MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

The Hypothesis Experiment Class approach to radioactivity: effects on conceptual understanding and student interest

verfasst von / submitted by

Maximilian Jeidler, BEd

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Education (MEd)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 199 520 523 02
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Lehramt Sek (AB) Unterrichtsfach Mathematik Unterrichtsfach Physik
Betreut von / Supervisor:	Univ.-Prof. Dr. Martin Hopf
Mitbetreut von / Co-Supervisor:	Micheal Malvern Hull, PhD

Abstract – English

Hypothesis Experiment Class (HEC) is an interesting teaching concept developed by Kiyonobu Itakura. At this point, however, little concrete research can be found in the literature on this teaching concept. Therefore, the aim of this master's thesis is to evaluate the concept under the aspects of its applicability within teaching. HEC is expressed in so-called Classbooks, which contain all the learning material. In particular, the Classbook "The Radiation Around Us" was examined, which deals with the everyday nature of radioactivity. This content focus is of additional interest because teaching concepts on the topic of radioactivity are few and far between, especially in German-speaking countries. To evaluate the concept, a pre-post study with a control group was conducted. A comparison was made with traditional physics lessons on radioactivity. The study was conducted on site in different schools and classes. The data suggest that participating students enjoyed working with the Classbook. In addition, students in the experimental group showed a better understanding of the everyday nature of radioactivity than students in the control group. The Classbook "The Radiation Around Us" and HEC in general therefore suggest themselves as a valuable supplement to existing teaching.

Abstract – Deutsch

Hypothesis Experiment Class (HEC) ist ein interessantes Unterrichtskonzept, dass von Kiyonobu Itakura entwickelt worden ist. Zu diesem Unterrichtskonzept finden sich in der einschlägigen Forschungsliteratur wenige konkrete Untersuchungen. Daher besteht das Ziel der vorliegenden Masterarbeit darin, dieses Konzept im Hinblick auf seine Anwendbarkeit im Unterricht zu evaluieren. HEC drückt sich in sogenannten Classbooks aus, in denen das ganze Lernmaterial enthalten ist. Speziell wurde dabei das Classbook „The Radiation Around Us“ untersucht, das sich mit der Alltäglichkeit von Radioaktivität auseinandersetzt. Dieses Unterrichtsthema ist zusätzlich interessant, weil es nur wenige Unterrichtskonzepte zum Thema Radioaktivität gibt, vor allem im deutschsprachigen Raum. Um das Konzept HEC zu evaluieren, wurde im Rahmen der vorliegenden Untersuchung eine Prä-Post-Studie mit einer Kontrollgruppe durchgeführt. Verglichen wurde dabei der Einsatz des HEC-basierten Lernmaterials mit traditionellem Physikunterricht zu Radioaktivität. Die Studie wurde direkt in verschiedenen Schulen und Klassen durchgeführt. Die gewonnen Daten legen nahe, dass den Schüler*innen die Arbeit mit dem Classbook gefallen hat. Außerdem war das Verständnis von Schüler*innen der Experimentalgruppe hinsichtlich der alltäglichen Natur von Radioaktivität besser als jenes der Kontrollgruppe. Das Classbook „The Radiation Around Us“ und können als fruchtbare Ergänzung zum eigenen Unterricht nahegelegt werden.

Contents

1 Literature review

To begin with, it is important to understand what Hypothesis Experiment Class (HEC) is, and how it defines "experiments" and "hypothesis" as well as the relationship between them. Hypothesis Experiment Class makes use of so called Classbooks. The Classbook presented in this paper is about radioactivity. Therefore, I will outline the usual way of teaching radioactivity in order to be able to compare it with HEC. In addition, I will also present other curricula that deal with radioactivity. Also, I will briefly outline why special attention should generally be paid to radioactivity in the classroom. When talking about teaching a subject in physics, student ideas are always of interest. The most important ones on radioactivity are also presented in this chapter. At the beginning, I will give an overview of what is meant by radiation and radioactivity. I will discuss important measurands and what role both natural and artificial radioactivity play in our lives.

