Faculty of Physics, Astronomy and Applied Computer Science

Jagiellonian University

Researching innovative strategies for enhancement of understanding and motivation in science

Daniel DZIOB

DISSERTATION FOR THE DEGREE DOCTOR OF PHILOSOPHY IN BIOPHYSICS

Supervisor: **prof. dr Mojca ČEPIČ**

Associate supervisor: **dr Dagmara SOKOŁOWSKA**

Kraków, 2020

Wydział Fizyki, Astronomii i Informatyki Stosowanej

Uniwersytet Jagielloński

Oświadczenie

Ja niżej podpisany *Daniel Dziob* (nr indeksu: 1041805) doktorant Wydziału Fizyki, Astronomii i Informatyki Stosowanej Uniwersytetu Jagiellońskiego oświadczam, że przedłożona przeze mnie rozprawa doktorska pt. „*Researching innovative strategies for enhancement of understanding and motivation in science*'' jest oryginalna i przedstawia wyniki badań wykonanych przeze mnie osobiście, pod kierunkiem prof. dr Mojcy Čepič oraz dr Dagmary Sokołowskiej. Pracę napisałem samodzielnie.

Oświadczam, że moja rozprawa doktorska została opracowana zgodnie z Ustawą o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (Dziennik Ustaw 1994 nr 24 poz. 83 wraz z późniejszymi zmianami).

Jestem świadom, że niezgodność niniejszego oświadczenia z prawdą ujawniona w dowolnym czasie, niezależnie od skutków prawnych wynikających z ww. ustawy, może spowodować unieważnienie stopnia nabytego na podstawie tej rozprawy.

Kraków, ………………………………… …………………………………

podpis doktoranta

Wszystkim tym,

bez których ta praca by nie powstała

i nie doszedłbym do miejsca, w którym jestem.

This dissertation has been written based on the scientific results previously reported in the following articles:

[A1] D. Dziob, D. Sokołowska, M. Młynarczyk and J. K. Mościcki (2019). *A simplified experimental method to study conductivity percolation.* Review of Scientific Instruments 90, 076105, https://doi.org/10.1063/1.5096969

[A2] D. Dziob and D. Sokołowska (2020). *Experiment on percolation for Introductory Physics Laboratories – a case study*. American Journal of Physics, 88(6), https://doi.org/10.1119/10.0000810

[A3] D. Dziob and D. Sokołowska (2020). *Water network percolation on yeast as an experiment proposal for advanced physics laboratories for bioscience students.* European Journal of Physics 41(2), https://doi.org/10.1088/1361-6404/ab4ed2

[A4] D. Dziob and M. Čepič (2020). *Simple method for measuring thermal conductivity.* Physics Education 55(4), https://doi.org/10.1088/1361-6552/ab8100

[A5] D. Dziob, Ł. Kwiatkowski and D. Sokołowska (2018). *Class tournament as an assessment method in physics courses: a pilot study.* EURASIA Journal of Mathematics, Science and Technology Education 14(4), https://doi.org/10.29333/ejmste/81807

[A6] D. Dziob (2020). *Board game in physics classes - a proposal for a new method of student assessment.* Research in Science Education 50(3), https://doi.org/10.1007/s11165-018-9714-y

Accordingly to the enclosed authors contributions forms, the personal contribution of Thesis author to the individual articles can be specified as follows:

[A1] Thesis author developed the method, conducted and validated the experiment, obtained the data, participated in data analysis and drafted the paper.

[A2] Thesis author designed and conducted the research, obtained the data, analyzed the data and participated in writing the paper.

[A3] Thesis author designed and conducted the research, obtained the data, analyzed the data and participated in writing the paper.

[A4] Thesis author designed and conducted the experiment, obtained the data, participated in data analysis and writing the paper.

[A5] Thesis author designed and conducted the research, obtained the data, participated in data analysis and in writing the paper.

[A6] Thesis author designed and conducted the research, obtained and analyzed the results and wrote the paper.

Thesis overview

Science in general is important for society and its further development. But when taught in school, it seems to be perceived by many students as irrelevant to their out-of-school experiences. The work described here attempted to address this rift by introducing contemporary physics topics and novel experiments to students at various educational levels as well as by developing new, active assessment methods. This was particularly interesting for biophysical topics, discussed in the thesis, since didactics of biophysics is relatively a new field and novel ways of introducing interdisciplinary content are still under development. As a contemporary physics topic, a percolation phenomenon was utilized. A novel, simplified experimental method was developed and used to prepare two laboratory activities for bioscience students at master level and physics students at bachelor level. Implementation of the method showed that it yielded results consistent with those of more advanced methods. Laboratory exercises with students showed that they were able to properly perform the experiment and analyze data. Their opinions were generally positive, but indicated a need for slight modifications. For high school students, a simple experiment involving quantitative measurements for determining the thermal conductivity was developed. Finally, two assessment methods in the form of games played in groups, namely a tournament and a board game, were developed and their levels of effectiveness were examined. Both methods created the opportunity to assess not only students' acquisition of content knowledge, but also development of their research skills as well as their awareness of a historical context. In each case, these methods improved students' achievements and were generally very well received by students.

Streszczenie

Nauka, rozumiana jako postęp wiedzy i rozwój technologiczny, jest ważna dla społeczeństwa i jego dalszego rozwoju. Jednak tak rozumiana nauka oraz nauka szkolna, zamknięta w ramach przedmiotów przyrodniczych, wydają się być z dwóch różnych światów. Nauka w wydaniu szkolnym jest odbierana przez wielu uczniów jako zbiór przestarzałych teorii i faktów, nieprzystających do ich codziennych doświadczeń spoza szkoły. Poniższa praca podejmuje próbę zmniejszenia tego rozdźwięku poprzez wprowadzenie tematów fizyki współczesnej oraz nowych eksperymentów na różnych etapach edukacji, jak również poprzez stworzenie nowych, aktywnych metod oceniania. Było to szczególnie interesujące w przypadku tematów biofizycznych, omawianych w pracy, ponieważ dydaktyka biofizyki jest stosunkowo nową dziedziną, a sposoby wprowadzania treści interdyscyplinarnych są wciąż w fazie rozwoju. Jako temat fizyki współczesnej zostało wykorzystane zjawisko perkolacji. Opracowano nową, uproszczoną metodę pomiarową, a następnie wykorzystano ją do przygotowania dwóch ćwiczeń dla studentów fizyki studiów licencjackich i studentów biofizyki studiów magisterskich. Przeprowadzone badania pokazały, że nowa metoda pomiarowa pozwala na uzyskanie wyników eksperymentalnych zgodnych z innymi, bardziej skomplikowanymi metodami pomiarowymi. Zajęcia laboratoryjne ze studentami pokazały, iż byli oni w stanie prawidłowo przeprowadzić eksperyment oraz poprawnie opracować i zinterpretować uzyskane dane. Uzyskane od studentów opinie były w większości pozytywne, ale wskazały możliwości pewnych modyfikacji przygotowanych ćwiczeń. Dla uczniów szkół średnich został przygotowany nowy, prosty eksperyment pozwalający na wyznaczenie wartości współczynnika przewodnictwa cieplnego różnych substancji. Ponadto zostały przygotowane i ocenione pod względem efektywności dwie metody aktywnego oceniania uczniów, w postaci mechanizmu grupowej gry, dokładnie turnieju, oraz gry planszowej. Obie metody stwarzają możliwość oceny nie tylko wiedzy uczniów, ale także rozwoju ich umiejętności badawczych, jak na przykład przeprowadzanie doświadczenia, i kontekstu historycznego. Obie metody były pozytywnie odebrane przez uczniów i zaobserwowano ich pozytywny wpływ na osiągnięcia uczniów.

Contents

1. Introduction

Biophysics stands at the crossroads of all (STEM) subjects, i.e. physics, biology, and chemistry, as well as mathematics, engineering and technology. This position is privileged for two reasons. First, biophysics benefits from knowledge and discoveries made in all of these subjects. Secondly, it naturally bridges different fields. As such, biophysics provides a more holistic view on science research and would thus be expected to improve attitudes towards science in general. This holistic approach is in line with actual trends in education, and puts forward the idea to teach STEM subjects combined together, without dividing science into more and more specialized disciplines. This helps to develop not only content knowledge, but also a wider range of attitudes and skills.

The ability to look beyond just one field or specialization in science is especially important in our increasingly technological world [1, 2], both for future developments [3] as well as for being able to actively participate in the society [4]. This is reflected by how young people assess their own interest toward science — notably for students, for whom science plays an important role in shaping their lives and future and generally makes them appreciate science professionals [5-10]. Paradoxically, this awareness does not mean that they enjoy science subjects during their education [1, 9, 11]. This is particularly notable for physics, although it is thought by some to be a key subject in science education [12]. This lack of engagement in science is often explained by students becoming disenchanted with science [13]. They perceive school science as irrelevant to their out-of-school experiences and for the society in which they live [11, 14-17].

The problem of “relevance” of school science has been discussed by many scholars. This problem is always connected with a question of how in general school curricula should be prepared and how in particular science education should be carried out [18]. Since the 1990s academic staff members from many universities have become involved in science education, both in theoretical considerations and in practical implementation [18]. Their involvement resulted in a development of various approaches and methods to teach science together with scientific validation of their effectiveness. The popular methods that were developed included inquiry-based learning [19-21], problem-based learning [22-24], and collaborative learning [25-27], just to mention a few. Nevertheless, most of these methods focus on the question "how to teach?", and omit the question "what to teach?". But seemingly every day, especially in current times, brings new discoveries in physics, biology and other disciplines. Therefore, a rift between novel science (physics) discoveries and science (physics) curricula, even at the university level, has occurred and still widens. This rift increases students' sense that physics taught to them and physics from scientific laboratories come from two different worlds.

The overall aim of the work described here is to build a bridge between these two worlds. The aim was specifically pursued by developing novel topics from contemporary physics, introducing these topics to students at various educational levels, validating the effects of introducing these topics on the attitudes of the students, and simplifying complicated experimental setups to address the needs of students. By contemporary physics we mean physics topics that (1) are not older than a few decades, (2) are used in everyday/industrial applications, and (3) are actively studied. The topic introduced to students was percolation, specifically introduced to university students in two forms, one for students seeking a bachelor degree in physics and the other for those in a master program in biophysics. A more in depth discussion of this choice is provided in the following paragraphs. In addition, a simplified experimental setup was prepared for a quantitative study of thermal conductivity, since this common phenomenon manifests itself in various fields such as medicine, physics, technology, and engineering. The development and evaluation of the topics are presented in four of the author's papers:

[A1] *A simplified experimental method to study conductivity percolation*, in which a novel, simplified experimental method is presented. The developed method was further used to devise a way to introduce the topic of percolation to students.

[A2] *Experiment on percolation for Introductory Physics Laboratories–A case study*, which describes an exercise on water network percolation in sand grains, with a lab unit prepared and implemented in Introductory Physics Laboratories, together with students' results and opinions.

[A3] *Water network percolation on yeast as a proposal of an experiment for advanced physics laboratories for bioscience students,* in which an experiment using a living organism, i.e., yeast cells, is presented together with students' results, as a proposal for advanced biophysics laboratories.

[A4] *Simple method for measuring thermal conductivity*, which describes a method enabling quantitative study of the thermal conductivity coefficient to be implemented at school settings.

We presumed that contemporary physics topics taught in an active way (e.g. laboratory experiments) could improve attitudes of students toward physics, therefore they required also novel approaches in assessment. Thus the second part of this work focused on the design, implementation, and verification of two assessment strategies, which enabled us to verify not only the pure content knowledge gained by students, but also their experimental skills, their understanding of historical context, etc. Two types of group-taken assessments are discussed, a tournament and a board game, as described in two other papers by the author:

[A5] *Class tournament as an assessment method in physics courses: a pilot study*. A description of a method involving a tournament for assessing student achievement, together with the results of this assessment and the opinions of the students.

[A6] *Board game in physics classes — a proposal for a new method of student assessment*. An investigation of the effectiveness of a group assessment in the form of a board game on the achievements of high school students and their attitudes towards physics.

Both aspects of our work, namely introducing contemporary physics topics and novel assessment strategies, have a common goal: making physics taught at school more like it is practiced in scientific laboratories, in order to enhance students' understanding and motivation in science. In this dissertation, three theses are addressed:

1. Selected contemporary physics topics could be effectively introduced to students at various educational levels.

2. Contemporary physics topics and novel experiments improve the attitudes of students towards physics and their motivation for learning.

3. Group taken assessment methods increase the amount and retention of knowledge gained by students and encourages them to pursue further learning.

2. Contemporary physics

Introducing contemporary physics topics to students at different educational levels poses several challenges. One challenge involves selecting the topic and its scope. What would be interesting for students and be understandable at their level of knowledge needs to be decided. Moreover, such topics usually require some simplifications. Thus novel topics would appear to be most effectively developed and implemented by scientists who actually are involved in the research on these topics.

Since for many years I have been involved in research on water network percolation in various biomimetic samples, I chose percolation to be the contemporary physics topic for the students in the current study. Percolation is a common transport phenomenon. In general, it describes the behavior of a system in which transport pathways may appear or be destroyed randomly. Since it is an extremely general model, it can be applied to a wide variety of cases. One such example is conductivity percolation in a metal grid connected to an external voltage source, in which bonds could be destroyed randomly (Fig. 1). Accordingly, as the number of bonds in the grid decreases, the current in the grid decreases, until the last bond that allows the current passing from one end to the other is destroyed. The density of undamaged bonds present in the network (or in other words: occupation level) at this point is called the percolation threshold.

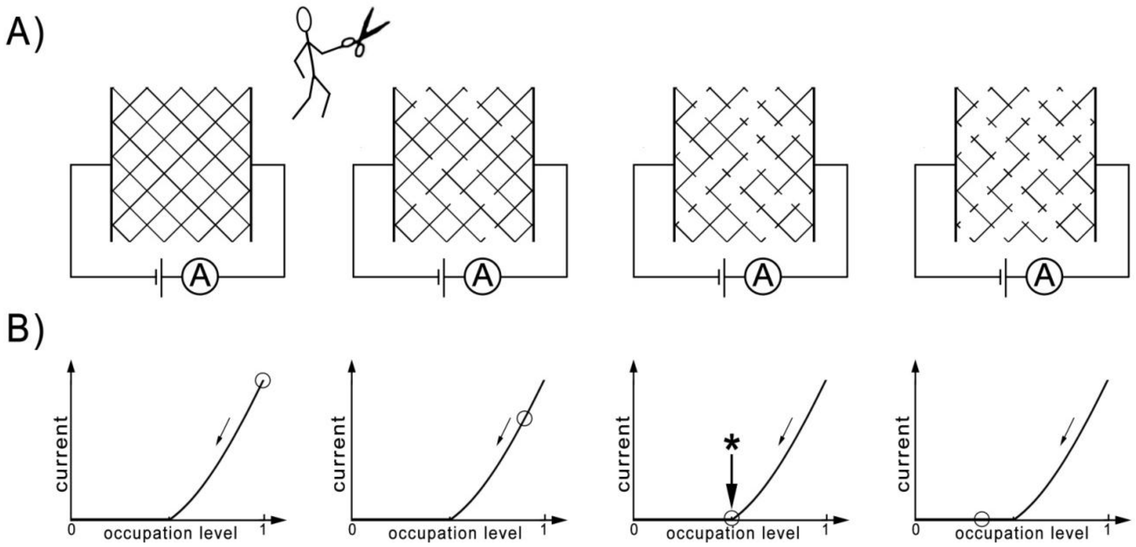


FIG. 1: Representation of the percolation phenomenon in a metal grid. A) Schematic pictures of metal lattices connected to the power source with different occupation levels, defined as the ratio of the number of conductive bonds to the total number of bonds. B) Illustrative graphs of an electric current in a lattice as a function occupation level, points marked with circles correspond to situations in panel A), and the asterisk indicates the percolation threshold. Initially the current passes through the lattice, but when single bonds are gradually and randomly removed, the current decreases until the whole system is disconnected, which means that the current stops. Retrieved from [A2].

Conductivity percolation can be mathematically described using a power-law dependence of the DC conductivity of the network σ and (concentration) occupation probability (level) p in the relationships

$$\begin{cases} \sigma(p) \propto (p - p^*)^t, \ p \geq p^* \\ \sigma(p) \propto (p^* - p)^{-s}, \ p \leq p^* \end{cases}, \tag{1}$$

where values marked with asterisks indicate values at the percolation threshold, and t and s are critical exponents, which reflect the dimensionality of the network. Occupation probability p is a general term, and applications of the above mathematical description to the experimental data require identification of p among measured quantities.

In the past, I was involved in the research of prof. J.K. Mościcki's group, at the Faculty of Physics, Astronomy and Applied Computer Science of the Jagiellonian University, that studied water network percolation on surfaces of biomimetic samples such as silicates, i.e., AEROSIL [28], mesoporous and microporous materials (SBA-15, MCM-41) [29]. The group started its research on percolation almost 20 years ago with studies on live yeast and algae [30, 31]. In that research, DC conductivity of a sample was measured while the sample was being dehydrated into the air. Therefore, as the occupation probability p, two quantities related to the amount of water were used. Namely, water mass fraction, ρ, defined as a ratio of the mass of water to the mass of the sample ($\rho = \frac{m_w}{m}$); and the hydration level, h, defined as the ratio of the mass of water to the mass of a completely dry sample ($h = \frac{m_w}{m_0}$). The equations describing the behavior of the system in the vicinity of the percolation thresholds could be expressed as

$$\sigma = \sigma_Y \frac{(\rho^*)^s}{(\rho^* - \rho)^s}, \quad \rho < \rho^* - \delta_1 \tag{2a}$$

and

$$\sigma = \sigma_W \frac{(\rho - \rho^*)^t}{(1 - \rho^*)^t}, \quad \rho > \rho^* + \delta_2 \tag{2b}$$

for three-dimensional (3D) percolation, and

$$(\sigma - \sigma^*) \propto (h - h^*)^{\mu}, \qquad h > h^* \tag{3}$$

for two-dimensional (2D) percolation. The symbols s and t stand for critical exponents, respectively, below and above the 3D percolation threshold; and μ denotes the critical exponent for 2D percolation. Theoretical and experimental works report t values in the range 1.9-2.2 and s values in the range 0.4-1.2 for 3D percolation, and μ values in the range 0.9-1.3 for 2D percolation. Other symbols in above equations: ρ^* stand for the water mass fraction at the percolation threshold, σ_w is the conductivity of intercellular water, σ_Y is the remnant conductivity. The sum of δ_1 and δ_2 indicates the "width" of the transition range Δ around the percolation threshold and can be related to σ_Y , σ_w , t and s [30].

Originally, conductance and capacitance measurements of a sample were done, using an impedance analyzer, in the whole range of frequencies for dielectric spectroscopy, i.e., 100 Hz to 2 MHz, at each time point. For each time point, the imaginary part of the dielectric loss spectrum ε'' in the whole range of frequencies was determined. The conductivity of a sample was calculated using the equation [28]:

$$\log[\varepsilon''(f)] = \log[\sigma(h)/\varepsilon_0] - (n-1)\log(2\pi f)\,, \qquad (4)$$

with data only from the low-frequency end of the spectrum, where the dielectric loss factor was predominantly determined by the static conductivity and where the spectrum was linear in log-log representation. ε_0 is the vacuum permittivity, f is a frequency of alternating current and n is close to zero in this case. The calculated conductivity was later used to determine the percolation threshold. Finding the percolation threshold required analyses of huge number of plots of conductivity versus hydration level on the log-log scale and linear fits of those plots using equations (2-3). The procedure was repeated until the best chi square was obtained for the largest number of data points. This method of data analysis was used in previously described research done in our group, and yielded results that improved our understanding of the behavior of water in various systems being subjected to dehydration.

While the research and results were interesting, the time-demanding and complicated nature of the data analysis method as well as the requirement of a quite advanced experimental setup limited the possibility of making the topic of percolation more accessible to students. The first obstacle was removed when a novel method of data analysis was developed in our group. In short, in the frequency range

100 Hz to 1 MHz, determination of the conductivity was replaced by an easily accessible monochromatic dielectric loss factor value, measured at any desired frequency in this range [32]. Moreover, the hydration level h was replaced by time-to-failure variable t [33]. Then, a percolation scaling equation was presented in the following form:

$$(\varepsilon_f'' - \varepsilon_f''^{*}) \sim (t^* - t)^{\mu_t};\ \ t \le t^*, \tag{5}$$

which significantly simplified the analysis process. $\varepsilon_f''^{*}$ and t^* are adequately dielectric loss factor and time at the percolation threshold and μ_t is a critical exponent.

The second obstacle, related to the necessity of the use of an impedance analyzer, which is a kind of a "black box" and also hardly accessible to students, encouraged us to develop a new measurement method. The development of such a method constituted the first work of this thesis and is described in the [A1].

2.1 A simplified experimental method to study conductivity percolation - paper [A1]

This paper describes a simplified experimental method developed to study conductivity percolation. The method was developed based on the idea of replacing the impedance analyzer with a more accessible piece of equipment. As taking direct measurements of the conductance of a sample was not possible, a circuit in which a capacitor with a sample inside was connected in series with a resistance decade box and AC generator was used (Fig. 2). The system acted as a simple voltage divider, measurements of selected amplitudes and the phase shift between them enabled for determination of the conductance for a given sample.

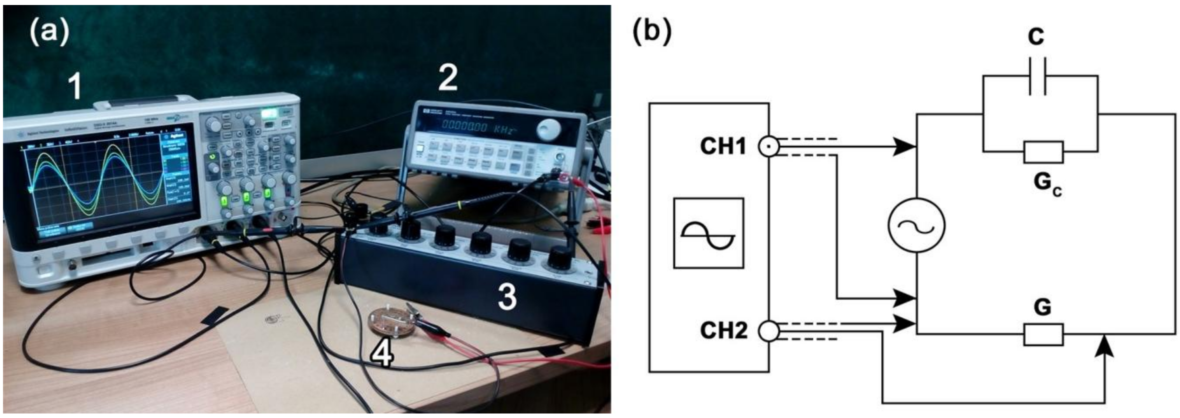


FIG. 2: (a) Photograph and (b) schematic diagram of the experimental setup: (1) oscilloscope, (2) AC generator, (3) resistance decade box, and (4) capacitor. Retrieved from [A1].

The effectiveness of the developed method was verified by comparing the results of water network percolation in a sample of sand grains when using this method with the results of the same measurements taken using the standard method i.e., using an impedance analyzer. The results obtained from the two methods were consistent, which proved that the new, simplified setup could also be used for scientific applications.

In addition to the development of the simplified measurement method for conductivity percolation studies, the development of the method also opened up an additional opportunity in percolation studies, one related to the phase shift and not discussed in the paper. It is related to the phase shift, we detected that changes during the measurements. The changes are obviously related to dehydration of the sample. In the experimental design, a phase shift φ between the voltage amplitude on a resistance decade box U_R and the voltage amplitude on an AC generator U_0 is measured. However, this phase shift is caused by the properties of the capacitor and could be easily found from the phasor diagram (with U_c denoting capacitor voltage). The capacitor is modeled as a parallel connection of the capacitance part and the resistance of a sample, and the phase of currents from both of them differs for $\pi/2$ (the currents I_C and I_{Rc} are perpendicular in the phasor diagram). Both currents have to be added to obtain a net current I_R. This net current is the same as the current through the resistance decade box, because they are connected in series, and, therefore, in phase with the voltage on the resistance decade box. Changes in the ratio of one current to the other in the capacitor would cause the observed change in the phase shift between measured voltages.

This reasoning could be supported with a simple consideration of phasors for wet and dry sample (Fig. 3). During the dehydration of the sample, the resistance part of the capacitor would increase, causing a decrease of the resistance current. At the same time, the capacitance also decreases, leading to an increase of capacitance reactance, and in that way to the decrease of the capacity current. However, during the process, the capacitance changes by two orders of magnitude, while the resistance of a capacitor by ca. six orders of magnitude.

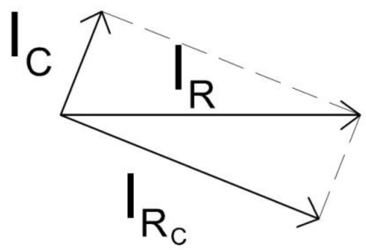
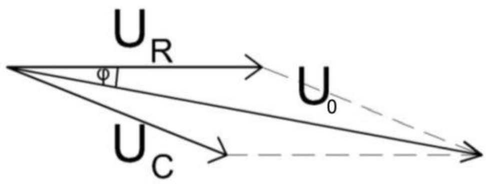
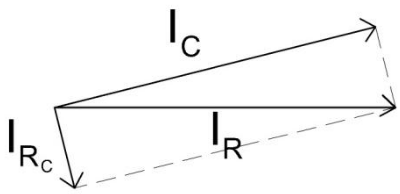
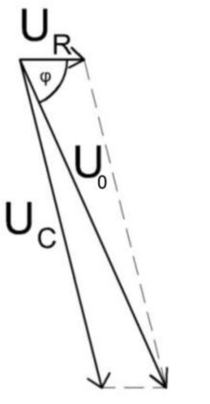
Phasors for wet and dry system. For the clarity of presentation diagrams are not drawn to scale.		
	Current phasor **for the capacitor**	Voltage phasor **for the whole system**
Beginnig of measurements "wet" system		
End of measurements "dry" system		

FIG. 3: Sketch of phasors for the described method for wet and dry system. I_C - capactive current, I_{RC} - resistive current, I_R - net current in the capacitor, U_R - voltage amplitude on a resistance decade box, U_0 - voltage amplitude on an AC generator, Uc - voltage amplitude on a capacitor, φ - phase shift measured in designed setup. Own work.

The crucial point is that the behavior of the system strongly depends on the initial choice of parameters: the resistance of the resistance decade box and the frequency of the current fixed at the AC generator. Depending on whether, for the same frequency, R is large or small, the shape of the phase-versus-time plot may differ. Figure 4 presents such plots obtained for various measurements. Note that for some conditions, the shape of the plot was observed to be characteristic — with the bend found just before the percolation threshold. This result was probably caused by the reorganization of water inside the sample. However, this hypothesis requires further investigation. Nevertheless, the development of the simplified experimental method opened a way to introduce the topic of percolation to students, while the presence of a characteristic shape for the phase change would be expected to significantly reduce the time needed for students to successfully conduct and analyze the experiment.

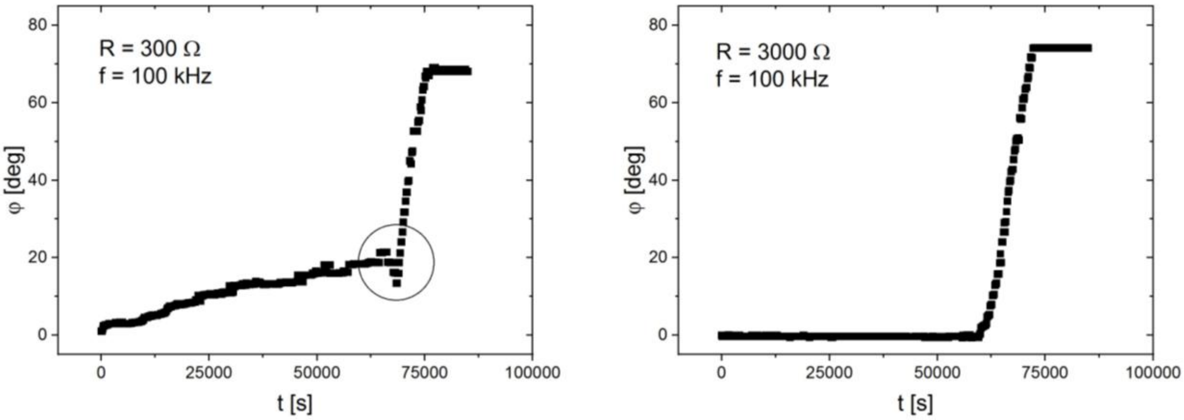


FIG. 4: Plots of measured phase versus time, with the two plots based on the same frequency of an alternating current but different values of resistance set at the resistance decade box. (a) When the resistance was 300 Ω, a characteristic bending (circled) was observed, just before the percolation threshold. (b) When the resistance was increased to 3000 Ω, such shape was not observed.

2.2 Percolation for students

The widely used percolation theory is a good candidate to demonstrate how a recent and interesting scientific topic can be introduced to physics and biophysics students. The aim was not to overwhelm students with technical details and the full potential of the cutting-edge method, but rather to introduce some simplifications without losing the big picture: that relatively basic physics can be used to explain percolation phenomena in the course of biomaterials dehydration. By using a hands-on approach, specifically by collecting and analyzing their own data, unlikely to appear in anybody else's measurements, students were expected to gain an understanding of the meaning of percolation threshold and the critical exponent. Understanding of the concept of percolation, gained through the examination of a simple system, may be extended to more complex ones, such as financial markets or, of greater current importance, the spread of epidemics.

The simplified method described above was used to prepare two different laboratory activities on percolation. The first one was an experimental module on the percolation for Introductory Physics Laboratories (described in work [A2]), in which water network percolation in a sample of moist sand was explored. However, since percolation is a phenomenon common to many fields, we chose another system, more familiar to biophysicists and life scientists, namely yeast (paper [A3]).

2.2.1 Experiment on percolation for Introductory Physics Laboratories – A case study - paper [A2]

The paper describes the design, implementation, and verification of a laboratory module on percolation at the Introductory Physics Laboratories (IPhL). The module was designed according to five goals of the IPhL listed by the American Association of Physics Teachers [34], namely (1) the art of experimentation, (2) experimental and analytical skills, (3) conceptual learning, (4) understanding the basis of knowledge in physics, and (5) developing collaborative learning skills. The module consisted of three main parts. In the first one, the students prepared themselves for the laboratory by using a script that described the phenomenon of percolation, as well as the measurement method to be used, and provided basic background on other relevant physics phenomena. The second part was a laboratory experiment conducted by students in pairs with supervision of a tutor. The laboratory was divided into two sessions: one in the morning and one in the afternoon of the same day. In the morning, pairs of students had to build a setup, choose and adjust parameters for measurements, and hydrate a sand sample with an adequate amount of water. In the afternoon of the same day, the students measured the amplitudes of the voltage on a generator (U_{IN}), on the resistance decade box (U_R) and the phase shift between them (φ), which were further utilized to find the dielectric loss factor (ε'') of the sample, following the equation [A2]:

$$\varepsilon'' = \frac{U_R G}{2\pi f\ C_o} \frac{U_{IN} cos\varphi - U_R}{U_{IN}^2 + U_R^2 - 2U_{IN}U_R cos\varphi}, \qquad (6)$$

where G is the conductance of the resistance decade box, f is the selected frequency and C_0 is the capacitance of the empty capacitor. These measurements had to be done just before the sample dried completely, in order to capture the percolation threshold. In the last part of the module, the students individually prepared scientific reports using the results of the experiment. These reports required data analysis together with the determination of the percolation threshold and dimensionality of the network.

This module was tested on a group of six first-year physics students. Data were collected in three steps: (1) the students were asked to fill out pre- and post-survey on percolation, (2) their reports were assessed by the tutor, and (3) the students were asked to evaluate the module and to comment on it after receiving the final grade. Scrutiny of the answers provided by the students in the percolation survey before and after the exercise showed that by participating in the module, the students acquired basic knowledge on percolation and its characteristics, and became familiar with a few applications of percolation theory. The laboratory reports provided by the students showed that, with use of a simplified experimental setup, they were able to analyze collected data in a correct way and determine the percolation threshold, as well as relate the obtained parameters to the dimensionality of the process. Note, though, that one group was unable to determine the percolation threshold from the data obtained during the experiment, since the sample did not dehydrate enough before the end of the class. These students prepared their laboratory reports based on their data and additionally in an appendix they found a percolation threshold with use of a data set provided by the tutor. This experience showed a weakness of the method of the proposed module — the risk of not capturing the percolation threshold in the allocated time. This risk was also one of few weak points of the module addressed by students in the questionnaire administered at the end of the implementation. However, in general, the students really appreciated the module, and found it interesting, understandable, and motivating. Most of them suggested that such experiment should be introduced in the IPhL for a wider group of students.

This research showed that it is possible to prepare and implement a laboratory module related to a contemporary physics topic like percolation. The proposed topic was understandable to students at their level of knowledge, and they participated in the laboratory willingly; and it could be found in their statements that such topics would increase their motivation for learning.

2.2.2 Water network percolation on yeast as a proposal of an experiment for advanced physics laboratories for bioscience students - paper [A3]

This work was motivated by research that showed that physics could be interesting for bioscience students when they found the studied physics phenomena relevant to their field [35-37]. In the presented paper, a water network percolation in a live sample, namely yeast, undergoing dehydration was proposed as such topic —

an example that binds physics with biology. Since water plays a crucial role in the functioning of living organisms, studying, describing and, to some extent, understanding its behavior seem to be interesting for scientists from different fields.

During the dehydration of yeast sample, the spatial distribution of water molecules inside the organisms has been demonstrated to change continuously [30, 38, 39]. According to these studies, water initially fills the entire space between yeast cells. But with ongoing evaporation, the amount of available water decreases - water forms thinner and thinner layers and water network breaks. This can be observed also in decreasing of sample conductivity. In the course of the dehydration, two rapid changes are observed. Physicists unmistakably recognize them as phase transitions that could be described by percolation theory. The first change is observed as a 3D percolation threshold (i.e., before this threshold water forms a 3D network), and the second one is a 2D percolation threshold (i.e., before this threshold electric current passes through the water layer on yeast surface, while after the phase transition there is no longer a percolation path and water forms only small, separated lakes on the surfaces of the cells, which do not transmit current anymore [40]).

The experiment prepared for biophysics students explored the above-mentioned phenomenon. In this case, the occupation probability p was related to the amount of water in a sample - water mass fraction ρ for 3D percolation and the hydration level h for 2D percolation (Eqs. 2-3). This experiment required measurements of the mass of the sample to be taken constantly. Since the conductivity of the sample also had to be measured during the course of the entire experiment, this measurement was automated. The software used during the experiment is described in the paper. The experiment was carried out by three graduate biophysics students and one graduate biochemistry student. They were asked to conduct the experiment, collect data, and analyze it in order to determine the 3D and 2D percolation thresholds. The students were able to determine both thresholds, and the results obtained were consistent with the earlier research done on yeast, using an impedance analyzer.

This case study showed that it is possible to effectively introduce to bioscience graduate students an exercise about percolation on a living organism. The students were able to conduct the experiment and analyzed the data correctly, which enabled them to find both 3D and 2D percolation thresholds and the percolation critical exponents. Unfortunately, it was not possible to implement the prepared laboratory during the regular course. Not included in the paper was a casual conversation by students in which they suggested that such experiments might be introduced to biophysics students instead of "pure physics" experiments and would be closer to their area of interest. However, such exploration could be a subject of further research.

2.3. Bringing advanced experiments into school laboratories

Despite the opportunity to introduce some contemporary physics topics into regular physics laboratories and therefore influencing students' perceptions of science, there are still topics in the curriculum that could be taught only as lectures because there is no available experiment allowing for quantitative measurements of the discussed phenomena at school laboratories. Surprisingly, one such topic is thermal conductivity. Probably most people are familiar with the experiment in which paper clips attached by a piece of wax to different metal rods heated at one end, clearly showing the dependence of heat flow on the type of material. However, determining the thermal conductivity values for materials is not possible in such experiment. Measurements that are taken to obtain quantitative data require a well-engineered experimental setup or advanced mathematical analysis of the data, which limit the possibility of utilizing them at school. Therefore, even though most people likely know intuitively that polystyrene has a different thermal conductivity than, for example, glass, it is impossible to measure exact values of these coefficients during the lessons. This gap was addressed in the work described in paper [A4].

2.4.1. Simple method for measuring thermal conductivity - paper [A4]

A novel, simple method for measuring thermal conductivity, adequate for school circumstances, was developed in the work presented in the paper. The experimental setup is presented in Figure 5. It consisted of a container (e.g., a plastic cup) filled with hot water and submerged into water with ice. Additionally, this container was covered by a polystyrene plate, which reduced the transfer of heat to the walls of the cup. Immersion of the container in water with ice made the amount of heat that radiated away negligible.

