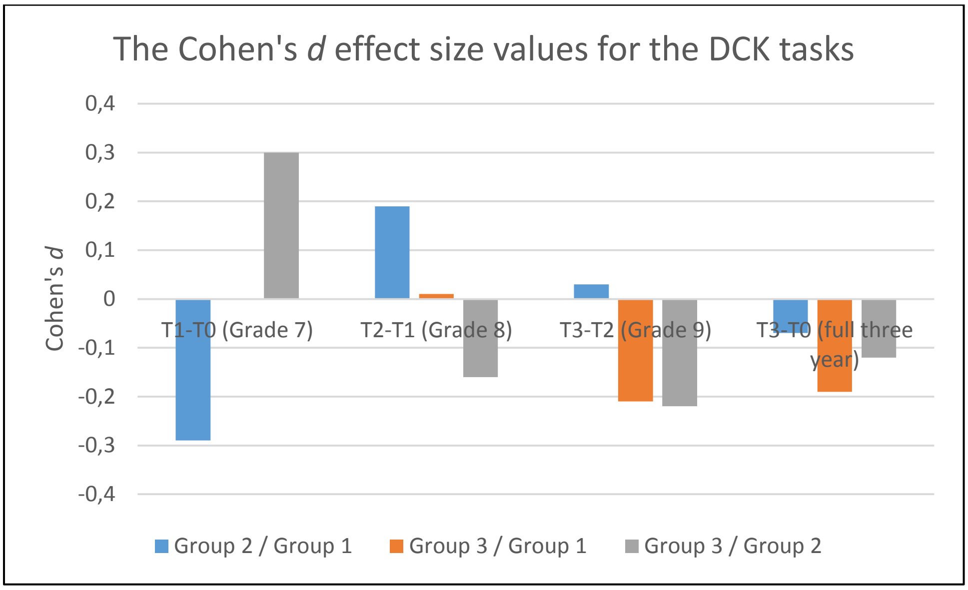


Figure 3

The Cohen's d effect size values calculated from the means and standard deviations of the differences between the test scores (T1 – T0; T2 – T1, T3 – T2 and T3 – T0) for the DCK tasks (N=603)



Since it was assumed that pupils' test scores were influenced by parameters other than the intervention, an ANCOVA analysis was conducted with test scores as the dependent variable. Group (the type of instruction methods), school ranking, mother's education and gender were the parameters. The covariate was prior knowledge (T0 test scores). This was also necessary because, once the sample was reduced by the number of pupils who missed any of the tests, the composition of the groups showed a significant difference in terms of school ranking. The Partial Eta Squared (PES) values for each parameter and the covariate effect sizes are shown in Table 2–4.

Figure 4

The Cohen's d effect size values calculated from the means and standard deviations of the differences between the test scores (T1 – T0; T2 – T1, T3 – T2 and T3 – T0) for the EDS tasks (N=603)

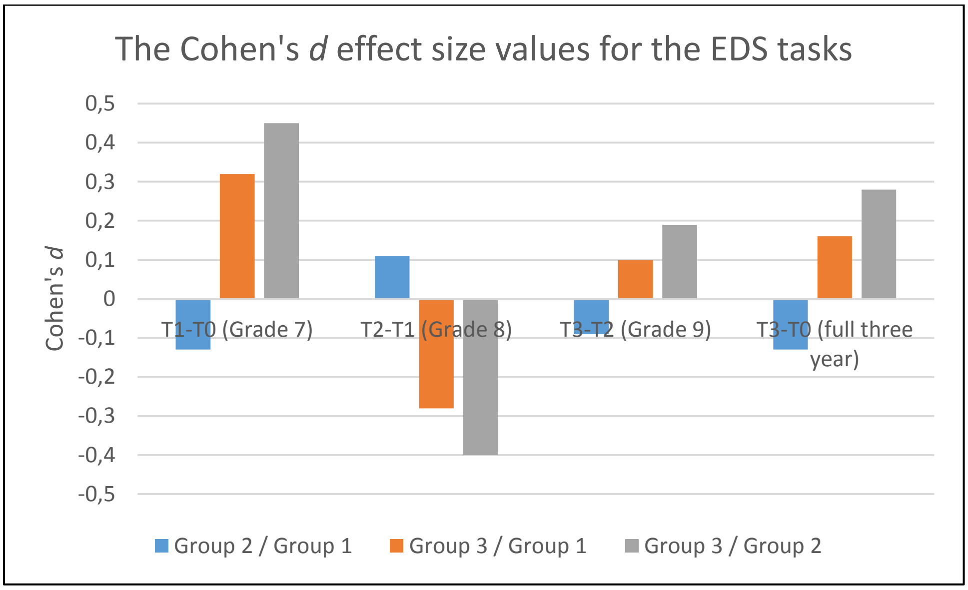


Table 2

The effects of the assumed parameters and the covariate (Prior knowledge, $T0_{TOTAL}$) on the changes for the whole test ('TOTAL') in the beginning of the project (T0) and in the end of Grade 7 (T1), Grade 8 (T2), Grade 9 (T3) (N=603)

Parameter (Source)	PES (Partial Eta Squared) $T0_{TOTAL}$	$T1_{TOTAL}$	$T2_{TOTAL}$	$T3_{TOTAL}$
Group	0.002	0.059*	0.004	0.002
School ranking	0.083*	0.052*	0.012	0.022*
Mother's education	0.008	0.002	0.000	0.000
Gender	0.003	0.004	0.000	0.001
Prior knowledge ($T0_{TOTAL}$)	-	0.132*	0.123*	0.089*

Note: * Significant at $p < 0.013$ level (Bonferroni correction)

Initially, as published earlier (Szalay et al., 2023), it was mainly the school ranking and, to a lesser extent, in the DCK tasks, the mother's education that had a significant effect on the scores. In the first year (in the Test 1), significant effect sizes (PES) were found for three parameters after the intervention for changes in the whole test and DCK tasks: type of intervention ('Group'), school ranking and prior knowledge. Of these, prior knowledge had the largest effect size in the whole test and school ranking in the DCK tasks. In the first year, only the type of intervention and prior knowledge seemed to have a significant effect on performance in the EDS tasks. Based on the results of the present analyses, it appears that at the end of the second school year (in the Test 2), only prior knowledge had a significant effect on changes in the whole test and both subtests, but school ranking also had a significant effect size in the EDS tasks. School ranking and prior knowledge influenced pupils' performance in the whole test and in both subtests in the third year of the study. However, the intervention also had a (barely) significant effect size in the EDS tasks.