1.1 What is radiation and how does it affect us?

As mentioned above, I would like to begin by giving a theoretical overview of the topic of radiation and radioactivity. First, I will clearly define a number of important theoretical terms. Then I will describe the most important types of decay. I will also provide the most relevant mathematical background, as well as a description of dose sizes. The Classbook relevant to this work is primarily concerned with natural radioactivity, which is why I will go into more detail about this, as well as artificial radioactivity. Finally, I will present a rough description of how radiation can be attenuated.

1.1.1 Terms related to radiation

There are different types of radiation. In general, radiation is the transport of energy and mass that is not bound to a medium. If electromagnetic energy is transported through space by electromagnetic waves, we speak of photon radiation. Photon radiation consists of individual radiation quanta, which are called photons. These include radio waves, light, X-rays and gamma rays. If the radiation transports any kind of matter, this is called corpuscular radiation. The so-called corpuscles can be electrons, protons, helium nuclei and even whole molecules. Because of the particle-wave duality, photons can also be regarded as particles. Both types of radiation, i.e. photon radiation and corpuscle radiation, are therefore summarised under the term particle radiation or simply radiation. In addition, radiation types can be distinguished according to whether they have the capacity to release electrons from the atomic shell of a "victim" atom. Radiation with this ability is called ionising, otherwise non-ionising. Electrically charged particles can ionise atoms directly. Electrically neutral particles can only ionise indirectly by first transferring their energy to another electrically charged particle (cf. Krieger, 2019, p. 17f). The present work refers to radiation from radioactive sources, which is ionising radiation that can consist of both corpuscles and photons. More on this below.

The term "ray" in this context only serves to illustrate the geometric course of the radiation. The space around a radiation source into which the respective particles are emitted is referred to as the radiation field (cf. Krieger, 2019, p. 41f). Relevant integral quantities that play a role in the description of radiation fields are, for example, the total energy and the number of particles. Quantities that change over time, such as the number of particles transported per unit of time, are referred to as flux, in this case as particle flux. If a quantity indicates something per unit area, it is generally referred to as fluence. This refers to the cross-sectional area of a unit circle that is perpendicular to the direction of propagation of the radiation field (Krieger, 2019, p. 44f). A quantity that is variable in terms of both area and time is called flux density (Krieger, 2019, p. 46).

1.1.2 Radioactivity and types of decay

"Atomic nuclei are called radioactive when they spontaneously transition from an unstable state to a more stable configuration or structure under radiation emission and energy release" (Krieger, 2019, p. 91). Such a transition is also called a decay. The stability of a nucleus has to do with its nucleons. Protons repel each other because they have the same electrical charge. The neutrons provide the necessary stability in the nucleus. For an atomic nucleus with a large atomic number to be stable, it usually requires a surplus of neutrons. If the corresponding balance between protons and neutrons is somehow disturbed, this leads to unstable nuclei. The nuclei that are produced in a decay are also called daughter nuclides. These can themselves be radioactive again, which leads to further transformations. If a nucleus transforms several times in this way, it is called a decay chain. With each transformation, the number of nuclear particles can change. We speak of a transformation when changes happen within the atomic nucleus. Such a change could be, for example, an increase or decrease in the number of protons or neutrons in the nucleus. If neither the mass nor the number of protons change during a transition, but the nucleus merely enters a state of lower energy via photon radiation, this is referred to as an isomeric transition (cf. Krieger, 2019, p. 91f).

In principle, a transformation can only take place if the mass of the initial nucleus is greater than the combined mass of the reaction products. The reaction products are the daughter nuclide and, depending on the type of decay, one or more particles or even additional atomic nuclei (cf. Demtröder, 2017, p. 34).

Common types of decay are alpha, beta and gamma decay. The latter is usually only the consequence of previous decays. They are named after the emitted particle that is produced during the decay. It should be mentioned at this point that an atomic nucleus only actually breaks into two pieces in the case of alpha decay. In the case of the other two types of decay, the term transformation may be more appropriate, because here only protons are converted into neutrons or vice versa. In each case, a particle is emitted. In addition to these three important transformations, there are others, such as electron capture, spontaneous nuclear fission, proton decay or internal conversion. All of these transformations can be divided into three groups, depending on which type of interaction plays the decisive role. Alpha decay, for example, has something to do with the strong interaction in the nucleus. The beta transformation and the electron capture concern the weak interaction. In transformations of this kind, nucleons in the nucleus transform into each other. The mass number does not change, which is why this conversion is called isobaric. Gamma transformations occur when the energy content of atomic nuclei changes by emitting a gamma quantum, which is why this type of transformation concerns the electromagnetic interaction. (cf. Krieger, 2019, p. 95ff).