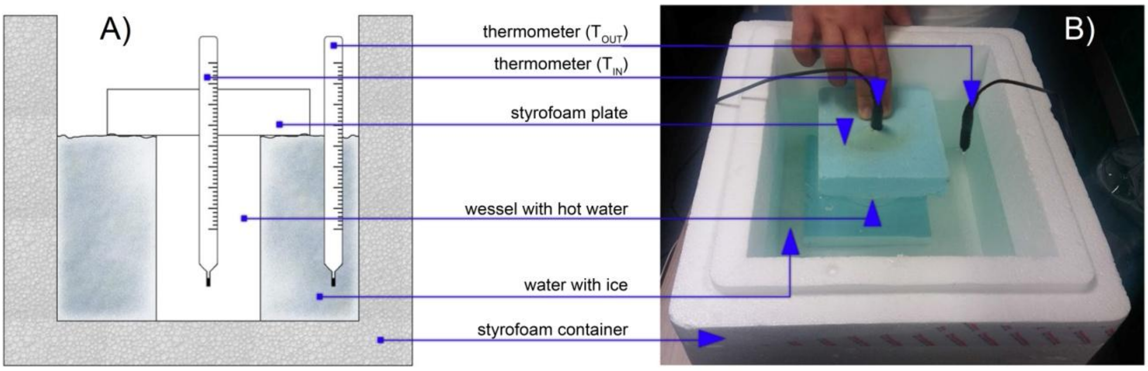


FIG. 5: A) Sketch and B) photograph of the experimental setup. Retrieved from [A4].

During the experiment, the temperature of water inside the cup was measured, with the results following Newton's law of cooling. However, a trick was done in the data analysis. In the approach, two temperature differences were calculated. The first one involving the change of the internal energy of water due to the absorbed or released heat, with the related change of the water temperature calculated using the equation $\Delta T_w(t) = T_{W0} - T_W(t)$, where T_{W0} is initial temperature of the water inside the container and $T_W(t)$ is the actual temperature of it. The second one being the magnitude of the difference between the temperature of the reservoir, T_R, and the temperature of the water in the container, $T_W(t)$, i.e., $\Delta T_Q(t) = T_W(t) - T_R$. Due to the linear dependence of the ratio $r(t) = {}^{\Delta T_w}\!/_{\Delta T_Q}$ on time, a simple reorganization of the data allowed for determination of the thermal conductivity for a material from which the cup was made of by carrying out a simple linear fitting. Under these conditions, the coefficient of thermal conductivity could be estimated directly from the slope of the plot of $r(t)$ versus time, as shown in Figure 6B. The results obtained for poor conductors were highly consistent with literature values, but for good conductors a discrepancy was observed. Possible reasons for this discrepancy are discussed in the paper. Nevertheless, the simplicity, intuitiveness, and the low cost of the experiment are the features that make the proposed experiment an excellent tool to use when introducing the topic of thermal conductivity at schools or even at IPhL. Having this experiment carried out by students also shows that the physical properties of different materials are not just abstract numbers measured or calculated by engineers, but are real entities that have consequences for the observed properties of matter.

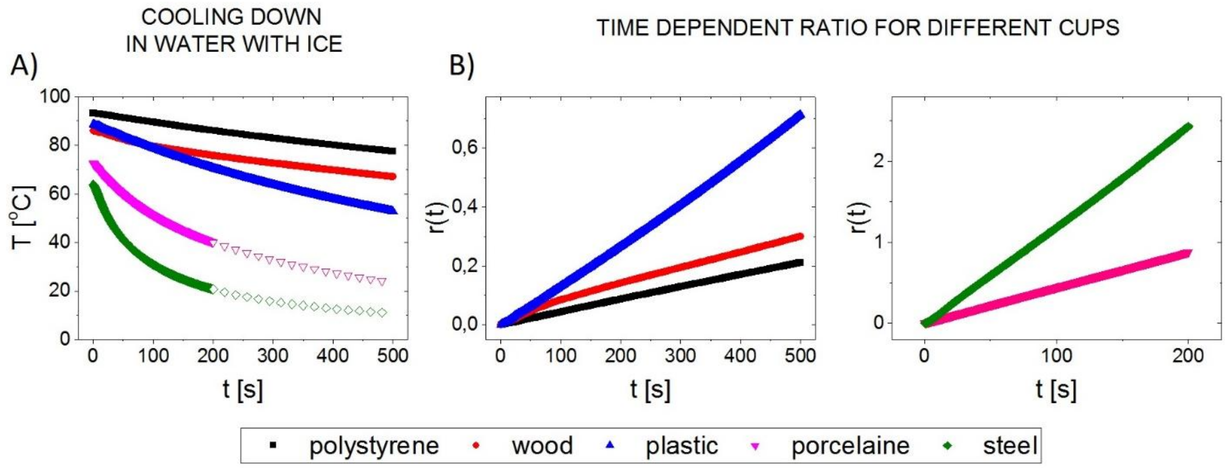


FIG. 6 A): Temperature of water in a cup immersed in a mixture of water and ice as a function of the duration of this immersion for each of various cups made of different materials as indicated. Circled points were not used in the analyses, but are shown to illustrate the course of further melting. B): The ratio $\boldsymbol{r(t)}$as a function of the duration of the immersion and calculated from the measured temperatures for the five studied cups. The data are presented in two separate graphs with different scales on the axes, as the changes of r(t) for porcelain and steel are much larger than for the other materials. Retrieved from [A4].

An additional finding not presented in the paper was that the experiment and the data analysis process were understandable for high-school students. A few of them were asked to perform the experiment at home, with a standard thermometer and cups they found at home. All of them conducted the experiment correctly (with time intervals between 15 and 120 s) and received values of thermal conductivities comparable with the ones in the literature (see Fig. 7).

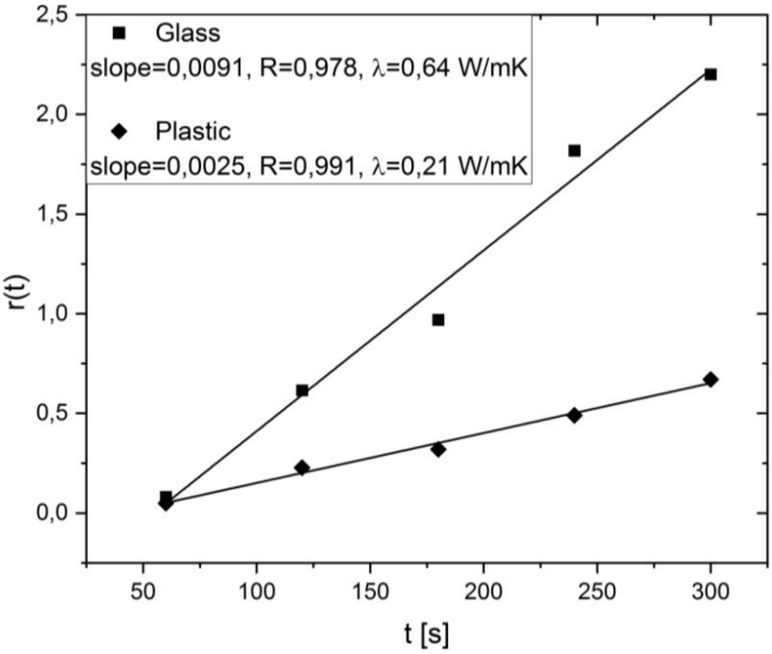


FIG. 7: Examples of students' measurements done with glass and plastic mugs in 1 min intervals together with calculated values of thermal conductivity.

3. Assessments in physics/science courses

Enhancement of motivation and interest requires not only novel topics and active methods of teaching, but is also close coupled to active assessment methods. Nowadays, most teaching methods focus not only on content knowledge, but also on the development of experimental skills as well as soft skills. However, despite the new approach used for instruction, at the end of the unit/course/semester, students are usually still assessed in a traditional way, with the use of a multiple-choice test or calculus tasks. This difference has been found to lead to a paradox already reported [41], namely a lack of a correlation between the achievements of students in Introductory Physics Laboratories and their results in final exams. Since most agree that active methods of learning have a lot of advantages, a natural consequence should be a change in assessment approaches, to bring them to a consonance with utilized methods.

Such assessments could be done in various ways, among others by the use of collaborative exams or summative-formative assessments [42-45]. A more detailed description of these assessment methods and their outcomes is provided in the paper [A5]. In the work for this dissertation, two assessments methods based on group-taken activities were prepared and implemented in physics classes. The first method used was similar to a tournament; while in the second one, a board game was utilized.

3.1 Class tournament as an assessment method in physics courses: a pilot study - paper [A5]

In this paper, a novel method for assessment is presented. The method was designed to be based on a tournament, in which all teams got through sets of various questions, solving them simultaneously. The assessment was implemented to one class of 30 high-school students after teaching the unit about electricity, in which relatively many experiments (taking measurements of voltage or current, constructing a proper circuit, conducting electrolysis of water) were performed by students and could not be assessed in a traditional way. Students formed groups randomly and were asked questions in order of increasing level of difficulty, i.e., simple open questions, multiple-choice questions, and more complicated open questions. After a few rounds of such questions, each group simultaneously received a different experiment to conduct and was asked to provide a correct explanation for their results, as well as being given a calculus task. At the end, an extra task, involving a complex maze of resistors, was prepared for all groups (see Fig. 8).

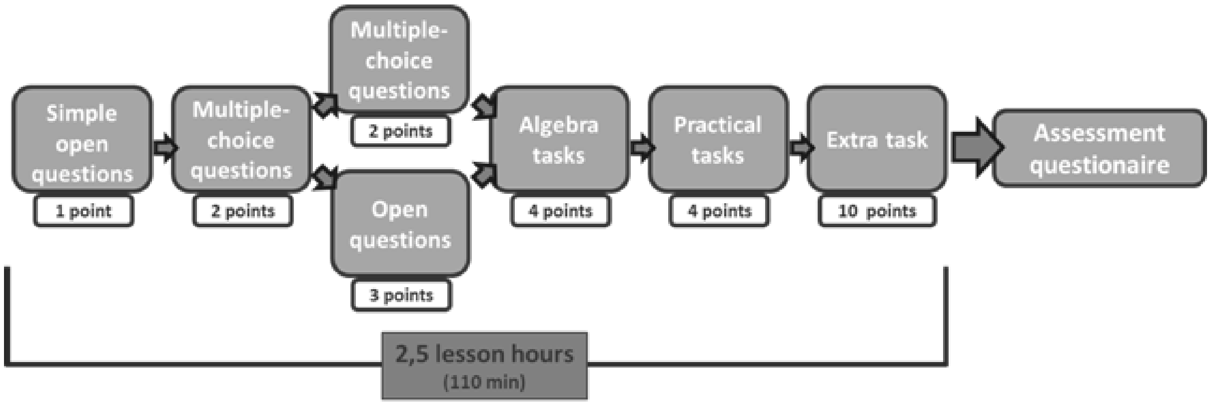


FIG. 8: Tournament testing sequence. Retrieved from [A5].

To assess the individual impact of every student on the overall result, a special assessment questionnaire was administrated to students after the tournament. Each student had to carry out a self-assessment and a peer-assessment of other fellow players from the same group, specifically addressing subject matter contribution and communication skills. This self-assessment and peer-assessment were included with an adequate weight in the final grade of each student, together with common results obtained by the group in the tournament. (A detailed description of this set of assessment tools is provided in the paper).

The effectiveness of the tournament assessment method was verified by comparing the students' results in the former tests with the results in the tournament and results in the unannounced traditional test on electricity taken one week later. Scores are shown in Fig. 9. Students received higher scores in the tournament than in the former tests, a difference that can be associated with their cooperation during the game. Also note that the post-test scores were higher than those of the former tests. Based on these differences, the tournament was concluded not only to assess the gain in knowledge by the students, but also to increase the retention of this knowledge. The superiority of the tournament was also reflected in the opinions collected from the students after the intervention was completed. Students appreciated the approach, emphasizing a positive impact of the cooperation in a team on their achievements and a motivating role of elements of rivalry. Moreover, they concluded, that such a way of assessment created an opportunity to revise or even learn something during the tournament, including what they had missed earlier.

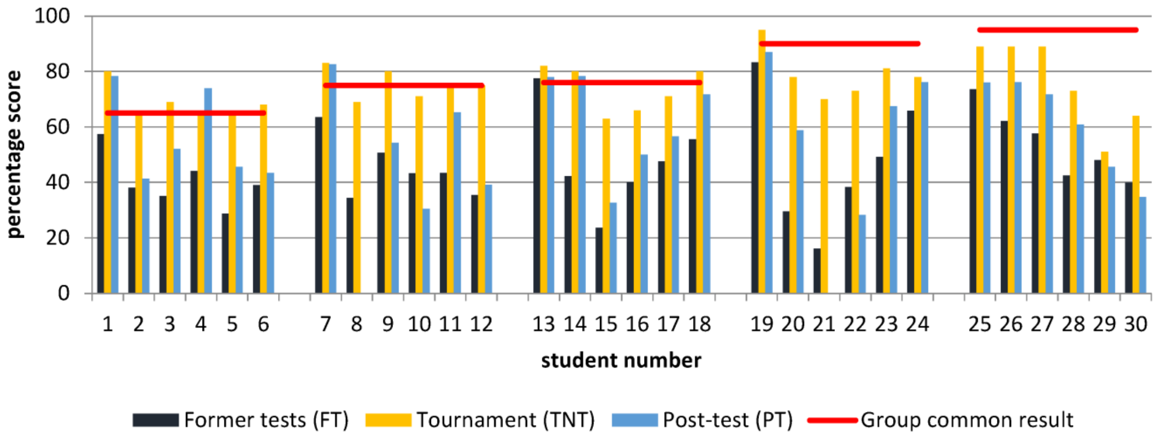


FIG. 9: Student test scores. For each student, the three vertical bars represent, from left to right, the average score in former tests (FT), the final score obtained in the tournament (TNT), and the mark gained in the traditional post-tournament test (PT), respectively. The horizontal lines represent the common result scored by each group in the tournament (based only on the first six stages, disregarding the qualitative component stemming from peer-assessment and self-assessment). Retrieved from [A5].

3.2 Board game in physics classes — a proposal for a new method of student assessment - paper [A6]

Another assessment method in the form of a board game played by groups of student is presented in paper [A6]. The approach was tested in two high schools on 131 students in total, divided into experimental and control groups. The board game used for this assessment strategy consisted of a circular game path composed of spots with various categories: physics phenomena charades, famous people, short-answer questions, multiple-choice questions, and simple experiments. The groups were made up of 4-5 students each, and the students moved their tokens according to the certain rules described in the paper. There were also two special lines marked on the board game. When each line was crossed by any group for the first time, the game was paused and all of the groups received simultaneously an algebra task and a complex experimental task to solve which were assessed after the game. The assessment in the proposed board-game form was done after implementation of two different units: 1) waves and vibrations, and 2) optics, because both of them provided an opportunity to verify not only the students' content knowledge, but also their experimental skills, socio-historical context and everyday life context, which together build their science literacy.

The method proposed in the paper was verified with respect to its influence on the students' content knowledge and their motivation. The influence on the content knowledge of the students was evaluated by comparing the achievements of the students in the former tests with their achievements in the board game (experimental groups) and with the achievements of the students in a traditional test (both, experimental and control groups). The traditional test was used instead of the board game by students from the control groups, while the students from the experimental groups took an unannounced test one week after the game. Comparison of the results showed that the achievements of students in the game were higher than those in the former tests. Moreover, students from experimental groups received on average better results in a standard test than did their colleagues from the control groups. Based on these observations, carrying out the assessment in the form of the board game taken by groups of students was concluded to lead to increased achievements of students both during and after the assessment process.

The influence of the assessment method on student motivation was verified by a short questionnaire based on a Likert-like scale, filled in by students after participation in the game. The questionnaire consisted of six questions, related to the students' pre-test preparation, engagement in a team work, difficulty in answering questions, test anxiety, final acquisition of knowledge, and motivation for future learning. There was also a comment box for general, unstructured student comments. In general, students very much appreciated the method, and they stressed that using the method increased their engagement, preparation, and final acquisition of knowledge. Moreover, following this method reduced stress related to the assessment and motivated them for future learning. While the students generally expressed enthusiasm in their free opinions, a few of them pointed out weaker parts of the method, such as a patchy distribution of types of questions or a statement that historical context should not be assessed. Nevertheless, such opinions, while very valuable, were in the minority.

4. Summary

There is a general feeling that physics, although important for understanding the world around us, remains unpopular among students. In this dissertation, two ways in which this perception could be changed were proposed: (1) introducing a contemporary physics topic and novel experiments for students and (2) developing two new, active method of assessment. I summarize the discussed issues in the form of three theses, which are explicitly presented in the articles.

Thesis 1. Selected contemporary physics topics could be effectively introduced to students at various educational levels.

The phenomenon of percolation was taken as an example of a contemporary physics topic that was introduced to the university students. First, a simplified experimental setup for water network conductivity percolation studies was developed [A1]. This step was crucial, since in general complicated and expensive experimental setups limit the possibility of introducing contemporary phenomena to students. The developed method was later used to prepare experiments on water network percolation in sand samples for physics students [A2] and water network percolation in living yeast for bioscience students [A3]. In both cases we concluded that the students were able to successfully conduct experiments and analyze obtained data to determine percolation thresholds. Moreover, the students generally appreciated the prepared experiments and stressed that these experiments were more interesting for them than were standard experiments.

So we conclude that it is possible to effectively implement selected contemporary physics topics at various educational levels. However, it requires a few simplifications and focusing only on some selected aspects, chosen to fit the given target group. Here, percolation was introduced as a topic for students pursuing master's and bachelor's degrees, but it would seem to be possible to introduce the phenomenon of percolation also at the high school level, at least for more advanced classes.

Thesis 2. Contemporary physics topics and novel experiments improve the attitudes of students towards physics and their motivation for learning.

An important aspect of the described implementations [A2, A3] was the topic of percolation itself, which is widely found and studied in many disciplines. Since students were given a chance to reproduce the actual experiments conducted in scientific laboratories, they had the feeling of participating in something "real", important, and useful. This feeling was in line with the findings already referenced to in paper [A2]: that those students who were confronted with more difficult, yet interesting problems, achieved better results. Other research showed that participation in laboratory sessions, during which students took more complicated and less predictable measurements, could motivate them and help them understand the introduced concepts better, as referenced to in paper [A2]. In our study, the students who took part in the laboratory sessions on percolation reported high levels of involvement and evaluated proposed experiments very positively. They emphasized that the experiment on percolation was much more interesting for them than were standard experiments conducted during their regular laboratory sessions. The obtained data also showed that introducing contemporary physics topics (or, in general, science topics) to students is beneficial, and could improve their attitude towards physics and motivation for learning the subject.

Another simple experiment, this time aimed at determining the coefficient of thermal conductivity, was proposed for high school students [A4]. Thermal conductivity is a phenomenon widely used in various fields (engineering, medicine, and others), but its quantitative measurements usually require a sophisticated set up. The proposed experiment utilized simple, easily accessible materials and allowed students to obtain reasonable values of the thermal conductivity for various materials. Again, this approach resulted in students achieving a better understanding of the physical problem at hand and a greater motivation for learning.

In summary, introduction of contemporary physics topics and simple, novel experiments, which allowed the students to take quantitative measurements of phenomenologically discussed phenomena, i.e. percolation and thermal conductivity, improved the attitudes of students towards physics and their motivation for learning. In all studied cases, the students emphasized that with support of novel topics and methods physics appeared as an active and lively discipline. They stressed that such topics could motivate them and others to learn physics. Of course, these findings should be verified using larger samples and with participants that are also less motivated from the start.

Thesis 3. Group taken assessment methods increase the amount and retention of knowledge gained by students and encourages them to pursue further learning.

New approaches to active teaching in physics need novel assessment methods. They need to concentrate not only on assessing the pure content knowledge, but also experimental and soft skills, which are important when preparing new laboratories or activities for students. The developed methods were based on playing a game in a group, and included a tournament [A5] and a board game [A6]. In both approaches, various types of questions and tasks were posed to students, which enabled verification of whether the students were able to conduct experiments presented during lessons. The results achieved by students in the games, as well as their results in a traditional test taken one week later, were better than their results in former tests as well as results of students from the control group. At the same time, the students generally appreciated such evaluation more than traditional methods, and they emphasized that besides testing knowledge, this form of assessment also gave them the opportunity to learn from one another and to verify their understanding during a group discussion.

This approach could be simply adapted to different topics and subjects, and could open an avenue to a more holistic assessment in general. Due to a peer-assessment component, everyone felt responsible for the group results and tried to prepare and do their best. This component seemed to be especially useful for weaker students, in particular those having problems with calculations, and by such an approach they were not doomed to failure, as in typical tests with algebra tasks. Therefore, based on the research results, the active methods of assessment were concluded to be able to improve student achievements in physics and motivate them for further learning.

Further issues

The work described in this dissertation can be a foundation for further research, focusing on combining both approaches, i.e., introducing contemporary physics topics to students together with verifying their achievements via an active method of assessment, like the proposed games played in a group. It is expected to be interesting to verify whether the positive effects of both approaches combine according to the superposition principle or whether instead there is a "saturation point". Additionally, long-term effects could be interesting to investigate – both for individuals and groups (classes), as well as both for separated and combined approaches.

Hopefully, further research will be continued based on the main message of this dissertation: that bringing physics closer to students, by introducing contemporary topics, novel experiments and active methods of assessment, could significantly improve their attitudes toward physics and motivation for further learning.

Bibliography

[1] Hofstein, A., Eilks, I. and Bybee, R. (2011). Societal issues and their importance for contemporary science education: A pedagogical justification and the state of the art in Israel, Germany and the USA. International Journal of Science and Mathematics Education, 9(6), 1459–1483.

[2] Bradley J. D. (2005). Chemistry education for development. Chemical education international 7. Retrieved 9 May 2020 at: https://old.iupac.org/publications/cei/vol6/index.html

[3] Burmeister, M., Rauch, F. and Eilks, I. (2012). Education for Sustainable Development (ESD) and secondary chemistry education. Chemistry Education Research and Practice, 13, 59-68.

[4] Roth, W.-M. and Lee, S. (2004). Science education as/for participation in the community. Science Education, 88, 263–291.

[5] Awan, R., Sarwar, M., Naz, A. and Noreen, G. (2011). Attitudes toward science among school students of different nations: A review study. Journal of College Teaching and Learning, 8(2), 43–50.

[6] Collins, A., Brown, J. S. and Newman, S. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L. B. Resnick (Ed.), Knowing, learning and instruction (pp. 453–494). Hillsdale, NJ: Erlbaum.

[7] Jenkins, E. W. and Nelson, N. W. (2005). Important but not for me: Students' attitudes towards secondary school science in England. Research in Science & Technological Education, 23, 41–57.

[8] Osborne, J. and Collins, S. (2001). Pupils' views of the role and value of the science curriculum: A focus-group study. International Journal of Science Education, 23, 441–467.

[9] Osborne, J. and Dillon, J. (2008). Science education in Europe: Critical reflections. London: King's College.

[10] Sjøberg, S. and Schreiner, C. (2005). How do learners in different cultures relate to science and technology? Asia-Pacific Forum on Science Learning and Teaching, 6, 1–17.

[11] Holbrook, J. (2008). Introduction to the Special Issue of Science Education International devoted to PARSEL. Science Education International, 19, 257–266.

[12] Teodoro, V. D. and Neves, R. G. (2010). Mathematical modelling in science and mathematics education. Computer Physics Communications, 182, 8–10.

[13] Aikenhead, G. (2006). Science education for everyday life. NewYork: Teachers' College Press

[14] Brickhouse, N. W. and Potter, J. T. (2001). Young women's scientific identity formation in an urban context. Journal of Research in Science Teaching, 38(8), 965–980.

[15] Calabrese Barton, A. (1998). Reframing science for all through the politics of poverty. Educational Policy, 12, 525–541.

[16] Dillon, J. (2009). On scientific literacy and curriculum reform. International Journal of Environmental & Science Education, 4, 201–213.

[17] Gilbert, J. K. (2006). On the nature of 'context' in chemical education. International Journal of Science Education, 28, 957–976.

[18] Stuckey, M., Hofstein, A., Mamlok-Naaman, R. and Eilks, I. (2013). The meaning of 'relevance' in science education and its implications for the science curriculum. Studies in Science Education, 49:1, 1-34.

[19] Khalick, A. (2004). Inquiry in science education: International perspectives. Science Education, 88(3), 397-419.

[20] Lazonder, A. and Harmsen, R. (2016). Meta-Analysis of Inquiry-Based Learning: Effects of Guidance. review of Educational Research, 86(3), 681-718.

[21] McDermott, L. (1996). Physics by Inquiry vols I and II (New York: John Wiley).

[22] Barron, B., Schwartz, D., Vye, N., Moore, A., Petrosino, A., Zech, L. and Bransford, J. (1998) Doing With Understanding: Lessons From Research on Problem- and Project-Based Learning. Journal of the Learning Sciences, 7(3), 271-311.

[23] Hmelo-Silver C. (2004). Problem-based learning: what and how do students learn? Educational Psychology Review, 16, 235–266.

[24] Mody, A. and Pradhan, H. (2011). Problem Based Learning in Basic Physics - I. School Science 49(3), 1-7.

[25] Reiner, M. (1998). Thought experiments and collaborative learning in physics. International Journal of Science Education 20(9), 1043-1058.

[26] Adolphus, T., Alamina1, J. and Aderonmu, T. (2013). The Effects of Collaborative Learning on Problem Solving Abilities among Senior Secondary School Physics Students in Simple Harmonic Motion. Journal of Education and Practice, 4(25), 95-100.

[27] Ismayati, E. (2018), The Design of Collaborative Learning for Teaching Physics in Vocational Secondary School. OP Conference Series: Materials Science and Engineering, 336, 012040.

[28] Sokolowska, D., Dziob, D., Gorska, U., Kieltyka, B. and Moscicki, J. K. (2013). Electric conductivity percolation in naturally dehydrating, lightly wetted, hydrophilic fumed silica powder. Physical Review E, 87(6).

[29] Nowak. J., Dziob, D., Rutkowska, M., Chmielarz, L., Gorska, U., Sokolowska, D. and Moscicki, J.K. (2015). DC conductivity percolation on drying moistened mesoporous silica SBA-15. Micropor. Mesopor. Mat., 206, 202-206.

[30] Sokolowska, D., Krol-Otwinowska, A. and Moscicki, J. K. (2004). Water-network percolation transitions in hydrated yeast. Phys. Rev. E, 70, 052901.

[31] Sokolowska, D., Krol-Otwinowska, A., Bialecka, M., Fiedor, L., Szczygiel, M. and Moscicki, J. K. (2007). Water-network percolation transition in hydrated blue-green algae in vivo. Journal of Non-Crystalline Solids, 353 (47-51), 4541-4545.

[32] Moscicki, J. K., Sokolowska, D., Kwiatkowski, L., Dziob, D. and Nowak, J. (2014). Note: Optimization of the numerical data analysis for conductivity percolation studies of drying moist porous systems. Review of Scientific Instruments, 85(2).

[33] Moscicki, J. K. and Sokolowska, D. (2013). Dehydration kinetics and conductivity scaling in time near the continuity failure of a water wetting film in granular media. Applied Physics Letters, 103(26).

[34] AAPT, American Association of Physics Teachers (1998). Goals of the introductory physics laboratory. American Journal of Physics, 66(6): 483–485.

[35] Redish. E., Bauer, C., Carleton, K., et al. (2014). NEXUS/Physics: An interdisciplinary repurposing of physics for biologists. American Journal of Physics, 82(5), 368-377.

[36] Meredith, D. and Redish E. (2013). Reinventing physics for life-sciences majors. Physics Today, 66(7), 38-43.

[37] Watkins, J. and Elby, A. (2017). Context dependence of students' views about the role of equations in understanding biology. CBE Life Science Education, 12(2), 274-286.

[38] Attfield, p., Kletsas, S., Veal, D., Van Rooijen, R. and Bell P. (2001). Use of flow cytometry to monitor cell damage and predict fermentation activity of dried yeasts. Journal of Applied Microbiology, 89(2), 207-214.

[39] Benley, L., de Maranon, I., Marechal, P. and Gervais, P. (2000). Influence of thermal and osmotic stresses on the viability of the yeast Saccharomyces cerevisiae. International Journal of Food Microbiology, 55(1-3), 275-279.

[40] Prof. J.K. Moscicki, private communication.

[41] Holmes, N. and Wieman, C. (2018). Introductory physics labs: we can do better. Physics Today, 71(1), 38.

[42] Duane, B. T. and Satre, M. E. (2014). Utilizing constructivism learning theory in collaborative testing as a creative strategy to promote essential nursing skills. Nurse Education Today, 34(1), 31–34.

[43] Kapitanoff, S. H. (2009). Collaborative testing: Cognitive and interpersonal processes related to enhanced test performance. Active Learning in Higher Education, 10(1), 56–70.

[44] Wininger, S. (2005). Using your tests to teach: Formative summative assessment. Teaching of Psychology, 32(3), 164–166.

[45] Wieman, C., Rieger, W. and Heiner, C. (2013). Physics exams that promote collaborative learning. The Physics Teacher, 52(1), 51-53.

Author's publications

All of them have been included in the respective formats of the journals in which they originally appeared.

A simplified experimental method to study conductivity percolation

Cite as: Rev. Sci. Instrum. 90, 076105 (2019); doi: 10.1063/1.5096969
Submitted: 21 March 2019 • Accepted: 18 June 2019 •
Published Online: 11 July 2019

Daniel Dziob,[1,2] Dagmara Sokołowska,[1,a)] Marcin Młynarczyk,[3] and Józef K. Mościcki[1]

AFFILIATIONS

[1]Smoluchowski Institute of Physics, Jagiellonian University, Lojasiewicza 11, 30-348 Krakow, Poland
[2]Department of Biophysics, Jagiellonian University Medical College, Sw. Lazarza 16, 31-530 Krakow, Poland
[3]Department of Computer Science and Econometrics, College of Economics and Computer Science, Sw. Filipa 17, 31-150 Krakow, Poland

[a)]**Author to whom correspondence should be addressed:** ufdsokol@cyf-kr.edu.pl.

ABSTRACT

A simplified experimental setup and a measurement method adapted to the simplified data analysis proposed in previous works are described in detail. The setup is intended for investigation of the conductivity percolation in the water network formed on granular materials and studied in the course of dehydration. The results show that the new, easily accessible experimental setup enables determination of the principal percolation parameters with the same accuracy as the more sophisticated equipment employing an impedance analyzer. Mean critical exponents obtained from data collected for the samples of moisturized sand grains sized 0.6–0.8 mm by means of both experimental methods give the same results within the limit of measurement uncertainty.

Percolation phenomenon relates to a broad range of transport phenomena occurring in various physical, social, or even market networks.[1] Their critical behavior in the course of a random network decay is described by the universal theory of percolation[2] that provides information of the network characteristics. Among others, the conductivity percolation enables the insight into the charge transport and electric properties of the conducting networks in powdered, granular, or porous media.[3–5] In the past several years, we have studied DC conductivity percolation in water networks on baker's yeasts,[6] algae,[7] and granular silica[8,9] on the basis of the impedance measurements carried out with the use of an impedance analyzer during the free dehydration of the sample through air in seclusion. In such systems, DC conductivity is a principal ion transport property that dominates dielectric spectra in the low frequency region (i.e., 100 Hz–100 kHz). Thus, in the vicinity of a critical point, denoted as a percolation threshold, DC conductivity, σ_{DC}, follows the universal law,

$$\sigma_{DC}(p) \sim \left(p - p^*\right)^{\mu}, \quad p \geq p^*, \tag{1}$$

where p is the density of occupations in the network, μ is a critical exponent informing about the dimensionality of the network,[10] and asterisks denote values at the percolation threshold. In the case of the conductivity percolation on 2D networks, p is represented by the sample hydration, h, defined as the ratio of the mass of water present in the sample to the mass of a dry sample. We showed[11] that the estimation of the percolation primary parameters, i.e., the percolation threshold and the value of the critical exponent, can be done if h is replaced by time variable, t, and that the time-consuming procedure of the data analysis can be even further optimized by considering the alteration of the percolation scaling Eq. (1),[12]

$$\left(\varepsilon_f'' - \varepsilon_f''^*\right) \sim \left(t^* - t\right)^{\mu_t}, \quad t \leq t^*, \tag{2}$$

where ε_f'' is the monochromatic dielectric loss factor value measured at the certain frequency f.

Conductivity percolation measurements of the moisture material in the course of dehydration require recording of dielectric permeability changes in the test sample over time. All our previous studies mentioned above were carried out either with the use of the Hewlett Packard 4192A impedance analyzer or the Agilent E4980A RLC precision meter, which allow the recording of capacitance (C) and conductance (G). However, both devices are quite expensive and not commonly available in research laboratories or student labs. Since the method of examination of the conductivity percolation

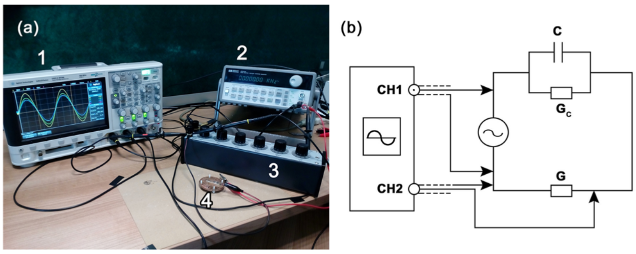


FIG. 1. (a) The configuration of the experimental setup *in situ*: (1) oscilloscope, (2) AC generator, (3) resistance decade box, and (4) capacitor and (b) a schematic diagram of the experimental setup.

has been enormously simplified, resulting in the analysis of data based on Eq. (2), we propose here the consequent simplification of the experimental setup based on the AC generator and oscilloscope, more often accessible in physics laboratories and thus enabling a broader deployment of the conductivity percolation studies.

The configuration of the experimental setup and its schematic diagram are shown in Fig. 1. The setup consists of an alternating current generator connected in series with a resistance decade box and a sample closed in a copper capacitor. An oscilloscope connected to the circuit records the input voltage (on generator), the voltage on the resistance decade box, and the phase shift, φ, between these two electronic signals, U_{IN} and U_{OUT}; see Fig. 2. The experimental setup is a typical voltage divider composed of the capacitor with the sample inside, having admittance Y_1, and the resistance decade box having admittance Y_2 [see Fig. 2(a)]. The capacitor with the sample inside can be modeled in a simplified manner as a pure capacitor with capacitance C connected in parallel with a resistor with conductance G_C [see Fig. 2(b)].

The complex admittance of the circuit is given as

$$Y = Y_1 + Y_2 = (G + G_C) + i2\pi fC, \tag{3}$$

where G is the conductance of the resistance decade box and f is the frequency of the AC voltage source electronic signal. Thus, the relationship between the input voltage U_{IN} and the output voltage U_{OUT} of the voltage divider is

$$\frac{U_{IN}}{U_{OUT}} = k = \frac{Y_1}{Y_1 + Y_2} = \frac{G_c + i2\pi fC}{G + G_c + i2\pi fC} = \frac{G_C^2 + GG_C + (2\pi fC)^2 + i2\pi fCG}{(G + G_C)^2 + (2\pi fC)^2}. \tag{4}$$

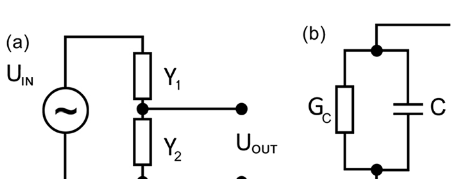


FIG. 2. (a) Experimental setup presented schematically as a voltage divider composed of a capacitor with a sample with admittance Y_1 and the resistance decade box with admittance Y_2 and (b) a sample with admittance Y_1 modeled as a pure capacitor with capacitance C and a resistor with conductance G_C.