Table 3

The effects of the assumed parameters and the covariate (Prior knowledge, $T0_{DCK}$) on the changes for the DCK tasks in the beginning of the project (T0) and in the end of Grade 7 (T1), Grade 8 (T2), Grade 9 (T3) (N=603)

Parameter	PES (Partial Eta Squared)			
(Source)	$T0_{DCK}$	$T1_{DCK}$	$T2_{DCK}$	$T3_{DCK}$
Group	0.000	0.035*	0.002	0.012
School ranking	0.022*	0.109*	0.002	0.034*
Mother's education	0.019*	0.000	0.002	0.000
Gender	0.007	0.002	0.001	0.002
Prior knowledge ($T0_{DCK}$)	-	0.044*	0.050*	0.016*

Note: * Significant at $p < 0.013$ level (Bonferroni correction)

Table 4

The effects of the assumed parameters and the covariate (Prior knowledge, $T0_{EDS}$) on the changes for the EDS tasks in the beginning of the project (T0) and in the end of Grade 7 (T1), Grade 8 (T2), Grade 9 (T3) (N=603)

Parameter	PES (Partial Eta Squared)			
(Source)	$T0_{EDS}$	$T1_{EDS}$	$T2_{EDS}$	$T3_{EDS}$
Group	0.003	0.050*	0.003	0.014(*)
School ranking	0.090*	0.012	0.056*	0.014(*)
Mother's education	0.000	0.005	0.007	0.001
Gender	0.000	0.004	0.000	0.004
Prior knowledge ($T0_{EDS}$)	-	0.076*	0.064*	0.066*

Note: * Significant at $p < 0.013$ level (Bonferroni correction)

The effect of the assumed parameters on the scores (absolute mean scores) estimated by the ANCOVA analysis model for the four tests for the whole test, the DCK tasks and the EDS tasks, as well as the significance of the differences, are presented in Tables A7–A9 in the Appendix. These data show that Group 3 performed better in the EDS task than the other two groups at the end of each school year, although the difference was not significant in the second year (Test 2). Among the other parameters, only in terms of the pupil's school ranking (low, medium, high) were there significant differences in EDS scores at the end of each year. In the DCK tasks, Group 2 in the first year and Group 3 in the third year scored significantly lower than the other two groups. The school ranking also caused significant differences in DCK scores, except at the end of the second year. Gender does not appear to have had much of an influence on EDS scores. The difference caused by the higher education degree of the pupil's mother was significant for EDS scores only in the second year.

Tables 5–7 show the relative estimated mean scores (the ratio of the estimated mean scores of the experimental groups compared to the control group) for the whole test and the subtests at the beginning of the project (Grade 7, T0) and at the end of each school year (Grade 7, T1; Grade 8, T2; Grade 9, T3).

The ratios in Table 5 suggest that the intervention did not have much impact. However, subtest data in Tables 6 and 7 show that this is due to the combined effects in the DCK and EDS tasks. Table 6 shows that in the DCK tasks, Group 2 scored lower than the control group at the end of the first year and Group 3 at the end of the third year. According to these data, the performance of Group 3 in the EDS tasks was higher than that of the control group at the end of each year, but the effect of the intervention appears to have been strongest in the first year and weakest in the second year.

Table 5

The estimated mean scores of the experimental groups divided by the estimated mean scores of the control group for the whole test ('TOTAL') in the four tests (N=603)

Ratio	$T0_{TOTAL}$	$T1_{TOTAL}$	$T2_{TOTAL}$	$T3_{TOTAL}$
Group 2 / Group 1	1.02	0.90	0.99	0.96
Group 3 / Group 1	0.99	1.09	1.04	1.00

Table 6

The estimated mean scores of the experimental groups divided by the estimated mean scores of the control group for the DCK tasks in the four tests (N=603)

Ratio	$T0_{DCK}$	$T1_{DCK}$	$T2_{DCK}$	$T3_{DCK}$
Group 2 / Group 1	1.01	0.85	0.96	0.95
Group 3 / Group 1	1.01	0.99	1.02	0.86

Table 7

The estimated mean scores of the experimental groups divided by the estimated mean scores of the control group for the EDS tasks in the four tests (N=603)

Ratio	$T0_{EDS}$	$T1_{EDS}$	$T2_{EDS}$	$T3_{EDS}$
Group 2 / Group 1	1.03	0.97	1.02	0.97
Group 3 / Group 1	0.97	1.19	1.04	1.10

Discussion

Based on the data presented, at the end of the third year of the project, it appears that of the two methods used to treat the experimental groups, the one applied to Group 3 has a small positive effect on the development of experimental design skills. Group 3 seems to have benefited from answering the series of questions before carrying out the experiments, which helped them to learn to design the experiments when they were 12–13-year-old. This may be the reason why this method had a significant positive impact on the development of EDS in the first year. However, the effect size of the changes in the EDS has decreased by the end of the three-year period. Group 2, on the other hand, only answer the questions on designing experiments after they have carried out the experiments step-by-step, following the recipe in their worksheet. This method does not seem to have produced any positive long-term results in terms of changes in EDS. The reasons are not known, but those pupils might have found this activity less interesting and important after the experiments than Group 3 did before the experiments. The treatment of Group 3 is similar to the 'critical thinking' pre-laboratory group, while the treatment of Group 2 resembles the 'paved road' pre-laboratory group described by van Brederode et al. (2020). The current results show that the intervention applied to Group 3 produced the slightly better results, similarly to the 'critical thinking' group in the above-mentioned study. These results are also in line with a study by Tseng et al. (2022), where evaluative reflection on peers' experimental designs improved pupils' science inquiry performance in designing experimental procedures more than recognising variables in the design of experimental procedures.

The present findings for Group 3 also support the need to provide learners from traditional learning styles with appropriate and adequate scaffolds to successfully complete an inquiry-based learning task (Seery et al., 2019). Scaffolding refers to the steps taken to reduce the degree of freedom in performing a task so that the learner can concentrate on the difficult skill to be learned (Bruner, 1978). The series of questions for identifying and controlling variables and constructing hypotheses play this role in the present research. Learners begin to think logically about abstract concepts and test hypotheses systematically at the Piagetian formal operational stage, which is expected from the age of 12 (Piaget, 1963). It is therefore understandable that seventh grade pupils can benefit from this intervention. However, the success of open-ended problem solving depends, among other things, on specific knowledge and understanding stored in long-term memory, so this type of problem solving is context-dependent (Reid & Amanat Ali, 2020, p. 471). Since each test contains different EDS tasks in different contexts, these measure whether far transfer has been achieved. This may be the reason why the effect sizes were found to be very limited in long term and not even significant in the second year. Nor did Cacciatore and Sevian (2009) find significant differences between groups in their performance in tasks not directly related to the content of the experiment.

However, it is difficult to understand what happened in the second year, when Group 2 developed better than Group 3. One possible reason might be that seeing the results of Test 1, which showed that the other two groups performed better, some teachers in Group 2 may have encouraged their pupils to work harder in the second year and/or to put more effort into Test 2 at the end of the second year. The opposite effect may have affected the performance of Group 3 in Test 2, who performed well in Test 1. However, these possible explanations are not supported by any evidence.

The DCK tasks included in the tests were intended to measure whether answering a set of questions aimed at developing EDS influences pupils' disciplinary content knowledge. This was investigated to see if the concern that the time needed to develop the EDS would reduce the time spent on learning content, which could reduce DCK, is justified. The decrease in the achievement of DCK tasks in Group 2 in the first year and in Group 3 in the third year indicates that this could not be ruled out.

The ANCOVA analysis showed that, in addition to prior knowledge, school ranking was the parameter that most often influenced pupils' test scores. It is reasonable to assume that in higher ranking schools, pupils' abstract thinking may develop faster in more challenging environments. On the other hand, these pupils are also likely to be more strongly encouraged by their teachers and parents to achieve higher levels of performance. Snook et al. (2009), in their commentary on Hattie's (2008) book, argued that in addition to social variables, 'school effects' are also important, and that of these, the teacher is considered the most important variable. Teachers from different schools work in different contexts, following different local curricula, which comply with the National Core Curriculum, but also have their own characteristics according to the institutions' own pedagogical programme. Siegler and colleagues (2010) also argued that school is the microsystem that, alongside the family environment, has the strongest influence on youngsters' development.