There are attractive forces between the nucleons. If one wants to break down a stable nucleus into its individual nucleons, energy must therefore be applied. This energy is also called the total binding energy of the nucleus (cf. Demtröder, 2017, p.24). The total binding energy of a nucleus is made up of the sum of the binding energies of the individual nucleons, whereby various effects must be taken into account in the calculation. For example, nucleons on the surface of a nucleus are less strongly bound than nucleons within a nucleus (cf. Krieger, 2019, p. 77ff). Total binding energies vary from nucleus to nucleus and depend on various factors. For instance, the number of particles in the nucleus plays a role, as does the ratio of protons and neutrons. If you divide the total binding energy of a nucleus by the number of nucleons in the nucleus, you get the average binding energy per nucleon, which varies between 6 - 9 MeV (cf. Krieger, 2019, p. 81f).

Following these preliminary clarifications, alpha, beta and gamma decay will briefly be described in more detail below.

Alpha decay

The alpha particle α emitted during alpha decay corresponds to a doubly ionised ^{4}He atom, i.e. a helium nucleus. The mass number of the parent nuclide is thus reduced by 4 during this decay, the neutron number by 2. No particle transformations occur. Compared to other light nuclides, alpha particles have a large binding energy. Inside the nucleus, an alpha particle can be excited by the clustering of several alpha particles and muster enough energy to leave the nucleus. A parent nuclide in the ground state can only decay into the ground state of a daughter nuclide if its total mass is higher than that of the daughter nuclide and the alpha particle together. Mostly, alpha decays occur in nuclides with atomic numbers greater than 82, which are all unstable. The spectra of alpha particles are always discrete. The kinetic energy of the alpha particle does not correspond exactly to the decay energy, since part of this is transferred to the daughter nucleus in the form of recoil energy (cf. Krieger, 2019, p. 98ff).

Beta decay

In beta decay, there are two cases that can be distinguished. Due to neutron excess in the nucleus, a neutron of the nucleus can transform into a proton (cf. Krieger, 2019, p. 108). This produces a negative charge in the form of an emitted electron, known as a β^{-}-particle. In addition, an electron antineutrino $\bar{\nu}_e$ is created in the process (cf. Krieger, 2019, p. 102). The atomic number of the atom increases by 1 through this process, while the neutron number decreases by 1 and the mass number remains unchanged. The decay in this case is called β^{-}-decay (cf. Krieger, 2019, p. 108). If instead a neutron deficit prevails in the nucleus, a proton can conversely transform into a neutron under suitable conditions (cf. Krieger, 2019, p. 110). In this case, a positive charge, the positron, is created. It is called a β^{+}-particle and is emitted together with an electron neutrino ν_e (cf. Krieger, 2019, p. 102). In such a transformation, one speaks of β^{+}-decay. The mass number remains unchanged again, while the atomic number decreases by 1 and the neutron number increases by 1. During this decay, a special interaction can also occur when the emitted positron hits a shell electron of the atom. In such a case, both particles annihilate each other, resulting in two gamma quanta (cf. Krieger, 2019, p. 110f). The spectrum of a beta decay is continuous in contrast to the alpha spectrum because the decay energy is distributed in the form of kinetic energy to 2 particles, the corresponding charge particle, and the associated neutrino. The spectrum thus follows a certain continuous distribution (cf. Krieger, 2019, p. 103).

Gamma decay

The gamma particle γ is a high-energy photon emitted by a parent nuclide with a certain excitation energy. Energetically excited nuclei are often the result of previous decays in which corpuscles were emitted, such as the alpha or beta decay described above. The gamma particles or gamma quanta count as electromagnetic radiation. Mass and atomic number therefore remain unchanged in this type of decay and no particle is converted. The mass of the atomic nucleus is only reduced by the mass-energy equivalent of the photon (cf. Krieger, 2019, p. 114f).