The module of expression in Eq. (4) equals the ratio of input and output voltage amplitudes, U_{IN_0} and U_{OUT_0}, respectively,

$$\frac{U_{IN_0}}{U_{OUT_0}} = |k| = \frac{\sqrt{\left(G_C^2 + GG_C + 2\pi f C^2\right)^2 + (2\pi fCG)^2}}{(G + G_C)^2 + (2\pi fC)^2}, \tag{5}$$

and the phase shift between the input and output signals is expressed by

$$\varphi = \mathrm{arctg}\left(\frac{2\pi fCG}{G_C^2 + GG_C + (2\pi fC)^2}\right). \tag{6}$$

In both Eqs. (5) and (6), G_C and C are the only unknown parameters, which can be derived after some calculation and expressed by measurable parameters,

$$C = \frac{|k|(G + G_C)}{\omega\sqrt{1 + \frac{1}{tg^2\varphi} - \frac{|k|}{tg\varphi}}} \tag{7}$$

and

$$G_C = \frac{\cos\varphi - |k|}{|k|^2 + 1 - 2\cos\varphi|k|}|k|G = \frac{U_{IN_0}\cos\varphi - U_{OUT_0}}{U_{IN_0}^2 + U_{OUT_0}^2 - 2U_{IN_0}U_{OUT_0}\cos\varphi}U_{OUT_0}G. \tag{8}$$

The value of the sample conductance G_C, calculated from experimental data gathered at the certain frequency of the electronic signal from an AC voltage source, leads to the value of the

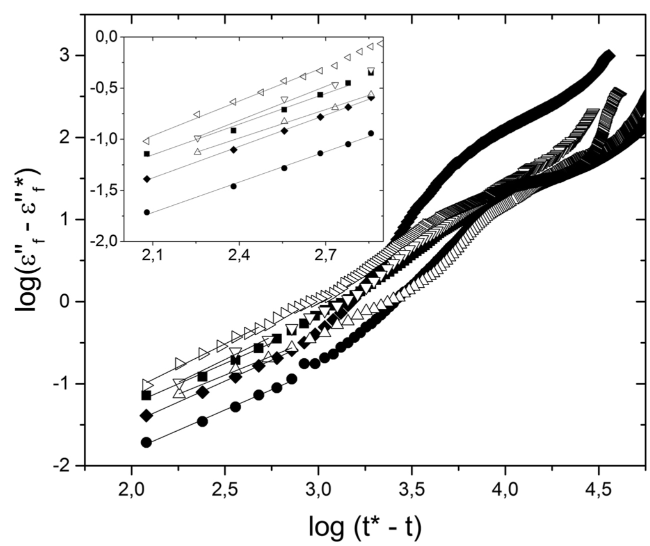


FIG. 3. Log-log plots of monochromatic dielectric loss factor $\left(\varepsilon_f'' - \varepsilon_f''^*\right)$ vs time-to-failure ($t^* - t$) in the vicinity of the conductivity percolation threshold for six samples of moisturized sand with diameter size in range 0.6–0.8 mm (asterisks denote the values at the percolation thresholds). Data fitted according to Eqs. (2), (8), and (9) for samples studied with the simplified experimental method (open symbols) and according to Eqs. (2) and (9) for samples studied with standard experimental method (closed symbols).

TABLE I. Mean critical exponents μ_t of conductivity percolation universal equation, obtained during the dehydration process of six sand samples of diameter size in range 0.6–0.8 mm. Three samples were studied with the use of the simplified experimental method and three others with the use of the standard experimental method.

	Simplified method	Standard method
Mean μ_t	1.069(48)	1.025(33)

monochromatic dielectric loss factor, ε_f'',

$$\varepsilon_f'' = \frac{G_c}{2\pi f C_0}, \tag{9}$$

where C_0 is the capacity of an empty capacitor.

In order to evaluate the experimental method described above, we conducted altogether six series of experiments on moisturized sand samples composed of sand grains with diameter size in range 0.6–0.8 mm, in the course of their dehydration. For the sake of the comparison of the results, three samples were studied utilizing the circuit with the impedance analyzer (a standard method) and three others were studied using the circuit with oscilloscope and the voltage divider (a simplified method). At the beginning of each series of experiments, the sand was irrigated with water, and the measurements were taken until the samples completely dried out. Measurements of electric parameters were taken every 5 min in the course of dehydration through air in seclusion. In a simplified measurement method, the AC generator frequency was set to 100 kHz in all three experiments and four parameters, i.e., $U_{IN_0}, U_{OUT_0}, \varphi,$ and t, were recorded. In a standard method, every 5 min the time of measurement was recorded as well as $C(f)$ and $G_C(f)$, while the electronic input signal frequency was swept through the range of 100 Hz–1 MHz. Nevertheless, for the purpose of the data analysis, only the values $C(f = 100 \text{ kHz})$ and $G_C(f = 100 \text{ kHz})$ were taken into consideration.

Figure 3 shows the results for all six samples, and Table I summarizes the mean values of the critical exponents, μ_t, obtained in the vicinity of the percolation threshold from three series of measurements utilizing a simplified experimental method by fitting the monochromatic dielectric loss factor ε_f'' according to Eq. (2) with the use of Eqs. (9) and (8) and from three other series of measurements utilizing a standard experimental method by fitting the data according to Eq. (2) with the use of Eq. (9).

The results obtained for the sand samples with the same sand range of grain diameters are consistent with each other within the limits of measurement uncertainty, and they have values typical to those reported by other authors for 2D conductivity percolation in hydrated granular samples.[6–9,11–13] This leads to the conclusion that the proposed simplified measurement method for studying the conductivity percolation phenomenon in moisturized granular samples in the course of dehydration to air is comparable with other more complex experimental methods, leading to the coherent and equivalent results, and opening the possibility to conduct such experiments without the need for specialized lab equipment.

REFERENCES

1. M. Sahini, *Applications of Percolation Theory* (CRC Press, 2003).
2. D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, 2nd ed. (Tylor and Francis, London, 1994).
3. G. Careri, A. Giansanti, and J. A. Rupley, Phys. Rev. A **37**, 2703 (1988).
4. F. Bordi, C. Cametti, and R. H. Colby, J. Phys.: Condens. Matter **16**, R1423 (2004).
5. G. A. Voth, Acc. Chem. Res. **39**, 143 (2006).
6. D. Sokolowska, A. Krol-Otwinowska, and J. K. Moscicki, Phys. Rev. E **70**, 052901 (2004).
7. D. Sokolowska, A. Krol-Otwinowska, M. Bialecka, L. Fiedor, M. Szczygiel, and J. K. Moscicki, J. Non-Cryst. Solids **353**, 4541 (2007).
8. D. Sokolowska, D. Dziob, U. Gorska, B. Kieltyka, and J. K. Moscicki, Phys. Rev. E **87**, 062404 (2013).
9. J. Nowak, D. Dziob, M. Rutkowska, L. Chmielarz, U. Gorska, D. Sokolowska, and J. K. Moscicki, Microporous Mesoporous Mater. **206**, 202 (2015).
10. C. Pennetta, L. Reggiani, and G. Tref, Phys. Rev. Lett. **84**, 5006 (2000).
11. J. K. Moscicki and D. Sokolowska, Appl. Phys. Lett. **103**, 263701 (2013).
12. J. K. Moscicki, D. Sokolowska, L. Kwiatkowski, D. Dziob, and J. Nowak, Rev. Sci. Instrum. **85**, 026102 (2014).
13. J. M. Luck, J. Phys. A: Math. Gen. **18**, 2061 (1985).

Experiment on percolation for Introductory Physics Laboratories—A case study

Daniel Dziob
Smoluchowski Institute of Physics, Jagiellonian University, Lojasiewicza 11, 30-348 Krakow, Poland and Department of Biophysics, Jagiellonian University Medical College, sw. Lazarza 16, 31-530 Krakow, Poland

Dagmara Sokołowska
Smoluchowski Institute of Physics, Jagiellonian University, Lojasiewicza 11, 30-348 Krakow, Poland

(Received 6 March 2019; accepted 6 February 2020)

In this study, a module for percolation is developed in light of five goals for the Introductory Physics Laboratories formulated by the American Association of Physics Teachers. The module was tested and validated on a group of six first-year physics students. The content was based on an experiment on the percolation of a water network in the course of the dehydration process that used a simplified method of measurement. The students' opinions of the module were recorded and analyzed. That they had learned about the percolation phenomenon was verified through their lab reports and a specially designed survey. The results show that the students had positive opinions of the implemented module, and that it had enabled them to acquire basic knowledge of percolation.
https://doi.org/10.1119/10.0000810

I. INTRODUCTION

Introductory Physics Laboratories (IPhL) is a common name for laboratories introduced predominantly in the first year of science courses at universities. They encompass the methodology of measurements and the calculations of measurement uncertainties alongside the empirical verification of the basic laws of physics. One of the most important works on IPhL is an article published in 1998 by the American Association of Physics Teachers,[1] which lists the goals of IPhL. It states the following five goals of designing and evaluating IPhL: the art of experimentation, experimental and analytical skills, conceptual learning, understanding the basis of knowledge in physics, and developing collaborative learning skills. The authors note that many of these goals are not explicit in IPhL programs.

Even after two decades of subsequent research and a number of articles emphasizing the role of experiments in understanding science,[2–4] in addition to publications on new experiments for introductory courses in physics,[5–7] some authors still point out that a lot of research on physics education at the undergraduate level still focuses on lectures and tutorials.[8,9] Introductory Physics Labs are studied less frequently, either as a separate course[9] or as a part of integrated courses,[10] where authors look at the learning in the whole course, but to the best of our knowledge, no past publication has described the preparation, implementation, and evaluation of a physics experiment for IPhL while engaging all goals for it as formulated by the AAPT. To address this issue, we formulated, implemented, and evaluated an experiment on the percolation phenomenon for IPhL that fulfills most aspects of all goals listed by the AAPT.[1] While implementing this experiment in IPhL classes, we examined how students performed and perceived this new laboratory experiment. We chose a topic that does not belong to the canon of the IPhL, but provides a deep insight into a phenomenon that is common to many areas of science and is not too demanding for first-year students.

II. THEORETICAL BACKGROUND

A. IPhL goals

The list of goals recommended for IPhL provides guidance on what the ideal laboratory experiment should look like. Thus, it is worthwhile analyzing and understanding these goals.

The first goal, "the art of experimentation," involves engaging students in the full experimental process, including helping them design their own experiments. This may be done by choosing the method of measurement, experimental tools, or scope of the measurements performed. However, most IPhL classes feature standard instructions that hinder the students' creativity and independent thinking.[11]

The second goal, "experimental and analytical skills," is the one most often associated with IPhL, and is used to define it. However, in many cases, classes in IPhL feature easy experiments that are well known even to high school students, or are easy to find on the Internet (often in the form of videos presenting the experiment together with a full description of it). Thus, students often complain that IPhL consists of boring and obsolete experiments with no connection to "real" science,[8] and that their results can be easily predicted even without performing any experiment. One way of encouraging students is by introducing ICT (Information and communications technology) to IPhL, such as by using data acquisition software or smartphones.[12,13] However, this simplifies most experiments by automating them and omitting the stage of designing the experiment. It has even been argued that the use of smartphones in labs discourages students from using "old-fashioned" experimental setups,[5] with which students can research the basics of physical phenomena while using simple measurement techniques (for example, connecting an ammeter to an electric circuit).

The third goal, "conceptual learning," states that laboratories should help students understand concepts in physics. However, recent research has suggested that IPhL classes, even those designed to support the learning content of associated lectures, do not have a measurable impact on the students' final results in introductory courses in physics.[11,14]

Nevertheless, the relevant authors have emphasized that the laboratories investigated in their research were mostly designed to present concepts in physics that had previously been introduced and discussed in lectures. Thus, the students participating in these laboratories already knew the laws being examined. On the contrary, other research has shown that students posed more difficult problems in the laboratory achieve better results.[15] It appears that IPhL can be more motivating, and can more effectively help students understand concepts in physics when used in lieu of simple experiments representing basic physics; phenomena using more complicated and less predictable measurements, which are more representative of contemporary science.[16,17]

The fourth goal, "understanding the basis of knowledge in physics," clearly points out that theories and experiments are intertwined in physics, and cannot be introduced without each other. On the one hand, theoretical considerations without supporting experiments are merely "a collection of equations and textbook problems."[1] On the other hand, the results of experiment have value only when they are supported by other results or may be explained by the theory. Otherwise, such results are essentially just numbers without meaning. Many authors have shown that students consider their experimental results to always be correct, and believe that performing a single appropriate measurement is enough for verification,[18,19] even without comparison with any other result or theory. Students do not understand the essence of experiments in physics, and focus more on performing measurements according to the given instructions. Other authors have shown that students carefully calculate the mean values while not describing them in any way further, without any comment or conclusion.[20,21] This behavior is encouraged in experiments that are easy to analyze and, in which, after several measurements, students simply have to present the mean value obtained, which is the result itself. The interplay between a theory and an experiment in physics results in a requirement that in IPhL, students should learn how to present the results, and to compare them with other results and theory.[22]

The fifth goal is "developing collaborative learning skills." Many recent studies have suggested that in addition to "hard" knowledge, soft skills are key to successful professional careers,[23–25] and should be developed at every stage of education. In the case of IPhL, using peer instructions,[26] collaborative lab reports,[27] and assessments that require the cooperation of the entire group[28] are exemplary of reaching this goal of IPhL. Thus, this last goal is the most commonly implemented one in IPhL.

The collection of goals mentioned above provides insight into the means of preparing IPhL experiments. Fulfilling all five goals is a worthwhile challenge because such a holistic approach can ensure a more complete and efficient development of students' research skills. Nevertheless, many of the goals are not explicit in traditional laboratory programs.[1] Below, we propose a module for percolation experiments in IPhL designed with respect to its goals, proposed by the AAPT. We chose a contemporary topic, normally not included in introductory courses, to study the feasibility of its implementation in IPhL and the students' perception of such a non-standard proposal.

B. Percolation phenomenon

The phenomenon of percolation concerns all the types of transport in the network, where the number of channels that enable this transport changes randomly. This is described by the theory of percolation, a probabilistic theory in which the key parameters are the percolation threshold and critical exponent. Percolation theory is used to mathematize the percolation phenomenon not only in physics (e.g., water transport in a net of pipes and conductivity in a net of wires),[29] but also in many other areas, such as medicine (e.g., spread of diseases and the functioning of brain connections),[30] public transport (airplane traffic),[31] and Internet network breakdown,[32] just to mention a few.

An illustrative explanation of the percolation principle is based on a description of the flow of water through a system of pipes or that of electric current through a metal grid. For example, one can imagine a metal grid connected to a power source at both ends (see Fig. 1(a)). Initially, current flows through the entire grid. However, when single connections are gradually and randomly removed, the amount of current flowing decreases until the entire system has been disconnected, which means that flow ceases. This point is called the percolation threshold.[29] In one implementation, the percolation theory can be used to investigate the random appearance or disappearance of electric conductivity in biological and biomimical systems.[33,34] In such systems, water enables conductivity and may freely dehydrate from the sample, reducing the occupational density of the lattice (observed as a decrease in conductivity) up to a point, called the percolation threshold, where long-distance connections no longer exist in the system. Thus, the percolation threshold is linked to the last configuration (or the lowest occupation density) in which the sample still conducts electricity—see Fig. 1. Near the percolation threshold, the conductivity σ depends on the probability that the electric charge finds an unrestricted path in the examined sample. This probability is directly proportional to the degree of hydration that can be parameterized, e.g., by the hydration level h, defined as the ratio of the mass of water in the sample to that of a dry sample.[35] Therefore,

$$(\sigma - \sigma^*) \propto (h - h^*)^{\mu}, \tag{1}$$

where (*) indicates values of the percolation threshold and μ is the critical exponent, the value of which is related to the dimensionality of the observed phenomenon.[36,37]

Conductivity percolation in a water network can be observed and measured by using an impedance analyzer. In such an approach, the examined hydrated sample is placed in a capacitor with holes in the upper plate to allow for evaporation.[35] The capacitor is placed on a balance and connected to the impedance analyzer, which measures the capacity and conductance of the filled capacitor at various frequencies in a defined period. Further analysis enables the recalculation of the measured conductivities of the examined sample, but the procedure is sophisticated. However, note that the change in frequency used in the above-described approach was limited to 100 Hz–2 MHz to avoid the electrode polarization effect observed at lower frequencies, and the Maxwell–Wagner effect observed at higher frequencies. In this range, the spectra of dielectric loss are dominated by contribution from DC conductivity σ ($\varepsilon'' \cong \sigma/2\pi\varepsilon_o f$, where ε_o is the permittivity of the vacuum), and the level of hydration is proportional to time. Thus, an analysis of conductivity percolation can be simplified substantially. To find the percolation threshold, one can use the variations in time of the dielectric loss factor, $\varepsilon''|_f$, at any fixed frequency in the considered range[38,39]

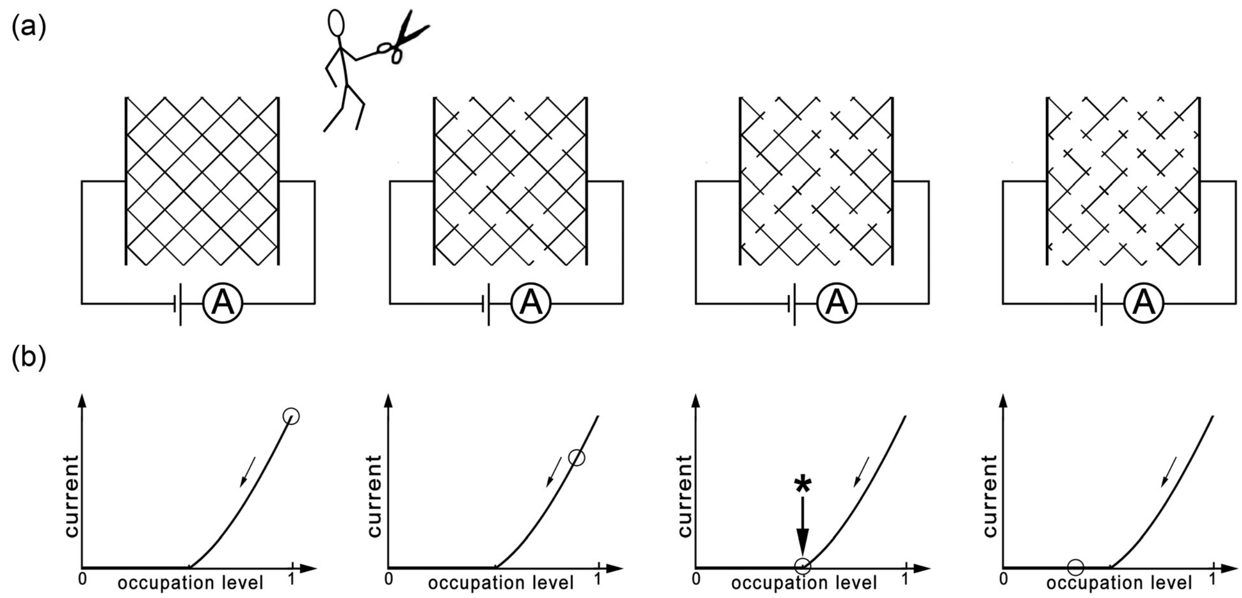


Fig. 1. Representation of percolation phenomenon in a metal grid. (a) Schematic pictures of a metal lattice connected to a power source with different levels of occupancy, defined as ratio of conductive connections to the total number of connections. (b) Schematic graphs of electric current in a lattice as a function of the level of occupancy. Points marked with circles correspond to situations in (a), and the asterisk indicates the percolation threshold. Initially, current flows through the lattice, but when single connections are gradually and randomly removed, it decreases until the system is disconnected, which means that flow ceases.

$$\left(\varepsilon''|_f - \varepsilon''|_f^{\,*}\right) \propto (t^* - t)^{\mu} \leftrightarrow \log\left|\varepsilon''|_f - \varepsilon''|_f^{\,*}\right| = \mu \log|t^* - t| + const. \quad (2)$$

III. EXPERIMENT ON PERCOLATION

We now describe an experiment on percolation and describe its relation to the five AAPT IPhL goals. A wide-ranging discussion of the implementation of goals of the AAPT is provided in Sec. VII.

A. Simplifying the experimental setup

Because an impedance analyzer is rarely available in IPhL, it is unfeasible to prescribe it for measurements in a student laboratory. To address this limitation, a simplified setup to measure conductivity percolation was recently developed.[40] It consists of a capacitor in serial connection with a resistance decade box and an AC generator. An oscilloscope is used to measure the data. In this setup (see Fig. 2), the capacitor and resistance decade box together act as a voltage divider and, therefore, measurements of the amplitude of voltage on a generator (U_{in}), its amplitude on the resistance decade box (U_R), and a phase shift (φ) between them enable the recalculation of the conductance of a sample placed in the capacitor (G_C), according to the following equation:

$$G_C = \frac{U_{in}\cos\varphi - U_R}{U_{in}^2 + U_R^2 - 2U_{in}U_R\cos\varphi} U_R G, \quad (3)$$

where G is the conductance of the resistance decade box ($G = 1/R$). Hence, the spectra of dielectric loss of the sample in the capacitor can be calculated as

$$\varepsilon'' = \frac{G_c}{2\pi f\, C_o}, \quad (4)$$

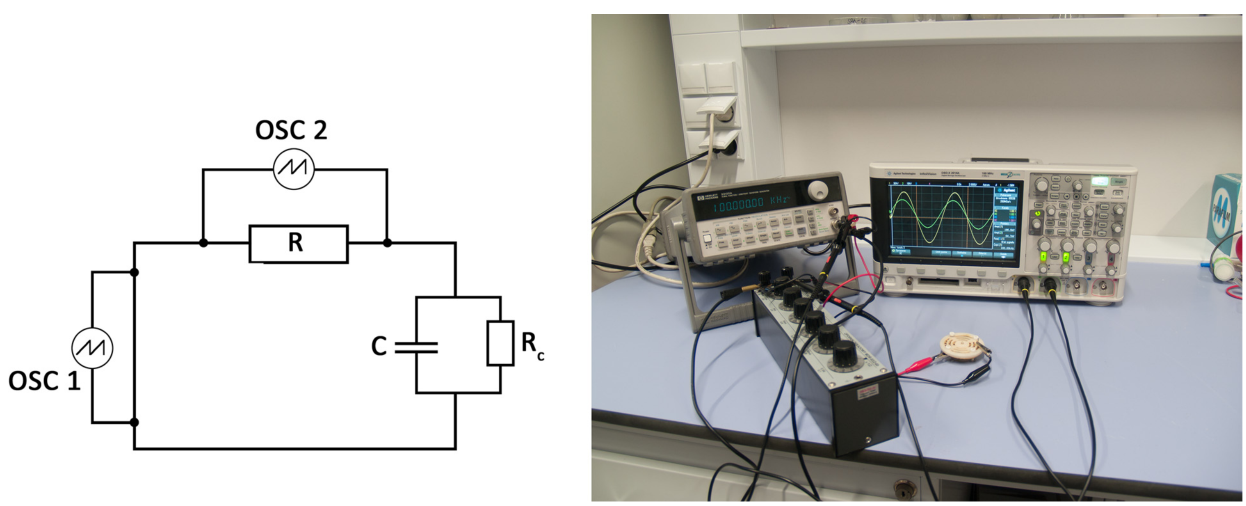


Fig. 2. Experimental setup. Left—schema; right—experiment.

where f is the selected frequency and C_0 is capacitance of the empty capacitor. Thus, it is possible to replace a complex RLC setup with a simpler one composed of instruments easily available in IPhL. Such an approach enables the introduction of the percolation experiment to IPhL.

B. Details of experiment and its connection to IPhL goals

Because the percolation theory is applicable to a broad range of transport phenomena originating in different scientific disciplines, the proposed experiment is an interesting lab investigation in IPhL that enables conceptual learning (AAPT IPhL goal III). In the proposed experiment, the percolation of the water network during the dehydration process can be examined. Unlike when using the RLC impedance analyzer, a simplified setup gives students the opportunity to easily manipulate the instruments and modify certain parameters to understand the consequences—for example, the resistance of the resistance decade box, the frequency of electric current, and the volume of water. In this way, the experiment does not consist only of simply turning on the setup and performing the experiment (the "turn up–turn on–measure" approach), it also requires asking questions and providing answers on the measuring process and modifications to it (AAPT IPhL goal I). Furthermore, the experimental method introduces basic knowledge of electronics, electric current, and measurements through an oscilloscope. Figure 3 shows plots of the spectra of dielectric loss as a function of time above, at, and below the percolation threshold of water conductivity in a 2D network built on surfaces of sand grains that served as a base. These plots show the substantial disturbance in conductivity when crossing the percolation threshold. Owing to the random nature of the phenomenon, the percolation threshold is not predictable *ad hoc*, but requires measurements in a full range of the hydration of the sample as well as further data analysis involving the careful detection of the percolation threshold (AAPT IPhL goal II). The experiment requires further analysis by students to determine the critical exponent, which can be used to interpret the dimensionality of the percolating network when its value is compared with those reported in the literature (percolation theory and experimental results; AAPT IPhL goal IV). Goal V is secured through the organization of the laboratory work in pairs.

IV. RESEARCH QUESTIONS

We prepared a module of percolation based on most aspects of the five IPhL goals listed by the AAPT.[1] The experiment described above was the main part of the designed module. Implementation was carried out using a group of undergraduate physics students who also evaluated the module. This case study reports findings related to the following research questions:

(1) To what extent is it feasible to design and implement an experiment on the percolation phenomenon based on the five IPhL goals in an introductory lab at our facility?
(2) To what extent does the proposed laboratory unit enable students to understand percolation theory?
(3) How was the lab module on percolation perceived by students?

V. METHODOLOGY

A. Laboratory module—Design and implementation

The percolation module proposed for IPhL is based on the conductivity percolation of a water network in a hydrated sand sample, freely evaporating into the air. The module is composed of three main parts: (1) student preparation, based on the delivered materials (script) and questions posed *a priori*; (2) the laboratory activity performed in pairs with some assistance from a tutor; and (3) individual work by the students to analyze data gathered in the experiment to determine the percolation threshold and prepare a lab report. The following describes these three parts in detail.

1. Student preparation

Before participating in the IPhL classes, the students received a script that provided an overview of the experiment along with the key objectives. The first part of the script consisted of a description of the percolation phenomenon together with examples of the implementation of percolation theory. This was followed by a description of the conductivity percolation of the water network and an introduction to the concept of level of hydration. The next part of the script described a method of measurement based on measurements in an RC circuit. It contained essential information on electronics (like the functioning of a voltage divider) and described the linear relation between the level of hydration and time (Fig. 4).

2. Laboratory activity guided by a tutor

Due to the nature of conductivity percolation, which requires the free dehydration of the sample into the air, the time spent by students in the laboratory was divided into two

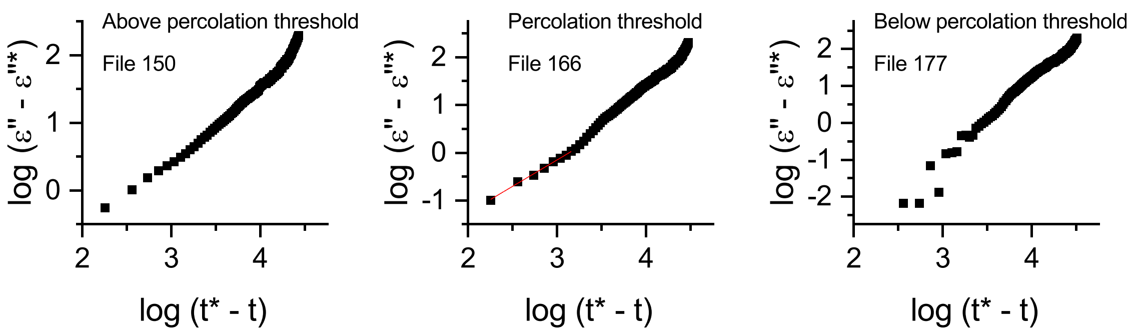


Fig. 3. Plots of dielectric loss spectra as a function of time above, at, and below the percolation threshold. Data obtained for a fixed frequency, $f = 100$ kHz. Own data.

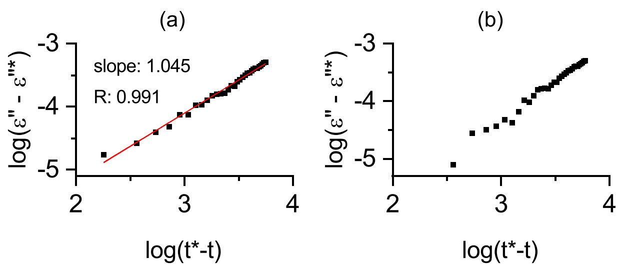


Fig. 4. Examples of graphs obtained by students (a) in the percolation threshold and (b) below the percolation threshold (both in the log–log scale). The solid line is a linear fit according to Eq. (2). The slope parameter corresponds to the value of the critical exponent.

parts. The first lasted approximately one hour and took place in the morning. In this hour, the students built an experimental setup (Fig. 2), and adjusted the frequency of the input current and resistance of the decade resistance box. They needed to choose a frequency in the range of 100 Hz–2 MHz, where this choice influenced the adjustment of the resistance that needed to be selected for the decade resistance box. The resistance needed to be adjusted to enable the observation of both measured voltage amplitudes U_{in} and U_R in the course of the measurements—from a hydrated to a completely dry sample.[40]

Subsequently, the students had to hydrate the sample with an adequate amount of water. For thin samples, the speed of dehydration was constant, which made it possible to calculate an approximate amount of water that needed to be added to the sample to ensure the observation of the percolation threshold (in the course of dehydration into the air) in a defined period. This stage determined the time to start the second part of the laboratory in the afternoon of the same day. For most groups, the second part began seven to eight hours after the first. In this part, the students measured the amplitude of the input voltage, that of the voltage on the resistor box, and the phase shift between them together in three-minute intervals. The rules of the IPhL at our faculty required that no experiment last longer than three hours, which limited the second part of the experiment to two hours.

In the course of the investigation, the students were allowed to modify the parameters of the experiment. Their choice did not influence the percolation parameters at the end, but the choice of parameters for the initial measurement setup resulted in a unique set of data and, consequently, unique data analysis by each group of students.

3. Individual student report

After the laboratory experiment, the students had one week to analyze the data, detect the percolation threshold, and determine the critical exponent and prepare a report. The lab reports were structured as a scientific paper consisting of an introduction/theoretical background, experimental setup, methodology, data analysis, discussion, and conclusions. To conclude their results, the students needed to compare their data with those reported in the literature.

B. Research sample

The module was tested by a group of six students in their first year of undergraduate physics studies at the Faculty of Physics, Astronomy, and Applied Computer Sciences of Jagiellonian University in Krakow, Poland. Prior to the implementation, each student had been asked to complete ten other experiments in IPhL during the spring semester. The percolation experiment substituted a standard experiment at the end of the term. However, owing to personal reasons, one of the students did not write his individual report on the experiment. Therefore, the data analysis involves the results provided by five students.

VI. INSTRUMENTS AND DATA COLLECTION

A. Percolation survey

Before the laboratory experiments, and before receiving the script, the students were asked to fill in a short survey on percolation. It consisted of four open questions, presented in Table I (left column). A similar survey was completed by the students after finishing the laboratories and receiving their grades (approximately 14 days later). The main difference between the surveys was that the question, "have you ever heard about percolation?," from the first run was replaced by the question "what is the percolation threshold and what information can it provide?"

B. Students' reports

Each student was obliged to prepare a lab report of the experiment. The report was verified and assessed by the tutor. It provided information of the extent to which the student had understood the idea of percolation, the method of measurement, the meaning of the critical exponent, and the process of determining the percolation threshold.

C. Questionnaire

The students were also asked to fill in a lab questionnaire after completing the laboratories, preparing lab reports, and being graded. This questionnaire sought to determine how the students had perceived the percolation experiment, its level of difficulty, attractiveness of the proposed module, and their opinions on the feasibility of including it in a regular IPhL set next to canonical experiments. The questionnaire contained five open-ended questions. In the first three, the students were asked to assign a score on a five-point Likert scale (1—definitely no; 5—definitely yes) and provide a brief explanation of each. The next two questions required descriptive answers. The questions are presented in Table II.

Table I. Surveys on percolation.

Pre-survey	Post-survey
Have you ever heard about percolation?	-
Please write what you know about percolation.	Describe, in your own words, what percolation is.
Which field of science do you associate percolation with?	To which fields of science does percolation theory apply?
Have you heard about any practical use of percolation?	Are there practical applications of percolation? What, if any?
	What is the percolation threshold and what information can it provide?

Table II. Questions in the questionnaire administered to students after completing the lab, preparing lab reports, and receiving their grades.

Question type	Questions
Likert scale and brief explanation	- Was the proposed topic interesting? - Was the tutor well prepared, and did they motivate students to engage in the lab? - Was the content of the provided script sufficient to prepare for the lab?
Description	- Do you think that the experiment on percolation is worth introducing to other students? Why or why not? - Please state changes you think are required to the experiment.

VII. RESULTS AND DISCUSSION

A. Percolation survey prior to and after implementation

When answering the first question of the survey prior to implementation, four of the six students indicated that they had never heard of percolation. One respondent answered "yes" but could not explain what percolation was. Another respondent associated percolation with the process of washing ingredients from the body using water. Of the fields of science associated with percolation, physics, geology, biophysics, and chemistry were mentioned, and one answer read "with liquids."

In the survey taken after the IPhL experiment on percolation, each student correctly described the percolation phenomenon—as a process of the random formation/breaking down of connections in networks/systems—alluding also to the theory of the critical phenomenon. Only one student failed to provide information on the random nature of the process. All students correctly defined the percolation threshold, and more than half properly related it to the critical exponents. Each student listed a few fields in which the percolation phenomenon takes place. The students listed different fields of physics and related sciences, but also such examples as epidemiology and pharmacokinetics.

The above comparison of the students' answers before and after the laboratory shows that after conducting the experiments, they could appropriately describe the percolation phenomenon, define the percolation threshold, and provide examples of the application of percolation theory. Considering that before the experiment, most students had not even heard of percolation, the knowledge they acquired in the course of the implementation is clear.

B. Students' reports

The lab reports were prepared by students separately and assessed by a tutor according to the faculty rules established for IPhL. The median results of all reports, including engagement and the work done while conducting experiments, presented on a scale of two to five, was five. This result, together with a comparison of the pre- and post-survey results on percolation mentioned in Subsection VII A, shows that the students had acquired significant knowledge of percolation through the proposed learning unit. However, note that these results should be approached with caution because of the small number of students tested. Table III contains a concise list of the assessed requirements and information on the extent to which each student fulfilled them. Examples of graphs obtained by students are presented in Fig. 4.

Note that one group of students did not capture the time at which the percolation threshold occurred during the experiment owing to the limited time of the laboratory classes (the sample did not dry in time). These students received from the tutor a set of other experimental data based on which they had to find the percolation threshold. With this intervention, each student correctly found the percolation threshold

Table III. Report requirements and the extent of their fulfillment by each student.

Requirement	Student 1	Student 2	Student 3	Student 4	Student 5
Introduction/theoretical background (appropriate theory, consistency).	+	+	+/− Lacking description of electronic part related to measurement method	+	+/− Curt introduction, a lack of connection between water and conductivity
Description of the experimental setup, methodology, and the course of measurements.	+	+	+	+	+/− A lack of description of the time course of measurements
Data analysis (data conversion and plotting adequate graphs, determining the percolation threshold, calculating the critical exponent).	+	+	+	+[a]	+[a]
Discussion, conclusions (comparing results with the literature, concluding 2D character of conductivity percolation in performed experiment, indicating possible mistakes and suggesting modifications to the measurement).	+	+/− A lack of comparison with the literature	+	+/− A lack of definition of percolation dimensionality	+

[a]In case a pair of students had collected data in the course of the experiment that did not yield the percolation threshold, the final analysis (as an appendix) was added by them on a set of data from a different experiment.

and determined the critical exponent for the phenomenon. The percolation thresholds calculated by the students concurred with the threshold found by the tutor. The critical exponents for all students pointed to a 2D percolation phenomenon, and their comparisons with the literature and theoretical values are listed in Table IV.

While the students' reports showed that they had determined the percolation threshold on their own using the instruction, the laboratories exposed a weakness in the proposed laboratory experiment: A risk that the percolation threshold will not be calculable during measurements limited in time due to, e.g., changes in humidity in the laboratory room or inaccurate admeasure of water by students prior to the experiment. This is, of course, an interesting aspect that introduces students to the reality of experimental work, but for the needs of the IPhL, the conditions of the experiment should be more specific. This can be achieved by ensuring constant humidity and temperature to control the dehydration process more precisely.

C. Questionnaire

The questionnaire was filled out by five students after completing the experimental work, writing the reports, and receiving the final grades. Their responses to each question and their medians for the first three questions, based on a five-point Likert scale, are shown below.

In response to the question "was the proposed topic interesting?," the median of student rating was 5. The students emphasized that it was an experience more interesting than most others proposed at the IPhL, with one stating: "Finally, at the IPhL, I did something that was not 100% predictable and was not easy to write on a piece of paper in 15 minutes." One student noted that being in the early stage of his studies, he considered the experiment to be a scientific curiosity, of sorts.

Responses to the question "was the tutor well prepared, and did they motivate students to engage in the lab?" had a median score of five. One of the comments was: "Yes, he knew a lot and made the time interesting by sharing this knowledge with us :)." Responses to the question "was the provided script sufficient to prepare for the lab?" had a median score of four. All students described the script as sufficient to prepare for the labs, suggesting small changes at the same time, such as a need for increasing or reducing the theoretical introduction, or a need for a more precise description of the data analysis. From the above, one can conclude that the students appreciated the preparation of the module and the experiment itself.

The second set of questions started with: "Do you think that an experiment on percolation is worth introducing to other students? Why or why not?" The students' answers to this question ranged from moderate optimism—"I do not consider this to be something that must necessarily be introduced," and "[…] it would be a useful minimum variation for students"—to very positive answers—"yes, no doubt. This is probably the only non-deadly [sic] boring experiment that could not be done at home and in a sensible way uses an oscilloscope, which in general is used very rarely in IPhL. It allows you to learn something interesting, get to know a new phenomenon." The opinions of the other students were also positive within the above range. However, they drew attention to the need to improve the procedure used to detect the percolation threshold so that one is always certain of "capturing it" in a class, without requiring data from the tutor. The solution to this problem has been discussed above. Changes suggested to the experiment included proposals to more rigorously control the mass of water (i.e., to capture the threshold of percolation in each experiment), and to extend the script, e.g., on the theory of RC circuits and phasors. However, the students, despite difficulties and problems, undoubtedly considered the experiment to be interesting, and did not regret the time they devoted to it.