The ANCOVA model calculations show that the mother's educational background had only rare and weak significant effects on the development of pupils' knowledge and skills in experimental design in this study. This seems to contradict the OECD report (2005), which finds that the most significant source of variation in pupils' learning is due to differences in the skills and attitudes pupils bring to school. The Education and Training Monitor (2020) report also shows that socio-economic background is a strong predictor of pupil achievement. However, the virtual contradiction with the results of the present study can be explained by the fact that the present sample is not representative of the cohort, as these pupils went through a very tough selection process when they took entrance exams to their current school.

Gender does not appear to have had much effect on test scores in the present research. This means that they are closer to the results published by authors who found no significant difference in pupils' acquisition of science process skills (SPS) with respect to gender (e.g. Walters & Soyibo, 2001) and under the circumstances of the present study do not support the finding reported by Tosun (2019) that gender is one of the most important predictors of SPS levels.

Conclusion

Summary of the Results and Answers to the Research Questions

Statistical analysis of the results of the four tests completed by the current sample of pupils showed that, in addition to the intervention, the prior knowledge and the school ranking had the biggest influence on scores in tasks measuring EDS. The intervention appeared to have more effect on performance in the first year, school ranking in the second year, and both had some effect in the third year. Prior knowledge seemed to have a bigger effect than the intervention at the end of each year. Prior knowledge and school ranking also had a significant effect on the DCK measured by end-of-school-year tests, except for school ranking in the second year.

The answers to the research questions are as follows.

RQ1: By the end of the first three years of the present project, the intervention resulted only in a very small significant positive change in the experimental design skills (EDS) of Group 3 compared to the control group (Group 1), as measured by the tests (Cohen's d: 0.16). Taking the mean scores of the EDS tasks estimated by the ANCOVA model, Group 3 seemed to have improved more than the other two groups by the end of each year, but the difference was only significant at the end of the first and third years. It can be assumed that the small positive change is still due to the fact that Group 3's worksheets contained questions supporting the experimental design. The change in Group 2's performance in the EDS tasks by the end of the third year was negative compared to that of the control group (Cohen's d: -0.13), although they appear to have made more progress in this respect than the other two groups in the second year of the project.

RQ2: By the end of the third year of the present project, both experimental groups had a negative change in disciplinary content knowledge (DCK) compared to that of the control group (Cohen's d for Group 2: -0.07 and Group 3: -0.19, respectively). Group 2 had a negative effect size in the first year (Cohen's d: -0.29) and Group 3 in the third year (Cohen's d: -0.21). Group 3 improved more than Group 2 in this respect in the first year (Cohen's d: 0.30), but less so in the second (Cohen's d: -0.16) and third years (Cohen's d: -0.22). Taking into account the three years, Group 3's performance is less good than that of Group 2 (Cohen d: -0.12) in DCK tasks.

RQ3: In terms of changes in scores in the EDS tasks, of the two experimental groups, Group 3 improved more than Group 2 in the first year (Cohen's d: 0.45) and in the third year (Cohen's d: 0.19), but less than Group 2 in the second year (Cohen's d: -0.40). Over the three years, Group 3 has a positive effect size in this respect compared to Group 2 (Cohen d: 0.28).

The intervention does not seem to have a strong positive effect, even in terms of EDS and even for the better performing Group 3. However, these tasks are designed to measure far transfer, as pupils will need to use their EDS in different contexts in their everyday lives. These problem-solving skills are context-dependent (Reid & Amanat Ali, 2020, p. 471) and not easy to measure. Prior knowledge and 'school effects', which are considered in the literature (e.g. Snook et al., 2009) as variables that influence performance, also had a significant impact on the results. Furthermore, test scores are also influenced by pupils' motivation to succeed in the test and their test-taking skills (Cannady et al., 2019).

Statistical analysis of the test results showed that from the beginning of the project to the end of the third year, Group 3 showed significantly, but only slightly better performance in EDS than the control group. The positive impact was strongest in the first year. It can therefore be assumed that significantly more pupils in Group 3 understood how to correctly perform a fair test than in Group 1 in the first year of the present project. No significant positive difference in the development of EDS was found between the experimental groups and control group in the first year of the previous longitudinal study, when no such scaffold was used, and at the end of its third year (Szalay et al., 2021). This difference was probably caused by the series of questions used in the present project, which helped pupils to develop their EDS. This is consistent with Baird's (1990) view that purposeful inquiry is not spontaneous, it must be learned. Reducing cognitive load can help in this process. Therefore, the use of a series of questions, a simplified version of the Experiment Design Diagram described by Cothron and colleagues (2000), still seems to positively influence the development of the EDS. A recently published meta-analysis by Arifin et al. (2025) also found that guided inquiry (N=24) had a statistically significant positive effect on critical thinking, and that the educational levels with the largest effects included lower secondary and upper secondary schools.

The decrease in DCK task performance in Group 2 in the first year and in Group 3 in the third year shows that the time spent developing the EDS might in some cases negatively affect the amount of disciplinary content learned. This may discourage some teachers from using the worksheets developed in this project. Especially in Hungary and in the present situation, where the National Core Curriculum requires the practice of experimental design, but EDS is not measured in the chemistry final exam, which serves as a written university entrance exam. Therefore, when training pre-service and in-service teacher students, the benefits of developing experiment design skills should be emphasised and different methods for doing so should be introduced.

Limitations

Pupils must remain in the same school for the four years of the project. Therefore, only schools that teach chemistry as a separate subject from Grade 7 to Grade 10 can participate in the research. Pupils of these schools must pass an entrance exam at the age of 12 and only the best pupils are admitted. This implies that the sample is not representative of this cohort, but highly selective. This is unfortunate, but there is no other way to conduct such a longitudinal study of chemistry teaching in this country.

The sample size for this research decreased each year, as the results of pupils whose class no longer participates in the project or who just miss any of the tests are not counted. Changes in the composition of the sample have caused significant differences in the composition of the groups in terms of school ranking. In theory, ANCOVA can handle this, but in statistical analyses there is always a chance that changes that appear significant are not in reality. The high attrition rate and non-representative sample (selective, exam-admitting schools) severely limit external validity.

Performance in tests is determined at least in part by pupils' motivation to succeed in the test and their test-taking skills (Cannady et al., 2019). Several studies have shown that for many school pupils, the motivation to learn chemistry is primarily to get good grades (Salta & Koulougliotis, 2015; Schumm & Bogner, 2016; Ardura & Pe´rez-Bitria´n, 2018; Komperda et al., 2020; de Souza et al., 2022; Zhang & Zhou, 2023). It is also known that pupils' motivation to learn science often declines as they are getting older (Schunk et al, 2014; Vedder-Weiss & Fortus, 2011; Vedder-Weiss & Fortus, 2013). Therefore, lack of motivation is a particular problem when it comes to measuring changes in pupil performance in such a longitudinal research study, and for methodological reasons it is not possible to reward pupil performance with marks. Although the context-based and systems thinking tasks on the worksheets (see Table 1), as well as the context of the EDS tasks, hopefully attracted the interest of many pupils, it is likely that not everyone was equally committed to performing to the best of their ability in the tests.

There is no research that can investigate all the variables and theoretical relationships that underpin an instructional model (Mack et al., 2019). In addition, 40-minute paper-based tests do not provide a complete picture of the impact of the intervention on pupils' knowledge and skill development. Reading comprehension is also important for science achievement (Neri et al., 2021), but its influence was not investigated in the present study.