1.1.3 Mathematical description and measurands for decays

When a particular radioactive atomic nucleus undergoes a transformation depends on chance and cannot be predicted. Decays are therefore described by statistical means. The decay constant λ corresponds to the probability with which a nucleus changes its state in the time interval dt. Statistical statements can therefore be made for a large number of radioactive nuclei, the accuracy of which increases with the number of nuclides considered. The decrease in a certain number of unstable nuclides N_o with time t can be described by the so-called **decay law** (cf. Krieger, 2019, p. 131f):

$$N(t) = N_0 \cdot e^{-\lambda t}$$

An important parameter in this context is the **activity** *A*, which indicates how many decays or transformations a material undergoes in a certain time interval. More precisely, activity is defined as "the statistical expected value of the quotient of the number of radioactive transformations that take place in a time interval [dN] and the time interval in which these transformations occur [dt]" (Krieger, 2019, p. 128).

$$A = \langle dN/dt \rangle$$

The unit is the becquerel, which corresponds to s^{-1} in SI units. If one forms the quotient of the activity of a substance and the mass of the substance, one obtains the so-called specific activity a, with the unit $Bq\ /\ kg$. In mixtures of non-radioactive and radioactive isotopes, the activity is related to the total mass of the mixture. If the activity is instead related to the volume in which the quotient of the two quantities is formed, the so-called activity concentration c_A is obtained (cf. Krieger, 2019, p. 128ff). If one does not consider the number of conversions, but the number of emitted radiation particles of a sample, one speaks of the **source strength**. Source strength refers to "the expected value of the quotient of the number of radiation particles of a certain type or photons of a certain energy dN_{ex} emitted from a radioactive sample per time interval dt and this time interval dt" (cf. Krieger, 2019, p. 130f). This quantity has units of $1/_s$. Although this is the same unit as becquerel, the unit Bq is avoided to reduce conflating the concepts of activity and source strength. Activity and source strength can assume the same numerical value under certain conditions. For example, there must be no absorption of radiation particles in the radioactive sample itself and only one radiation particle may be produced per decay (cf. Krieger, 2019, p. 130).

The time course of the activity can be specified in analogy to the number of particles as follows:

$$A(t) = A_0 \cdot e^{-\lambda t}$$

A_0 stands for the activity of a substance at time 0. The activity of a sample decreases by half after equal time intervals. This time is called the physical half-life $T_{1/2}$. It can be calculated quickly from the decay constant λ using the following formula:

$$T_{1/2} = \frac{\ln 2}{\lambda}$$

Finally, the so-called **mean lifetime** τ is important, which corresponds to the reciprocal of the decay constant λ. After the duration indicated by τ, the activity has decreased to its (1/e)-th part, which corresponds to about 37% (cf. Krieger, 2019, p. 132f).

A distinction is made between the physical half-life and the biological half-life, which can always be specified if the metabolism and excretion of a substance follow an exponential function. If a radioactive substance enters the body, i.e. if it is incorporated, it decreases through radioactive decay and excretion (cf. Krieger, 2019, p. 137f).

1.1.4 Dose quantities

In order to quantify the negative effects of radiation on the human body, so-called dose quantities are used. On a fundamental level, one speaks of absorbed dose in terms of the energy absorbed per mass element. It has the unit J/kg, which is also referred to as Gray (Gy) in this context. Dose quantities based on this absorbed does are called equivalent doses. In any case, dose quantities always refer to the mass of the irradiated volume. If a dose quantity also takes into account the stochastic risk associated with a person's radiation exposure, it is called a protective quantity. In order to be able to distinguish it from other dose quantities with regard to the unit, the unit of protective dose quantities is called sievert (Sv) (cf. Krieger, 2019, p. 318f).