Finally, we address the realization of goals of the IPhL in the proposed module. The first goal, the selection of the topic, i.e., the percolation phenomenon, is related to the goal III, "conceptual learning." Percolation has many applications in various branches of science, but not all of them can be described and explained by the same theory of percolation. Thus, to introduce the percolation phenomenon, we chose an experimental problem concerning the conductivity of percolation in a water network, easily accessible in the IPhL, highlighting at the same time other applications of percolation that are discussed in our module only at the conceptual level. To be more concrete, this wide range of applications has a common base, which is the universality of the critical exponents. Independently of the problem of percolation under investigation, the value of the critical exponent provides information about the dimensionality of the process. Thus, the critical exponents obtained in experiment provide the specific meaning of and information about the investigated network only when they are verified through theory. In this way, it becomes clear to the students that experiment and theory are intertwined—the AAPT IPhL goal IV.

The percolation experiment, although designed by us, provides students significant autonomy in approaching the experimental procedure. They can not only change parameters, but also need to adjust them, e.g., connecting the experimental setup properly, adjusting settings in a way to enable the measurement of all necessary quantities, and finding a percolation threshold in the designated time. Thus, the percolation experiment transforms the role of students from that of "passive reproducer" to that of "active co-creator" by involving them into the art of experimentation (the AAPT IPhL goal I). Conducting the experiment itself develops their experimental skills, e.g., when students can modify aspects of the experiment and use equipment that they are unfamiliar with from secondary school (oscilloscope). As each

Table IV. The average values of critical exponents obtained by students, and in other measurements.

	Results obtained by students	Measurements using a simplified setup (Ref. 40)	Measurements with RLC impedance analyzer[a]	Theory of conductivity percolation in 2D network
Critical exponent for water network percolation in moist sand sample	0.98–1.11	1.069	1.025	1.3[29]

[a]Experimental values for 2 D percolation are usually lower than those resulting from the theoretical models (Refs. 34 and 41).

experiment is unique and yields a unique dataset, it always requires a unique analysis. This allows for the development of analytical skills, which is part of the AAPT's IPhL goal II. Finally, the AAPT's IPhL goal V, "developing collaborative skills," was secured by the students working in pairs.

VIII. CONCLUSIONS

In this study, we prepared a laboratory module that engages most aspects of all five goals of the IPhL listed by the AAPT. It was implemented on a group of six first-year physics students during a three-hour class. The students worked in pairs and prepared separate reports that were reviewed by the tutor afterward. They were asked not only to take part in the lab, but also to evaluate the module on percolation prepared for the IPhL. Because the sample size was small, the conclusions drawn from this case study should be treated with due caution.

The students were engaged in the experiment on percolation, an interesting topic unknown to them from physics lectures that nonetheless has a wide range of application in different fields. Through preparation prior to the lab and participation in the experiment, they acquired knowledge on percolation that was verified by the reports assigned to them, and the surveys administered prior to and after the implementation. The students worked collaboratively during the lab, which helped strengthen their experimental skills. The subsequent exercise helped bolster their individual analytical skills afterward. Their motivation for the IPhL was hence enhanced to a greater extent than by other, standard experiments, which are perceived by them as too simple to be interesting and engaging.

We found that the module in the proposed form appeared feasible to the students. Slight modifications, suggested by them, are worth considering in further developing pedagogical material. Nevertheless, not a single student noted any significant difficulty in understanding the experiment or the phenomenon itself. Finally, based on the students' opinions, we found that their perception of the module was positive, and they described their participation in this lab as an interesting and valuable experience. It can be concluded that an experiment that goes beyond the basic content and laws of physics discussed in lectures can encourage students to participate more actively in the IPhL.

[1]AAPT, American Association of Physics Teachers, "Goals of the introductory physics laboratory," Am. J. Phys. **66**(6), 483–485 (1998).

[2]A. Albers, M. Rollnick, and F. Lubben, "First year university students' understanding of validity in designing a physics experiment," Afr. J. Res. SMT Educ. **12**(1), 31–52 (2008).

[3]T. Gregory and M. Al, "Students' perceptions of changes to the learning environments of undergraduate physics laboratories: An empirical study," Interactive Technol. Smart Educ. **15**(2), 165–180 (2018).

[4]B. Zwickl, N. Finkelstein, and H. Lewandowski, "The process of transforming an advanced lab course: Goals, curriculum, and assessments," Am. J. Phys. **81**, 63–70 (2013).

[5]F. Saraiva da Rocha, F. Fajardo, M. Grisolía, J. Benegas, R. Tchitnga, and P. Laws, "Using "student technology" in introductory physics: A comparison of three tools to study falling objects," Phys. Teach. **49**(3), 165–167 (2011).

[6]J. Henningsen, "Teaching laser physics by experiments," Am. J. Phys. **79**, 85–93 (2011).

[7]M. Parappilly and C. Hassam, "Race to improve student understanding of uncertainty: Using LEGO race cars in the physics lab," Am. J. Phys. **86**, 68–76 (2018).

[8]B. Wilcox and H. Lewandowski, "Students' views about the nature of experimental physics," Phys. Rev. Phys. Educ. Res. **13**, 0201110 (2017).

[9]C. Walsh, K. Quinn, C. Wieman, and N. G. Holmes, "Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking," Phys. Rev. Phys. Educ. Res. **15**, 010135 (2019).

[10]L. McDermott, "Melba Newell Phillips Medal Lecture 2013: Discipline-based education research—A view from physics," Am. J. Phys. **82**, 729–741 (2014).

[11]N. Holmes and C. Wieman, "Introductory physics labs: We can do better," Phys. Today **71**(1), 38–45 (2018).

[12]C. Countryman, "The educational impact of smartphone implementation in introductory mechanics laboratories," Paper presented at Physics Education Research Conference 2015, College Park, MD, retrieved from <https://www.compadre.org/Repository/document/ServeFile.cfm?ID=13957&DocID=4344>, accessed on October 24, 2019.

[13]M. Pierri-Galvao, "Restructuring the introductory physics lab with the addition of computer-based laboratories," J. Nat. Sci., Biol. Med. **2**(2), 164–167 (2011).

[14]C. Wieman and N. Holmes, "Measuring the impact of an instructional laboratory on the learning of introductory physics," Am. J. Phys. **83**, 972–978 (2015).

[15]L. McDermott, "Guest comment: How we teach and how students learn—A mismatch?," Am. J. Phys. **61**(4), 295–298 (1993).

[16]R. Blonder and M. Dinur, "Teaching nanotechnology using student-centered pedagogy for increasing students' continuing motivation," J. Nano Educ. **3**(1), 51–61 (2011).

[17]M. Cepic, "Introduction of current scientific results to education: Experiences from the case of liquid crystals In: In key competences in physics teaching and learning," in *Proceedings of the International Conference GIREP EPEC 2015*, Poland (2015), pp. 41–54.

[18]E. Etkina, S. Murthy, and X. Zou, "Using introductory labs to engage students in experimental design," Am. J. Phys. **74**(11), 979–986 (2006).

[19]R. L. Kung and C. J. Linder, "University students' ideas about data processing," Nord. Stud. Sci. Educ. **4**, 40–53 (2006).

[20]S. Allie, A. Buffler, L. Kaunda, B. Campbell, and F. Lubben, "First year physics students' perceptions of the quality of experimental measurements," Int. J. Sci. Educ. **20**(4), 447–459 (1998).

[21]D. Deardorff, "Introductory physics students' treatment of measurement uncertainty," Ph.D. thesis, North Carolina State University, 2001 (unpublished).

[22]A. Buffler, F. Lubben, and B. Ibrahim, "The relationship between students' views of the nature of science and their views of the nature of scientific measurement," Int. J. Sci. Educ. **31**(9), 1137–1156 (2009).

[23]M. M. Robles, "Executive perceptions of the top 10 soft skills needed in today's workplace," Bus. Commun. Q. **75**(4), 453–465 (2012).

[24]J. J. Heckman and T. Kautz, "Hard evidence on soft skills," Labour Econ. **19**(4), 451–464 (2012).

[25]B. Shulz, "The importance of soft skills: Education beyond academic knowledge," NAWA: J. Lang. Commun. **2**(1), 146–154 (2008).

[26]A. Cox and W. Junkin, "Enhanced student learning in the introductory physics laboratory," Phys. Educ. **37**(1), 37–44 (2002).

[27]M. Wood, "Collaborative lab reports with Google Docs," Phys. Teach. **49**, 158–159 (2011).

[28]A. Leung, B. Pour, D. Reynolds, and S. Jerzak, "New assessment process in an introductory undergraduate physics laboratory: An exploration on collaborative learning," Assess. Eval. Higher Educ. **42**(2), 169–181 (2017).

[29]D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, Washington, DC, 1991).

[30]R. Parshani, S. Carmii, and D. Havlin, "Epidemic threshold for the susceptible-infectious-susceptible model on random networks," Phys. Rev. Lett. **104**(25), 258701 (2010).

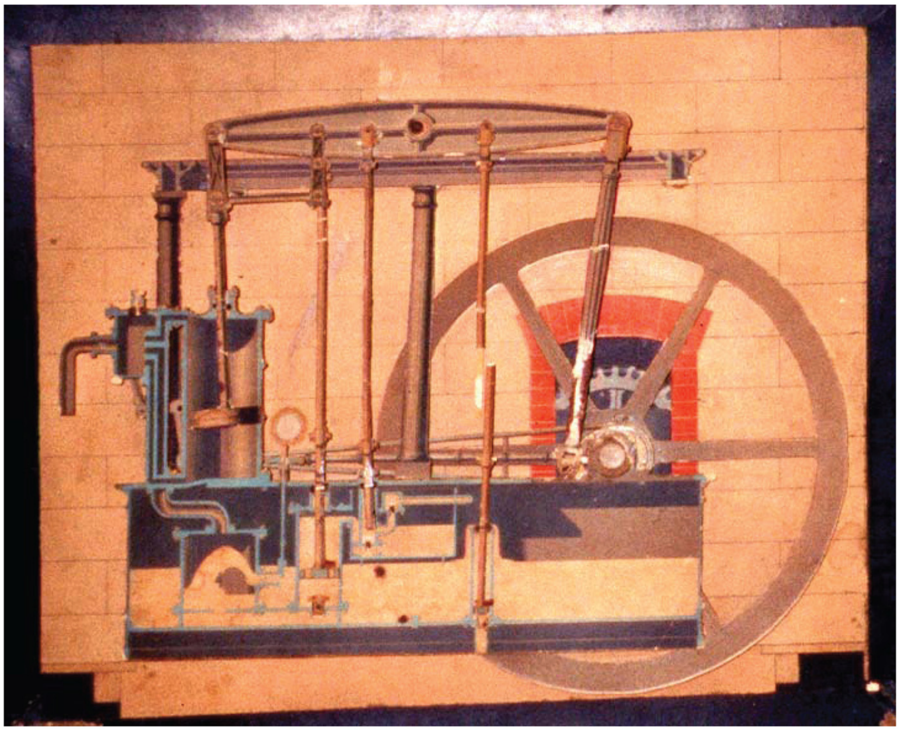
[31]S. Amor, T. Huy, and M. Bui, "A percolation based model for ATC simulation," in *Proceedings of the 4th International Conference on Computer Sciences, Research Innovation and Vision for the Future*, Vietnam (2006), pp. 17–22.

[32]R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, "Resilience of the Internet to random breakdowns," Phys. Rev. Lett. **85**, 4626–4628 (2000).

[33]F. Bordi, C. Cametti, and R. Colby, "Dielectric spectroscopy and conductivity of polyelectrolyte solutions," J. Phys.: Condens. Matter **16**, R1423–R1463 (2004).

[34]D. Sokolowska, A. Krol-Otwinowska, and J. K. Moscicki, "Water-network percolation transitions in hydrated yeast," Phys. Rev. E **70**, 052901 (2004).

[35]D. Sokolowska, D. Dziob, U. Gorska, B. Kieltyka, and J. K. Moscicki, "Electric conductivity percolation in naturally dehydrating, lightly wetted, hydrophilic fumed silica powder," Phys. Rev. E **87**(6), 062404 (2013).
[36]C. Pennetta, L. Reggiani, and G. Tref, "Scaling and universality in electrical failure of thin films," Phys. Rev. Lett. **84**, 5006–5009 (2000).
[37]K. Sieradzki, K. Bailey, and T. Alford, "Agglomeration and percolation conductivity," Appl. Phys. Lett. **79**, 3401–3403 (2001).
[38]J. K. Moscicki and D. Sokolowska, "Dehydration kinetics and conductivity scaling in time near the continuity failure of a water wetting film in granular media," Appl. Phys. Lett. **103**(26), 263701 (2013).
[39]J. K. Moscicki, D. Sokolowska, L. Kwiatkowski, D. Dziob, and J. Nowak, "Note: Optimization of the numerical data analysis for conductivity percolation studies of drying moist porous systems," Rev. Sci. Instrum. **85**(2), 026102 (2014).
[40]D. Dziob, D. Sokolowska, M. Mlynarczyk, and J. K. Moscicki, "A simplified experimental method to study conductivity percolation," Rev. Sci. Instrum. **90**, 076105 (2019).
[41]J. Rupley, L. Siemiankowski, G. Careri, and F. Bruni, "Two-dimensional protonic percolation on lightly hydrated purple membrane," Proc. Natl. Acad. Sci. U. S. A. **85**, 9022–9025 (1988).



Steam Engine Model

This French model of a steam engine was made of cardboard. Pivots allowed the action of the walking beam to be demonstrated to students, and the action of the cylinder and piston combination to be observed, as well as the valves that admitted steam into the system. The model is mid-nineteenth century and was on display at the museum of the University of Mississippi when I photographed it in 2006. (Picture and text by Thomas B. Greenslade, Jr., Kenyon College)

IOP Publishing European Journal of Physics
Eur. J. Phys. 41 (2020) 025801 (9pp) https://doi.org/10.1088/1361-6404/ab4ed2

Water network percolation on yeast as an experiment proposal for advanced physics laboratories for bioscience students

Daniel Dziob[1,2] and **Dagmara Sokołowska**[1]

[1] Jagiellonian University, Smoluchowski Institute of Physics, Lojasiewicza 11, 30-348 Krakow, Poland
[2] Jagiellonian University Medical College, Department of Biophysics, Sw. Lazarza 16, 31-530, Krakow, Poland

Received 31 July 2019, revised 5 October 2019
Accepted for publication 17 October 2019
Published 16 January 2020

Abstract

Water is a crucial element of every living system, but its importance reaches further than biology. Thus, studying properties of water in different setups creates opportunities to bring together students of many disciplines. Here we propose a laboratory experiment on water network percolation in hydrated yeast, which enables description of the behavior of the water network surrounding living organisms during the dehydration process. Since the problem is interesting from a physical as well as biological point of view, the experiment can be introduced to student labs of both disciplines. In the experiment a simple RC circuit is used to observe 3D and 2D percolation phenomena in a sample of yeast. The parameters characterizing the phenomenon, such as percolation threshold and critical exponent, derived from the experimental data, provide information about the spatial organization of the water network surrounding yeast cells. The results obtained by four bioscience students using a simplified experimental setup are comparable with those presented in the literature and obtained by utilization of much more complex experimental methods.

Keywords: physics education, biophysics, percolation, physics laboratories

(Some figures may appear in colour only in the online journal)

Introduction

Physics can be perceived differently by different scientists. One example is a difference in perception of physics between biologists and physicists, especially regarding how and what to teach in physics courses for biology or life-science students in general [1, 2]. Biology students argue that describing biology in terms of numbers and variables is unappealing to them and does not aid understanding. However, at the same time they are able to conduct some physical experiments and calculations, which allow them to understand some physical phenomena, processes and relationships [3, 4]. Thus, physics could be interesting and even useful for biology students; however, it has to incorporate authentic biological applications, preferably based on contemporary research. This can be done not only by careful selection of physics topics, e.g. fluid flow instead of heat engines, but also by development of novel experiments in which physics supports understanding or describes some biological mechanisms, which are hard to consider only at the biological level [3, 5]. This could also be easily applied to other life-science students, especially biophysicists.

In this contribution we propose a laboratory experiment on water network percolation in hydrated yeast, which enables description of the water network and its evolution in a sample of living organisms during the dehydration process. It is well known that water is a 'molecule of life' [6] and that lack of water causes death in most living organisms [7, 8]. Thus, describing (and to some extent understanding) the behavior of the water network during the dehydration of living organisms could be interesting and valuable for life-science students. For this, percolation could serve as a flagship experiment. Percolation theory is utilized in various fields: science (e.g. modeling of brain connections [9]), industry and economics (flight planning, financial modeling [10]) or even global networks [11]. It is also widely used in biology in different contexts. It is worth mentioning, for example, the propagation of a forest fire [12], the spread of the plague virus [13], the development of animal habitats [14] or the fragmentation of the hepatitis B virus [15]. However, none of the above examples can be transferred to the student lab, due to the scale of the phenomenon, security, experimental method or complexity of the analysis. Thus, the experiment proposed by us may be an interesting, easily applicable alternative, enabling investigation in the student laboratory and thus understanding of the principles of an universal phenomenon. The activity described in this paper stands at the crossroads of physics and biology, and could be used in teaching both, especially in regard to students studying cross-disciplinary subjects.

In this contribution we present only an extract of the percolation theory, introducing relevant equations in their final form. We refer readers more interested in the details to the cited works.

Theoretical background

By its definition, the percolation phenomenon describes transport in any network in which the number of channels enabling this transport changes randomly [12]. In such a network a sharp phase transition is observed in the course of the continuous and random change of the number (or, more generally, density) of channels. An example of such a network is a grid formed by water molecules in a sample of living organisms [16, 17]. A characteristic feature of all percolation phenomena is their critical and universal behavior in the proximity of a well-defined threshold density of transport channels, called the percolation threshold. In any percolating system the transport-describing variable has a power-law dependence on the departure from the percolation threshold. The value of a critical exponent in this power-law dependence well defines the dimensionality of the percolating network [12, 18].

The percolation phenomenon has been extensively studied, both theoretically and experimentally, in regard to different systems and transport processes. One of the first described examples was the conductivity percolation on a grid of random resistor networks [19]. In such a grid the connections consist of electrical resistors that are present with probability p (which can be described in terms of density of resistor bonds) in the network. If the grid is undamaged ($p = 1$) and connected to the battery, the electric current flows through the network. However, if one starts replacing resistors with insulators (or cutting the grid) at random, probability p decreases and the current stops flowing at a certain probability p^* (or at a certain density of resistors present in the network), called the percolation threshold. Just before reaching the percolation threshold, the conductivity has a power-law dependence, with critical exponent μ:

$$\sigma \sim (p - p^*)^{\mu}. \tag{1}$$

In this study we focus on conductivity percolation, which can be observed in porous and granular materials, including biological samples, in the course of their dehydration [16, 20–23]. In such systems conductivity is enabled by the presence of water molecules forming a water network. If the number of water molecules decreases, the conductivity also decreases. Since dehydration is a random process of breaking water connections, conductivity percolation is observed in such systems.

In biological materials, like the yeasts used in this study, one can observe two percolation phenomena: three- and two-dimensional [21]. Dimensionality refers to the water network in which percolation occurs.

3D percolation

3D percolation occurs first, when the amount of water in a sample is relatively high. To describe it, we introduce a water mass fraction $\rho = \frac{m_w}{m}$, defined as the ratio of mass of water to mass of the sample. For a classic case, the conductivity of the network (σ) follows a power-law dependence [12] with the water mass fraction as a variable. In the vicinity of the percolation threshold [21]:

$$\sigma = \sigma_{\mathrm{Y}} \frac{(\rho^*)^s}{(\rho^* - \rho)^s},\ \rho < \rho^* - \delta_1 \tag{2a}$$

$$\sigma = \sigma_{\mathrm{W}} \frac{(\rho - \rho^*)^t}{(1 - \rho^*)^t},\ \rho > \rho^* + \delta_2 \tag{2b}$$

where ρ^* is the water mass fraction at the percolation threshold, σ_{W} is the conductivity of intercellular water, and σ_{Y} is the remnant conductivity. Symbols s and t stand for critical exponents, respectively below and above the percolation threshold. For 3D percolation theoretical and experimental works indicate t in a range of 1.9–2.2, and s in a range of 0.4–1.2 [12, 21]. The sum of δ_1 and δ_2 indicates the 'width' of the transition range Δ around the percolation threshold and can be related to σ_{Y}, σ_{W}, t and s [21, 24].

2D percolation

In the course of further dehydration of the sample the amount of water decreases, which results in a further decrease in conductivity, until the 2D percolation threshold at which point the conductivity disappears. This is the last point at which we can observe the long-range connection of the water network [22, 23]. In the examined system, the conductivity in the vicinity of the percolation threshold can be described by the relation:

$$(\sigma - \sigma^*) \propto (h - h^*)^{\mu},\ h > h^* \tag{3}$$

where h is the hydration level, defined as the ratio of mass of water to mass of a completely dry sample $\left(h = \frac{m_w}{m_0}\right)$, μ is the critical exponent for 2D percolation ($0.9 < \mu < 1.3$)

[12, 22, 25], and asterisks indicate the values at the percolation threshold. Note that the water fraction level (ρ) in a 3D network is replaced by the hydration level (h) in a 2D network. This is because of the fact that close to the 3D percolation threshold the sample is a wetted system (similar to microemulsions [26]), while close to the 2D percolation threshold the system is moistened and thus similar to humidified silicates [21, 22] or hydrated lysozyme powders [27].

Proposed activity

Biological sample

In the proposed activity students study the electric behavior of a sample of hydrated yeast, mostly because in such a sample both 3D and 2D conductivity percolation phenomena can be observed in one course of dehydration. Another reason is that it is easily accessible in comparable samples at a very low price in any supermarket. Last but not least, yeast gives the life-science students a biological context that can be more motivating for them than in the case of using a non-living material.

Industrial baker's yeast forms an amorphous matrix, in which globular yeast cells each of diameter c.a. 5 μm are tightly packed. The space between them is filled with water, which is called intercellular water. This water forms a conductive network, spatial at the beginning of the dehydration process. During the dehydration, firstly water slowly evaporates from this space until the point at which this network is destroyed. At this point, the remaining water in the sample covers the cell surfaces (and is called sometimes 'skin water'). This transition between a 3D and 2D water network is observed as a 3D percolation, and the well-defined density at which it occurs is called the 3D percolation threshold. From this point, water conducts only on the cell surfaces. However, it still evaporates, so this skin network also deteriorates until the point at which the last connection between two ends of the sample breaks. This is observed as a 2D percolation phenomenon at the 2D percolation threshold, well defined in terms of so-called hydration level h (see below). The water in the sample is still present, but, figuratively speaking, only as small lakes on the yeast cell surfaces, unconnected with each other and therefore non-conducting. Note that at the 2D percolation threshold the yeast is still alive [21, 28]; however, the determination of the exact relationship between viability of the yeast sample and moisture content is beyond the scope of this lab experiment, mostly due to the equipment constraints in typical biophysics student laboratories. High viability of the yeast sample in the course of the study is secured by very slow dehydration executed via still air and by keeping the surrounding temperature at 23 °C [29, 30].

Experimental setup

Similar studies have been done previously [21], but they required advanced equipment and therefore were not transferable for utilization in student laboratories. To achieve this, we designed an experiment with a simplified experimental setup [31], in which a thin sample of yeast is placed inside a parallel-plate capacitor. The upper plate has 56 small holes, which enable free dehydration of the sample to the air. This capacitor is connected in series with a resistance decade box, both connected to an AC power supply. This system, presented in figure 1, is a voltage divider, in which as a result of drying the sample, the capacitance and resistance of the sample change. This is observed as modulation in voltage amplitudes measured across each element and the phase shift between them. A simple oscilloscope is

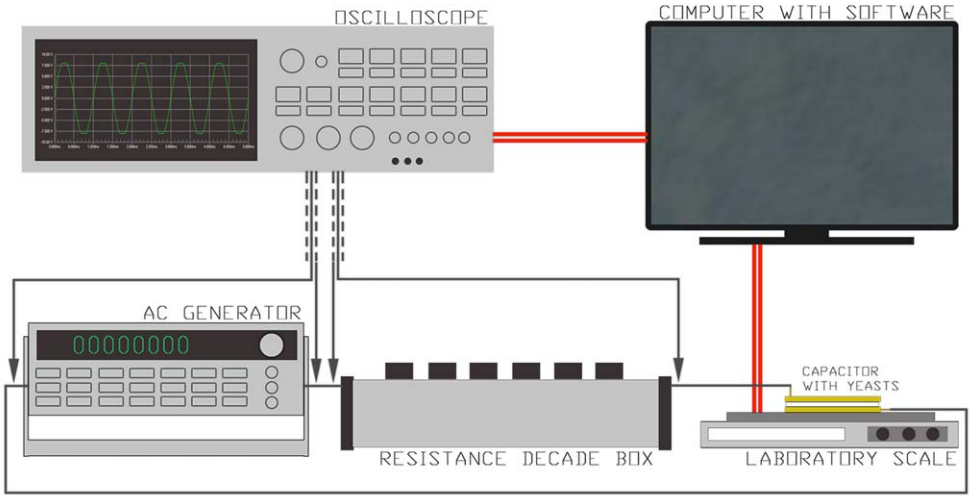


Figure 1. Experimental setup schema.

used to conduct measurements. Two signals need to be measured simultaneously: the voltage amplitude across the resistance decade box (U_R) and the voltage amplitude across the AC generator (U_{IN}), together with the phase shift between them (φ). Typically conductivity percolation studies require measurements over a huge range of frequencies at each time point [21, 22]. Such data are further analyzed in quite a sophisticated and time-consuming way in order to find the conductivity of the sample. However, the latest research showed [32] that for frequencies no higher than 100 kHz (low-frequency region) this could be substantially simplified by conducting the experiments and data analysis at one selected frequency, due to the fact that in this range of frequencies the dielectric loss spectrum is dominated by the conductivity contribution. The comparative results of the conductivity percolation parameters obtained with use of both methods of data analysis did not show any statistically significant differences in the values of the studied parameters over a range of different samples [32]. Thus, it was concluded that as long as the frequency dependence of electrical quantities is not being investigated, both the simplified and the extended data analysis carry the same information for investigation of the percolation phenomenon. Therefore, if the resistance of a resistance decade box (or conductance, $G = 1/R$) is known, and measurements are conducted at a constant frequency of alternating current, one can calculate the conductivity of a sample:

$$\sigma = \frac{U_{IN}\cos\varphi - U_R}{U_{IN}^2 + U_R^2 - 2U_{IN}U_R\cos\varphi} \cdot \frac{U_R G \varepsilon_0}{C_0} \tag{4}$$

where C_0 is the capacitance of an empty capacitor and ε_0 is the vacuum permittivity.

The second parameter needed for examining the percolation is the amount of water in the yeast sample (then recalculated into water mass fraction, ρ or hydration level, h). To define both, the mass of the sample is measured until it dries completely. The mass is measured on a scale with accuracy 0.001 g. It is worth noting that since the capacitor is connected to other parts of the experimental setup with cables, it will influence scale readings, and this has to be taken into account during the analysis of the data.

Since the percolation threshold could not be predicted *a priori*, all measured quantities have to be read from the beginning until the sample is completely dry. We propose measurements with a constant time step, e.g. 3–5 min. To this end, automatization of the process

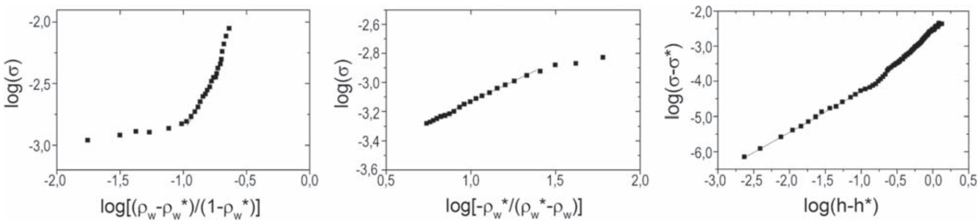


Figure 2. Examples of students' results. From left: Above 3D percolation, below 3D percolation, above 2D percolation threshold.

is required, since measurement can take a few days. For recording the electric quantities we have used KeySight BenchVue Oscilloscope software, and for storage of the mass readings we have used a simple program we designed ourselves, which records data from the scale and saves it to a text document at the programmed measurement pace.

In our group of students all measurements were conducted independently at the same experimental setup with the same external conditions ($T = 23$ °C and humidity around 60%). The frequency of the alternating current was set at 100 kHz, and the resistance of the resistance decade box was set to one of 300 Ω, 1 kΩ, 3 kΩ and 10 kΩ, in order to verify whether this change influences the final results. Measurements were conducted at a 3 min interval. The students' main tasks during the experiment were as follows:

- filling the capacitor with the sample of fresh baker's yeast;
- assembling the experimental setup;
- setting up experimental parameters (frequency, resistance of the resistance decade box, time step);
- recording data;
- data analysis: identification of the percolation threshold, determination of critical exponents and the width of the 3D percolation transition; and
- drawing conclusions.

The entire experiment usually lasts 3–4 days, mostly due to very slow dehydration of the yeast sample via still air and the initial form of the sample containing as much as 70% water. In the case of other samples, like sand or aerosil, the duration of dehydration would be much shorter, i.e. 5–8 h. Since it is not necessary to intervene after only 4 days, data recording can easily last an entire week and the experiment can be treated as a two-session-long laboratory project, which is quite common at our university for labs with bachelor's and master's students.

Examples of students' results

Below we present the results of four master's students who conducted the proposed activity. Three studied biophysics and one studied biochemistry.

Data analysis is required at first recalculation of the measured quantities in order to obtain the conductivity of the sample and water mass fraction or hydration level. In the next steps, the students prepared graphs illustrating the experimental data on a log–log scale, according to equations (2) and (3). On this scale, the slope of the fitted linear function is equal to a critical exponent which describes the dimensionality of the network in which the observed phenomenon occurs. In figure 2 examples of students' results at both percolation thresholds

Table 1. Comparison of students' results of percolation parameters with data obtained by one of us (DS) with use of a different experimental setup and data analysis method [21].

Percolation 3D			Percolation 2D		
Quantity	Students	Literature	Quantity	Students	Literature
ρ^*	0.589(17)	0.595	h*	0.298(14)	0.302
s	0.59(07)	0.57(01)	μ	1.12(06)	1.08(02)
t	1.99(12)	1.94(11)			
Δ	0.089(11)	0.09(1)			

are shown. Note that the process of finding the percolation threshold is a kind of 'hunting'. It requires preparation of many graphs, each for a different water mass fraction or hydration level and their corresponding conductivity. The best parameter estimation is obtained in a step-by-step procedure, consisting of picking a discrete value of ρ or h as a potential value ρ^* or h^* together with the corresponding value of σ^*, and by fitting the respective equations (2*a*), (2*b*) or (3) in the vicinity of ρ^* or h^* (this sequence is repeated over the entire range of discrete ρ or h values). Successive linear regression analyses on the log–log scale yield the fitting parameters ρ^* or h^* and the exponent t, s or μ, of which we choose those resulting from the best fit within the largest range of ρ or h in the vicinity of ρ^* or h^*.

Students' results are gathered in a table 1 together with the literature values. The uncertainties of results were calculated as standard deviations. As one can see, students' results are consistent with the literature values for both percolation thresholds, which shows that the proposed simplified experimental and data analysis method is suitable for such measurements. Students' results were similar to each other, so *a priori* adjusted resistance of a resistance decade box did not influence the final results. The values of critical exponents obtained in the experiment confirm identification of both 3D and 2D percolations with similar mass fractions and hydration levels among all students.

Extension of the laboratory experiment

Due to the limited lab time spanning only two sessions separated by a week-long period of sample dehydration, the percolation conductivity study was limited to investigation of a ready-to-use sample of baker's yeast. However, if the lab class could last longer, the following modifications and extensions of the proposed lab module could be implemented:

- rehydration of the sample done just after reaching the percolation threshold and re-examination of the conductivity percolation parameters with subsequent comparison of the results from both courses of dehydration;
- measurement of the yeast sample viability in the course of dehydration by utilization of one of the methods described for example in [33];
- preparation of other living hydrated biological samples (e.g. algae) and comparative study of the conductivity percolation parameters in samples containing different species.

The abovementioned alterations to the laboratory activity on conductivity percolation can be also utilized as a second cycle of the percolation lab, delivered in a more open-inquiry manner.

Conclusions

Percolation theory is widely used in different physical, industrial, economical, medical, biological and social contexts, therefore it seems valuable and important to encourage students to explore it. The experiment proposed in this paper studies the conductivity percolation phenomenon in a sample of living, wetted yeasts, in which conductivity changes are related to the amount of water in the sample. The sample is studied with use of a simplified experimental setup and on the basis of a simplified data analysis method, enabling implementation of the activity in most student labs. The results of the experiments provide information about the spatial organization of the water in the yeasts. We conclude that the experiment on percolation conductivity proposed in this article could be interesting and valuable for bioscience students, giving them the opportunity to conduct physical investigation on biological samples, providing physical meaning in a biological context and showing the students the beauty of universality of some mathematical constructs, which they usually treat with reserve.

ORCID iDs

Daniel Dziob https://orcid.org/0000-0003-2065-600X
Dagmara Sokołowska https://orcid.org/0000-0002-3993-1588

References

[1] Hoskinsona A, Couch B, Zwickl B, Hinko K and Caballerob M 2014 *Am. J. Phys.* **82** 434
[2] Redish E F *et al* 2014 *Am. J. Phys.* **82** 368
[3] Meredith D and Redish E 2013 *Phys. Today* **66** 38
[4] Watkins J and Elby A 2013 *Cell Biol. Educ.-Life Sci. Educ.* **12** 274
[5] Gouvea J S, Sawtelle V, Geller B D and Turpen C 2013 *Cell Biol. Educ.-Life Sci. Educ.* **12** 187
[6] Franks F 2000 *Water: A Matrix of Life* 2nd edn (London: Royal Society of Chemistry) New edition
[7] O'Brien E M 2006 *J. Biogeogr.* **33** 1868
[8] Rodrigues D F and Tiedje J M 2008 *Appl. Environ. Microbiol.* **74** 1677
[9] Parshani R, Carmii S and Havlin S 2010 *Phys. Rev. Lett.* **104** 258701
[10] Amor S, Huy T and Bui M 2006 A percolation based model for ATC simulation *Proc. 4th Int. Conf. on Computer Sciences, Research Innovation and Vision for the Future* (Vietnam) p 17–22
[11] Grabowski A, Kruszewska N and Kosinski R A 2008 *Eur. Phys. J.* **B66** 107–13
[12] Stauffer D and Aharony A 1991 *Introduction to Percolation Theory* (Washington, DC: Taylor & Francis)
[13] Davis S, Trapman P, Leirs H, Begon M and Heesterbeek J A P 2008 *Nature* **454** 634
[14] Boswell G P, Britton N F and Franks N R 1998 *Proc. Roy. Soc. B: Biol. Sci.* **265** 1921
[15] Brunk N, Siang L, Glazier J, Butske W and Zlotnick A 2018 *Phys. Biol.* **15** 5
[16] Bordi F, Cametti C and Colby R 2004 *J. Phys. Condens. Matter* **16** R1423
[17] Sokolowska D, Krol-Otwinowska A, Bialecka M, Fiedor L, Szczygiel M and Moscicki J K 2007 *J. Non-Cryst. Solids* **353** 4541
[18] Pennetta C, Reggiani L and Tref G 2000 *Phys. Rev. Lett.* **84** 5006
[19] Fisch R and Harris A B 1978 *Phys. Rev.* B **18** 416
[20] Voth G A 2006 *Acc. Chem. Res.* **39** 143
[21] Sokolowska D, Krol-Otwinowska A and Moscicki J K 2004 *Phys. Rev.* E **70** 052901
[22] Sokolowska D, Dziob D, Gorska U, Kieltyka B and Moscicki J K 2013 *Phys. Rev.* E **87** 062404
[23] Nowak J, Dziob D, Rutkowska M, Chmielarz L, Gorska U, Sokolowska D and Moscicki J K 2015 *Micropor. Mesopor. Mater.* **206** 202
[24] Saidi Z, Mathew C, Peyrelase J and Boned C 1990 *Phys. Rev.* A **42** 872

[25] Rupley J, Siemiankowski L, Careri G and Bruni F 1988 *Proc. Natl Acad. Sci. USA* **85** 9022
[26] van Dijk M A 1985 *Phys. Rev. Lett.* **55** 1003
[27] Careri G, Giansanti A and Rupley J A 1988 *Phys. Rev.* A **37** 2703
[28] Attfield P V, Kletsas S, Veal D A, van Rooijen R and Bell P J 2000 *J. Appl. Microbiol.* **89** 207
[29] Benley L, de Maranon I M, Marechal P-A and Gervais P 2000 *Intl. J. Food Microbiol.* **55** 275
[30] Gervais P, Marechal P-A and Molin P 1992 *Biotechnol. Bioeng.* **40** 1435
[31] Dziob D, Mlynarczyk M and Sokolowska D 2019 *Rev. Sci. Instrum.* **90** 076105
[32] Moscicki J K, Sokolowska D, Kwiatkowski L, Dziob D and Nowak J 2014 *Rev. Sci. Instrum.* **85**
[33] Kwolek-Mirek M and Zadrag-Trecza R 2014 *FEMS Yeast Res.* **14** 1068–79

PAPER

Phys. Educ. 55 (2020) 045004 (6pp) iopscience.org/ped

Simple method for measuring thermal conductivity

Daniel Dziob[1,2,4] and **Mojca Čepič**[3]

[1] Jagiellonian University, Smoluchowski Institute of Physics, Krakow, Poland
[2] Jagiellonian University Medical College, Department of Biophysics, Krakow, Poland
[3] University of Ljubljana, Faculty of Education, Ljubljana, Slovenia

E-mail: daniel.dziob@uj.edu.pl

Abstract

Heat transfer is a phenomenon well known from everyday life. It is intuitively connected to the properties of materials, that is, to the physics concept of thermal conductivity relevant for cooking or maintaining the constant temperature in rooms, even without being familiar to the underlying physics. However, measurement of thermal conductivity is usually demanding, but here we present a simple, quick, and almost hands-on method that yields quite accurate values for thermal conductivity of insulating and semi-insulating materials, appropriate for a classroom setting.