The findings may be also influenced by a number of random events. Although the relatively large sample size should compensate for these in a statistical sense, we can never be absolutely certain (Lawrie, 2021).

Recommendations

The use of Group 3 worksheets and similar experimental design exercises, which guide students through the experimental design process using a series of questions, can be recommended to practising teachers as one possible way to develop EDS. However, teachers need to be aware that the use of these worksheets might reduce the amount of disciplinary content knowledge acquired.

Acknowledgements

This study was supported by Research Programme for Public Education Development of the Hungarian Academy of Sciences under project number SZKF-6/2021. The authors would like to express their gratitude for the financial assistance provided, which played a crucial role in the successful completion of this research. The authors would like to thank the participating schools, teachers and pupils for their invaluable contributions to this work.

Conflicts of Interest

There are no conflicts of interest to declare.

References

Abraham, M. R., Craolice, M. S., Graves, A. P., Aldahmas, A. H., Kihega, J. G., Palma Gil J. G., & Varghese, V. (1997). The nature and state of general chemistry laboratory courses offered by colleges and universities in the United States. *Journal of Chemical Education*, *74*(5), 591–594. https://doi.org/10.1021/ed074p591

Akuma, F. V., & Callaghan, R. (2019). A systematic review characterizing and clarifying intrinsic teaching challenges linked to inquiry-based practical work. *Journal of Research in Science Teaching*, *56*(5), 619–648. https://doi.org/10.1002/tea.21516

Ardura D., & Pe´rez-Bitria´n A. (2018). The effect of motivation on the choice of chemistry in secondary schools: adaptation and validation of the Science Motivation Questionnaire II to Spanish pupils, *Chemistry Education Research and Practice, 19*(3), 905–918. https://doi.org/10.1039/C8RP00098K

Arifin, Z., Sukarmin, Saputro, S., & Kamari, A. (2025). The effect of inquiry-based learning on pupils' critical thinking skills in science education: A systematic review and meta-analysis, *EURASIA Journal of Mathematics, Science and Technology Education*, 2025, *21*(3), em2592, https://doi.org/10.29333/ejmste/15988

Arnold, J. C., Boone, W. J., Kremer, K., & Mayer, J. (2018). Assessment of competencies in scientific inquiry through the application of Rasch measurement techniques, *Education Sciences, 8*(4), 184–203. https://doi.org/10.3390/educsci8040184

Arnold, J. C., Mühling, A., & Kremer, K. (2021). Exploring core ideas of procedural understanding in scientific inquiry using educational data mining. *Research in Science & Technological Education, 41*(1), 1–21., https://doi.org/10.1080/02635143.2021.1909552

Baird, J. R. (1990). Metacognition, purposeful inquiry and conceptual change, In E. Hegarty-Hazel (Ed.), *The pupil laboratory and the science curriculum* (pp. 183−200). Routledge.

Banchi, H., & Bell R. (2008). The many levels of inquiry. *Science and Children*, *46*(2), 26–29., The-Many-Levels-of-Inquiry-NSTA-article.pdf

Bloom B. S., Engelhart M. D., Furst E. J., Hill W. H., & Krathwohl D. R. (1956). *Taxonomy of Educational Objectives*: Part I, Cognitive Domain.

Bolte, C., Holbrook, J., & Rauch, F. (Eds.). (2012). *Inquiry-based Science Education in Europe: Reflections from the PROFILES Project* (pp. 9–67). Berlin: Freie Universität Berlin, Division of Chemistry Education. http://www.profiles-project.eu

Bruner, J. S. (1978). The role of dialogue in language acquisition. In A. Sinclair, R. J. Jarwelle, & W. J. M. Levelt (Eds.). *The child's concept of language* (pp. 241–256.). Springer

Bunce, D. M., & Cole, R. S. (Eds.). (2007). *Nuts and Bolts of Chemical Education Research*. ACS SYMPOSIUM SERIES 976. American Chemical Society.

Burke, K. A., Greenbowe, T. J., & Hand, B. M. (2006). Implementing the Science Writing Heuristic in the chemistry laboratory. *Journal of Chemical Education*, *83*(7), 1032–1038. https://doi.org/10.1021/ed083p1032

Cacciatore, K. L., & Sevian, H. (2009). Incrementally approaching an inquiry lab curriculum: Can changing a single laboratory experiment improve pupil performance in general chemistry? *Journal of Chemical Education*, *86*(4), 498–505. https://doi.org/10.1021/ed086p498

Cannady, M. A., Vincent-Ruz, P., Chung, J. M., & Schunn, C. D. (2019). Scientific sensemaking supports science content learning across disciplines and instructional contexts. *Contemporary Educational Psychology, 59*(10), 101802. https://doi.org/10.1016/j.cedpsych.2019.101802

Chen Y.-Ch., Kimberley Wilson K. & Lin H.-sh., (2019). Identifying the challenging characteristics of systems thinking encountered by undergraduate students in chemistry problem-solving of gas laws, *Chemistry Education Research and Practice, 20*(3), 594–605. https://doi.org/10.1039/C9RP00070D

Chen, L., & Xiao, S. (2021). Perceptions, challenges and coping strategies of science teachers in teaching socioscientific issues: A systematic review. *Educational Research Review*, *32*(2), 100377. https://doi.org/10.1016/j.edurev.2020.100377

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Routledge. https://doi.org/10.4324/9780203771587

Cooper, M. M. (2013). Chemistry and the Next Generation Science Standards. *Journal of Chemical Education*, *90*(6), 679–680. https://doi.org/10.1021/ed400284c

Cothron, J. H., Giese, R. N., & Rezba, R. J. (2000). *Pupils and Research: Practical Strategies for Science Classrooms and Competitions*, 3rd ed. Dubuque, IA: Kendall/Hunt Publishing Company.

Crocker, L., & Algina, J. (2006). *Introduction to Classical and Modern Test Theory*, 2nd ed., Wadsworth Publishing Company.

Csíkos, Cs., Korom, E., & Csapó, B. (2016). Tartalmi keretek a kutatásalapú tanulás tudáselemeinek értékeléséhez a természettudományokban. *Iskolakultúra*, *26*(3), 17–29. https://doi.org/10.17543/ISKKULT.2016.3.17

de Souza R., da Silva, M. d S. B., Barbato, D. M. L., de Guzzi, M. E- R., & Kasseboehmer, A. C. (2022). Motivation to learn chemistry: a thorough analysis of the CMQ-II within the Brazilian context, *Chemistry Education Research and Practice*, *23*(4), 799–810. https://doi.org/10.1039/D2RP00107A

del Mar López-Fernández, M., González-García, F., & Franco-Mariscal, A. J. (2022). How can socio-scientific issues help develop critical thinking in chemistry education? A reflection on the problem of plastics. *Journal of Chemical Education, 99*(10), 3435–3442. doi: 10.1021/acs.jchemed.2c00223.