The dose quantity most frequently mentioned in connection with radiation protection and radiation exposure and at the same time the most complicated dose quantity, is the so-called effects dose E. The effect dose is a protective quantity, refers to the entire human body or individual areas of it and is intended to be a measure of the stochastic risk associated with radiation exposure. It is calculated from the individual organ doses together with the corresponding weighting factors of the individual organs (cf. Krieger, 2019, p. 332). Organ equivalent doses, in turn, are a measure of how a particular radiation affects the corresponding organ of an idealised reference person. Therefore, it is also a protective quantity (cf. Krieger, 2019, p. 329). For this purpose, so-called "radiation weighting factors" are assigned to different types of radiation. These factors are assigned as a lump sum. Photons and electrons receive a weighting factor of 1, protons one of 2 and alpha particles one of 20. The weighting factor of neutrons depends strongly on their energy, whereby the average value corresponds to about 10 (cf. Krieger, 2019. p. 239f).

1.1.5 Natural and artificial radioactivity

Nuclear transformation processes can take place without any external influence. If this is the case with a nuclide, it is generally referred to as natural radioactivity. Radioactivity is called artificial on the other hand, if an energetic imbalance is induced in a nuclide from the outside. This can happen, for example, through nuclear reactions or nuclear fission (cf. Krieger, 2019, p. 91).

1.1.5.1 Natural radioactivity and natural radiation exposure

Natural radioactivity primarily includes two types of nuclides: the so-called primordial radionuclides and the cosmogenic radionuclides. The former are nuclides that were created during the time when the Earth itself was formed. They usually have very long half-lives, which is why they have lasted so long.

Today, there are 30 primordial radionuclides still in existence. Among them are some that directly decay into stable daughter nuclides. The others are components of one of the three natural decay series that still exist today (cf. Krieger, 2019, p. 150). A decay series includes the parent nuclide and all subsequent daughter nuclides. In the case of primordial nuclides, alpha decays, which reduce the mass number by four, and beta decays, which leave the mass number unchanged, predominate. In a natural decay series, the mass numbers of the atomic nuclei that occur in it therefore always differ by exactly four. For this reason, there are only four different natural decay series that are offset from each other. Of these, as mentioned above, three still exist in nature today, all of which end in a stable lead isotope (cf. Krieger, 2019, p. 153). In addition, all natural decay series contain a radioactive radon nuclide, which enters the air we breathe via the soil and thus contributes to the natural radiation exposure of humans (cf. Krieger, 2019, p. 155).

The second category of natural radionuclides - cosmogenic nuclides - are continuously produced in the Earth's atmosphere by cosmic radiation. This ionising radiation has a high energy between 10^9 and 10^{20} eV and constantly hits the Earth's atmosphere. Interactions with air atoms produce a number of high-energy particles, including protons, neutrons and gamma quanta. These directly contribute to the natural radiation exposure of humans. At the same time, they transform atomic nuclei in the atmosphere through interactions, which can turn them into radionuclides. These are then also called cosmogenic nuclides and contribute to radiation exposure as well. (cf. Krieger, 2019, p. 157).

One cosmogenic radioactive nuclide worth mentioning that is produced in this process is the carbon nuclide ^{14}C. It is produced relatively constantly and is therefore found in a certain constant proportion in every living being. The absorption of this nuclide ends when a living being dies. Since the half-life of the nuclide is known precisely, conclusions about its age can be drawn from the remaining proportion in a fossil find. This procedure is called the radiocarbon method (cf. Krieger, 2019, p. 159).

Both categories of radionuclides described above contribute to the natural radiation exposure of humans. A distinction is made between external radiation, by which is meant any radiation that reaches the body from the outside, and internal radiation, which is produced by incorporated radionuclides, i.e. radionuclides that have penetrated the body. If the radiation also comes from nuclides that are found on earth, whether in the ground or in the air, we speak of terrestrial radiation. If the radiation instead stems from outer space, it is called cosmic radiation. In any case, external radiation exposure is very dependent on a person's location (cf. Krieger, 2019, p. 517).