Keywords: thermal conductivity, physics education, experiments

Supplementary material for this article is available online

1. Introduction

Heat transfer is a well-known everyday phenomenon. People intuitively connect it to the properties of materials, for example, to cooking or insulation of houses, even without being familiar with the underlying physics. Heat transfer strongly depends on the material, characterized by its coefficient of thermal conductivity λ. In the school setting, it is unfortunately rather demanding to find its approximate value, because experiments usually require rather sophisticated equipment not available in schools [1–4]. An additional problem is the heat transfer due to convection, which is difficult to eliminate, or at least to control and estimate [3, 5, 6]. Several experiments, however, provide experience and scaffold the concept to students on the semi-quantitative level [7], but teachers are often forced to limit the consideration of thermal conductivity to comparison of materials with significantly different conductive properties, such as metals and polystyrene. We believe that this paper bridges this gap, as we present a simple, quick, and almost hands-on method for measuring the thermal conductivity of insulating and semi-insulating materials, appropriate for a classroom setting.

The suggested measurement of thermal conductivity is based on measuring the temperature's dependence on time of a cooling water immersed in a mixture of ice and water for a relatively short period of time. The setup for the suggested experiment is very simple and straightforward. For the analysis of the measurements we suggest

[4] Author to whom any correspondence should be addressed.

a simple trick for reorganization of the measured data, which straightforwardly leads to the coefficient of thermal conductivity. The method gives relatively accurate thermal conductivities for poor conductors and for conductors that have smaller thermal conductivity than still water, which is used as a cooling medium. The method has some limitations, however. Although the setup eliminates the influence of convection and radiation, it is still not appropriate for very good conductors like metals. Possible reasons for this limitation are discussed in the paper as well.

The method is actually a modernized and simplified version of the approach presented more than two decades ago [2]. The experimental setup of this method is much simpler and easier to assemble. The estimation of the transferred heat, which was based on determination of the mass of melted ice [2] is replaced by a continuous measurement of the temperature, and the analysis of the measured data can be simplified for time periods during which the measured medium cools down for a few tens of degrees only.

2. Thermal conductivity and cooling of water

Let us consider a closed container with conductive walls filled with water. The walls are made of a material with the coefficient of thermal conductivity λ having a thickness d, and the surface area of the container S. The water in the container has a mass m_W and the specific heat c_W. Initially, the container is filled with hot water close to the boiling point. The container with the water having the initial temperature T_{W0} is quickly immersed into a large reservoir with a mixture of water and ice having temperature T_R. Temperature of the water in the container $T_W(t)$ is measured for a few minutes. For described circumstances, the interfaces between the container and the water in its interior and between the container and the water-ice mixture in the reservoir have the same temperature as the liquid in contact, therefore the transfer of heat by radiation is negligible (see also supplementary materials 1, available online (stacks.iop.org/PhysED/55/045004/mmedia)). In addition, the liquid inside the container is slightly mixed, which ensures equal temperatures of the water near the walls and in the middle of container. The convection in the reservoir with the mixture of water and ice is negligible. Therefore, to describe the temperature dependence of the water in the container on time $T_W(t)$, only the analysis of heat conduction through the walls of the container satisfies.

Let us define the temperature difference between the time dependent water temperature and the (constant) reservoir temperature as θ, i.e. $\theta(t) = T_w(t) - T_R$. The temperature of water in the container decreases due to the heat transfer through the walls of container:

$$m_W c_W \frac{d\theta}{dt} = -\lambda \frac{S}{d}\theta, \qquad (1)$$

which yields the time dependence of the temperature difference as

$$\theta(t) = \theta_0 e^{-\frac{t}{\tau}}, \tau = \frac{m_W c_W}{\lambda}\frac{d}{S}. \qquad (2)$$

Here θ_0 is the initial temperature difference $T_w(0) - T_R$. The characteristic time, τ, is the time when the temperature difference decreases to $1/e$ (approximately 37%) of the initial value. Equations (1) and (2) are classical expressions students meet at exercises in introductory physics courses. For times smaller than characteristic times $\Delta t < \tau$, the dependence (2) can be linearized, which leads to the same expression as in (1) in a non-differential form. In the time Δt the temperature of water in the container changes for $T(t) - T(0) = \theta(t) - \theta_0$ due to the transfer of heat to the container. The transfer of heat is a consequence of temperature difference $\theta(t)$ between the water in the container and in the reservoir. The linearized equation (1) becomes

$$m_W c_W (\theta(t) - \theta_0) = -\lambda \frac{S}{d}\theta(t)\,\Delta t. \qquad (3)$$

Let us use a shorter notation for the time dependent temperature difference $\theta(t) = \theta$ in continuation, to further simplify the expression (3) to

$$\frac{(\theta_0 - \theta)}{\theta} = \frac{\lambda S}{m_W c_W d}\Delta t. \qquad (4)$$

Introducing the ratio between the magnitude of the temperature change and the temperature difference as $r(t) = \frac{(\theta_0 - \theta)}{\theta}$ and see that it depends linearly on time

$$r(t) = \frac{\lambda S}{m_W c_W d} \Delta t, \tag{5}$$

with a gradient $\frac{\lambda S}{m_W c_W d}$, allows for straightforward determination of thermal conductivity λ.

3. The experimental setup

The here-presented experimental setup consisted of a reservoir with inner dimensions 24 cm times 25 cm and the height about 20 cm containing several liters (between 5 and 7 in figure 1(b)) of water with crushed ice, in ratio 1:1 approximately. For the setup, a polystyrene box for transport of cooled vegetables was used. As containers for the warm water, we used cups from different materials. Each cup was filled with hot water directly out of the kettle, it was covered by a thick polystyrene cover, and the thermometer connected to Vernier interface was inserted into the hole in the cover. The cup prepared in this way was immersed into the reservoir as shown on the schematic drawing in figure 1(a). The whole procedure, from pouring the almost boiling water to the cup to inserting the cup into the reservoir, should be as quick as possible in order to achieve the highest possible temperature difference between the water in the cup and in the reservoir. The temperature of the water in the cup was measured by a thermometer for the duration of a few minutes. The temperature of the mixture of ice and water in the reservoir was also controlled and it was slightly stirred once per minute to prevent local increases of the temperature close to the cup's walls.

4. The measured coefficients of the thermal conductivity

To measure the thermal conductivity of different materials, the time dependent temperature of the cooling water in five different cups made of different materials was measured. The cups used for this experiment were made of polystyrene, wood, plastic, porcelain, and stainless steel. Properties of the cups, together with calculated characteristic times τ from the thermal conductivities available in literature [8] are presented in table 1. Temperature was measured by the Coach6 system external thermometer with an accuracy of 0,1 °C in 1 s intervals. Diameters were obtained by a caliper and for the specific heat of water, the value 4 200 J kg^{-1} K^{-1} was used. More properties of cups are given in the supplementary material 2.

Figure 2(a) shows the temperature dependence of water in the studied cups during cooling. Figure 2(b) shows the time dependence of the ratio $r(t)$ calculated from the measured temperatures for a time period of 500 s (polystyrene, wood, plastic) or of 200 s (porcelain, steel). The coefficient of the linear fit was extracted for each material and a thermal conductivity was calculated from the fitted slope (5). Results are presented in table 2.

The obtained values of thermal conductivity are consistent with the ones given in literature[11] for all used materials except for steel. Therefore, one can conclude that for insulators like polystyrene or weak insulators like wood, plastic, or porcelain, the method is quite reliable and accurate. However, comparing the measured thermal conductivity for steel and its value in literature, one can observe a severe discrepancy. How can this result be interpreted? The calculated characteristic time for a steel cup is 4 s, but the one measured according to the method is more than 70 s. Even more, the time dependence of the ratio $r(t)$ within the measured characteristic time span is almost linear, suggesting that also the cooling behavior is very similar to poor and weak conductors and, therefore, the method could be used. But, it is far from true. We investigated this phenomenon into more detail. We measured several 'cups', that is, empty cans from various metals like cola can and cans previously filled with conserved vegetables made of different metals. This discrepancy occurred for each good thermal conductor. We speculate that the origin of this discrepancy is the following. Cans from good thermal conductors transfer the heat very efficiently through the walls of the can. As the water in the can close to the walls cools almost uniformly, the convection is probably small. One therefore cannot consider that the whole water in the can cools uniformly, but the transfer of heat to the water in the interior of the cup has to be considered as well. The thermal conductivity of water without a convection is 0.6065 W m^{-1} K^{-1} which is more than ten times less than that of steel. Layers of water close

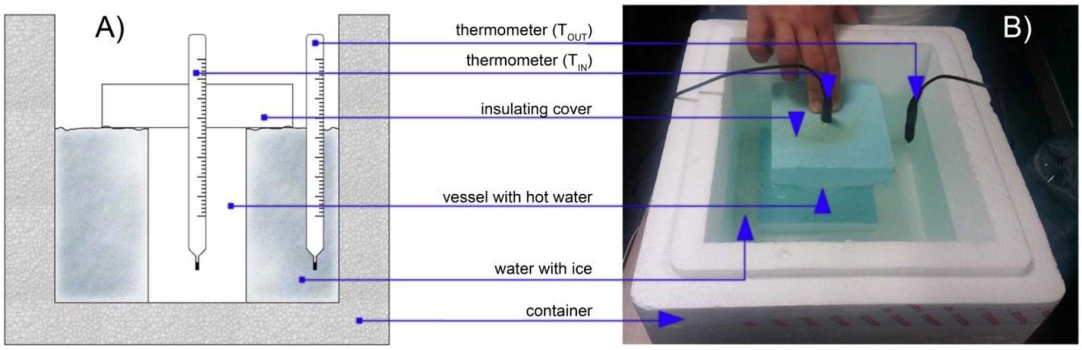


Figure 1. (a) The sketch of the experimental setup. (b) The actual setup.

Table 1. The cups and their properties with calculated characteristic times τ.

	Polystyrene	Wood	Plastic	Porcelain	Metal (stainless steel)
Photo					
S (cm^2)	213 ± 8	196 ± 7	307 ± 9	203 ± 7	230 ± 8
d (mm)	2.0 ± 0.1	5.0 ± 0.1	2.0 ± 0.1	4.9 ± 0.1	0.6 ± 0.1
m_w(g)	260 ± 10	270 ± 10	653 ± 10	348 ± 10	584 ± 10
λ (W m^{-1} K^{-1})[c]	0.03	0.16[a]	0.3[b]	1.5	16
τ (s)	3400 ± 430	1807 ± 168	596 ± 56	240 ± 20	4.0 ± 0.9

[a] The thermal coefficient of wood depends on a type of wood and the direction with respect to fibers and it varies from 0.12 to 0.17.
[b] The thermal coefficient of plastic depends on the type of polymers and the temperature and it varies from 0.17 to 0.50, type of plastic in the mug was unknown.
[c] The values for the thermal conductivity were taken from [8]. Additional data are available in supplementary material 2 online.

to the walls with intensive heat transfer actually act like an inner 'insulation' in the cup. If the data for good conductors are analyzed in the same way as for poor and weak conductors, what is actually measured is an effective thermal conductivity that includes the layers of water close to the walls. This explanation is further supported by closer examination of the initial time dependence of the ratio $r(t)$ of the wooden cup. It is not linear at the beginning of the measurement. Why not? Wood is a relatively good insulator but also has a significant thermal capacity. Therefore, after the wooden cup with the hot water was set into the reservoir, the cup's walls started to cool. The conduction of heat could not have been described by the simple equation given in (2) as long as the temperature profile in the wooden walls of the container was not linear.

The presented methodology is therefore limited to materials with the thermal conductivity comparable or lower than water, which was already mentioned earlier [5]. As the thermal conductivity is calculated from the slope of the $r(t)$, one remains on the safe side if the water temperature does not change more than 10 °C to 20 °C and the temperature difference to the reservoir is still above 50 °C. To avoid unexpected transition phenomena like the time needed for warming or cooling the cups, it is advised to start the measurement

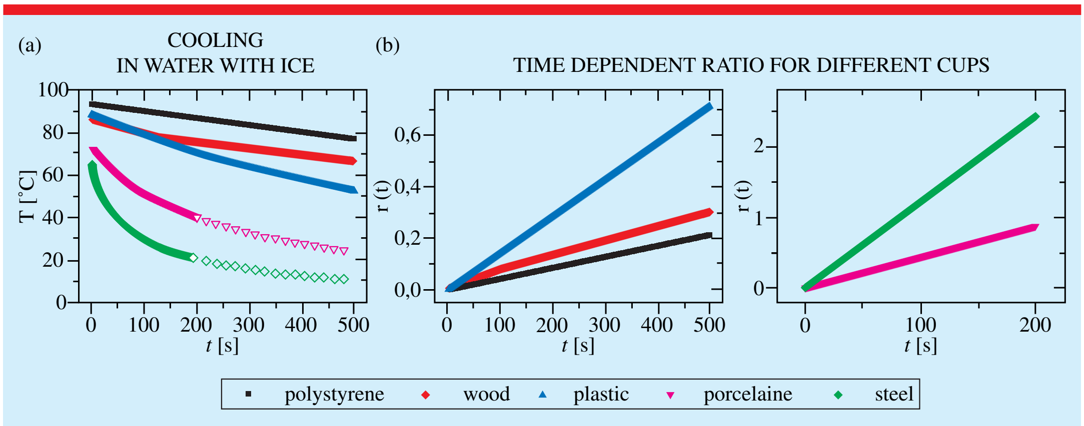


Figure 2. (a) Temperature dependence of the water in the cup immersed in the mixture of water and ice in the analyzed range. (b) The time dependent ratio $r(t) = \frac{(\theta_0 - \theta)}{\theta}$ calculated from the measured temperatures for the studied cups. The data is presented in two separate graphs with different time and ratio scales as the slopes of *r(t)* for porcelain and steel are much larger. Measurements presented by open symbols in (a) are not included in the second graph in (b).

Table 2. Results of linear regression to acquired data and calculated thermal conductivity coefficients. Theoretical values retrieved from [8]. Uncertainties of experimental values are between 20% and 30% in spite the simplicity of the method.

		Polystyrene	Wood	Plastic	Porcelain	Metal (stainless steel)
Slope k ($\Delta k/k = 0.02$)		0.00042	0.00053	0.00143	0.0047	0.014
Characteristic time $\tau(\Delta\tau/\tau = 0.02)$		2380	1890	700	212	71.5
$\lambda(\mathrm{W\ m^{-1}W\ K^{-1}})$	Literature	0.03	0.12–0.17	0.17–0.50	1.5	16
	Exp. value	0.04 ± 0.01	0.15 ± 0.03	0.25 ± 0.05	1.57 ± 0.30	4.4 ± 1.4

about 10–20 s after the cup is immersed to the reservoir.

In spite of the simplicity, the method gives surprisingly accurate thermal conductivities of weak and good thermal insulators. Because the water temperature in the cup for materials that could be measured by this method does not change dramatically, the computerized measurement as was used here is not crucial. The temperature could also be measured manually still resulting in relatively accurate values for thermal conductivities.

5. Conclusions

We presented a simple and efficient method for a relatively fast measurement of thermal conductivity appropriate for insulators or semi-insulators, that is, poor conductors. The presented setup could be used as a demonstration experiment in the classroom setting or as a rather simple and straightforward laboratory experiment. Simplicity, intuitiveness, and the low cost of the experiment are features that make the proposed experiment an excellent support to introduction of thermal conductivity. Moreover, the suggested method for the measurement of thermal conductivity links together the heat flow and its consequences, which additionally develops the awareness of processes that are common reasons for changes of temperatures and their dynamics. It also shows that the physical properties of different materials are not just abstract numbers calculated by engineers but real features, the consequences of which one can observe and measure their values.

ORCID iDs

Daniel Dziob https://orcid.org/0000-0003-2065-600X
Mojca Čepič https://orcid.org/0000-0003-0951-5302X

Received 20 January 2020, in final form 24 February 2020
Accepted for publication 18 March 2020
https://doi.org/10.1088/1361-6552/ab8100

References

[1] Lu L, Yi W and Zhang D L 2001 3ω method for specific heat and thermal conductivity measurements *Rev. Sci. Ins.* **72** 2996–3003
[2] Siviter J P 1993 A quick and simple way to measure the thermal conductivity of poor conductors *Phys. Ed.* **28** 254–5
[3] Talpe J H, Bekeris V I and Acha C E 1990 Measurement of thermal conductivity and heat capacity in an undergraduate physics laboratory *Am. J. Phys.* **58** 379–81
[4] Ortuno M, Marquez A, Gallego S, Neipp C and Belendez A 2011 An experiment in heat conduction using hollow cylinders *Eur. J. Phys.* **32** 1065–75
[5] Bee G, Ballentine K and Thomsen M 2008 Realistic problems involving thermal conductivity *Am. J. Phys.* **76** 970–4
[6] Ashegi M, Kurabayashi K, Kasnavi R and Goodson K E 2002 Thermal conduction in a doped single-crystal silicon films *J. App. Phys.* **91** 5079–88
[7] Planinšič G 2011 Soda cans aid teaching of thermal conductivity *Phys. Ed.* **46** 143
[8] https://www.engineeringtoolbox.com/thermal-conductivity (Accessed: 23rd November 2019)

Mojca Ćepič teaches future teacher of physics at University of Ljubljana, Faculty of Educations, both, theoretical subjects and methodology of teaching physics. Her research is focused on two fields: on physics education and on theoretical soft matter physics, more precisely on polar smectic liquid crystals. She is mostly occupied with designing and testing of simplified experiments, which introduce contemporary physics results to pre-university education.

Daniel Dziob is completing a doctorate in biophysics at the Faculty of Physics, Astronomy and Applied Computer Sciences, Jagiellonian University, Krakow. His research focuses on two fields: physics education and cell migration. He also works and teaches biophysics for medical students at the Jagiellonia University Medical College. As a hobby, he teaches physics to high school students, hikes and annoys his friends and collaborators with new ideas.

MODESTUM
OPEN ACCESS
EURASIA Journal of Mathematics, Science and Technology Education
ISSN: 1305-8223 (online) 1305-8215 (print)
2018 14(4):1111-1132 DOI: 10.29333/ejmste/81807

Class Tournament as an Assessment Method in Physics Courses: A Pilot Study

Daniel Dziob [1,2*], Lukasz Kwiatkowski [3], Dagmara Sokolowska [1]

[1] Smoluchowski Institute of Physics, Jagiellonian University, Krakow, POLAND
[2] Department of Biophysics, Jagiellonian University Medical College, Kracow, POLAND
[3] Department of Econometrics and Operations Research, Cracow University of Economics, Krakow, POLAND

Received 23 August 2017 ▪ Revised 9 December 2017 ▪ Accepted 11 December 2017

ABSTRACT

Testing knowledge is an integral part of a summative assessment at schools. It can be performed in many different ways. In this study we propose assessment of physics knowledge by using a class tournament approach. Prior to a statistical analysis of the results obtained over a tournament organized in one of Polish high schools, all its specifics are discussed at length, including the types of questions assigned, as well as additional self- and peer-evaluation questionnaires, constituting an integral part of the tournament. The impact of the tournament upon student improvement is examined by confronting the results of a post-test with pre-tournament students' achievements reflected in scores earned in former, tests written by the students in experimental group and their colleagues from control group. We also present some of students' and teachers' feedback on the idea of a tournament as a tool of assessment. Both the analysis of the tournament results and the students' and teachers' opinions point to at least several benefits of our approach.

Keywords: team work, cooperative learning, collaborative testing, K-12 physics, assessment methods, assessment as learning

INTRODUCTION

Testing knowledge is an integral part of educational assessment, the latter being a process of documenting content knowledge, skills, attitudes and beliefs, usually focused on an individual learner or a learning community as a whole. The most popular distinction in types of assessment is founded upon the difference between formative and summative assessment (Black & Wiliam, 1998; Garriso & Ehringhaus, 2007; Harlen & James, 1997; McTighe & O'Connor, 2005; Wiliam & Black, 1996 and references therein). In general, the formative assessment is carried out throughout a unit (course, project), whereas the summative one - at the end of a unit (course, project) (Harlen & James, 1997; McTighe & O'Connor, 2005). Some authors seem to distinguish between these types of assessment arguing that the summative assessment is "assessment *of* learning", while the formative one is "assessment *for* learning" (Black et al., 2004; Earl, 2004; Looney, 2011; Taras, 2005).

Focusing on the summative assessment (SA), we can point to three major criteria defining it: i) SA is used to determine whether students have learned what they were expected to learn (Earl, 2004; Harlen & James, 1997; Torrance & Pryor, 1998); ii) SA is carried out at the end of a specific teaching period, and therefore it is generally of an evaluative nature, rather than diagnostic one (Earl, 2004; Harlen & James, 1997; Torrance & Pryor, 1998); iii) SA results are often recorded as scores or grades that are then factored into a student permanent academic record (Biggs, 1998; Bloom et al., 1971; Earl, 2004).

Summative assessment can be performed in many ways (Black et al., 2010, 2011; McTighe & O'Connor, 2005; Scriven, 1967), though written tests are still the most prevalent (Talanquer et al., 2015; Taras, 2009; Vercellati et al., 2013). However, in different fields a few researchers have come up with an idea of carrying out assessment in some alternative manners (Dochy et al., 1999; Rebello, 2011; Schuwirth & Vleuten, 2004). These include, in particular, different forms of a written test, extensively described and compared in the literature, such as free- and multiple-response tests (Wilcox & Pollock, 2014), concept tests (such as the Test of Understanding Graphs in Kinematics

✉ daniel.dziob@uj.edu.pl **(*Correspondence)** ✉ kwiatkol@uek.krakow.pl ✉ ufdsokol@cyf-kr.edu.pl

Contribution of this paper to the literature

- The study suggests and appraises a new method for evaluation, combining summative assessment with elements of formative one in a form of a tournament game taken in groups and being an example of "assessment as learning".
- The tournament is very flexible for inclusion of theoretical and practical tasks in different formats and may also comprise self- and peer-assessment questionnaires, as well as evaluation of attitude, motivation and interest.
- The analysis of the tournament results and students' opinions about the implementation in physics classes points out academic benefits for students and equal opportunities of improvement both for low- and high-performers.

(Maries & Singh, 2013), Force Concept Inventory (Hestenes et al., 1992) or Brief electricity and magnetism assessment (Ding et al., 2006) and others (Hitt et al., 2014; Wilcox et al., 2015), constructed-response tests (Slepkov & Shiell, 2014), essay tests (Kruglak, 1955), laboratory skills tests (Doran et al., 1993) and others. Also, many modifications and extensions of these tests have already been proposed in the literature, improving upon their original form (Ding, 2014; Docktor et al., 2015; Wooten et al., 2014; Zwolak & Manogue, 2015). On the other hand, some authors propose to blend formative and summative assessment techniques. According to (Wininger, 20015), such a combination, named "formative summative assessment", entails reviewing exams with students so that they get feedback about their comprehension of concepts. Nowadays, we can find different proposals of combining these two types of assessments (Fakcharoenpohl & Stelzer, 2014; Pawl et al., 2013; Wilcox & Pollock, 2015; Yu & Li, 2014), and the boundaries between them become more and more vague. One example of such an approach is "collaborative testing" – an idea of giving students the opportunity for working in groups during an exam (Guest & Murphy, 2000), at the end of an individual exam (Lusk & Conklin, 2003) or, more often, after the first, but before the second exam taken individually (Cortright et al., 2003; Ives, 2014; Rao et al., 2002) (the last two are sometimes named "two-stage exams"). Research has shown that there are many benefits of utilizing collaborative testing as a constructivist learning method. They are described in detail in (Duane & Satre, 2014; Gilley & Clarkston, 2014; Kapitanoff, 2009) and references therein.

In our study, we use a tournament – a competitive game between groups in the classroom – as a tool for summative assessment with formative evaluation elements. On the one hand, applying the mechanics of a game to make the process more appealing can be considered a gamification (Apostol et al., 2013; Deterding et al., 2011). Although the idea of introducing games in teaching is not new (Ifenthaler et al., 2012 and references therein; Moncada & Moncada, 2014), the very term of gamification has been coined only a few years ago, and has been gaining more and more popularity since then (Dicheva et al., 2015; Sadler et al., 2013; Sung & Hwang, 2013). The benefits of gamification (or, in more broad terms, game-based learning (e.g. Ifenthaler et al., 2012)) in the educational context are widely described in the literature (Banfield & Wilkerson, 2014; Dicheva et al., 2015; Hanus & Fox, 2015; Seaborn & Fels, 2015; Sung & Hwang, 2013).

Moreover, a tournament can be also considered as a kind of "collaborative testing", but unlike the forms mentioned above, we first conduct a group exam (distinguishing individual students' marks through their involvement and contribution in the group work), and, secondly, provide a control, individual test (only for the purpose of research, not influencing students' final marks). Following (Earl, 2004), where also the idea of "assessment *as* learning" is introduced (and in which student self-assessment, and, thereby, self-motivation are brought into focus (Hickey et al., 2012)), we design an alternative form of testing knowledge, combining the assessment with learning and a game at the same time. And by learning we mean not only the subject matter itself, but also acquiring and developing other skills, as well as stimulating positive, both intra- and interpersonal dispositions, such as self-motivation, language skills and group work in the form of cooperative learning (Jolliffe, 2007; Kagan, 1990; Slavin, 2000).

RESEARCH DESIGN

In this section we provide details on the tournament itself, including its organization, questions assigned and relevant evaluation procedures.

Tournament Schema

The tournament was performed in a high school in Wolbrom (a small town of ca. 9 000 inhabitants, in the South of Poland), and it involved 30 students in their final class (K-12). At the time the class had just accomplished a 22-hour course on electricity.

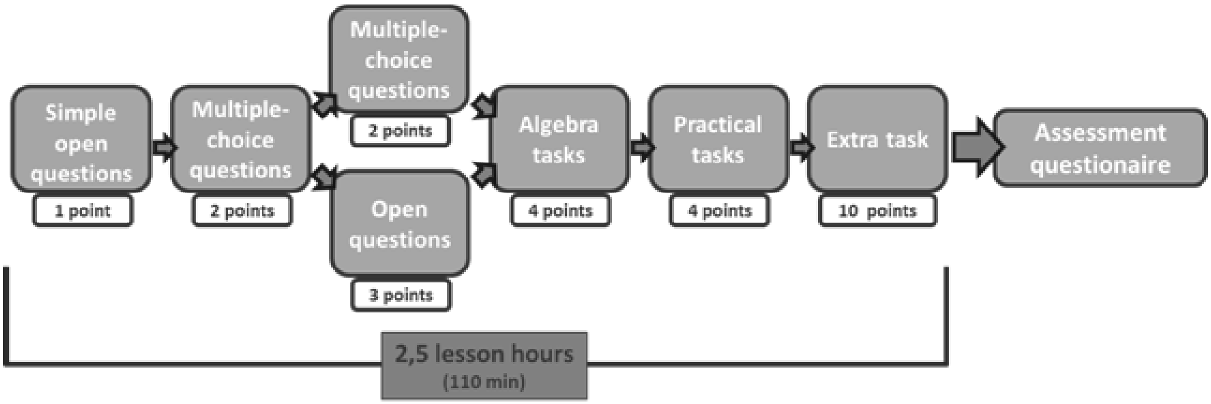


Figure 1. Tournament testing sequence

At the beginning of the actual event (lasting for 2.5 lesson hours), the students were divided randomly into 5 groups, by lottery, drawing out lots with names of fairytale heroes upon entering the classroom. After drawing a card with the name of a hero, every student held a seat at one of the five tables, each grouping heroes of one of five fairy tales. Then, actual tournament started. **Figure 1** presents the scheme of the entire process, which was designed on the basis of the former experience with utilizing different assessment formats by both the researchers and the teachers teaching in the school where the tournament was implemented. The overriding goal while formulating the scheme was to make the assessment more holistic by including tasks oriented not solely on the content-matter itself, as it often happens in typical tests, but also ones related to everyday life, and allowing the teacher to evaluate students' experimental skills as well.

The tournament began with open questions and multiple-choice questions with an increasing level of difficulty, and, therefore, an increasing number of available points (which were the reward for every correct answer). Further, calculus and some practical tasks were assigned. Finally, all groups faced an extra, common task, with the elements of time competition (the winning team was the first one that rang the bell and provided the correct solution to the problem faced by all teams at the same time). At the end, the students were asked to fill in a special self- and peer-assessment questionnaire. After a week, a post-test was performed. At each stage, the whole process was monitored by two independent teachers (apart from the major teacher of the class), who were responsible for the verification of verity and integrity of the student evaluation. The assistant teachers were not allowed to involve themselves in the tournament itself, with their scope of duties largely limited to overall student supervision and taking notes on the engagement and behavior of the participants. However, their role was also to aid the major teacher in assessing those students, who – at the stage of student self- and peer-evaluation (explained below) – would be found to appraise themselves or their teammates erroneously or unjustifiably.

The first two stages were organized in multiple rounds. The first phase comprised three rounds, and the second – two rounds. At the first three stages, in each round the teams attempted a task in consecutive turns. At the third phase the students faced a choice of undertaking either a 3-point open question or a 2-point multiple-choice one. It was intended to introduce some element of decision-making risk, thereby facilitating students' sense of responsibility for the choices made. The following three stages (i.e. the calculus, practical and extra tasks) were single-rounded and at each of them all teams were challenged with their tasks at the same time. The students were already familiar with the forms of all the assignments, for similar had been administered to them during previous class tests.

In the first three types of questions students from the currently "active" group were required to choose the number of a question, and then the team had the appointed time (respectively 30 seconds, 1 minute or 2 minutes) to deliver the answer. If they did not succeed or their answer was incorrect, other groups could take over the question and score extra points by ringing the bell and providing the correct answer. Allowing for such a possibility was meant to ensure attention and an active interest of each group in the question currently dealt with by any other team. During this part of the tournament, questions were projected onto the wall screen so as to make it available for all teams at the same time. In the calculus and practical tasks all groups worked simultaneously over different, randomly selected problems, received on sheets of paper. For providing the correct solution each group could earn maximally 4 points, and there was no possibility of intercepting unsolved problems by other teams. The practical task score included: 1 point for building a properly working experimental setup, 1.5 points for providing a valid explanation, and the remaining 1.5 points for answering the teacher's question on "*What would happen if…?*" The extra task was the same for all groups, and, again, it was projected on the wall screen so as to make it available to all teams at the same time. The first group which solved the problem won (according to the rule "first-come, first-

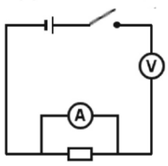
Q4. What is electric current? What is the conventional direction of a current flow?
Q6. What is an electrolyte?
Q8. What is the unit of electric current?
Q9. Give three examples of using Joule heat in everyday life.
Q12. Suggest a formula describing the relation between temperature and the resistance of conductors.

Figure 2. Examples of simple open questions (1 point)

Q2. Ah, this heat
For all metals, an increase of temperature causes an increase in resistance. What is the effect of temperature increase?
A) increase in the oscillation speed of both atoms and ions
B) increase in the density of the electron gas
C) affecting the binding of/from the metal to the valence
D) deterioration of contact between the microcrystals

(Correct answer: A)

Q7. Absent-minded electrician
When building an electric circuit somebody swapped a voltmeter with an ammeter. What happens when you turn on the voltage source?



A) ammeter burns out
B) voltmeter burns out
C) both voltmeter and ammeter burn out
D) meters do not burn out, and the current decreases to almost zero

(Correct answer: D)

Figure 3. Examples of multiple-choice questions (2 points)

win”), and scored extra 10 points, which were also added to the maximum number of points possible to obtain by the team.

It should be clarified here that, at each stage, the correct solution along with a proper explanation to each question were delivered – either by the contestants (with or without the teacher assistance) or by the teacher himself (in those cases where the students were found incapable of delivering a valid solution on their own). Such a practice served as a means of an immediate clarification and refinement of the students’ understanding of the underlying physical concepts.

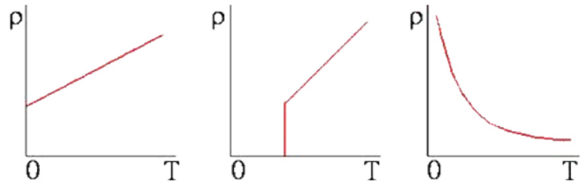
Simple open questions (1 point)

Open questions, each 1-point worth, were meant “to warm up” the students. The tasks were related to some basic knowledge from the curriculum, requiring the students to provide correct simple formulae, units etc., and also examining their basic context knowledge (see **Figure 2**).

Multiple-choice questions (2 points)

Then, two rounds ensued of multiple-choice scientific reasoning questions (each worth 2 points). The students were requested not only to point out the correct answer, e.g. “C”, but also to provide a proper explanation of their choice (see **Figure 3**).

Q1. The figures below show the relationship between the resistivity and the temperature for three different materials. Which graph corresponds to: metal, superconductor, semiconductor? Assign and justify.



Q5. Draw a scheme of an electrochemical cell. What determines the voltage obtained from a given cell?

Figure 4. Examples of open questions (3 points)

Q1. Calculate the voltage across the ammeter in the following circuit:

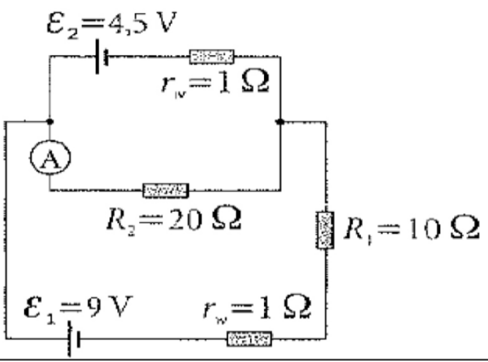


Figure 5. Example of a calculus task

Open questions (3 points)

At the third stage of the tournament, each group faced a choice between a multiple-choice question worth 2 points and an open question for 3 points. The latter was more challenging, requiring broader knowledge and ability of connecting facts (see **Figure 4**).

Calculus tasks (4 points)

After this part, calculus tasks followed. Each group had to choose a different problem (see **Figure 5** for an example) and was given 10 minutes to provide the correct solution. As previously mentioned, this time all groups worked simultaneously. As a result each team could receive maximally 4 points, with a lower score given upon delivery of either an incomplete or partially faulty solution.

Practical tasks (4 points)

The penultimate challenge was a practical task. Each group had to pick randomly an experimental task on one of the following six themes: galvanization, electrochemical cell, electrolysis of water, Ohm's law, building a circuit according to an assigned scheme, and voltage measurement in a designated point. Each team was requested to build a proper circuit, carry out the experiment and give valid description and explanation of the phenomenon at hand. All necessary equipment in each case, with some redundant materials mixed in, was available on a table. Then, students had to decide by themselves which objects were indispensable to accomplish the task.

Table 1. Student self-assessment questionnaire

Question	1-6 scale
Were you involved in the work group?	Communication skills
Did you communicate adequately in the group?	
Did you take part in the discussion on the problem?	
Did you take into account the opinions of others?	
Did you prepare for the test beforehand?	Subject matter contribution
Did you take part in solving problems and tasks?	
Did you have sufficient knowledge to solve the issues?	
Did you contribute to the final result of the group?	

Extra task (10 points)

The final and most demanding task was common to all groups. For all groups at the same time, a slide with the "Monstrous maze of resistors" adopted from (Halliday et al., 2001, p.728, task 8) was displayed on the wall screen. The first group which found the solution and gave the correct answer received 10 points.

Assessment questionnaires

After the tournament each student was asked to fill in individually special self- and peer-assessment questionnaires, aimed at evaluating himself/herself and other fellow players from the same group in various aspects. Each of eight questions required allotting a score between 1 and 6. Four of them were related to the "communication skills", whereas the other four were focused on assessing the "subject matter contribution". In **Table 1** we present the self-assessment questionnaire. The peer-assessment questions were designed analogously.