Domin, D. S. (1999). A Review of laboratory instruction styles. *Journal of Chemical Education*, *76*(4), 543–547. https://doi.org/10.1021/ed076p543

Dumitrescu, C., Olteanu, R. L., Gorghiu, L. M., & Gorghiu, G. (2014). Learning chemistry in the frame of integrated science modules - Romanian pupils' perception, *Procedia - Social and Behavioral Sciences, 116*, 2516–2520. https://doi.org/10.1016/j.sbspro.2014.01.603

Education and Training Monitor. (2020), Luxembourg: Publications Office of the European Union. https://doi.org/10.2766/984100

Farell, J. J., Moog, R. S., & Spencer, J. N. (1999). A guided-inquiry general chemistry course. *Journal of Chemical Education*, *76*(4), 570–574. https://doi.org/10.1021/ed076p570

Fay M. E., Grove N. P., Towns M. H. and Bretz S. L., (2007), A rubric to characterize inquiry in the undergraduate chemistry laboratory. *Chemistry Education Research and Practice, 8*(2), 212–219. https://doi.org/10.1039/B6RP90031C

Goodey, N. M., & Talgar, C. P. (2016). Guided inquiry in a biochemistry laboratory course improves experimental design ability. *Chemistry Education Research and Practice, 17*(4), 1127–1144. https://doi.org/10.1039/C6RP00142D

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Howell, D. C. (2012). *Statistical methods for psychology*. Cengage Learning.

Johnstone, A. H., & Wham, A. J. B. (1982). The demands of practical work. *Education in Chemistry*, *19*(3), 71–73.

Karplus, R., & Thier, H. (1967). *A New Look at Elementary School Science*. Rand-McNally.

Klemeš, J. J., Fan, Y. V., & Jiang, P. (2021). Plastics: friends or foes? The circularity and plastic waste footprint. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, *43*(13), 1549−1565. https://doi.org/10.1080/15567036.2020.1801906

Komperda, R., Hosbein, K. N., Phillips, M. M., & Barbera, J. (2020). Investigation of evidence for the internal structure of a modified science motivation questionnaire II (mSMQ II): a failed attempt to improve instrument functioning across course, subject, and wording variants, *Chemistry Education Research and Practice, 21*(3), 893–907. https://doi.org/10.1039/D0RP00029A

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Lawrie, G. (2021). Considerations of sample size in chemistry education research: numbers do count but context matters more! *Chemistry Education Research and Practice, 22*(4), 809–812. https://doi.org/10.1039/D1RP90009A

Lawrie, G. A., Graulich, N., Kahveci, A., & Lewis, S. E. (2021). Ethical statements: a refresher of the minimum requirements for publication of Chemistry Education Research and Practice articles, *Chemistry Education Research and Practice, 22*(2), 234–236. https://doi.org/10.1039/D1RP90003J

Lawson, A. E., Abraham, M. R., & Renner, J. W. (1989). *A Theory of Instruction Using the Learning Cycle to Teach Science Concepts and Thinking Skills*. Monograph, Number One; National Association for Research in Science Teaching: Kansas State University, Manhatten, KS.

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681–718. https://doi.org/10.3102/0034654315627366

Lee, H.-S., Gweon, G.-H., Webb, A., Damelin, D., Dorsey, C. (2024). Measuring epistemic knowledge development related to scientific experimentation practice: A construct modeling approach, *Science Education*, *108*(2), 383-411. https://doi.org/10.1002/sce.21836

Legjobbiskola. (n.d.). *Ranking of Hungarian schools*. Retrieved October 4, 2021, from Rangsorok | legjobbiskola.hu

MacDonald, R. P., Pattison, A. N., Cornell, S. E., Elgersma, A. K., Greidanus, S. N., Visser, S. N., Hoffman, M., & Mahaffy, P. G. (2022). An interactive planetary boundaries systems thinking learning tool to integrate sustainability into the chemistry curriculum. *Journal of Chemical Education, 99*(10), 3530−3539. https://doi.org/10.1021/acs.jchemed.2c00659

Mack, M. R., Hensen, C., & Barbera, J. (2019). Metrics and methods used to compare pupil performance data in chemistry education research articles. *Journal of Chemical Education, 96*(3), 401–413. https://doi.org/10.1021/acs.jchemed.8b00713

Matthews, M. R. (Ed.). (2018). *History, Philosophy and Science Teaching – New Perspectives*, Springer. https://doi.org/10.1007/978-3-319-62616-1

McLoughlin, E., Finlayson, O., McCabe, D., & Brady, S. (2015). IBSE Teaching & Learning Units, Volume 2, Chemistry, ESTABLISH Project. http://www.establish-fp7.eu/sites/default/files/general/Chemistry.pdf

MTA-ELTE Research Group. (n.d.). *MTA-ELTE Research Group on Inquiry-Based Chemistry Education, Inquiry-Based Chemistry Education and Systems Thinking Project*. Retrieved September 17, 2025, from Természettudományos Oktatásmódszertani Centrum

National Core Curriculum of Hungary. (2020). 5/2020. (I. 31.) Korm. rendelet A Nemzeti alaptanterv kiadásáról, bevezetéséről és alkalmazásáról szóló 110/2012. (VI. 4.) Korm. rendelet módosításáról, Magyar Közlöny, 2020. jan. 31., 17, pp. 290–446. Retrieved March 5, 2025, from: A Kormány 5/2020. (I. 31.) Korm. rendelete a Nemzeti alaptanterv kiadásáról, bevezetéséről és alkalmazásáról szóló 110/2012. (VI. 4.) Korm. rendelet módosításáról – eGov Hírlevél

Neri, N. C., Guill, K., & Retelsdorf, J. (2021). Language in science performance: Do good readers perform better? *European Journal of Psychology of Education*, *36*(1), 45–61. https://doi.org/10.1007/s10212-019-00453-5

OECD. (2005). Teachers matter: Attracting, developing and retaining effective teachers. Overview. OECD, p.2. Teachers Matter | OECD

Pedaste, M., Maeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review, 14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Piaget, J. (1963). *The Origins of Intelligence in Children*. Norton.

Reed, J. J., & Holme, T. A. (2014). The Role of Non-Content Goals in the Assessment of Chemistry Learning. In: L. K. Kendhammer, & K. L. Murphy (Eds.), *Innovative Uses of Assessment for Teaching and Research*, American Chemical Society (pp. 147–160.) https://doi.org/10.1021/BK-2014-1182.CH009

Reid N. & Amanat Ali A. (2020). *Making Sense of Learning*, Springer Nature Switzerland, AG.

Restucia, D., Taska, B., & Bittle S. (2018, March 14.). *Different Skills, Different Gaps: Measuring and Closing the Skills Gap,* Burning Glass Technologies, U.S. Chamber of Commerce Foundation. Different Skills, Different Gaps: Measuring and Closing the Skills Gap | U.S. Chamber of Commerce Foundation

Reynders, G., Suh, E., Cole, R. S., & Sansom, R. L. (2019). Developing pupil process skills in a general chemistry laboratory. *Journal of Chemical Education, 96*(10), 2109–2119. https://doi.org/10.1021/acs.jchemed.9b00441

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H., & Hemmo, V. (2007). *Science Education NOW: A Renewed Pedagogy for the Future of Europe*. European Commision. Directorate-General for Research. rapportrocardfinal.pdf (europa.eu)

Rodriguez, J. M. G., & Towns, M. H. (2018). Modifying laboratory experiments to promote engagement in critical thinking by reframing prelab and postlab questions. *Journal of Chemical Education, 95*(12), 2141–2147. https://doi.org/10.1021/acs.jchemed.8b00683