Both certain primordial radionuclides and their decay chains, as well as certain cosmogenic radionuclides, contribute to so-called external terrestrial radiation exposure. However, the contribution of cosmogenic radionuclides is negligible. The largest contribution to external terrestrial radiation exposure is made by primordial radionuclides. In any case, it is primarily gamma radiation that contributes to external radiation exposure (cf. Krieger, 2019, p. 517). Radioactive nuclides are found, for example, in various rocks such as granite, limestone, sandstone and clay slate. In addition, in Germany they are still found in various types of soil, such as grey earth or black earth (cf. Krieger, 2019, p. 518). External radiation exposure is therefore mainly determined by so-called ground radiation. An important primordial radionuclide in this context is the potassium isotope ^{40}K (cf. Krieger, 2019, p. 519). Compared to radiation exposure outdoors, exposure indoors should not be underestimated, as people usually spend long periods of time indoors. Although buildings made of natural stone, concrete or brick provide good shielding against radiation from outside, these materials are themselves sources of radiation. Since one is surrounded by radiation sources on all sides in a building, the local dose rate inside buildings may even be greater than outside (cf. Krieger, 2019, p. 523). If all these factors are taken into account, an average annual effective dose for adults can be estimated, which amounts to 0.84 mSv/yr (cf. Krieger, 2019, p. 524).

Moreover, the radionuclides mentioned above can also enter the body through the consumption of food and drink as well as through inhalation. In this way, they contribute to the so-called internal radiation exposure. The body metabolises radioactive nuclides in the same way as non-radioactive isotopes. Therefore, radioactive potassium, for example, can accumulate throughout the body. However, certain nuclides may also prefer certain organs and thus can be found there more often (cf. Krieger, 2019, p. 535). In addition, radionuclides are found in tobacco plants and therefore ultimately also in cigarette smoke (cf. Krieger, 2019, p. 542).

External cosmic rays are divided into primary radiation coming directly from space, the resulting secondary radiation produced in the Earth's atmosphere, and solar radiation coming from the sun. The primary galactic radiation consists mainly of protons, originates predominantly from our own galaxy, and is probably produced during super novae. It is impossible to estimate exactly where the particles come from, as they are strongly deflected several times in the Earth's magnetic field. In any case, the flow of particles is constant in time and isotropic. As already mentioned above, further, so-called secondary radiation is produced by nuclear reactions of the primary particles with air molecules. This radiation consists mainly of neutrons, whereby the exact composition varies greatly with altitude. At an altitude of 10 km, which is relevant for air traffic, electrons and neutrons predominate. From there, the particle flux decreases sharply down to sea level, where the radiation consists mainly of muons (cf. Krieger, 2019, p. 526ff). Moreover, the radiation emanating from the sun, known as solar radiation or solar wind, consists largely of protons and is absorbed in the upper layers of the atmosphere. It therefore plays more of a role for space travel (cf. Krieger, 2019, p. 529). The average annual effective dose rate due to cosmic radiation is estimated at 380 µSv/yr, taking into account the factors mentioned above. In this, an increase with altitude should also be noted, which is particularly important for intercontinental flights, which are conducted at altitudes between 7 and 12 km (cf. Krieger, 2019, p.

532). Thus, the effective dose rate during such a flight is estimated at 5 - 8 µSv/h (cf. Krieger, 2019, p. 534).

Finally, cosmogenic radionuclides can also enter the body via food and breath and thus contribute to internal radiation exposure. Representatives of these are mainly tritium, beryllium-7, sodium-22 and the already mentioned carbon-14 (cf. Krieger, 2019, p. 536f).

1.1.5.2 Artificial radioactivity and artificial radiation exposure

Artificial radioactivity is essentially about making nuclides radioactive that were previously stable. The aim is to disturb their neutron-proton ratio in such a way that their stability suffers. This can be done in various ways (cf. Krieger, 2019, p. 161).

On the one hand, nuclides can be irradiated with particles. In a process called neutron activation, the particles used are neutrons. Nuclides can capture a neutron in a reaction. The resulting neutron surplus makes them beta-minus active. Typically, this produces nuclides with lifetimes that are long enough for them to be used in technical or medical applications. A representative example of this usage would be the cobalt isotope ^{60}Co, which is used in radiation therapy as well as in materials testing. If a nuclide is instead bombarded with protons, hydrogen isotopes, or alpha particles, there is an excess of protons in the nucleus as a result. This makes the nuclide beta-plus active, however, a very large amount of energy is required to break through the Coulomb barrier (cf. Krieger, 2019, p. 161f).