Evaluation Process

The final note granted to each student consisted of three components:

I. the group percentage result from the tournament questions (the first six stages) – with a weight of 0.5,

II. the questionnaire-based assessment result (in percentage terms) for the "subject matter contribution" – with a weight of 0.3,

III. the questionnaire-based assessment result (in percentage terms) for the "communication skills" – with a weight of 0.2.

The percentage score for each team was obtained through dividing the number of points accumulated by the group by the maximal number of points possible to obtain. The points scored for answering the questions taken over from other groups were not included in the maximal number of possible points.

The questionnaire-based assessment results were included in the final score according to the authors' own approach presented below. For each person, the algorithm proceeded as follows:

1) Firstly, the median score was calculated of "subject matter contribution" and, separately, "communication skills" points in the self-assessment results (S).
2) Secondly, the median score was calculated of "subject matter contribution" and, separately, "communication skills" points attributed to the student by all other members of the group (the peer-assessment, P).
3) Then, the "subject matter contribution" and "communication skills" scores were obtained separately according to the rule:
   - If $|S - P| \leq 1$ (a consistent evaluation): take P as the final score,
   - else (an inconsistent evaluation): take P – 0.5 as the final score.

There are three premises behind the above algorithm. Firstly, we choose to represent the "average" (benchmark) score (in both S and P) by a median rather than a mean, for the previous – as opposed to the latter – is robust to extremities. Secondly, the assumed value of "1" as a tolerable discrepancy between S and P still ensuring a consistent evaluation is our arbitrary choice that appears justifiable in view of the 6-point scale employed in the questionnaires. Note that under such a scale, a tolerable deviation span of 2 points (i.e. *plus*/*minus* 1 point) constitutes ca. 33% of the entire 6-point range. Finally, in the case of an inconsistent evaluation (i.e. $|S - P| \geq 1$) we penalize the P result with an arbitrary value of 0.5. Note that regardless of the precise relation between the S and P assessments, the penalization is always downward, which is intended to reduce a risk of „collusion" among the students, and to stimulate honest and reasonable both self- and peer-assessments (the students had been familiarized with the algorithm prior to the tournament). In a broader perspective, such an approach should work both ways – preventing the participants from an unduly high as well as too low self-esteem. We address this issue to a greater extent in Subsection IV.C. The final score in the tournament, calculated according to the algorithm above, is henceforth denoted as “TNT”.

In addition, the students’ and teachers’ opinions about the tournament as an assessment method were collected just after the implementation. All students were asked to express their reflections in an open-descriptive form, whereas the teachers took part in a semi-structured interview based on three items: (1) general perception of the activity, (2) opinion on feasibility of use in other subjects, and (3) the added value of a tournament comparing to traditional assessment methods. We discuss the results in Subsection III.F.

DATA COLLECTION AND ANALYSIS

In this section we provide details about the pre-tournament test and the post-tournament test, to assess students’ progress (attributable to the tournament) with respect to their former achievements. To this end, a statistical analysis of relevant results is further performed.

Post-test

The post-test was prepared in a traditional, written form, and conducted one week after the tournament, with neither a prior review of the relevant content knowledge during regular classes nor a post-tournament discussion of the tournament problems and results (let us recall, however, that all tasks administered to students during the competition were then elucidated in the process either by the students providing the solution or by the teacher).

The test was unannounced, so the students have not been induced to make any additional efforts to prepare for it. In 60% the test comprised tasks utilized during the tournament, and in the remaining 40% it was based on problems totally new to the students, though similar to the ones given in the tournament. The post-test score is expressed in percentage terms, and, henceforth, denoted as “PT”.

Former Tests

Each student, during the school year and before implementation of the tournament, participated in three tests: on thermodynamics, gravitation and electrostatics. All tests were taken individually. They contained mixed problems, including content knowledge and scientific reasoning tasks, multiple choice, open-response and calculus problems. To measure each student’s achievements prior to the tournament, we used the average of his/her results on the three tests. The quantity obtained (expressed in percentage terms) is further referred to as the “former tests score” and denoted as “FT”.

Basic Statistical Analysis

Figure 6 presents each student’s three individual scores: on the former tests (FT), the tournament (TNT), and the post-test (PT), along with horizontal bars indicating the common (for each group) result gained from the tournament. All scores are provided in percentage terms. Note that the discrepancies between the group common result and the group members’ individual scores stem from the outcomes obtained in the assessment questionnaires. Notice that, incidentally, the final marks assigned to each student within the fifth team were all lower than the common result of ca. 95%. This observation can be explained by the fact that nobody in the group was perceived as a leader, and all the team members were clearly aware of the fact that their final result was the effect of their cooperation (rather than attributable to the knowledge of a single leading person).

It can be easily noticed that the TNT marks were predominantly way above the FT results. What appears far more justifiable, however, is the comparison of the students’ achievements and skills prior to and after the tournament, reflected in the FT and PT results, respectively. In that regard, however, we still observe a systematic (i.e. for almost all tournament participants) increase in score, with the result hinting at a positive impact of the tournament on the students’ improvement.

In what follows, to explore the results in more detail, we conduct statistical analysis. As far as the sample size is concerned, since two students (no. 8 and 21; see **Figure 6**) were absent from the post-test, we exclude them from further considerations, and carry out the necessary calculations based on the sample of n = 28 students. Note that according to such a limited sample size the statistical inferences presented below should be approached with some reserve.

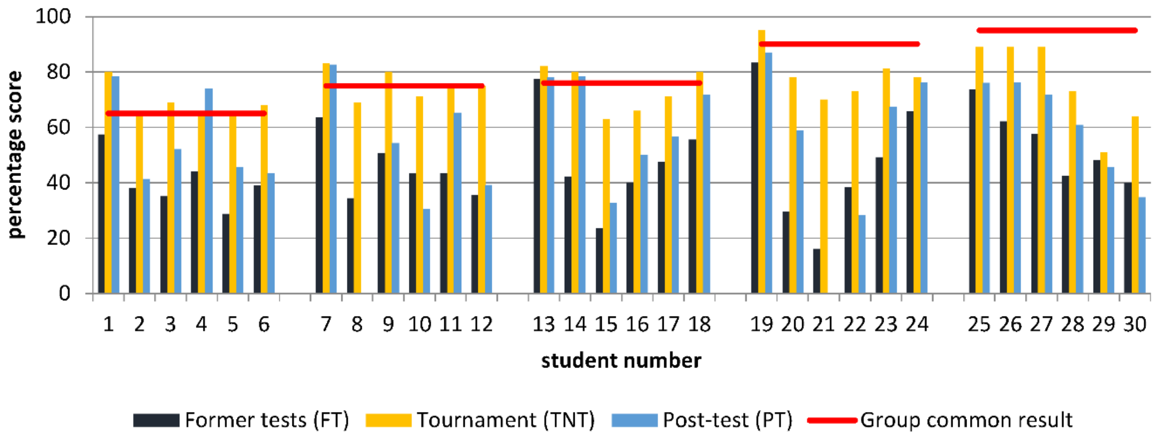


Figure 6. Student scores. For each student three vertical bars represent (starting with the leftmost one): the average score in former tests (FT), the final score obtained in the tournament (TNT), and the mark gained in the post-tournament test (PT). The horizontal lines represent the common result scored by each group in the tournament (based only on the first six stages, disregarding the qualitative component stemming from peer- and self-assessment)

Table 2. Basic statistics of the student results, including: the average score in the former tests (FT), the final score in the tournament (TNT), the result in the post-tournament test (PT), and differences between PT and FT (PT – FT). The last row contains statistics for the modified gain factor (MGF)

Variable	Characteristics						
	Mean	95%-confidence interval for mean	Median	Lower quartile	Upper quartile	Interquartile range	Standard deviation
FT [%]	48.39	(42.64; 54.14)	43.79	38.65	57.48	18.83	14.83
TNT [%]	74.96	(71.17; 78.76)	75.00	67.00	80.50	13.50	9.78
PT [%]	59.16	(52.29; 66.04)	59.78	44.57	76.09	31.52	17.73
PT – FT [pp]	10.77	(6.22; 15.32)	10.11	3.44	18.22	14.78	11.73
MGF	0.22	(0.13; 0.3)	0.23	0.07	0.37	0.30	0.22

In **Table 2** and **Figure 7** we present basic descriptive statistics and empirical distributions (histograms, with normality tested by the Lilliefors and the Shapiro-Wilk tests) for several variables, including the FT, TNT and PT scores, as well as the differences: TNT – FT and PT – FT (the latter measuring the "absolute" gain in student content knowledge). Moreover, we also examine a modified gain factor (MGF), which is our adaptation of the normalized gain (or the g-factor) (Hake, 1998), originally proposed in (Gery, 1972). The MGF measure is meant to relate the "absolute" gain in a student PT score to the points missed on FT, and is therefore calculated according to the formula:

$$MGF = \frac{PT-FT}{100-FT}.$$

From **Table 2** it can be inferred that the students scored, on average, ca. 48.4% upon the former tests, with the standard deviation hovering around 14.8 percentage points (henceforth, pp). Half of the students recorded the FT result below ca. 43.8%, whereas the other half – above that number. (The means and medians differ on account of positive skewness of the empirical distribution; see **Figure 7(a)**). On the other hand, results obtained during the tournament are distinctive on two counts. Firstly, the average TNT score is much higher as compared to FT. Arguably, the difference can be attributed to the team work and cooperation among the students. Note, however, that ultimately these two scores should not be compared *per se*, since calculation of the TNT results include a strong "qualitative" component. Secondly, the TNT scores are more concentrated (as compared with FT) around the mean, with a drop in standard deviation of ca. 5 pp. Moreover, the TNT distribution is far more symmetrical than its FT counterpart (see **Figure 7(b)**), thereby closing the gap between the mean and median (both equal around 75%; see **Table 2**). In general, the TNT scores are more regularly, symmetrically distributed and strongly shifted rightwards as compared to the FT results. (Note, however, that for all but one the analyzed variables, with PT being the exception, despite more or less conspicuous irregularities such as skewness and multimodality, the null hypothesis of normality is not rejected, which, admittedly, is largely due to the low sample size. Still, as implied by the corresponding p-values, the TNT distribution is far closer to normal than actually any of the others; see **Figure 7**).

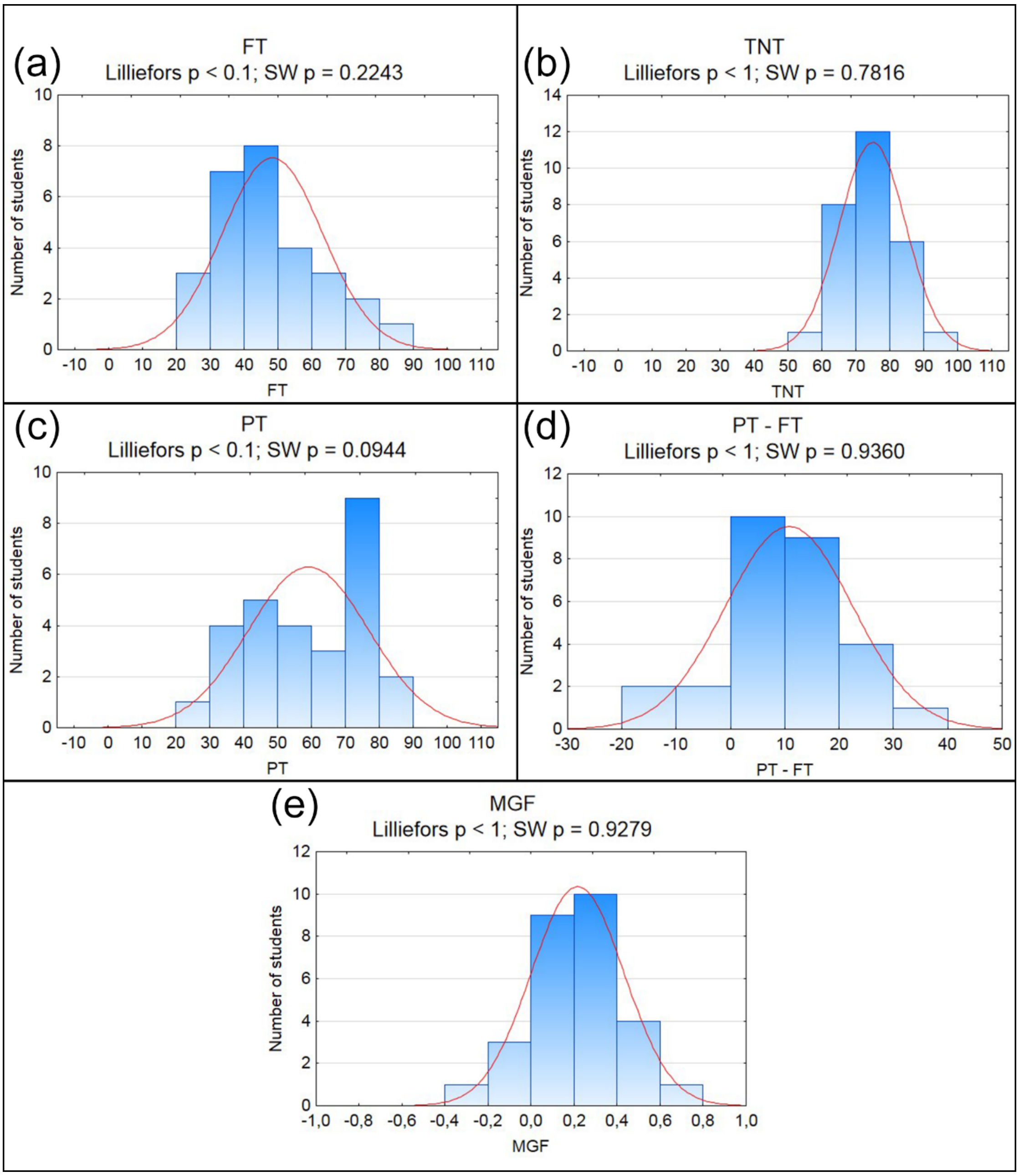


Figure 7. Histograms of the students' results: (a) the average score in the former tests (FT), (b) the final score in the tournament (TNT), (c) the result in the post-tournament test (PT), (d) differences between PT and FT (PT – FT). Panel (e) displays the histogram for the modified gain factor (MGF). In each case the normal density is fitted (solid line), accompanied by p-values for testing normality through the Lilliefors and the Shapiro-Wilk tests (denoted as "Lilliefors p" and "SW p", respectively)

Moving on to the PT results, it appears, interestingly, that these are somehow less regular than FT, on two counts. Firstly, the PT distribution has a higher dispersion, as implied by both standard deviation and, in particular, interquartile range (see **Table 2**). Secondly, as long as the FT distribution features only a single mode (somewhere between 40 and 50%), the PT histogram exhibits a pronounced bimodality. Apparently, the two PT modes correspond with the ones present in the FT and the TNT distributions, with the global PT mode (between 70 and 80%) coinciding with the TNT one, and the second, a local one (between 40 and 50%) – with the FT mode. In statistical terms, one could argue that the PT distribution is a mixture of the FT and TNT distributions. Practically speaking, it could be inferred that the PT scores are formed as a confluence of student prior physical expertise (measured by FT) and the knowledge and skills acquired during the tournament.

Finally, we proceed with the analysis of the results scored in the post-test in relation to student content knowledge and skills prior to the tournament (FT results). The average difference between the PT and FT scores

Table 3. Testing positive means for: difference between the results gained in the post-test and the former tests (PT – FT), and the modified gain factor (MGF). In the second column values of the Student-t statistics are displayed for testing a positive mean. The last column presents corresponding p-values

Characteristics	Test statistics	p-value
PT – FT	4.86	2.20×10^{-5}
MGF	5.30	6.80×10^{-6}

Table 4. Basic statistics of the student results in the control and the experimental groups, including: the average score in the former tests (FT), the final score in the tournament (TNT), the result in the post-tournament test (PT) and differences between PT and FT (PT – FT). The last row contains statistics for the modified gain factor (MGF). Results for the control group are indicated with letter "c" in the superscript

Variable	Characteristics						
	Mean	95%-confidence interval for mean	Median	Lower quartile	Upper quartile	Interquartile range	Standard deviation
FT^c [%]	48.50	(41.43; 55.56)	47.12	38.21	58.63	20.42	15.94
FT [%]	48.39	(42.64; 54.14)	43.79	38.65	57.48	18.83	14.83
PT^c [%]	47.68	(40.79; 54.57)	48.91	35.87	58.7	22.83	15.54
PT [%]	59.16	(52.29; 66.04)	59.78	44.57	76.09	31.52	17.73
$PT^c – FT^c$ [pp]	-0.82	(-5.15; 3.51)	-2.88	-6.36	3.58	9.94	9.77
PT – FT [pp]	10.77	(6.22; 15.32)	10.11	3.44	18.22	14.78	11.73
MGF^c	-0.05	(-0.16; 0.06)	-0.05	-0.17	0.11	0.28	0.25
MGF	0.22	(0.13; 0.3)	0.23	0.07	0.37	0.30	0.22

totals ca. 10.8 pp (see **Table 2**), and it is statistically significant, regardless of the α level (see **Table 3**). (Note, however, that four out of 28 students scored lower in PT than in FT, so negative increments were also reported). Improvement of the student performance is also indicated by the results obtained for the modified gain factor. A test of positive MGF mean indicated that it was significantly positive at any typical α level (see **Table 3**). Note, however, that the MGF histogram exhibits two pronounced and equivalent modes, which may question the use of the mean as a measure of central part of the distribution. Nevertheless, both modes are positive. Furthermore, almost 86% of the probability mass in the histogram is localized to the right of zero, which implies that a learner positive gain was reported for a predominant number of students (i.e. 24 out of 28; see **Figure 7(e)**).

Control Group

In order to validate a statistical approach to examining the influence of the tournament on students' achievements, we formed a control group of 22 students also attending a K-12 class. The control group students took the same former tests and the same post-test as the experimental group students (i.e. the ones analyzed in the previous subsection), but did not participate in the tournament. The former tests results, the post-test scores and the modified gain factor for the control group, which are analyzed below, are calculated in the same manner as in the case of the experimental group, and denoted analogously, i.e. FT^c, PT^c and MGF^c, respectively.

Table 4 summarizes basic statistics of the results gained by the students of each group (i.e. the control and the experimental one), whereas **Figure 8** depicts the histograms of the control group's outcomes along with the normality tests. With regard to the latter, it appears that only MGF^c features some slight departures from the normal distribution, which is attributable to the heavy left tail of the histogram. Statistics presented in **Table 4** reveal a very close similarity of the former test results in both groups not only in terms of means, but other characteristics as well, thereby indicating the validity of the control group at hand for our "experiment". A battery of statistical tests for the equality of: means, variances and the very distributions of FT and FT^c, corroborate this presumption (see **Table 5**).

A visual inspection of the mean values displayed in **Table 4** may indicate a non-negligible positive effect of participating in the tournament on student achievements. As long as there is no significant discrepancy (at $\alpha = 0.1$) between the control group's former and post-tests scores (p-value ≈ 0.7; see **Table 6**), it turns out that the tournament participants scored significantly higher on the post-test than their control group counterparts, both in terms of a simple difference between PT and PT^c (p-value ≈ 0.01; see **Table 6**), and the modified gain factor (p-value ≈ 0.0001; see **Table 6**).

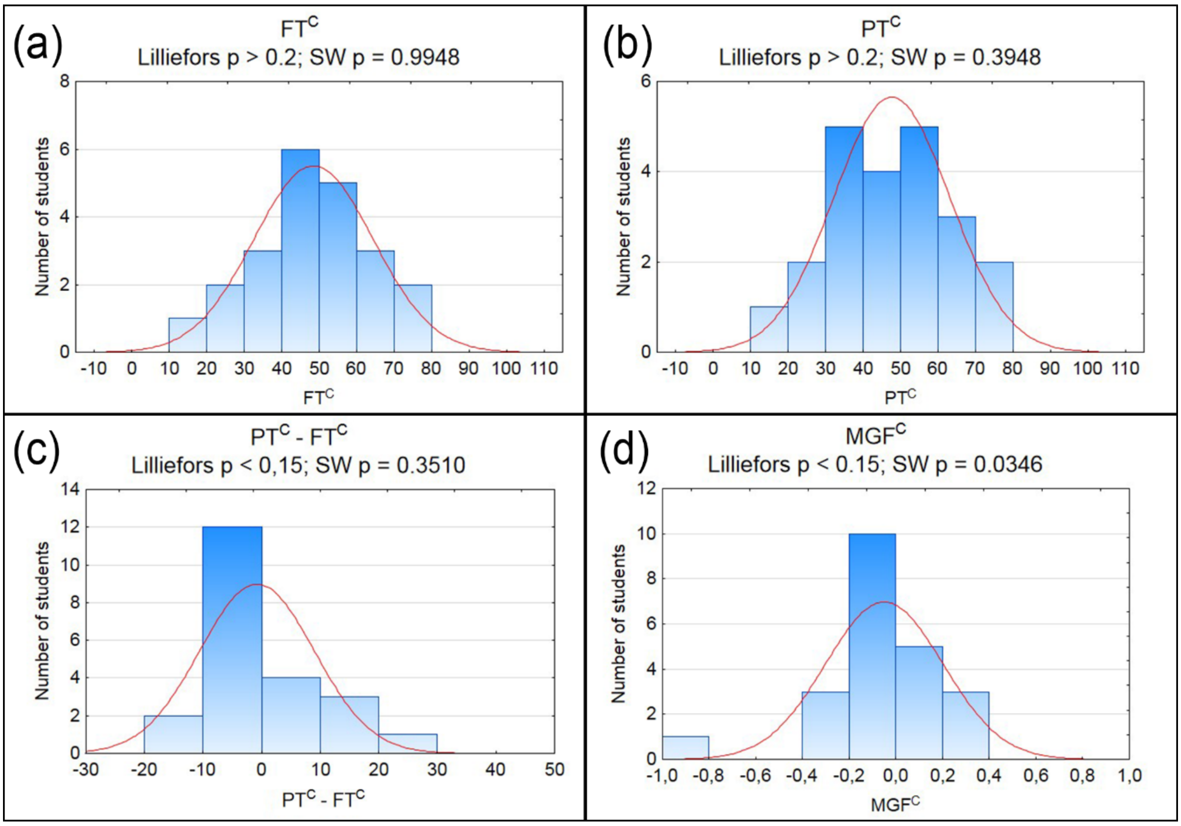


Figure 8. Histograms of the students' results in the control group: (a) the average score in the former tests (FTc), (b) the result in the post test (PT^c), (c) differences between and PT^c and FT^c (PT^c – FT^c). Panel (d) displays the histogram for the modified gain factor (MGFc). In each case the normal density is fitted (solid line), accompanied by p-values for testing normality through the Lilliefors and the Shapiro-Wilk tests (denoted as "Lilliefors p" and "SW p", respectively)

Table 5. Testing the control group for its compatibility with the experimental group, by means of examining the equality of means, variances and distributions of the former tests results obtained in each group (FT^c and FT, respectively). The null (the alternative, correspondingly) hypothesis in each testing procedure states the equality (the inequality) of a given characteristics of the former test results in both groups. For the Mann-Whitney test two statistics are considered: U, the original one, and Z, following approximately the standard normal distribution

Equality of FT and FT^c's …	Testing procedure	Test statistics	p-value
Means	*t*-test	-0.0238	0.9811
Variances	*F*-test	1.1558	0.7141
	Levene	0.1363	0.7136
	Brown-Forsythe	0.1985	0.6579
Distributions	Kolmogorov-Smirnov	0.1266	> 0.10
	Mann-Whitney	300 (*U*) 0.1466 (*Z*)	0.8835 (*U*) 0.8845 (*Z*)

Table 6. Testing the effect of the tournament in terms of: the mean difference between the former and the post-test results in the control group (PT^c – FT^c); inequality between the mean post-test results in the experimental and the control group (PT and PT^c, respectively); inequality between the mean modified gain factors in the experimental and the control group (MGF and MGF^c, respectively)

Characteristics	Test statistics	The alternative hypothesis	p-value
PT^c – FT^c	-0.3930	Mean (PT^c – FT^c) different from zero	0.6982
PT and PT^c	2.4370	Mean PT different from mean PT^c	0.0186
		Mean PT higher than mean PT^c	0.0093
MGF and MGF^c	3.9543	Mean MGF different from mean MGF^c	0.0003
		Mean MGF higher than mean MGF^c	0.0001

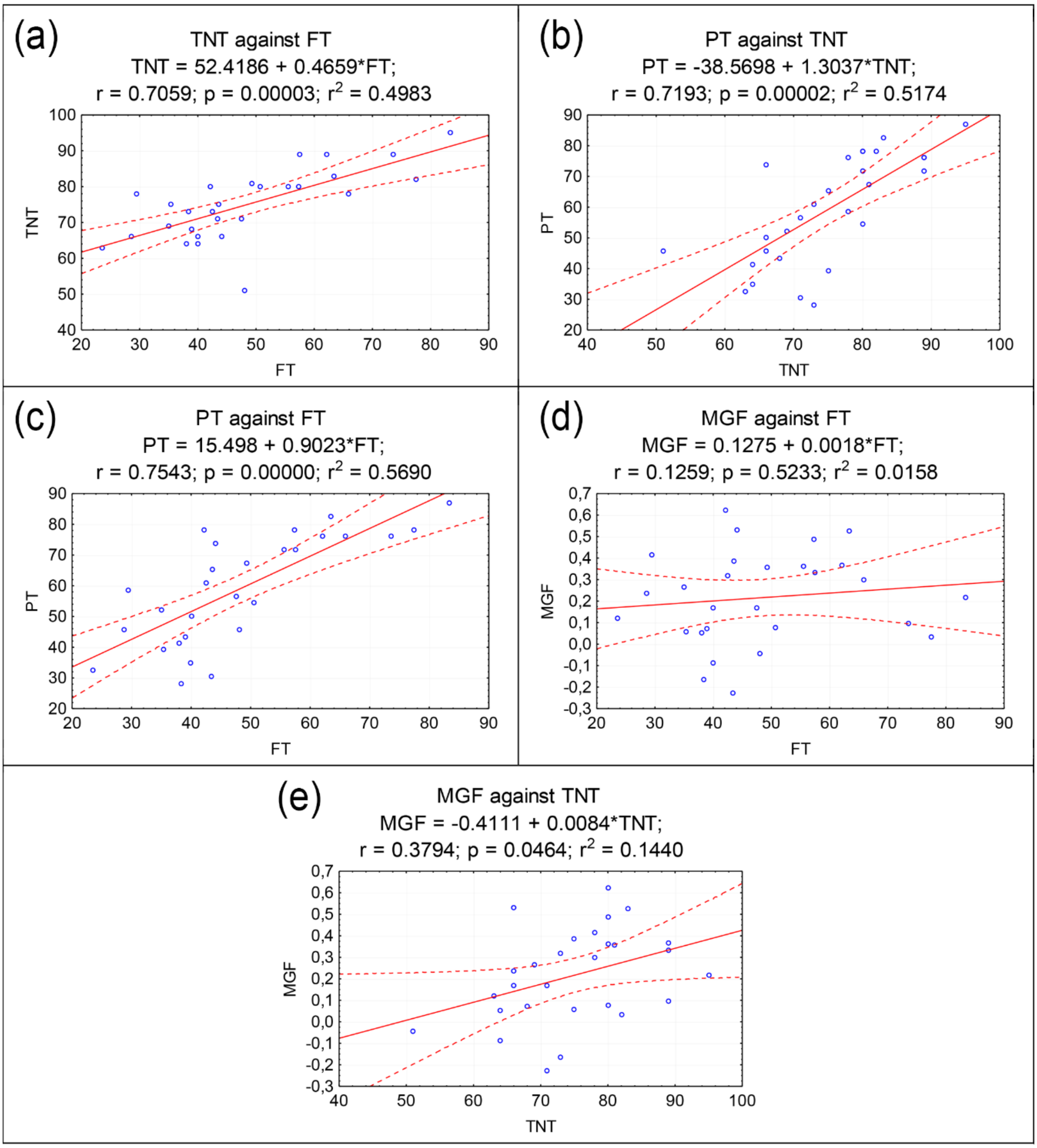


Figure 9. Scatter plots for selected pairs of the students' results. Apart from data points, in each plot a linear regression is fitted and the 95%-confidence band is marked. Below each regression equation we provide: Pearson's linear correlation coefficient (r), p-value for testing a non-zero correlation coefficient, and determination coefficient (r^2)

Correlation and Regression Analysis

Below, the analysis of correlations between selected pairs of the considered variables is performed. **Figure 9** displays relevant scatter plots, along with fitted linear regressions, 95%-confidence bands, linear correlation coefficients (r), p-values for testing their significance, and, at last, the square of correlation coefficients (r^2), which coincide with determination coefficients in the fitted regressions. Based on **Figure 9**, the following general conclusions can be formulated:

1) A positive and statistically significant correlation between FT and TNT implies that students who performed better prior to the tournament, also scored higher in the tournament (see **Figure 9(a)**). It is worth underlining that the value of correlation coefficient ($r = 0.7059$) is negatively affected by a single outlying TNT score equal 51 (obtained by student no. 29), exclusion of which raises the coefficient value to $r = 0.8021$.

2) Similarly, the TNT and the PT results are positively and statistically significantly interrelated (r = 0.7193), indicating that better (likewise, worse) performance in the post-test coincides with a higher (respectively, lower) score in the tournament (see **Figure 9(b)**). Exclusion of the outlier represented by student no. 29 slightly raises the correlation coefficient to r = 0.7466.
3) As expected on the basis of the two above observations, there also occurs a positive and significant relation between the PT and FT scores, indicating that high (likewise, low) notes in the post-test were mostly obtained by those who already performed high (low, respectively) in the former tests (see Figure **9(c)**).
4) Some slight (r = 0.1259), yet statistically insignificant correlation is observed between MGF and FT, hinting at no dependence of a student gain upon his/her previous performance (see **Figure 9(d)**).
5) On the other hand, it appears that student improvement (as measured by MGF) is significantly and positively influenced by the tournament performance, though the correlation coefficient is only ca. 0.38 (see **Figure 9(e)**). The result suggests that the learner gain is generally higher in the case of those who scored higher on TNT.

The inferences formulated in items no. 1-3 point to an intuitive relation according to which the better a student has fared so far, the higher his/her performance in the tournament, and, eventually, in the post-test. Further, result no. 4 implies generally that the student gain, arguably attributable to the tournament, hardly depends on his/her former achievements. In broad terms, it would follow that the tournament provided equal opportunities of improvement to all students. Nevertheless, conclusion no. 5 would still indicate that those who performed better in the tournament (as the effect of their active involvement in cooperative work), actually improved slightly (yet significantly) more than the others.

The results presented above provided us an incentive to build two simple bivariate linear regression models in order to jointly evaluate the impact of the former tests and the tournament results on the post-test score and the modified gain factor. The two models take the following form:

$$Y = \beta_0 + \beta_1 FT + \beta_2 TNT + \varepsilon,$$

with Y representing the dependent variable (i.e. either PT or MGF), and ε denoting normally distributed random errors with zero mean and satisfying typical assumptions of a standard linear regression model. In **Table 7** the following estimation results are presented: determination coefficient (R^2), point estimates, standard errors, p-values against the alternative of a non-zero coefficient (i.e. H_1: *coefficient* ≠ 0), p-values against the alternative of either a positive or negative coefficient (i.e. H_1: *coefficient* > 0, or H_1: *coefficient* < 0), depending on the sign of the point estimate. (Though not reported in the paper, the Lillierfors and Shapiro-Wilk tests do not reject the normality of residuals in any of the regression models considered below, therefore validating testing the regression coefficients by means of a standard Student's t-test).

As regards regressing PT against FT and TNT, it appears that both regressors positively influence the PT score. More specifically, if a student scored higher in FT (likewise, TNT) by 1 pp, then he/she would score also higher in PT by ca. 0.59 pp (0.68, respectively). The results are (positively) significant at α equal 0.01 and 0.05, correspondingly. With respect to the determination coefficient, we note that about 64% of the post-test results is explained by the former tests and the tournament performance.

In view of these results, it is worth noting that also in the control group there is a positive correlation between the post-test and the former tests results. The correlation coefficient between PT^c and FT^c equals 0.81, while its counterpart in the experimental group: 0.75. A slightly higher value in the control group indicates that the PT^c and FT^c scores are more similar to each other than the corresponding results obtained by those students who participated in the tournament, which is also evidenced by the basic statistics presented in **Table 4**. Such an observation may be simply attributed to the lack of intervention in the control group (so that PT^c and FT^c are largely similar), and, at the same time, the (positive) impact of the tournament modifying the students' former achievements so that their PT scores differ more from FT than in the case of the control group. Nevertheless, one should bear in mind that comparing the two correlation coefficients at hand should be made with caution, because in the case of the experimental group the TNT score is yet another variable that is positively correlated with both: FT and PT. Hence, measuring correlation between PT and FT by means of a simple correlation coefficient, which – by construction – fails to take TNT explicitly into account, appears inadequate. Therefore, in order to disentangle the effect of the students' former achievements and the tournament upon their post-test results, we resorted to the multiple regression analysis discussed in the previous paragraphs.

With respect to the regression for MGF, we note that as long as the student gain depends positively on the tournament performance (at α = 0.05), it is not determined by FT (see **Table 7**). As already mentioned above, it would follow that the student improvement, arguably attributable to the tournament, does not depend on their former achievements, and, in broad terms, that the tournament provided equal opportunities of improvement to all students. Note, however, that only about 18% of the modified gain factor can be explained by the former achievements and the tournament performance.

Table 7. Regression results for PT and MGF. Asterisks indicate statistical significance (a non-zero coefficient): ** for α = 0.01, * for α = 0.05. Note that ε equals 0 in the estimated model

	Dependent variable: PT			Dependent variable: MGF		
Regressor	**Constant**	**FT**	**TNT**	**Constant**	**FT**	**TNT**
Parameter	β_0	β_1	β_2	β_0	β_1	β_2
Point estimate	-19.8827	0.5878**	0.6750*	-0.5421	-0.0041	0.0128*
Standard error	17.6894	0.2031	0.3077	0.3234	0.0037	0.0056
p-value against a non-zero coef.	0.2717	0.0078	0.0378	0.1062	0.2778	0.0320
p-value against a positive/negative coef.	0.1358	0.0039	0.0189	0.0531	0.1389	0.0160
Determination coefficient (R^2)		0.6386			0.1841	

Qualitative Analysis of Students' and Teachers' Opinions

Students' opinions

After the post-test, and before getting informed about their final marks, the students were asked to express anonymously their opinions about a tournament as a tool of assessment. The participants were encouraged to formulate their comments in an open, descriptive form, with no predefined questionnaire to follow. Such an approach was meant to induce student openness and spontaneity, with no intent on our part to perform any further (quantitative) analysis of the answers. Some examples of the comments are cited below:

Student A:

I think that this form of a test is good, because we can share our knowledge with others and vice versa, helping each other. We can memorize more and learn new things.

Student D:

This is a better form of consolidation and verification of our knowledge and skills.

Student E:

This is a good idea, because it was performed in the form of a game. A student can show what he or she knows without being stressed.

Student K:

Fabulous! We can integrate, everybody who had any idea but wasn't sure about it had an opportunity to consult/discuss it with other members of the group.

Student O:

I really liked explanation of each answer given afterwards. This way it was possible to understand more.

Student W:

Everybody wanted to receive a good note and knew that there is "collective responsibility" and tried to do his/her best.

Student Z:

I suggest a different way of intercepting questions. Frankly, the bell was getting on my nerves and caused me a headache.

It is worth noting that, except for the last one (regarding the bell ringing), all the students' opinions were positive and enthusiastic.

Teachers' opinions

Just after the tournament two assistant teachers and the teacher conducting the lesson were asked to take part in a semi-structured interview about their perception of the intervention. The common agreement was that the method positively influenced the engagement of the students and raised their interest in physics. All of them also admitted that the method seems to be largely universal and feasible to extend to other topics and school subjects. They also pointed out that, contrary to the traditional assessment methods, oriented mostly on the content knowledge itself, the tournament evaluated also practical and soft skills. One of the teachers said: "*... this is a good opportunity to inure students to the way they might be assessed in their future study and work where not only knowledge and individualism counts, but also cooperation skills.*" The other teacher indicated "*... the method is attractive to young students, sharpens their focus and develops positive attitudes towards science, so much emphasized in the curriculum.*"

DISCUSSION

Social Benefits

Based on the ones delivered above we proceed with a short discussion about students' social benefits arising from participation in the tournament. Note, however, that in our study we did not measure any of the effects mentioned below, including diminishing students' anxiety, improving their social skills and the ability of critical thinking. Although a relevant quantitative analysis of these psychological phenomena appears worthwhile, it is beyond the scope of the current research. Therefore, in this subsection we draw our conclusions solely on the students' and teachers' feedback, relating them to the findings commonly presented in the literature on collaborative testing and gamification.