Rodriguez, J-M. G., Hunter, K. H., Scharlott, L. J., & Becker, N. M. (2020). A Review of research on Process Oriented Guided Inquiry Learning: Implications for research and practice. *Journal of Chemical Education*, 97(10), 3506–3520. https://doi.org/10.1021/acs.jchemed.0c00355

Rodríguez-Berrios, R. R., & Rodríguez-Vargas, J. A. (2025). Development of an accessible guided-inquiry laboratory experiment for undergraduate food chemistry pupils through qualitative analysis of lipid oxidation in fats and oils. *Journal of Chemical Education*, *102*(2), 737–745. https://doi.org/10.1021/acs.jchemed.4c00514

Salta, K. & Koulougliotis, D. (2015). Assessing motivation to learn chemistry: adaptation and validation of Science Motivation Questionnaire II with Greek secondary school pupils, *Chemistry Education Research and Practice, 16*(2), 237–250. https://doi.org/10.1039/C4RP00196F

Savec, V. F., & Devetak, I. (2013), Evaluating the effectiveness of pupils' active learning in chemistry. *Procedia - Social and Behavioral Sciences*, 106, 1113 – 1121. https://doi.org/10.1016/j.sbspro.2013.12.125

Schafer, A. G. L., Kuborn, T. M., Cara, E., Schwarz, C. E., Deshaye, M. Y., & Stowe, R. L. (2023). Messages about valued knowledge products and processes embedded within a suite of transformed high school chemistry curricular materials. *Chemistry Education Research and Practice, 24*(1), 71–88. https://doi.org/10.1039/D2RP00124A

Schoffstall, A. M., & Gaddis, B. A. (2007). Incorporating guided-inquiry learning into the organic chemistry laboratory. *Journal of Chemical Education, 84*(5), 848. https://doi.org/10.1021/ed084p848

Schumm M. F. & Bogner F. X. (2016). Measuring adolescent science motivation, *International Journal of Science Education, 38*(3), 434–449. https://doi.org/10.1080/09500693.2016.1147659

Schunk, D. H., Pintrich, P. R. & Meece, J. L., (2014), Motivation in education: theory, research and applications, Upper Saddle River, Pearson.

Science Olympiad (2020), *Experimental design Division C Checklist, Part I – Design and Construction of the Experiment.* https://www.soinc.org/sites/default/files/uploaded_files/Experimental_Design_Checklist_Division_C.pdf

Seery, M. K., Jones, A. B., Kew, W., & Mein, T. (2019). Unfinished recipes: Structuring upper-division laboratory work to scaffold experimental design skills. *Journal of Chemical Education, 96*(1), 53–59. https://doi.org/10.1021/acs.jchemed.8b00511

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental designs for Generalized Causal Inference*. Houghton Mifflin and Company.

Siegler, R. S., Deloache, J. S., & Eisenberg, N. (2010). *Study Guide for How Children Develop*. Worth Publishers.

Sirum, K., & Humburg, J. (2011), The Experimental design Ability Test (EDAT), *Bioscene: Journal of College Biology Teaching, 37*(1) 8–16. Bioscene

Snook, I., Clark, J., Harker, R., O'Neil, A.-M., & O'Neil, J. G. (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. (PDF) Invisible Learnings: A commentary on John Hattie's book visible learning: A synthesis of over 800 metaanalyses relating to achievement

Szalay, L., Tóth. Z., & Borbás, R. (2021). Teaching of experimental design skills: results from a longitudinal study. *Chemistry Education Research and Practice, 22*(4), 1054–1073. https://doi.org/10.1039/D0RP00338G

Szalay, L., Tóth, Z., & Kiss, E. (2020). Introducing students to experimental design skills. *Chemistry Education Research and Practice, 21*(1), 331–356. https://doi.org/10.1039/C9RP00234K

Szalay, L., & Tóth Z. (2016). An inquiry-based approach of traditional 'step-by-step' experiments. *Chemistry Education Research and Practice*, *17*(4), 923–961. https://doi.org/10.1039/C6RP00044D

Szalay, L., Tóth, Z., Borbás, R., & Füzesi, I. (2023). Scaffolding of experimental design skills. *Chemistry Education Research and Practice, 24*(2), 599–623. https://doi.org/10.1039/D2RP00260D

Szalay, L., Tóth, Z., Borbás, R. & Füzesi, I. (2024) Progress in developing experimental design skills among junior high school learners; *Journal of Turkish Science Education. 21*(3), 484-511. https://doi.org/10.36681/tused.2024.026

Szozda, A. R., Bruyere, K., Lee, H., Mahaffy, P. G., & Flynn, A. B. (2022). Investigating educators' perspectives toward systems thinking in chemistry education from international contexts. *Journal of Chemical Education, 99*(7), 2474–2483. https://doi.org/10.1021/acs.jchemed.2c00138

Tosun, C. (2019). Scientific process skills test development within the topic ''Matter and its Nature'' and the predictive effect of different variables on 7th and 8th grade pupils' scientific process skill levels, *Chemistry Education Research and Practice, 20*(1), 160–174. https://doi.org/10.1039/C8RP00071A

Tseng, Y-J., Hong, Z-R., & Lin, H-s. (2022). Advancing pupils' scientific inquiry performance in chemistry through reading and evaluative reflection. *Chemistry Education Research and Practice, 23*(3), 616–627. https://doi.org/10.1039/D1RP00246E

Underwood, S. M., Posey, L. A., Herrington, D. G., Carmel, J. H., & Cooper, M. M. (2018). Adapting assessment tasks to support three-dimensional learning. *Journal of Chemical Education, 95*(2), 207–217. https://doi.org/10.1021/acs.jchemed.7b00645 van Brederode, M. E., Zoon, S. A., & Meeter, M. (2020). Examining the effect of lab instructions on pupils' critical thinking during a chemical inquiry practical. *Chemistry Education Research and Practice, 21*(4), 1173–1182. https://doi.org/10.1039/D0RP00020E

Vedder-Weiss, D., & Fortus, D. (2011). Adolescents' declining motivation to learn science: Inevitable or not? *Journal of Research in Science Teaching, 48*(2), 199–216. https://doi.org/10.1002/tea.20398

Vedder-Weiss, D., & Fortus, D. (2013). School, teacher, peers, and parents' goals emphases and adolescents' motivation to learn science in and out of school. *Journal of Research in Science Teaching, 50*(8), 952–988. https://doi.org/10.1002/tea.21103

Vélemények az MTA-ELTE (n.d.) *Vélemények az MTA-ELTE Kutatásalapú Kémiatanítás Kutatócsoport által a 2021/2022. tanévben készített kísérlettervező feladatlapokról*. Retrieved September 17, 2025, from https://docs.google.com/forms/d/e/1FAIpQLScSv8PJKOYe7ArEI-ClhdCC4VT2Eg-ju4BGaX_kWL-3o--YoQ/viewform

Vorholzer, A., Von Aufschnaiter, C., & Boone, W. J. (2020). Fostering upper secondary pupils' ability to engage in practices of scientific investigation: A comparative analysis of an explicit and an implicit instructional approach. *Research in Science Education, 50*(1), 333–359. https://doi.org/10.1007/s11165-018-9691-1

Walker, J. P., & Sampson, V. (2013). Learning to argue and arguing to learn: Argument-Driven Inquiry as a way to help undergraduate chemistry pupils learn how to construct arguments and engage in argumentation during a laboratory course. *Journal of Research in Science Teaching*, *50*(5), 561–596. https://doi.org/10.1002/tea.21082

Walker, M. (2007). *Teaching inquiry-based science.* Lightning Source, LaVergne, TN.