Transuranium elements in particular are also artificially produced radionuclides. One way to produce them is through repeated neutron capture. This method is called breeding. Alternatively, atoms can also be bombarded with so-called heavy ions. In this way, elements up to atomic number Z = 118 can be produced (cf. Krieger, 2019, p. 162).

Another important area that falls under the term artificial radioactivity is the production of radionuclides through nuclear fission. The artificially produced transuranium nuclei mentioned above usually fission spontaneously. If a certain amount of excitation energy is added to a nucleus to cause it to fission, this is called induced nuclear fission. The necessary energy is usually supplied by bombardment with neutrons. When actinide nuclei fission, 2 - 3 neutrons are produced as fission products in addition to two daughter nuclides. These neutrons can induce further nuclear fissions when they hit corresponding nuclei. Nuclei that emit neutrons during fission are also called neutron sources. Such neutron sources find various applications in medicine and technology. (cf. Krieger, 2019, p. 162ff).

If people are exposed to radiation from artificially produced radionuclides in any form, this is referred to as civilisation-induced or artificial radiation exposure. Utilisation of this type of radiation exposure includes applications in medicine, the production of building materials, the use of fossil fuels and a number of occupational fields. However, careless use of radioactive materials causes radiation damage to people, which has been a major health hazard in the past. One of the best-known examples of this is probably Marie Curie, who experimented with radioactive material without protection. The radiation exposure to which she was exposed led to lethal anaemia in this case. (cf. Krieger, 2019, p. 548f).

In medicine, X-rays play the most significant role in terms of radiation exposure. In 2008, the mean effective dose from medical X-rays corresponded approximately to the mean of natural radiation exposure. On the one hand, attempts are being made to reduce radiation exposure through applications with X-rays, but at the same time, the number of examinations is constantly increasing. Therefore, an increase in the mean effective dose from this area is also to be expected (cf. Krieger, 2019, p. 550). In addition, gamma and beta emitters are also used in medicine for diagnostic or therapeutic purposes. Such radioactive materials are used in a process called scintigraphy, an imaging procedure that can be applied to various organs. Nuclides are used that have a short physical and biological half-life. The effective dose in such examinations is between 4 and up to 20 mSv per examination, depending on the organ, which corresponds to several natural annual doses (cf. Krieger, 2019, p. 554).

In the context of nuclear energy, there are several possibilities for a person to be exposed to radiation. From the extraction and processing of the ores used in nuclear power plants to the intermediate or final storage, a certain amount of radiation exposure is to be expected at each step. This varies greatly depending on the actual area of activity. During the mining of uranium ores and the operation of nuclear reactors, certain amounts of radon gas are also released into the environment, leading to increased radiation exposure of about 40 and 10 µSv/yr, respectively, in the corresponding areas. In the case of nuclear accidents, the local and regional radiation exposure is much greater (cf. Krieger, 2019, p. 561ff).

The use of fossil fuels also contributes to artificial radiation exposure. Hard coal, for example, contains radioactive components that are released during combustion. In large plants, effective doses of up to 0.7 mSv/yr can be expected. In contrast to hard coal, the combustion of other fossil fuels such as natural gas or petroleum does not contribute significantly to artificial radiation exposure (cf. Krieger, 2019, p. 564).

Finally, radioactive substances are also found in more everyday objects. For example, in luminous paints, smoke detectors or certain electronic components. Alpha emitters, whose radiation is absorbed by the devices' casings, are mainly used for this purpose (cf. Krieger, 2019, p. 565).

1.1.6 Attenuation of radiation

The attenuation of radiation is mainly due to two factors: increasing distance and absorption when passing through matter. The "square law of distance" states that the intensity, the number of particles, and the energy density of radiation coming from a point source decrease quadratically with distance. The reason for this lies in the geometry of the matter. Beams of radiation emerging from point sources of radiation diverge, i.e. they run apart. The area through which a certain beam of radiation passes is four times as large at twice the distance (cf. Krieger, 2019, p. 41f).