The tournament was organized in the form of a team game, but with elements of rivalry. In this way it can be perceived as a form of activity in which group work skills, desirable in some academic areas and also by employers, are naturally activated, playing crucial role in accomplishing tasks (Dallmer, 2004; Dicheva et al., 2015; Kapitanoff, 2009; Lusk & Conklin, 2003; Sandahl, 2010; Seaborn & Fels, 2015; Shindler, 2003). Simultaneously, the tournament induced far less test anxiety (as compared with traditional, individually taken written test) by giving students a sense of being supported by the other team members tasks (Banfield & Wilkerson, 2014; Kapitanoff, 2009; Lusk & Conklin, 2003; Sandahl, 2010; Zimbardo et al., 2003). Working together may improve communication skills as well. Students learn to listen to each other, share information, and respond to ideas proposed in discussions, which stimulate knowledge assimilation (Hanus & Fox, 2015; Jolliffe, 2007). What is worth noting is that vocabulary and concepts used in group and class discussions may provide retrieval cues that help students recall relevant information. Moreover, the requirement of providing not only the answer to a question, but also the explanation for it, necessitated that the students should be able to understand and present their lines of reasoning and reconsider them, if needed. Therefore, a tournament may also yield an improvement in students' ability of critical thinking as well as facilitate their intrinsic motivation tasks (Banfield & Wilkerson, 2014; Kapitanoff, 2009; Lusk & Conklin, 2003; Shindler, 2003). Finally, an active involvement in the self- and peer-assessment process may improve student confidence and adequate self-esteem (Hendrix, 1996), thereby enhancing retention of knowledge (Sawtelle et al., 2012). Taking all the above into consideration, cooperative testing of knowledge may become a significant part of the learning process.

Academic Benefits

The main purpose of this research was to examine the impact of taking an exam in the form of a tournament on student achievements. Firstly, a statistically significant increase is observed in students' achievements in the tournament as compared to their former tests results (the average difference amounted to ca. 26 pp, in favor of the tournament scores, being positively significant at any typical α level). Secondly, we also find evidence for improvement of student content knowledge and problem solving skills, as indicated by the results of the post-test taken by the students a week after the tournament (the average difference between marks in the post-test and former tests scored ca. 11 pp; the mean of modified gain factor totaled 0.22; both results are positively significant at any typical α level). Our findings remain in accordance with much research on positive impact of collaborative testing. Studies presented in (Bloom, 2009; Haberyan & Barnett, 2010; Kapitanoff, 2009; Lusk & Conklin, 2003), focused on the effects of taking exams in a collaborative way for numerous groups with various numbers of students and of different subject/specialization, indicated higher students' achievements as compared with traditional ways of individual testing of knowledge. Moreover, in (Bloom, 2009) it was found that collaborative exam scores were also higher than the ones earned in individually taken exams during which students were allowed to use course textbooks and their notes. Further, some researchers show that students' performance also improved in a longer perspective, as indicated by post-tests taken some time after the collaborative exam (Cortright et al., 2003; Jensen et al., 2002; Simpkin, 2005). Notice that in our research we established a positive and statistically significant impact of participation in the tournament on students' achievements in the post-test.

Finally, in the context of the tournament organization, let us emphasize that the event was not preceded by any traditional, individually taken test on the subject matter (i.e. electricity), though, conceivably, it would be worth contrasting the post-test results with the ones obtained in a typical pre-test on the same content. In our approach we followed conclusions formulated in (Dahlström, 2012), who suggested that the learning gain due to taking a collaborative final exam might be higher if the students had no previous individual encounter with relevant tasks. In the cited paper it was found that in the post-test the students scored higher on new problems (i.e. the ones that had not been used in the pre-test) than on the questions they had already been given previously. A possible logic behind this observation is that the lines of reasoning followed by a student during an individually taken exam tend to persist afterwards, therefore hindering acquiring new ways of thinking and solving the problem, even after participating in a collaborative activity. It would follow then that, as claimed in (Dahlström, 2012), "it might be preferable to collaborate without first deciding on questions individually." Taking this as well as our findings into account, we infer that a class tournament is a well-justifiable and effective learning activity, in which the three approaches to assessment (i.e. *of*, *for* and *as* learning (Earl, 2004)) merge together.

Comments on TNT Grading

In our study, a tournament is proposed as a form of summative assessment with formative elements, since it served us to evaluate students' content knowledge and practical skills in a particular physics area (though, obviously, a contest-based evaluation procedures are readily adaptable to other areas of education). As implied in the previous section, the tournament assessment yielded significantly higher final scores in comparison with the results obtained in former, classical and individually written tests. On the one hand, to some, such an outcome may cast doubt on a tournament as a valid means of student evaluation, for no longer only the content knowledge is subjected to scrutiny, but also other aspects of student performance, particularly group work skills. However, as mentioned in the previous section, in view of a voluminous literature on collaborative work and group work assessment strategies, the apparent discrepancy between the TNT and FT results is perfectly justifiable. Nevertheless, let us also note, however, that addressing the issue of what a student grade should reflect actually requires settling on what exactly should be subjected to assessment. This, in turn, is often a matter of national educational regulations and curricula, differing across countries. It should be highlighted that a tournament as a tool of student evaluation leaves the teacher much space for modifications in terms of the formulation and the difficulty level of tasks, the way of calculating composite and final scores, etc.

Organizational Considerations

Another issue that may arise among teachers searching for a practically valid and feasible alternative to classical forms of student assessment is the question of the organizational effort behind it. As regards a tournament itself, we admit that it may (though need not) be a more demanding and time-consuming endeavor than preparing and conducting an individually written test. Even setting the issue of the time cost aside, the idea of a tournament may still be approached by some with reluctance due to the need of an active and ceaseless involvement of a teacher during the event itself. Nevertheless, there are manifest benefits of this additional effort, among which the most obvious one is the online feedback between students and teacher. This allows the teacher to elicit constantly, during the process, and to monitor not only the students' content knowledge, but also their ways of understanding (Stang & Roll, 2014). Once the teacher spots some deficiencies in either the content or the reasoning, he/she is enabled to straighten them out online. Obviously, a typical written test does not allow for such a possibility (Franklin & Hermsen, 2014). Therefore, during a tournament, by listening attentively to students' responses, understanding students' lines of reasoning, and addressing them relevantly, the teacher has a unique opportunity to assess the participants in a most formative manner.

Self- vs. Peer-assessment

Other doubts may arise with respect to the self- and peer-assessment evaluation procedure implemented in our study. There are many papers in the literature on assessing student engagement in a broadly defined group work, with many different strategies and ideas (e.g. Fernandezbreis et al., 2009; Moccozet et al., 2013). It may appear to some that the algorithm implemented in our research, primarily designed by us to encourage truthfulness in the contestants, tends to affect only those students who appraise themselves too high as compared to the evaluation by his/her teammates. However, it should be stressed that the formula hinges upon the absolute value of the difference between the self- and peer-evaluation scores, thereby equally penalizing unduly over- as well as underestimated self-assessments. Hence, we regard the scheme proposed in this paper – obviously remaining open to further enhancements and suitable adaptations – as developing a student's sense of need to provide honest evaluations, both with respect to themselves and the other members of his/her team. To this aim we deem it of utmost importance for the teacher to provide the participants – before the tournament – with an explanation of how possible discrepancies between the self- and peer-assessments are going to affect their final scores, making then clear indications that, in view of the implemented algorithm, honesty is the best policy – also for those students who tend to underestimate their achievements, skills or abilities. Then, while contrasting the self- and peer-assessments results after the tournament, the teacher is able to pinpoint those participants whose scores are overly divergent. In such cases the teacher should be prompted to take proper measures, such as discussing individually the noticed discrepancy with each of the selected students in order to trace its origins. Depending on the teacher's judgment, for some students it may emerge advisable to further seek a professional psychological advice so as to eventually develop in them a proper overall subjective emotional evaluation of his/her own worth. In addition, we also believe that performing tournaments cyclically would enable the teacher to track each student's dynamics in

Table 8. Analysis of the differences between the self- and peer-assessment scores by gender. "S" and "P" stand for the self- and peer-assessment scores, respectively. The table reports on the number of students for whom a given inequality between S and P occurred. Note that S – P > 0 (S – P < 0, respectively) indicates that a student overestimated (underestimated) his/her contribution in his/her teammates' opinion. The cases of S – P > 1 and S – P < –1 are regarded as an inconsistent evaluation, resulting in a penalization of the final score (see Sec. II B, the third point of the algorithm of the questionnaire-based part of assessment)

	Subject matter contribution		Communication skills	
No. of cases	**Boys**	**Girls**	**Boys**	**Girls**
S – P < 0	16	9	10	7
S – P > 0	2	2	9	4
S – P = 0	1	0	0	0
S – P < –1	3	2	1	2
S – P > 1	0	1	1	1

terms of their self-esteem (Lindsey & Nagel, 2015). Incidentally, let us note that all the tournament participants, although not used to self- and peer-assessment, embraced unequivocally the practice of mutual evaluation.

It may also be interesting to analyze the number of participants for whom the discrepancy between the self- and the peer-assessments scores (denoted as "S" and "P", respectively) was too large, resulting in a penalty for an inconsistent evaluation (which is the case when $|S - P| > 1$; see Subsection II.B). In **Table 8** we report relevant quantities by gender (there were 11 girls and 19 boys participating in the tournament; note that the total number of participants is 30, including also the two students who did not take the post-test). Overall, the students of both sexes tend to evaluate consistently both their subject matter contribution and communication skills. However, the gaps between the numbers of males and females who undervalue their contribution (i.e. $S - P < 0$) and the ones that overrate it ($S - P > 0$) are far more evident for the subject matter involvement than for the communication skills. Moreover, an overwhelming majority of students of both sexes is inclined to underestimate (rather than overestimate) their subject matter contribution. Exceptions of students overly underrating their contribution (i.e. $S - P < -1$) include three boys and two girls with respect to the subject matter, and only one boy and two girls in terms of the communication skills. Interestingly, only in one of these cases the tournament participant (a boy) underestimated himself on both counts. All the other students under consideration scored $S - P < -1$ only in one of the analyzed aspect. On the other hand, the cases of overvaluation of one's involvement were relatively rarer. In terms of the subject matter no boys and only one girl overrated their contribution, whereas with regard to communication two students (each of a different sex) evaluated their performance too enthusiastically.

Gender Differences

The final issue we would like to raise here, and the one that quite naturally spins off from the previous subsection, is an analysis of the major results (i.e. FT, TNT and PT) by gender, so as to identify and characterize possible sex-specific effects and dependencies, collectively termed as a gender gap (Kost et al., 2009; Madsen et al., 2013; Pollock et al., 2007). Basic descriptive statistics, presented in **Table 9** for the male and female students separately, imply no relevant gender discrepancies in terms of the mean and median scores, with the boys performing slightly better than the girls. However, we refrain from testing statistical significance of these differences, for under such low sample sizes no valid conclusions could be obtained. Further, we also notice that despite the similarities between the groups' means, the boys' scores are more diffused for typically-written tests (FT and PT). Although the TNT results appear more evenly dispersed in both groups (as indicated by the standard deviations and the ranges between maxima and minima), the "innermost" 50% of the scores (i.e. those between the lower and the upper quartile) obtained by the females are more scattered than for the male students.

Table 9. Basic statistics of the students' results by gender, including: the average score in the former tests (FT), the final score in the tournament (TNT), the result in the post-tournament test (PT), differences between TNT and FT (TNT – FT), as well as PT and FT (PT – FT). The last row contains statistics for the modified gain factor (MGF)

	Sex	Mean	Median	Minimum	Maximum	Lower quartile	Upper quartile	Interquartile range	Standard deviation
FT	Boys	49.08	43.79	23.57	83.33	38.33	57.58	19.25	16.99
	Girls	47.15	45.12	35.08	65.87	38.97	50.63	11.67	10.57
TNT	Boys	76.72	75.00	63.00	95.00	71.00	82.00	11.00	9.30
	Girls	71.80	73.50	51.00	83.00	64.00	80.00	16.00	10.33
PT	Boys	60.01	63.04	28.26	87.00	45.65	76.00	30.35	18.45
	Girls	57.61	53.26	34.78	82.61	43.48	76.09	32.61	17.19
TNT - FT	Boys	27.64	27.28	4.51	48.51	22.62	34.67	12.04	10.65
	Girls	24.65	27.49	2.97	37.78	19.55	31.71	12.17	10.56
PT - FT	Boys	10.93	11.99	-12.90	29.83	3.67	18.33	14.66	11.75
	Girls	10.46	7.36	-5.14	36.04	3.25	18.11	14.86	12.35
MGF	Boys	0.22	0.23	-0.23	0.53	0.09	0.37	0.28	0.21
	Girls	0.21	0.17	-0.09	0.62	0.05	0.36	0.30	0.24

CONCLUDING REMARKS

In this paper we present both, quantitative and qualitative results of a tournament as a method of assessing student performance in physics classes on electricity. Based on students' results and students' and teachers' opinions we can come up with the following conclusions:

I. As compared with the control group results, the tournament proved to significantly enhance the experimental group students performance.

II. For most learners in the experimental group their results got in an individually written post-test (taken a week after the intervention) were higher than their average performance beforehand.

III. Scores obtained during the tournament were higher than in traditionally performed tests.

IV. The alternative method of testing analyzed in our paper appears to provide equal opportunities of improvement both for low- and high-performers through the tournament approach.

V. Both students and teachers appreciated the method very much because it enabled students to help each other in solving problems in a more cooperative, less stressful way and develop soft skills.

In general, there are several advantages of such a form of examination that outweigh organizational difficulties mentioned earlier in this study. These include: supporting weaker students by collaboration with others, setting a framework of cooperative-learning among students, development of group-work skills, stress-free testing, and, in addition to these, integration of the class.

Finally, let us note that our approach can be easily transferred and adapted to testing achievements in fields other that physics, particularly the natural sciences. Nevertheless, the subject which we had chosen for testing out method was physics, which is largely due to its obvious feature of combining algebraic calculations with both the description and explanation of real-world phenomena. Implementations of the tournament as an assessment method in other areas could be the subject of further studies.

ACKNOWLEDGEMENTS

Łukasz Kwiatkowski would like to acknowledge partial financial support by the Foundation for Polish Science (within the START 2014 program), and partial support from research funds granted to Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

Daniel Dziob would like to acknowledge the Polish Ministry of Science and Higher Education financial support under program 7150/E-338/M/2017.

REFERENCES

Apostol, S., Zaharescu L., & Aleze, I. (2013). Gamification of Learning and Educational Games. *eLearning & Software for Education, 2*, 67.

Banfield, J., & Wilkerson, B. (2014). Increasing Student Intrinsic Motivation and Self-Efficacy through Gamification Pedagogy. *Contemporary Issues in Education Research (CIER) CIER, 7*(4), 291. doi:10.19030/cier.v7i4.8843

Biggs, J. (1998). Assessment and Classroom Learning: A role for summative assessment? *Assessment in Education: Principles, Policy & Practice, 5*(1), 103-110. doi:10.1080/0969595980050106

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. doi:10.1080/0969595980050102

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice, 17*(2), 215-232. doi:10.1080/09695941003696016

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the Black Box: Assessment for Learning in the Classroom. *Phi Delta Kappan, 86*(1), 8-21. doi:10.1177/003172170408600105

Black, P., Harrison, P., Hodgen, C., Marshall, J., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice, 18*(4), 451-469. doi:10.1080/0969594X.2011.557020

Bloom, B. S., Hastings, J. T., Madaus, G. F., & Baldwin, T. S. (1971). *Handbook on the formative and summative evaluation of student learning*. New York: McGraw-Hill.

Bloom, D. (2009). Collaborative Test Taking. *College Teaching, 57*(4), 216-220.

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., & Dicarlo, S. E. (2003). Student Retention Of Course Content Is Improved By Collaborative-Group Testing. *Advances in Physiology Education, 27*(3), 102-108. doi:10.1152/advan.00041.2002

Dahlström, O. (2012). Learning during a collaborative final exam. *Educational Research and Evaluation, 18*(4), 321-332. doi:10.1080/13803611.2012.674689

Dallmer, D. (2004). Collaborative test taking with adult learners. *Adult Learning, 15*, 4-7. doi:10.1177/104515950401500301

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification: Using game-design elements in non-gaming contexts. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* (p. 2425). Vancouver, US. doi:10.1145/1979742.1979575

Dicheva, D., Dichev C., Agre, G., & Angelova, G. (2015). Gamification in Education: A Systematic Mapping Study. *Educational Technology & Society, 18*(3), 75–88.

Ding, L. (2014). Seeking missing pieces in science concept assessments: Reevaluating the Brief Electricity and Magnetism Assessment through Rasch analysis. *Physical Review Special Topics - Physics Education Research, 10*(1). doi:10.1103/PhysRevSTPER.10.010105

Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics - Physics Education Research, 2*(1). doi:10.1103/PhysRevSTPER.2.010105

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*(3), 331-350. doi:10.1080/03075079912331379935

Docktor, J. L., Strand, N. E., Mestre, J. P., & Ross, B. H. (2015). Conceptual problem solving in high school physics. *Physical Review Special Topics - Physics Education Research, 11*(2). doi:10.1103/PhysRevSTPER.11.020106

Doran, R. L., Boorman, J., Chan, F., & Hejaily, N. (1993). Alternative assessment of high school laboratory skills. *Journal of Research in Science Teaching, 30*(9), 1121-1131.

Duane, B. T., & Satre, M. E. (2014). Utilizing constructivism learning theory in collaborative testing as a creative strategy to promote essential nursing skills. *Nurse Education Today, 34*(1), 31-34. doi:10.1016/j.nedt.2013.03.005

Earl, L. M. (2004). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.

Fakcharoenphol, W., & Stelzer, T. (2014). Physics exam preparation: A comparison of three methods. *Physical Review Special Topics - Physics Education Research, 10*(1). doi:10.1103/PhysRevSTPER.10.010108

Fernandezbreis, J., Castellanosnieves, D., & Valenciagarcia, R. (2009). Measuring individual learning performance in group work from a knowledge integration perspective. *Information Sciences, 179*(4), 339-354. doi:10.1016/j.ins.2008.10.014

Franklin, S. V., & Hermsen, L. M. (2014). Real-time capture of student reasoning while writing. *Physical Review Special Topics - Physics Education Research, 10*(2). doi:10.1103/PhysRevSTPER.10.020121

Garriso, C., & Ehringhaus, M. (2007). *Formative and summative assessments in the classroom.* Retrieved on March 25, 2016 from http://www.amle.org/Publications/WebExclusive/Assessment/tabid/1120/Default.aspx

Gery, F. W. (1972). *Does mathematics matter? In Research Papers in Economic Education*, A. Welsh (Ed.), 142–157. New York: Joint Council on Economic Education.

Gilley, B., & Clarkston, B. (2014). Research and Teaching: Collaborative Testing: Evidence of Learning in a Controlled In-Class Study of Undergraduate Students. *Journal of College Science Teaching, 043*(03), 83.

Guest, K. E., & Murphy, D. S. (2000). In support of memory retention: A cooperative oral final exam. *Education, 121*, 350-354.

Haberyan, A. & Barnett, J. (2010). Collaborative testing and achievement: are two heads really better than one? *Journal of Instructional Psychology, 37*(1), 32-41.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74. doi:10.1119/1.18809

Halliday, D., Resnick, R. & Walker, J. (2001). *Fundamentals of Physics*, *3*. New Jersey: John Wiley & Sons, Inc.

Hanus, M., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80*, 152-161.

Harlen, W., & James, M. (1997). Assessment and Learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice,* 4(3), 365-379. doi:10.1080/0969594970040304

Hendrix, J. C. (1996). Cooperative Learning: Building a Democratic Community. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 69*(6), 333-336.

Hestenes, D., Wells, M., & Swackhammer, G. (1992). Force Concept Inventory. *The Physics Teacher, 30*, 141-158. doi:10.1119/1.2343497

Hickey, D. T., Taasoobshirazi, G., & Cross, D. (2012). Assessment as learning: Enhancing discourse, understanding, and achievement in innovative science curricula. *Journal of Research in Science Teaching, 49*(10), 1240-1270.

Hitt, G. W., Isakovic, A. F., Fawwaz, O., Bawa'Aneh, M. S., El-Kork, N., Makkiyil, S., & Qattan, I. A. (2014). Secondary implementation of interactive engagement teaching techniques: Choices and challenges in a Gulf Arab context. *Physical Review Special Topics - Physics Education, 10*(2). doi:10.1103/PhysRevSTPER.10.020123

Ifenthaler, D., Eseryel, D., & Ge, X. (2012). *Assessment in game-based learning: Foundations, innovations, and perspectives*. New York: Springer.

Ives, J. (2014). Measuring the Learning from Two-Stage Collaborative Group Exams. In *2014 PERC Proceedings* (p. 123). Minneapolis, US. doi:10.1119/perc.2014.pr.027

Jensen, M., Moore, R., & Hatch, J. (2002). Cooperative Learning: Part I: Cooperative Quizzes. *The American Biology Teacher, 64*(1), 29-34. doi:10.1662/0002-7685(2002)064[0029:CLPICQ]2.0.CO;2

Jolliffe, W. (2007). *Cooperative learning in the classroom: Putting it into practice*. London: Paul Chapman.

Kagan, S. (1990). The structural approach to cooperative learning. *Educational Leadership, 47*, 12-15.

Kapitanoff, S. H. (2009). Collaborative testing: Cognitive and interpersonal processes related to enhanced test performance. *Active Learning in Higher Education, 10*(1). http://psycnet.apa.org/doi/10.1177/1469787408100195

Kost, L. E., Pollock, S. J., & Finkelstein, N. D. (2009). Characterizing the gender gap in introductory physics. *Physical Review Special Topics - Physics Education Research, 5*(1). doi:10.1103/PhysRevSTPER.5.010101

Kruglak, H. (1965). Experimental Study of Multiple-Choice and Essay Tests. I. Am. J. Phys. *American Journal of Physics, 33*(12). doi:10.1119/1.1971143

Lindsey, B. A., & Nagel, M. L. (2015). Do students know what they know? Exploring the accuracy of students' self-assessments. *Physical Review Special Topics - Physics Education Research, 11*(2). doi:10.1103/PhysRevSTPER.11.020103

Looney, J. (2011). Integrating Formative and Summative Assessment. *OECD Education Working Papers, 58*(5).

Lusk, M. & Conklin, L. (2003) Collaborative testing to promote learning. *Journal of Nursing Education*, *42*(3), 121-124.

Madsen, A., McKagan, S. B., & Sayre, E. C. (2013). Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap? *Physical Review Special Topics - Physics Education Research Phys. Rev. ST Phys. Educ. Res., 9*(2). doi:10.1103/PhysRevSTPER.9.020121

Maries, A., & Singh, C. (2013). Exploring one aspect of pedagogical content knowledge of teaching assistants using the test of understanding graphs in kinematics. *Physical Review Special Topics - Physics Education Research, 9*(2). doi:10.1103/PhysRevSTPER.9.020120 · Source: arXiv

McTighe, J., & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership, 63*(10), 10+17.

Moccozet, L., Tardy, C., Opperecht, W., & Leonard, M. (2013). Gamification-based assessment of group work. In *Proceedings of the Interactive Collaborative Learning (ICL) Conference* (p. 171). Kazan, RU.

Moncada, S., & Moncada, T. (2014). Gamification of Learning in Accounting Education. *Journal of Higher Education Theory and Practice, 14*, 9.

Pawl, A., Teodorescu, R. E., & Peterson, J. D. (2013). Assessing class-wide consistency and randomness in responses to true or false questions administered online. *Physical Review Special Topics - Physics Education Research, 9*(2). doi:10.1103/PhysRevSTPER.9.020102

Pollock, S. J., Finkelstein, N. D., & Kost, L. E. (2007). Reducing the gender gap in the physics classroom: How sufficient is interactive engagement? *Physical Review Special Topics - Physics Education Research, 3*(1). doi:10.1103/PhysRevSTPER.3.010107

Rao, S. P., Collins, H. L., & DiCarlo, S. E. (2002). Collaborative testing enhances student learning. *Advances in Physiology Education, 26*(1). doi:10.1152/advan.00032.2001

Rebello, N. S. (2011). Comparing students' performance on research-based conceptual assessments and traditional classroom assessments. In *2011 PERC Proceedings* (pp. 66-68). Omaha, US.

Sadler, T. D., Romine, W. L., Stuart, P. E., & Merle-Johnson, D. (2013). Game-Based Curricula in Biology Classes: Differential Effects among Varying Academic Levels. *Journal of Research in Science Teaching, 50*(4), 479-499. doi:10.1002/tea.21085

Sandahl, S. S. (2010). Collaborative testing as a learning strategy in nursing education. *Nursing Education Perspectives, 31*(3), 142-147.

Sawtelle, V., Brewe, E., & Kramer, L. H. (2012). Exploring the relationship between self-efficacy and retention in introductory physics. *Journal of Research in Science Teaching, 49*(9), 1096-1121. doi:10.1002/tea.21050

Schuwirth, L. W., & Cees P M Van Der Vleuten. (2004). Different written assessment methods: What can be said about their strengths and weaknesses? *Med Educ Medical Education, 38*(9), 974-979. doi:10.1111/j.1365-2929.2004.01916.x

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*, 39-83. Chicago, IL: Rand McNally.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies, 74*, 14-31.

Shindler, J. V. (2003). "Greater Than the Sum of the Parts?" Examining the Soundness of Collaborative Exams in Teacher Education Courses. *Innovative Higher Education, 28*(4), 273-283.

Simpkin, M. G. (2005). An Experimental Study of the Effectiveness of Collaborative Testing in an Entry-Level Computer Programming Class. *Journal of Information Systems, 16*, 273-280.

Slavin, R. E. (2000). *Cooperative Learning: Theory, research, and practice*. New Jersey: Prentice-Hall.

Slepkov, A. D., & Shiell, R. C. (2014). Comparison of integrated testlet and constructed-response question formats. *Physical Review Special Topics - Physics Education Research, 10*(2). doi:10.1103/PhysRevSTPER.10.020120

Stang, J. B. & Roll, I. (2014). Interactions between teaching assistants and students boost engagement in physics labs. *Physical Review Special Topics - Physics Education Research, 10*(2). doi:10.1103/PhysRevSTPER.10.020117

Sung, H. & Hwang, G. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education, 63*, 43-51.

Talanquer, V., Bolger, M., & Tomanek, D. (2015). Exploring prospective teachers' assessment practices: Noticing and interpreting student understanding in the assessment of written work. *Journal of Research in Science Teaching, 52*(5), 585-609. doi:10.1002/tea.21209

Taras, M. (2005). Assessment – Summative and Formative – Some Theoretical Reflections. *British Journal of Educational Studies, 53*(4), 466-478.

Taras, M. (2009). Summative assessment: The missing link for formative assessment. *Journal of Further and Higher Education, 33*(1), 57-69. doi:10.1080/03098770802638671

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Buckingham: Open University Press.

Vercellati, S., Michelini, M., Santi, L., Sokolowska, D., & Brzezinka, G. (2013). Investigating MST curriculum experienced by eleven-year-old Polish and Italian pupils. In *E-Book Proceedings of the ESERA 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning* (Vol. 10, pp. 180-189). Cyprus.

Wilcox, B. R., & Pollock, S. J. (2014). Coupled multiple-response versus free-response conceptual assessment: An example from upper-division physics. *Physical Review Special Topics - Physics Education Research, 10*(2). doi:10.1103/PhysRevSTPER.10.020124

Wilcox, B. R., & Pollock, S. J. (2015). Upper-division student difficulties with the Dirac delta function. *Physical Review Special Topics - Physics Education Research, 11*(1). doi:10.1103/PhysRevSTPER.11.010108

Wilcox, B. R., Caballero, M. D., Baily, C., Sadaghiani, H., Chasteen, S. V., Ryan, Q. X., & Pollock, S. J. (2015). Development and uses of upper-division conceptual assessments. *Physical Review Special Topics - Physics Education Research, 11*(2). doi:10.1103/PhysRevSTPER.11.020115

Wiliam, D., & Black, P. (1996). Meanings and Consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537-548.

Wininger, S. (2005). Using Your Tests to Teach: Formative Summative Assessment. *Teaching of Psychology, 32*(3), 164-166.

Wooten, M. M., Cool, A. M., Prather, E. E., & Tanner, K. D. (2014). Comparison of performance on multiple-choice questions and open-ended questions in an introductory astronomy laboratory. *Physical Review Special Topics - Physics Education Research, 10*(2). doi:10.1103/PhysRevSTPER.10.020103

Yu, H., & Li, H. (2014). Group-based Formative Assessment: A Successful Way to Make Summative Assessment Effective. *TPLS Theory and Practice in Language Studies, 4*(4), 839-844. doi:10.4304/tpls.4.4.839-844

Zimbardo, P. G., Butler, L. D., & Wolfe, V. A. (2003). Cooperative College Examinations: More Gain, Less Pain When Students Share Information and Grades. *The Journal of Experimental Education, 71*(2), 101-125. doi:10.1080/00220970309602059

Zwolak, J. P., & Manogue, C. A. (2015). Assessing student reasoning in upper-division electricity and magnetism at Oregon State University. *Physical Review Special Topics - Physics Education Research, 11*(2). doi:10.1103/PhysRevSTPER.11.020125

http://www.ejmste.com

Res Sci Educ (2020) 50:845–862
https://doi.org/10.1007/s11165-018-9714-y

Board Game in Physics Classes—a Proposal for a New Method of Student Assessment

Daniel Dziob [1,2]

Published online: 27 March 2018

Abstract The aim of this study was to examine the impact of assessing students' achievements in a physics course in the form of a group board game. Research was conducted in two groups of 131 high school students in Poland. In each school, the research sample was divided into experimental and control groups. Each group was taught by the same teacher and participated in the same courses and tests before the game. Just after finishing the course on waves and vibrations (school 1) and optics (school 2), experimental groups took part in a group board game to assess their knowledge. One week after the game, the experimental and control groups (not involved in the game) took part in the post-tests. Students from the experimental groups performed better in the game than in the tests given before the game. As well their results in the post-tests were significantly higher statistically than students from the control groups. Simultaneously, student's opinions in the experimental groups about the board game as an assessment method were collected in an open-descriptive form and in a short questionnaire, and analyzed. Results showed that students experienced a positive attitude toward the assessment method, a reduction of test anxiety and an increase in their motivation for learning.

Keywords Assessment methods · Board game · Collaborative testing · Gamification · High school physics

Introduction

In recent years, gamification—which refers to the use of game-based elements, such as game mechanics, esthetics, and game thinking in non-game contexts aimed at engaging people,

✉ Daniel Dziob
daniel.dziob@uj.edu.pl

[1] Smoluchowski Institute of Physics, Jagiellonian University, Lojasiewicza 11, 30-348 Krakow, Poland

[2] Department of Biophysics, Jagiellonian University Medical College, Lazarza 16, 31-530 Krakow, Poland

motivating action, enhancing learning, and solving problems—has become increasingly popular (Apostol et al. 2013; Deterding et al. 2011). Admittedly, the idea of introducing games in teaching is not a new concept. People have been using digital games for learning in formal environments since the 1960s (Ifenthaler et al. 2012; Moncada and Moncada 2014). However, the term of gamification was coined only a few years ago, and since then has been gaining more and more popularity (Dicheva et al. 2015; Sung and Hwang 2013). The benefits of gamification (or, in more broad terms, game-based learning, e.g., Ifenthaler et al. 2012) in educational contexts are widely described in the literature. Among them are increasing student intrinsic motivation and self-efficacy (Banfield and Wilkerson 2014; Seaborn and Fels 2015), motivation effect and improvement of the learning process (Dicheva et al. 2015; Sadler et al. 2013), as well as improving the positive aspects of competition (Burguillo 2010; Conklin 2006).

Games can reinforce knowledge and bridge the gap between what is learned by creating dynamic, fun, and exciting learning environments (Royse and Newton 2007). They are a powerful teaching strategy, and they challenge and motivate students to become more responsible for their own learning (Akl et al. 2013). However, this requires having the game to be well-designed and structured clearly with a framework that provides effective outcomes (Allery 2004). The review presented in Dicheva et al. (2015) suggests that early adopters of gamification are mostly Computer Science/IT educators. This is in line with the rising popularity of computer games, which have become prominent in the last decade. Nowadays, many articles can be found, in which using computer games in the teaching process are introduced and evaluated (Eskelinen 2001; Ko 2002; Rieber and Noah 2008). Nevertheless, not all of them are proper for school circumstances. Zagal et al. (2006) points out that some of the designed games are highly opaque and complex in rules, and did not include players collaborating to play the game: Therefore, these games did not affect students peer-learning. Through peer collaboration, students build on each other's knowledge to develop new attitudes, cognitive skills and psychomotor skills (Adams 2006; Damon and Phelps 1989). The same authors suggest that for such a purpose board games could be used due to their transparency regarding the core mechanics. Moreover, board games provide the teachers with an opportunity to guide or direct children to meet specific educational goals by extending their learning during and after playing the game (Durden and Dangel 2008; Wasik 2008). Teachers can also facilitate communication amongst children, build understanding about games, discuss concepts, and provide feedback to one another (Griffin 2004).

Board games are also successfully used in early childhood education (Ramani and Siegler 2008; Shanklin and Ehlen 2007) as a pedagogical tool that reinforces a positive environment for learning (Dienes 1963). Games also appear to build positive attitudes (Bragg 2003) and self-esteem, and enhance motivation (Ernest 1986). They have been found to be also effective in promoting mathematical learning (Bright et al. 1983), mathematical discussion (Ernest 1986; Oldfield 1991), social interaction (Bragg 2006), and risk taking abilities (Sullivan 1993). Some types of board games were also used in medical education and have been found as useful methods for conveying information and promoting active learning (Neame and Powis 1981; Richardson and Birge 1995; Saunders and Wallis 1981; Steinman and Blastos 2002). In the present study, a board game—a competitive game between groups of students in a classroom—was used as an assessment in order to examine if it could increase high school students' achievements and retention of knowledge in physics.

To assess student achievements in general, and as a result of the board game specifically, there are two main formats of assessment distinguished and widely discussed in the literature, namely formative and summative assessment (Harlen and James 1997; Wiliam and Black 1996). In general, formative assessment is carried out throughout a unit (course, project) and its purpose is to provide feedback to students about the learning process. Summative assessment is given at the end of a unit (course, project) and is used to summarize students' achievements usually in the form of grades (Harlen and James 1997; Looney 2011; McTighe and O'Connor 2005). Even though summative assessment could be performed in many ways, some authors pointed to a lack of post examination feedback for students as a weakness (Leight et al. 2012; Talanquer et al. 2015). In our study, the board game was used essentially as a tool for summative assessment, although it also includes some elements of formative evaluation. Such a combination was dubbed by Wininger (2005) as a *formative summative assessment*. It entails reviewing exams with students so that they get feedback about their knowledge comprehension. One example of this approach is collaborative testing that aims to give students an opportunity to work in groups during or at the end of an exam (Guest and Murphy 2000; Lusk and Conklin 2003). Research has shown that there are many benefits to collaborative testing. These are described in detail by Duane and Satre (2014), Gilley and Clarkston (2014), Kapitanoff (2009), and based also on literature about the positive impact of group testing (Millis and Cottell 1998; Michaelson et al. 2002; Hodges 2004) and peer-learning (Slusser and Erickson 2006; Meseke et al. 2008; Ligeikis-Clayton 1996), which are parts of collaborative learning. Among others, the most important benefits of collaborative learning are increasing students' achievements (Bloom 2009; Haberyan and Barnett 2010), reduction of test anxiety (Zimbardo et al. 2003), improvement of critical thinking ability (Shindler 2003), and collaboration skills (Lusk and Conklin 2003; Sandahl 2010).

The assessment in the form of a game employed in the current research is based on the authors' previous experiences and research (Dziob et al. 2018). It evaluates not only the content matter knowledge itself, as in typical tests, but combines a few different aspects together, as schematically shown in Fig. 1. It assesses the relationship between content knowledge and everyday life, as well as socio-historical context. Moreover, it gives the opportunity to assess research skills required to conduct experiments. The form of the board game enables development of social and entrepreneurial skills in the form of a *challenge-yourself* competition, which allows students to surpass individual limitation (Doolittle 1997).

This study reports on the efficacy of assessing students' knowledge by means of a group board game approach and measuring its effects on students' learning outcomes. The research questions are as follows:

1) What is the effect and influence of the board game assessment on student learning outcomes when compared with student prior results in physics?
2) What is the effect and influence of the board game assessment on student learning outcomes when compared with a traditional teaching approach?