Walters, Y. B., & Soyibo, K. (2001). An analysis of high school pupil's performance on five integrated science process skills. *Research in Science and Technological Education, 19*(2), 133–143. https://doi.org/10.1080/02635140120087687

Watts, F. M., & Finkenstaedt-Quinn, S. A. (2021). The current state of methods for establishing reliability in qualitative chemistry education research articles. *Chemistry Education Research and Practice, 22*(3), 565–578. https://doi.org/10.1039/D1RP00007A

Wren, D., & Barbera, J. (2013). Gathering evidence for validity during the design, development, and qualitative evaluation of thermochemistry concept inventory items. *Journal of Chemical Education, 90*(12), 1590–1601. https://doi.org/10.1021/ed400384g

York, S., & Orgill, M. (2020). ChEMIST Table: A tool for designing or modifying instruction for a Systems Thinking Approach in chemistry education. *Journal of Chemical Education, 97*(8), 2114–2129. https://doi.org/10.1021/acs.jchemed.0c00382

Zhang, J., & Zhou, Q. (2023). Chinese Chemistry Motivation Questionnaire II: Adaptation and validation of the Science Motivation Questionnaire II in high school pupils. *Chemistry Education Research and Practice, 24*(1), 369–383. https://doi.org/10.1039/D2RP00243D

Appendix

Table A1.a

Topics and context of the worksheets no. 13-15 and teacher guides used in the school year 2023/2024.

No.	Topic	Experiments that Group 1 and Group 2 pupils had to do following step-by-step instructions, but Group 3 pupils had to design before doing the experiment	Context and elements of systems thinking in the "Let's think!" parts for motivation purposes. These are the same on the pupil worksheets of all the three groups.
13.	Excitation energy and flame tests	Pupils carry out flame tests. They are given a table showing the wavelength ranges of the colours. They are also explained that the energy of a photon is inversely proportional to its wavelength. Group 3 pupils have to design experiments to find out whether the excitation energy of the metal atoms of two unknown metal salts they find on their tray is lower or higher than the excitation energy of the sodium of the table salt (sodium chloride).	Although fireworks are an eye-catching sight, they have many side effects. When they explode, they emit air pollutants and a sudden burst of sound and light, which can scare animals, and they can cause injuries. Pupils have to group particles, labelled with their chemical symbols and formulae, according to their role in the use of fireworks (combustible substances, combustion promoters, substances responsible for the colour effect, air polluting combustion products).
14.	Inter-molecular forces	Pupils are explained that the stronger the interaction between the particles of a liquid, the harder it is to separate them, so the harder it is to increase the surface area, and so the larger are the droplets of the liquid. Surfactants (such as dishwashing detergents) reduce surface tension because they make it easier to increase the surface area of the liquid, e.g. because the polar part of their particles is in the water and the apolar part is in the air. Pupils should compare the surface tension of pure water and water containing detergent. To do this, Group 3 pupils need to work out how to measure the volume of a single drop of two liquids.	The pond skater ("Jesus bug") can walk on water because the interactions that hold water molecules together are strong enough to keep the pond skater on the surface. However, the pond skater cannot run on the surface of the surfactant-contaminated water. Pupils should identify sources of natural and artificial surfactants that can pollute water. They should also identify the different types of direct environmental damage caused by the presence of surfactants and the resulting foaming.
15.	Rates of chemical reactions	Pupils are explained that we should always control the rate of chemical reactions according to our particular needs. They are then given materials and equipment to measure the rate of reaction between sodium thiosulphate solution and hydrochloric acid of different concentrations at different temperatures. Group 3 pupils have to design experiments to slow down the reaction and then to accelerate it.	The pupils have to imagine that their friend fell while cycling and they used Betadine containing iodine to disinfect a nasty bruise on his skin. But some of the brown Betadine spilled onto their friend's favourite light-coloured shorts. Iodine is not soluble in water, but according to the internet, "*the Betadine solution ... can be removed with a solution of sodium thiosulphate*", which is available from the chemist's. The pupils have to decide whether the sodium thiosulphate solution should be dilute or concentrated and whether it should be cold or hot to get the fastest result.

Table A1.b

Topics and context of the worksheets no. 16-18 and teacher guides used in the school year 2023/2024.

16.	Chemical equilibrium	Using a model experiment, pupils have to decide if it is true that a combination of Diet Coke and Mentos candy caused the death of a boy. They are explained that carbon dioxide gas is added to the solution during the production of carbonated soft drinks. In a closed bottle, carbon dioxide is involved in the physical and chemical processes that lead to the equilibrium between gas and liquid. Group 3 pupils have to design the model experiment.	In beer and champagne, carbon dioxide is produced as a by-product of fermentation. Sometimes the beer behaves like a geyser, but in extreme bad cases the beer bottle can explode. The pupils have to explain by circling the appropriate arrows on a diagram how a change in temperature affects the solubility of carbon dioxide, i.e. the amount of gaseous carbon dioxide in the bottle and therefore the pressure in the bottle.
17.	Acid base reactions	Pupils have to measure which bottles contain 10% and 20% vinegar. They can use red cabbage juice as an acid-base indicator. They are also given a solution of drain cleaner with sodium hydroxide as the main ingredient. Group 3 pupils have to design an experiment to find out which glass contains the more dilute and which the more concentrated vinegar solution.	If a corrosive chemical (e.g. descaler, rust remover, drain cleaner) gets on the skin, it should be washed off immediately with plenty of water. In the case of acid burns, a weak and dilute base solution can be used to neutralise, but not concentrated and strong alkaline solutions, which would also be corrosive. In the case of an injury caused by an alkali, a weak and dilute acid solution should obviously be used. The pupils have to choose which neutralising solution to use for the different types of injuries.
18.	Redox reactions	Starting with the fake news about hydrogen peroxide, which claims that it can cure almost any disease, even cancer, pupils are told that hydrogen peroxide is only an external disinfectant and should not be ingested. It is then used as an oxidising or reducing agent in experiments (reacting with sodium hypochlorite and potassium iodide). Group 3 pupils are asked to design an experiment, using matches and a wooden splint, to find out whether the hydrogen peroxide solution acts as an oxidising or reducing agent in the experiments. (They should recognise a pattern that if hydrogen peroxide is a reducing agent, it is oxidised to elemental oxygen, which ignites the glowing splint.)	Like hydrogen peroxide, hypochlorite is used for disinfection and decolourisation. However, the use of hypochlorite has its own dangers, as toxic chlorine gas is formed when hypochlorite is used in combination with acidic cleaning agents like hydrochloric acid. Pupils are asked to write down the reaction equation for that chemical reaction. They then have to show on a diagram, by circling the appropriate arrows, how the addition of acid shifts the equilibrium towards an increase in the amount of chlorine gas released into the air.