Methodology

Participants

The research was conducted on a group of 131 students in total from two high schools in Poland. Students were divided into experimental (of $n = 37$ and $n = 36$ in school 1 and 2, respectively) and control groups ($n = 31$ and $n = 26$). Each group was taught by the same

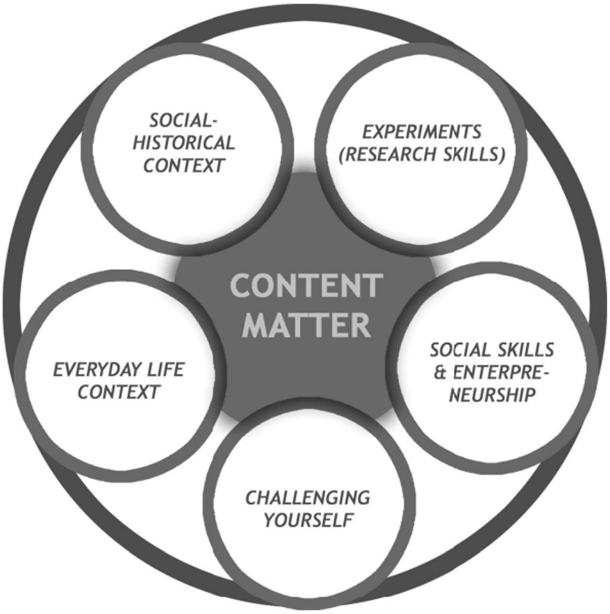


Fig. 1 Assessment strategy components. In addition to the content matter knowledge, all other expressed elements were involved in the assessment process. Own work

teacher and followed the same curriculum during their education. Just before the experiment, the students had accomplished a 25-h course on vibrations and waves (in school 1) and on optics (school 2). After finishing the unit, experimental groups took part in the assessment in the form of a group board game (described below, hereinafter the intervention) and 1 week after in a traditional test. Students from control groups participated only in the traditional test the same as experimental group, but without the intervention. In each group, the ratio of males to females was similar (3:2).

Intervention

The section below contains a detailed description of the intervention: a game which students from experimental groups played once at the end of the unit together with the evaluation process. The description includes procedure and examples of questions used in students' assessment.

Intervention Organization

The game lasted approximately 2.5 lesson hours (approx. 110 min). At the beginning, students were divided randomly into groups of 4 to 5 people each, and asked to take seats around the game board table. Each group began with questions concerning some physics phenomena and students moved their tokens (one per group) forward by the number of right answers or correctly named concepts. At the end of the game (when allocated time ended), students were asked to fill in a self- and peer-assessment questionnaire. At each stage of the game after students had attempted, a scientifically accepted answer to each question was provided together with a proper explanation by students or, if needed (when students didn't pass), by the teacher. Thus, this approach allows the teacher to immediately rectify and clarify students' misconceptions.

Game Board—Organization

The game board consisted of a circular path, and the participants moved their group token along this path. The path was made up of a random assortment of five potential categories, or activities, to land on. These included physics phenomena charades, famous people, short answer questions, multiple-choice questions, and simple experiments. All questions required the students to perform different types of activities and allow them to obtain a different number of points. Because the number of points obtained at each stage was identical with number of spots the token was moved, the scoring system was identical with the movement system as in typical board game. Additionally, there were also two special lines on the board. Whenever any group reached one of them, the members of both groups received special algebraic tasks or complex experimental tasks to solve. Figure 2 presents the game board with the fields of different type indicated on it.

Physics Phenomena Charades

Upon reaching this field, one representative of a given group received six cards with names of various physics phenomena related to waves and vibrations (school 1) or optics (school 2; see Fig. 3). Their aim was to describe each concept, without using the words given, so that the rest of the team could guess the name. The time for this task was limited to 1 min (measured by a small

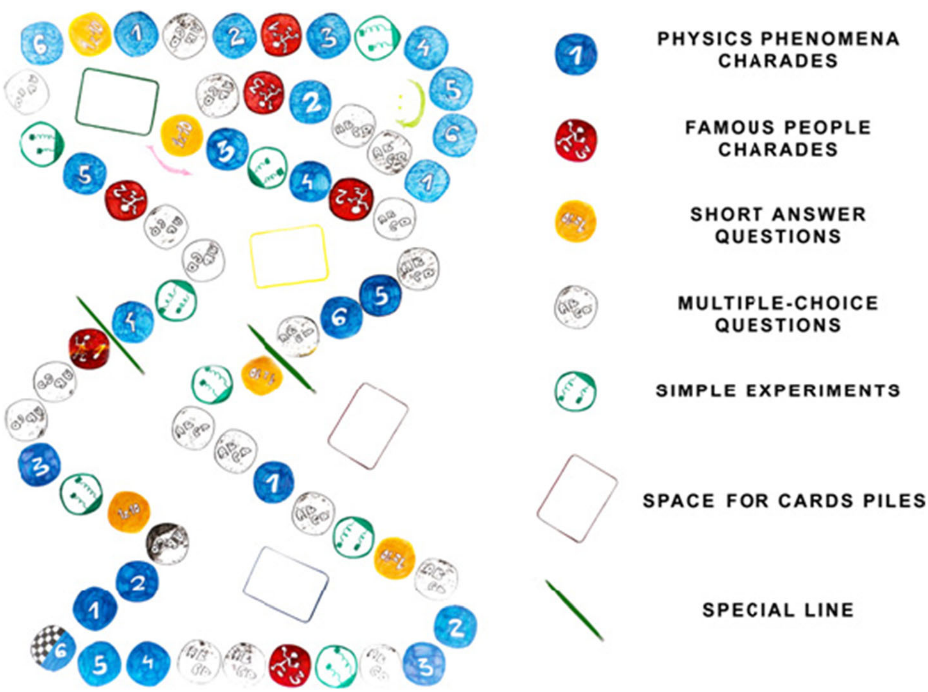


Fig. 2 The design of the board game

hourglass). After the end of the round, tokens were moved forward by the number of fields equal to the number of correctly guessed charades.

Famous People Charades

These questions were similar to the previous ones, but they related to important people connected with the concepts of waves, vibrations, and acoustics (physicists, musicians, etc.) or optics (Fig. 4). The scoring system was identical to the one employed in the physics phenomena charades.

Short Answer Questions

The short answer questions differed with respect to their level of difficulty, but usually they required only a true/false answer (see Fig. 5). The questions were asked by the teacher and the time of each group's round was 1 min. Within that time, all members of the currently active group could answer as many questions in a row as they managed, without passing their turn to another group. If the provided answer was wrong, the next group took over and had an opportunity to answer other questions. At the end of each round, the groups moved their token forward by the number of the correctly answered questions divided by 2 and rounded up.

Multiple-Choice Questions

Upon reaching a field of this category, a group received multiple-choice questions related to scientific reasoning (Fig. 6). Students had to point out the correct answer and provide comprehensive argumentation for their choice. By providing the right answer together with the correct explanation, students could move forward by 2 fields on the board. Otherwise, no move was allowed.

Simple Experiments

Upon reaching a field of this category, the students had to conduct some simple experiments in order to prove relevant phenomena (Fig. 7). The equipment necessary for each experiment, with some extra materials among them, were available to students. An important part of the task was the necessity to make a decision which objects were essential. The other groups taking part in the game were enabled to ask the currently active team detailed questions about the conducted experiment and ask for additional explanations. Having carried out the experiment and addressed the questions properly, the group was allowed to move forward by 2 fields.

Waves and vibrations:	**Optics:**
Amplitude	*Incident ray*
Isochronism	*Diffuse reflection*
Wave front	*Interference pattern*
Superposition principle	*Ray diagram*
Fundamental frequency	*Focal point*
Ear canal	*Light absorption*

Fig. 3 Examples of physics phenomena charades

Waves and vibrations:	Optics:
Ernst Chladni	*Willebrord Snell*
Robert Hooke	*Zacharias Janssen*
Christian Huygens	*Thomas Young*
Alexander Graham Bell	*Hans Lippershey*
Wolfgang Amadeus Mozart	*Galileo Galilei*

Fig. 4 Examples of famous people charades from games on waves and vibrations and optics

Algebra Tasks

When one of the groups reached the first special line on the game board after the end of the round, all competing groups simultaneously received three algebra tasks. They had 10 min to solve them. For accomplishing this task, each group could receive a maximum of 4 points and moved their token forward by 4 fields. Incorrect or incomplete solutions, assessed by the teacher, reduce the amount of points.

Experimental Task

When one of the groups reached the second special line on the board at the end of the round, all competing groups simultaneously received one experimental task, which was neither discussed nor solved during any previous class. The students had to come up with an experimental design to examine the effect of damping the movement of the harmonic oscillator (school 1) or examine the surrounding medium refractive index on the glass lens focal length (school 2). The groups received special worksheets prepared in accordance with an inquiry-based methodology. Students had to formulate a proper hypothesis, describe the plan of the experiment, draw the experimental setup, write down their observations, analyze the results, and draw the conclusions. This part took up to 20 minutes. For this task, students could receive a maximum of 10 points.

Instruments and Data Collection

Former Achievements

Before the intervention, students from each group were tested individually in four tests throughout the school year; on kinematics, energy, gravitation, and rigid body rotational motion (school 1);

Waves and vibrations:

1. *In which state of matter does a sound wave have a maximum speed?*
2. *Mechanical waves can be divided into transverse and ...*
3. *Can harmonic motion be described by a tangent function?*
4. *A standing wave does not change its frequency over time. True or false?*

Optics:

1. *If Ann is standing 3 feet in front of a plane mirror, how far from her will her image be located?*
2. *The white light consists only of red, green and blue colors. True or false?*
3. *Can a real image be obtained in the concave mirror?*
4. *Lenses can be divided into converging and...*

Fig. 5 Examples of short answer questions

Waves and vibrations:

Q: Which oscillatory motion is not a harmonic motion?

A motion of reflecting ball

B. motion of the weight on the spring

C. vibrating motion of the liquid in the tube-shaped U

D. movement of the tube floating on the liquid surface after partially submerged and released

Why?

Optics:

Q: Which quantity always changes when light is passes from one medium to another with a different speed of light within?

A. Frequency
B. Color
C. Direction
D. Wavelength

Why?

Fig. 6 Examples of short answer questions

and electrostatics, current, magnetic field, and induction (school 2). They comprised mixed problems: content knowledge and scientific reasoning tasks, multiple-choice, open-response, and algebra problems. Tests were the same for experimental and control groups. The average of each student's percentage results on the four tests was used to measure his/her achievement prior to the game, henceforth called *average former achievements* and denoted as *FA*.

Assessment Questionnaires

When the game ended, each student was asked to fill in individually two questionnaires of self- and peer-assessment in order to evaluate themselves and other fellow players from the same group under various sides. Each questionnaire was composed of eight questions designed on a 6-point Likert scale. Half of the questions focused on the students' communication skills, while the rest on subject matter contribution. In Table 1, the self-assessment questionnaire is presented. The peer-assessment questions were designed in the similar way.

Evaluation Process

The questionnaire-based assessment results were included in the final score according to the author's own approach presented below and described in detail in Dziob et al. (2018).

Waves and vibrations:

1. *Create a standing wave in the air.*
2. *Experimentally demonstrate the superposition principle.*
3. *Demonstrate that a wave does not cause matter to move towards it distribution.*

Optics:

1. *Experimentally demonstrate the law of reflection.*
2. *Show the refraction of light.*
3. *Demonstrate that the light always travels via a straight line.*

Fig. 7 Examples of simple experiments tasks

Table 1 Student self-assessment questionnaire

Question	1–6 scale
Were you involved in the group work?	Communication skills
Did you communicate adequately with the group?	
Did you take part in the discussion on the problem?	
Did you take into account the opinion of other students?	
Did you prepare for the test beforehand?	Subject matter contribution
Did you participate in solving problems and tasks?	
Did you have sufficient knowledge to solve the issues?	
Did you contribute to the final result of the group?	

1. The mean average score was calculated based on the "subject matter contribution" and, separately, "communication skills" points in the self-assessment results (S).
2. The average score was calculated based on the "subject matter contribution" and, separately, the "communication skills" points ascribed to the student by the other members of the group (the peer-assessment, P).
3. At the end, the "subject matter contribution" and "communication skills" scores were obtained separately as follows:
1. If $|S - P| \leq 1$ (a consistent evaluation): take P as the final score
2. if not (an inconsistent evaluation): take $P - 0.5$ as the final score

The percentage score for each team was calculated by dividing the number of points (number of fields) accumulated by the group by the maximum number of points available to obtain. The final overall score given to each student consisted of three parts:

1. the group common percentage result from the board game—with the weight of 0.5,
2. the questionnaire-based assessment percentage result for the "subject matter contribution"—with the weight of 0.3,
3. the questionnaire-based assessment percentage result for the "communication skills"—with the weight of 0.2.

The final score for each student after the game, calculated according to the algorithm above and expressed in a percentage form, is henceforth referred to as *game score* (*GS*).

Post-test

An unannounced post-test was conducted in the experimental groups 1 week after the game. The same test was given to students from the control groups, just after finishing the unit. It was prepared in a traditional written form. There was neither a review of the relevant content knowledge during regular classes nor a post-game discussion of the game problems and results before this test. The *post-test* (*PT*) score is expressed in percentage terms.

Students' Opinions Questionnaire

Students from the experimental groups received an anonymous short evaluation questionnaire a week after the game (just after the post-test). It consisted of six questions asking "How the knowledge assessment method influences your…", and each answered on a linear point scale ranging from −5 to +5, with the numbers indicating the most negative (−5), through none (0) to the most positive (+5) impact. The evaluated aspects covered pre-test preparation, engagement in team work, answer difficulty, test anxiety, final acquisition of knowledge, and motivation for future learning. It was also a space for students to present opinions on the game. Exact questions are presented in the results section together with students' answers.

Data Analysis

Basic Statistical Analysis

Below, a statistical analysis of the data is carried out, firstly for the experimental groups, and then in comparison with the control groups. In Table 2, we present basic descriptive statistics and empirical distributions (in the form of histograms, with normality tested by the Shapiro-Wilk tests) for each set of results, i.e., the FA, GS, and PT, for both experimental groups. The numbers 1 and 2 in superscripts indicate the schools.

All examined variables have normal distribution. Student t test showed that the differences among each variable means are statistically significant (in each case p value < 0.05). This allowed for comparison of the students' results in different tests. On average, the students from experimental groups scored 47%/59% (school 1/school 2) in the former test, 70%/80% in the game, and 58%/68% in post-test. An increase of almost 23 percentage points (pp.) between FA and GS in both experimental groups might emerge as the result of student cooperation during the game. The PT results are lower than GS. However, they are still statistically significantly higher than the FA results ($p < 0.05$), which may suggest a positive impact of game-based assessment on students' achievements. It should be noticed, however, that at each stage the results of students from the first school are lower than those from the second. This is consistent with the author's observation about the educational standards in each school. Therefore, in what follows, both groups will be analyzed independently, in comparison to adequate control groups from the same schools.

In both schools, control groups were formed from the students who studied with the same teacher and who completed the same courses. In Tables 3 and 4, basic descriptive statistics for control and experimental groups former achievements (FA) and results in post-test (PT) are presented. In each school, average former achievements of the students from both the

Table 2 Basic statistics of the results obtained by students from experimental groups

Variable	Characteristics						
	Mean	95% Confidence interval for mean	Median	Lower quartile	Upper quartile	Standard deviation	
FA^1 [%]	47.05	(44.38; 49.72)	45.41	40.31	54.67	8.00	School 1
GS^1 [%]	69.74	(66.46; 73.03)	69.71	64.58	77.47	9.86	
PT^1 [%]	58.44	(54.29; 62.56)	57.58	48.48	69.70	12.44	
FA^2 [%]	58.58	(53.07; 64.07)	55.37	46.68	68.54	16.25	School 2
GS^2 [%]	80.01	(76.27; 83.76)	79.81	71.56	87.84	11.07	
PT^2 [%]	67.76	(62.56; 72.97)	68.36	56.78	76.82	15.38	

Superscripts indicate the schools

FA average former achievements, *GS* the final score in the game, *PT* the result in the post-game test experimental and the control group are similar. As proved by the *t* test (each data are normally distributed), there are no statistically significant differences ($p > 0.05$) between FA in experimental and control group within one school.

The former achievements and post-test results within the control groups were tested in the same way. Results indicate ($p > 0.05$) that there is no statistically significant difference between former achievements (FA^C) and post-test results (PT^C) in control groups. It implies that the post-test can be considered as a reliable tool, neither harder nor simpler than the former test. It allows the comparison of the PT results between experimental and control group. In school 1, the difference between mean results is close to 8 pp. ($p = 0.0000$), and in school 2, it is a little bit above 10 pp. ($p = 0.0003$). It clearly shows that experimental groups obtained statistically higher results in PT than their colleagues from the control groups. In other words, students from the experimental groups gained significantly more knowledge than their colleagues from the control groups.

Students' Opinions

Students' opinions about the board game were collected just after the post-test, but before providing them with the information about their final marks. Participants filled in a questionnaire (on −5 to +5 scale, where 0 means no impact) and expressed their comments anonymously in an open, descriptive form. Mean results for the questionnaire questions in both schools are provided in Table 5.

Because the answers for each were normally distributed (tested by Shapiro-Wilk test), the comparisons of the H_0: mean against zero was calculated using the Student *t* test. The test showed ($p < 0.05$) that in each question students' answers differ significantly (were higher or lower) than "0" value, which means no impact. In other words, for each question, students state significant influence of the board game on the tested issue. Students in both schools judged the assessment in the form of a group board game beneficial for their preparation, and pre-test preparation was marked by students from each school as positive. This means that students would spend more time preparing for the game, as opposed to spending time on preparation for a traditional test. Both experimental groups agreed that the level of engagement of their team-mates was high and that answering questions was easier than in traditional, individually taken tests. It corresponds with the students' opinions that this new form of assessment prompts them to give answers, even if they feel uncertain about their correctness. Anxiety during the test was assessed at − 3.1 and − 2.3 in both experimental groups, respectively, which means that this form of assessment reduces the

Table 3 Basic statistics of the students from the first school

Variable	Characteristics						
	Mean	95% Confidence interval for mean	Median	Lower quartile	Upper quartile	Standard deviation	*t* test
FA^1 [%]	47.05	(44.38; 49.72)	45.41	40.31	54.67	8.00	$p = 0.3120$
FA^{1C} [%]	49.43	(45.31; 53.55)	48.28	42.07	58.12	11.23	
PT^1 [%]	58.44	(54.29; 62.56)	57.58	48.48	69.70	12.44	$p = 0.0000$
PT^{1C} [%]	50.67	(46.43; 54.92)	49.34	41.03	61.03	11.57	

Superscripts differentiate experimental (^{1}) and control group (^{1C}) within first school

FA average former achievements, *PT* the result in the post-game test

Table 4 Basic statistics of the students from the second school

Variable	Characteristics						
	Mean	95% Confidence interval for mean	Median	Lower quartile	Upper quartile	Standard deviation	*t* test
FA^2 [%]	58.58	(53.07; 64.07)	55.37	46.68	68.54	16.25	$p = 0.6539$
FA^{2C} [%]	56.81	(51.21; 62.41)	56.49	51.78	64.11	14.16	
PT^2 [%]	67.76	(62.56; 72.97)	68.36	56.78	76.82	15.38	$p = 0.0003$
PT^{2C} [%]	57.65	(51.43; 63.87)	58.52	44.04	65.76	15.72	

Superscripts differentiate experimental (^{2}) and control group (^{2C}) within second school

FA average former achievements, *PT* the result in the post-game test

anxiety normally associated with traditional exams. Students also indicated, both in the questionnaires and open opinions, that the board game improved their final level of knowledge. Students also felt motivated by the game to continue learning.

A few examples of the opinions are presented below:

- Student A:

This is a good option to test for people who are weaker in calculation. Not everyone is able to solve a complex task, but anyone can learn theory.

- Student E:

This form of the test was very good, because you could learn also during the test. It teaches cooperation in the way you could have fun.

- Student K:

I think that we learned and invented more during this game than during a written test. It was a very good possibility for integration.

- Student O:

Each group should get all kinds of questions. Then it would be more fair. Questions should be more focused on physics, without connections to history.

Table 5 Mean results for each question in the questionnaire for both experimental groups

How the form of the assessment influences your		School 1	School 2
1.	Pre-test preparation	3.2 (1.1)	2.8 (1.6)
2.	Engagement into team work	3.6 (0.9)	3.7 (0.8)
3.	Easy of answering	2.9 (1.8)	3.4 (1.2)
4.	Test anxiety	− 3.1 (1.6)	− 2.3 (2.1)
5.	Final acquire of knowledge	2.4 (1.8)	3.9 (1.2)
6.	Motivation for future learning	3.5 (0.9)	3.4 (1.7)

Numbers in brackets denote standard deviations

The vast majority of students' opinions were positive and enthusiastic. A few of them used the feedback to provide helpful and insightful comments for improving the assessment. In the discussion section, we relate them to the findings commonly presented in the literature on collaborative testing and gamification.

Discussion

The main purpose of this research was to investigate the influence of assessing students' achievements in the form of a group board game in comparison to their former achievements and traditional tests. The first important finding is a statistically significant increase in students' achievements in the game in comparison to their former achievements. This result is consistent with research on the positive impact of collaborative testing, which shows that students' results obtained in collaborative taken exams are higher than in individual ones (Bloom 2009; Haberyan and Barnett 2010; Kapitanoff 2009; Lusk and Conklin 2003). Some authors controvert, however, the ability of collaborative testing to improve content retention (Leight et al. 2012; Woody et al. 2008), pointing out that only their performance during the collaborative exam is higher. Our second results addressed this problem. We found that students from experimental groups gained higher results in the post-test taken 1 week after the game with respect to the results obtained by the control groups. In other words, the students assessed by the game obtained not only high performance in the game but also in a knowledge test taken after the game. This finding is encouraging with respect to other research that shows improvement in students' achievement after the collaborative exam in the long run (Cortright et al. 2003; Jensen et al. 2002; Simpkin 2005). The results show also that the assessment method is efficient independently of the level of students' performance.

The students' opinions were encouraging and supported findings in the literature. Board games can be perceived as a form of activity in which group work skills are exploited and play an essential role in accomplishing tasks (Dallmer 2004; Kapitanoff 2009; Lusk and Conklin 2003; Sandahl 2010; Seaborn and Fels 2015; Shindler 2003). Some researchers (Dicheva et al. 2015; Sadler et al. 2013) suggested that *gamification* could improve the learning process, which can be inferred from the increase in students' results in post-test. By playing the game, the students learn to listen to everybody else's answers, provide fellow players with their know-how, and respond to ideas proposed in discussions. According to Hanus and Fox (2015) and Jolliffe (2007), the above can stimulate knowledge assimilation. In the students' opinions expressed in the questionnaire and open-descriptive form, the board game assessment has a positive impact on their motivation and social interactions, which also corresponds to the literature findings (Banfield and Wilkerson 2014; Bragg 2006; Seaborn and Fels 2015). Furthermore, the assessment in the form of a game induces far less test anxiety by giving students a sense of being supported by the other team members (Banfield and Wilkerson 2014; Kapitanoff 2009; Lusk and Conklin 2003; Sandahl 2010; Zimbardo et al. 2003). Similar results can be found in other research, in which results from student's attitude surveys confirm that collaborative testers have more positive attitudes towards the testing process in general compared to students who take assessments individually (Bovee and Gran 2005; Giuliodori et al. 2008; Meseke et al. 2009). Finally, an active involvement in the self- and peer-assessment process may increase the students' self-assurance and adequate self-esteem (Hendrix 1996), thereby enhancing retention of knowledge (Sawtelle et al. 2012).

Comments on Organization of the Game

Preparation of the board game has a few important aspects, which should be described in order to give the reader the impression of how to adapt the idea to her/his own purpose. The most important is to decide on the topic, which should give considerable benefits for assessing knowledge in a non-standard form. A board game has to have clear rules, provide a sufficient rationale for collaboration, be a challenge for participants, and provide different types of activities and experiences. It is connected with the next important step, which is to precisely define the goals of the event, regarding prepared tasks. Chosen activities should allow assessment not only the content knowledge but also all other aspects (e.g., Science as a Human Endeavor and Science Inquiry Skills) chosen by the teacher. Breedlove et al. (2004) reported that the effects of collaborative testing were directly related to the level of cognitive processing required by the test questions. The activities, rules, and scoring system should be modified and matched to the groups. Particularly, in our research, one student claimed that there was a possibility to guess the proper word in the charades without physics knowledge. This can be improved by additional rules or modifying the charades questions with other types of activities. Because the effectiveness of the collaborative testing may depend on students earlier teaching strategies and improve over time as students become more familiar with the collaborative process (Castor 2004), the modification of the game seems to be natural consequence.

Further Issues

The study examined only the short-term effect on students' retention knowledge. One can suppose that because the initial level of the forgetting curve was higher in experimental groups than in control groups, after a few months, experimental groups should also obtain better results. This assumption has to be, however, tested in future work. The method could be also implemented and verified in subjects other than physics as well as in a wider spectrum of school levels. Even though literature findings about collaborative testing and board games in many science subjects are very enthusiastic, only few of them focus on the distinction between high- and low-performing students (Giuliodori et al. 2008). This question should be also examined in future work.

Another issue is the claim that a teacher could influence the results, e.g., by focusing on the experimental group or neglecting control groups. In our research, control group results in the post-test were similar to their results in all other tests taken before, taken in average as former achievements. This approach, unlike the typical pre-test, allows us to address this remark. Comparison between pre- and post-test provides clear information about students gain in the examination topic, but can be easily influenced by the teacher, which will be found only in the lower achievements for the control group. In our approach, we assumed that an uninfluenced teaching style will have an effect in unchanged students' results in the post-test, which is assumed to be allowed. However, the implementation of the method under different circumstances could also provide worth worthwhile information.

Future research could also examine collaborative testing as a more effective standard assessment strategy across a curriculum (Meseke 2010). Because the game always has to be a challenge for students, some modifications should be introduced in the type of questions and rules or the board game should be used interchangeably with another collaborative method.

Concluding Remarks

This paper studied the influence of a board game as an assessment method on high school students' achievement. Students from experimental groups performed better in the game than in the former tests. Simultaneously their achievement in a traditional test taken 1 week after was significantly higher than for students from control groups. It implies that assessing students' achievement in the form of a game may improve their performance and short-term achievements.

The improvement of students' achievements may result in combining collaborative testing with gamification. Apart from quantitative results, the students' enthusiastic opinions are also indicative of the social benefits of the approach, such as the development of group work skills, supporting weaker students through collaboration with others, and, in addition to these, integration of the class. It appears that game-based assessment enhances students' retention of knowledge and provides opportunities for improvement for each student, regardless of their former performance. Moreover, it helps to improve students' attitudes towards their learning and add valuable collaborative learning experience to enhance the school curriculum.

The approach can be easily modified and adapted as a testing method in fields other that physics, especially natural sciences, in which assessing the experimental skills and socio-historical context are also under consideration.

References

Adams, P. (2006). Exploring social constructivism: Theories and practicalities. *Education, 34*(3), 243–257.

Akl, E. A., Kairouz, V. F., Sackett, K. M., Erdley, W. S., Mustafa, R. A., Fiander, M., Gabriel, C., & Schunemann, H. (2013). Educational games for health professionals. *Corhane Database of Systematic Reviews, 3*, 1–46.

Allery, L. (2004). Educational games and structured experiences. *Medical Teacher, 26*(6), 504–505.

Apostol, S., Zaharescu, L., & Aleze, I. (2013). Gamification of learning and educational games. *eLearning & Software for Education, 2*, 67.

Banfield, J., & Wilkerson, B. (2014). Increasing student intrinsic motivation and self-efficacy through gamification pedagogy. *Contemporary Issues in Education Research CIER, 7*(4), 291.

Bloom, D. (2009). Collaborative test taking. *College Teaching, 57*(4), 216–220.

Bovee, M., & Gran, D. (2005). Effects of collaborative testing on students satisfaction survey. *Journal of Chiropractic Education, 19*(1), 47.

Bragg, L. A. (2003). Children's perspectives on mathematics and game playing. In L. Bragg, C. Campbell, G. Herbert, & J. Mousley (Eds.), *Mathematics Education Research: Innovation, Networking, Opportunity: Proceedings of the 26th Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. Vol. 1, pp. 160–167). Geelong: Deakin University.

Bragg, L. A. (2006). *The impact of mathematical games on learning, attitudes, and Behaviours*. Bundoora: La Trobe University.

Breedlove, W., Burkett, T., & Winfield, I. (2004). Collaborative testing and test performance. *Academic Exchange Quarterly, 8*(3), 36–40.

Bright, G. W., Harvey, J. G., & Wheeler, M. M. (1983). Use of a game to instruct on logical reasoning. *School Science and Mathematics, 83*(5), 396–405.

Burguillo, J. C. (2010). Using game theory and competition-based learning to stimulate student motivation and performance. *Computers & Education, 55*(2), 566–575.

Castor, T. (2004). Making student thinking visible by examining discussion during group testing. *New Directions for Teaching and Learning, 100*, 95–99.
Conklin, A. (2006). Cyber defense competitions and information security education: An active learning solution for a capstone course. In System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on (Vol. 9, pp. 220b–220b). IEEE.
Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., & Dicarlo, S. E. (2003). Student retention of course content is improved by collaborative-group testing. *Advances in Physiology Education, 27*(3), 102–108.
Dallmer, D. (2004). Collaborative test taking with adult learners. *Adult Learning, 15*, 4–7.
Damon, W., & Phelps, E. (1989). Critical distinctions among three methods of peer education. *International Journal of Educational Research, 13*, 9–19.
Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification: Using game-design elements in non-gaming contexts. In Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (p. 2425). Vancouver, Canada.
Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Educational Technology & Society, 18*(3), 75–88.
Dienes, Z. P. (1963). *An experimental study of mathematics learning*. New York: Hutchinson.
Doolittle, P. (1997). Vygotsky’s zone of proximal development as a theoretical foundation for cooperative learning. *Journal on Excellence in College Teaching, 8*(1), 83–103.
Duane, B. T., & Satre, M. E. (2014). Utilizing constructivism learning theory in collaborative testing as a creative strategy to promote essential nursing skills. *Nurse Education Today, 34*(1), 31–34.
Durden, T., & Dangel, J. (2008). Teacher-involved conversations with young children during small group activity. *Early Years, 28*, 251–266.
Dziob, D., Kwiatkowki, L., & Sokolowska, D. (2018). Introducing a class tournament as an assessment method of student achievements in physics courses. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(4), 1111–1132.
Ernest, P. (1986). Games: A rationale for their use in the teaching of mathematics in school. *Mathematics in School, 15*(1), 2–5.
Eskelinen, M. (2001). Towards computer game studies. *Digital Creativity, 12*(3), 175–183.
Gilley, B., & Clarkston, B. (2014). Research and teaching: Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching, 043*(03), 83.
Giuliodori, M. J., Lujan, H. L., & DiCarlo, S. E. (2008). Collaborative group testing benefits high- and low-performing students. *Advances in Physiology Education, 32*(4), 274–278.
Griffin, S. (2004). Building number sense with number worlds: A mathematics program for young children. *Early Childhood Research Quarterly, 19*, 173–180.
Guest, K. E., & Murphy, D. S. (2000). In support of memory retention: A cooperative oral final exam. *Education, 121*, 350–354.
Haberyan, A., & Barnett, J. (2010). Collaborative testing and achievement: Are two heads really better than one? *Journal of Instructional Psychology, 37*(1), 32–41.
Hanus, M., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80*, 152–161.
Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice, 4*(3), 365–379.
Hendrix, J. C. (1996). Cooperative learning: Building a democratic community. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 69*(6), 333–336.
Hodges, L. (2004). Group exams in science courses. *New Dir Teach Learn, 2004*, 89–93.
Ifenthaler, D., Eseryel, D., & Ge, X. (2012). *Assessment in game-based learning: Foundations, innovations, and perspectives*. New York: Springer.
Jensen, M., Moore, R., & Hatch, J. (2002). Cooperative learning: Part I: Cooperative quizzes. *The American Biology Teacher, 64*(1), 29–34.
Jolliffe, W. (2007). *Cooperative learning in the classroom: Putting it into practice*. London: Paul Chapman.
Kapitanoff, S. H. (2009). Collaborative testing: Cognitive and interpersonal processes related to enhanced test performance. *Active Learning in Higher Education, 10*(1), 56–70.
Ko, S. (2002). An empirical analysis of children’s thinking and learning in a computer game context. *Educational Psychology, 22*(2), 219–233.
Leight, H., Saunders, C., Calkins, R., & Withers, M. (2012). Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class. *CBE Life Sciences Education, 11*(4), 392–401.
Ligeikis-Clayton, C. (1996). Shared test taking. *J NY State Nurses Assoc, 27*(4), 4–6.
Looney, J. (2011). Integrating Formative and Summative Assessment. OECD Education Working Papers, 58(5).

Lusk, M., & Conklin, L. (2003). Collaborative testing to promote learning. *Journal of Nursing Education, 42*(3), 121–124.
McTighe, J., & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership, 63*(10), 10–17.
Meseke, C., Nafziger, R., & Meseke, J. (2008). Student course performance and collaborative testing: A prospective follow-on study. *Journal of Manipulative and Physiological Therapeutics, 31*(8), 611–615.
Meseke, C. A., Bovee, M. L., & Gran, D. F. (2009). The impact of collaborative testing on student performance and satisfaction in a chiropractic science course. *Journal of Manipulative and Physiological Therapeutics, 32*, 309–314.
Meseke, C., Nafziger, R., & Meseke, J. (2010). Student attitudes, satisfaction, and learning in a collaborative testing environment*. *Journal of Manipulative and Physiological Therapeutics, 24*(1), 19–29.
Michaelson, L. K., Knight, A. B., & Fink, D. L. (2002). *Team-based learning: A transformative use of small groups*. Westport, CT: Preager.
Millis, B. J., & Cottell, P. G. (1998). *Cooperative learning for higher education faculty.* Phoenix: American Council on Education and Oryx Press.
Moncada, S., & Moncada, T. (2014). Gamification of learning in accounting education. *Journal of Higher Education Theory and Practice, 14*, 9.
Neame, R., & Powis, D. (1981). Toward independent learning: Curricular design for assisting students to learn how to learn. *Journal of Medical Education, 56*(11), 886–893.
Oldfield, B. J. (1991). Games in the learning of mathematics—Part 1: Classification. *Mathematics in School, 20*(1), 41–43.
Ramani, G. B., & Siegler, R. S. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science, 11*(5), 655–661.
Richardson, D., & Birge, B. (1995). Teaching physiology by combined passive (pedagogical) and active (andragogical) methods. *The American Journal of Physiology, 268*, S66–S74.
Rieber, L., & Noah, D. (2008). Games, simulations, and visual metaphors in education: Antagonism between enjoyment and learning. *Educational Media International, 45*(2), 77–92.
Royse, M. A., & Newton, S. E. (2007). How gaming is used as an innovative strategy for nurse education. *Nursing Education Perspectives, 28*(5), 263–267.
Sadler, T. D., Romine, W. L., Stuart, P. E., & Merle-Johnson, D. (2013). Game-based curricula in biology classes: Differential effects among varying academic levels. *Journal of Research in Science Teaching, 50*(4), 479–499.
Sandahl, S. S. (2010). Collaborative testing as a learning strategy in nursing education. *Nursing Education Perspectives, 31*(3), 142–147.
Saunders, N. A., & Wallis, B. J. (1981). Learning decision-making in clinical medicine: A card game dealing with acute emergencies for undergraduate use. *Medical Education, 15*(5), 323–327.
Sawtelle, V., Brewe, E., & Kramer, L. H. (2012). Exploring the relationship between self-efficacy and retention in introductory physics. *Journal of Research in Science Teaching, 49*(9), 1096–1121.
Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies, 74*, 14–31.
Shanklin, S. B., & Ehlen, C. R. (2007). Using the monopoly board game as an in-class economic simulation in the introductory financial accounting course. *Journal of College Teaching & Learning, 4*(11), 17–22.
Shindler, J. V. (2003). "Greater than the sum of the parts?" examining the soundness of collaborative exams in teacher education courses. *Innovative Higher Education, 28*(4), 273–283.
Simpkin, M. G. (2005). An experimental study of the effectiveness of collaborative testing in an entry-level computer programming class. *Journal of Information Systems, 16*, 273–280.
Slusser, S. R., & Erickson, R. (2006). Group quizzes: An extension of the collaborative learning process. *Teaching Sociology, 34*(3), 249–262.
Steinman, R. A., & Blastos, M. T. (2002). A trading-card game teaching about host defence. *Medical Education, 36*(12), 1201–1208.
Sullivan, P. (1993). Short flexible mathematics games. In J. Mousley & M. Rice (Eds.), *Mathematics of primary importance* (pp. 211–217). Brunswick, Victoria: The Mathematical Association of Victoria.
Sung, H., & Hwang, G. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education, 63*, 43–51.
Talanquer, V., Bolger, M., & Tomanek, D. (2015). Exploring prospective teachers' assessment practices: Noticing and interpreting student understanding in the assessment of written work. *Journal of Research in Science Teaching, 52*(5), 585–609.
Wasik, B. A. (2008). When fewer more: Small groups in early childhood classrooms. *Early Childhood Education Journal, 35*, 515–521.

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537–548.

Wininger, S. (2005). Using your tests to teach: Formative summative assessment. *Teaching of Psychology, 32*(3), 164–166.

Woody, W., Woody, L. & Bromley, S. (2008). Anticipated group versus individual examinations: a classroom comparison. *Teach Psychol. 35*, 13–17.

Zagal, J., Rick, J., & Hsi, I. (2006). Collaborative games: Lessons learned from board games. *Simulation & Gaiming, 37*(1), 24–40.

Zimbardo, P. G., Butler, L. D., & Wolfe, V. A. (2003). Cooperative college examinations: More gain, less pain when students share information and grades. *The Journal of Experimental Education, 71*(2), 101–125.