Table A2

The number of pupils in low, medium and high-ranking schools (N=603)

Group	Low ranking schools	Medium ranking schools	High ranking schools	Total
Group 1	49	78	31	158
Group 2	60	79	90	229
Group 3	72	47	97	216
Total	181	204	218	603

Table A3

The means of scores and their SD-s for the whole test ('TOTAL'), the DCK and EDS tasks of T0 and the results of the ANOVA analysis (N=603)

Group	$T0_{TOTAL}$ (SD)*	$T0_{DCK}$ (SD)**	$T0_{EDS}$ (SD)**
Group 1	10.87 (3.70)	5.37 (1.75)	5.49 (2.60)
Group 2	11.48 (3.28)	5.56 (1.79)	5.92 (2.41)
Group 3	11.18 (3.50)	5.56 (1.69)	5.63 (2.58)
$F(2, N = 603)$	1.477	0.667	1.508
p	0.229	0.514	0.222
Sign.	-	-	-

*Note. *Maximum scores: 18; **: Maximum scores: 9*

Table A4

The means of scores and their SD-s for the whole test ('TOTAL'), the DCK and EDS tasks of T1 and the results of the ANOVA Analysis (N=603)

Group	$T1_{TOTAL}$ (SD)*	$T1_{DCK}$ (SD)**	$T1_{EDS}$ (SD)**
Group 1	8.83 (3.53)	4.42 (2.08)	4.41 (2.09)
Group 2	8.45 (3.09)	3.98 (1.69)	4.47 (2.04)
Group 3	10.0 (3.36)	4.62 (1.97)	5.39 (2.10)
$F(2, N = 603)$	13.27	6.651	14.61
p	0.000	0.001	0.000
Sign.	1, 2 < 3	2 < 1,3	1, 2 < 3

*Note. *Maximum scores: 18; **: Maximum scores: 9*

Table A5

The means of scores and their SD-s for the whole test ('TOTAL'), the DCK and EDS tasks of T2 and the results of the ANOVA Analysis (N=603)

Group	$T2_{TOTAL}$ (SD)*	$T2_{DCK}$ (SD)**	$T2_{EDS}$ (SD)**
Group 1	9.33 (3.18)	3.52 (2.04)	5.81 (1.91)
Group 2	9.59 (3.07)	3.48 (1.94)	6.11 (1.81)
Group 3	9.90 (3.20)	3.74 (2.19)	6.16 (1.86)
$F(2, N = 603)$	1.526	0.961	1.853
p	0.218	0.383	0.158
Sign.	-	-	-

*Note. *Maximum scores: 18; **: Maximum scores: 9*

Table A6

The means of scores and their SD-s for the whole test ('TOTAL'), the DCK and EDS tasks of the T3 and the results of the ANOVA Analysis (N=603)

Group	$T3_{TOTAL}$ (SD)*	$T3_{DCK}$ (SD)**	$T3_{EDS}$ (SD)**
Group 1	7.99 (3.51)	3.41 (1.85)	4.58 (2.30)
Group 2	8.10 (3.73)	3.44 (2.10)	4.66 (2.47)
Group 3	8.34 (3.63)	3.15 (1.88)	5.19 (2.42)
$F(2, N = 603)$	0.458	1.443	3.858
p	0.633	0.237	0.022
Sign.	-	-	1,2 < 3

*Note. *Maximum scores: 18; **: Maximum scores: 9*

Table A7

The effects of the assumed parameters estimated by the model of the ANCOVA analysis (absolute mean scores) for the whole test ('TOTAL') and the significance of their differences for the four tests (N=603)

Group	$T0_{TOTAL}$	$T1_{TOTAL}$	$T2_{TOTAL}$	$T3_{TOTAL}$
1. Group 1	10.83	8.98	9.47	8.21
2. Group 2	11.04	8.11	9.42	7.90
3. Group 3	10.74	9.78	9.84	8.18
Significant difference*	-	2 < 1 < 3	-	-
School ranking	$T0_{TOTAL}$	$T1_{TOTAL}$	$T2_{TOTAL}$	$T3_{TOTAL}$
1. Low	9.79	8.14	9.15	7.74
2. Medium	10.53	8.79	9.56	7.69
3. High	12.29	9.95	10.02	8.86
Significant difference*	1 < 2 < 3	1 < 2 < 3	1 < 3	1, 2 < 3
Mother's education	$T0_{TOTAL}$	$T1_{TOTAL}$	$T2_{TOTAL}$	$T3_{TOTAL}$
1. No degree in higher education	10.46	8.77	9.54	8.03
2. Degree in higher education	11.28	9.14	9.61	8.16
Significant difference*	1 < 2	-	-	-
Gender	$T0_{TOTAL}$	$T1_{TOTAL}$	$T2_{TOTAL}$	$T3_{TOTAL}$
1. Boy	11.06	9.13	9.60	8.02
2. Girl	10.68	8.79	9.55	8.18
Significant difference*	-	-	-	-

Note. * $p<0.05$

Table A8

The effects of the assumed parameters estimated by the model of the ANCOVA analysis (absolute mean scores) for the DCK tasks and the significance of their differences for the four tests (N=603)

Group	$T0_{DCK}$	$T1_{DCK}$	$T2_{DCK}$	$T3_{DCK}$
1. Group 1	5.24	4.54	3.69	3.54
2. Group 2	5.30	3.84	3.56	3.38
3. Group 3	5.29	4.49	3.78	3.03
Significant difference*	-	2 < 1, 3	-	3 < 1, 2
School ranking	$T0_{DCK}$	$T1_{DCK}$	$T2_{DCK}$	$T3_{DCK}$
1. Low	5.09	3.51	3.72	3.20
2. Medium	5.09	4.24	3.55	2.95
3. High	5.64	5.11	3.76	3.81
Significant difference*	1, 2 < 3	1 < 2 < 3	-	1, 2 < 3
Mother's education	$T0_{DCK}$	$T1_{DCK}$	$T2_{DCK}$	$T3_{DCK}$
1. No degree in higher education	4.94	4.26	3.82	3.30
2. Degree in higher education	5.61	4.32	3.54	3.34
Significant difference*	1 < 2	-	-	-
Gender	$T0_{DCK}$	$T1_{DCK}$	$T2_{DCK}$	$T3_{DCK}$
1. Boy	5.42	4.37	3.73	3.40
2. Girl	5.13	4.21	3.63	3.23
Significant difference*	2 < 1	-	-	-

Note. * $p<0.05$

Table A9

The effects of the assumed parameters estimated by the model of the ANCOVA analysis (absolute mean scores) for the EDS tasks and the significance of their differences for the four tests (N=603)

Group	$T0_{EDS}$	$T1_{EDS}$	$T2_{EDS}$	$T3_{EDS}$
1. Group 1	5.60	4.40	5.76	4.61
2. Group 2	5.75	4.25	5.86	4.48
3. Group 3	5.44	5.24	6.00	5.09
Significant difference*	-	1, 2 < 3	-	1, 2 < 3
School ranking	$T0_{EDS}$	$T1_{EDS}$	$T2_{EDS}$	$T3_{EDS}$
1. Low	4.70	4.46	5.26	4.40
2. Medium	5.44	4.46	5.95	4.65
3. High	6.65	4.95	6.40	5.13
Significant difference*	1 < 2 < 3	1, 2 < 3	1 < 2 < 3	1, 2 < 3
Mother's education	$T0_{EDS}$	$T1_{EDS}$	$T2_{EDS}$	$T3_{EDS}$
No degree in higher education	5.52	4.42	5.67	4.62
Degree in higher education	5.67	4.83	6.08	4.84
Significant difference*	-	-	1 < 2	-
Gender	$T0_{EDS}$	$T1_{EDS}$	$T2_{EDS}$	$T3_{EDS}$
Boy	5.64	4.74	5.86	4.59
Girl	5.55	4.51	5.88	4.86
Significant difference*	-	-	-	-

Note. * $p<0.05$