The effective cross-section plays a decisive role in absorption. If we look at corpuscular radiation, this is understood, in simplified terms, as a measure of the "hit area" that opposes the radiation. The actual object that is hit can be electrons or whole atoms. The effective cross-section per target electron is obtained as the quotient of the atomic effective cross-section and the atomic number Z. A large effective cross-section means a high probability of a radiation particle hitting the object under consideration. In the case of photon radiation, this descriptive image no longer applies because we are dealing with quantum mechanical systems. The cross-section of action is therefore defined by the reaction rate R and the current density of the incident particles j. The former refers to the number of reactions of any kind that take place per unit of time, the latter to the number of incident particles per area and unit of time (cf. Krieger, 2019, p. 47f).

Cross sections usually refer to individual particles. In order to find out how much a specific amount of a certain material attenuates a given radiation, the (linear) attenuation coefficient μ is taken into account. It is closely related to the atomic cross-section. The attenuation coefficient is the product of the atomic cross-section ${}_a\sigma$, the mass density ρ and the atomic number density n_a (cf. Krieger, 2019, p. 48f).

$$\mu = {}_a\sigma \cdot \rho \cdot n_a$$

The atomic number density indicates how many atoms a particular material contains per unit mass and is formed as the quotient of the Avogadro constant N_A and the molar mass M of the material (cf. Krieger, 2019, p. 48).

$$n_a = \frac{N_A}{M}$$

Which interaction processes cause photon radiation to weaken varies greatly and depends on the energy of the photons as well as the atomic number of the material through which they pass. For example, with heavy elements such as tungsten or lead, the photoelectric effect is predominant. With smaller atomic numbers, the Compton effect is important. The attenuation coefficient for photon radiation can also be formed as the sum of the individual interaction coefficients. In addition, the density of the material plays a decisive role in all interactions. Attenuation coefficients for different materials therefore usually do not show a simple course. Since all interactions of photons with matter depend on the density of the material, the quotient of the attenuation coefficient μ and the density ρ of the material is often formed. $\boldsymbol{\mu/\rho}$ is then referred to as the mass attenuation coefficient (cf. Krieger, 2019, p. 199ff).

1.2 Student ideas about radiation

According to Schecker and Duit (2018), when teaching a topic in physics, not only the subject content is relevant, but also the learning prerequisites of the students (p. 2). How students think about a topic based on previous experience is very central. It is therefore worthwhile to take a look at the so-called student conceptions or student ideas, and also at conceptions that circulate in the general public. Especially when it comes to radioactivity, it is interesting to look at any previous experiences students might have had with the subject. According to Hopf and Schecker (2018), children typically have no personal experience with radioactivity, which is mainly due to the fact that small doses of nuclear radiation cannot be perceived by the human body in a noticeable way. At the same time, this does not mean that children have no experience with the topic at all. The media in particular contribute to children coming into contact with information on radioactivity (p. 231). Linsje et al. (1990) compared media reports on the Chernobyl disaster with statements and ideas of pupils who had not yet learned anything about radioactivity at school. In the resulting data, the similarity between the two factors was very clear (p. 77). Pupils therefore can be said to have ideas about radioactivity even before they are taught about it. These are influenced by the media. In the following section, I would like to summarize all student ideas or naïve ideas from the literature I was able to find that are relevant to this study.

First off, Hopf and Schecker (2018) make it clearthat few students are aware of natural radiation exposure. Many assume that any exposure to nuclear radiation is man-made. According to the authors, one reason for this could be that artificial radiation exposure is so strongly dominated by X-rays. These man-made X-rays, however, are produced by accelerating electrons (not by using radioactive sources). As a result, the share of natural radiation exposure from radioactive sources in the total radiation exposure, both natural and artificial, appears low. However, if one compares only the radiation from different radioactive sources (that is, leaving out the X-ray machines), the impression would be different. If one wants to make the significance of natural radiation from radioactive sources clear to pupils, one should make it clear that man-made x-rays are not radioactive in origin. In addition, the authors state that many students would assume that radioactivity is exclusively dangerous or deadly (cf. p. 231f).

The second relevant idea mentioned by Hopf and Schecker (2018) refers to a term often used in the German language that leads to confusion. The term is "radioaktive Strahlung", literally meaning “radioactive radiation”, which is used to refer to nuclear radiation or radiation from radioactive materials. Students would confuse radiation with the transport of radioactive material. This would subsequently lead to learners assuming that materials exposed to ionising radiation become radioactive themselves and emit radiation. In this conception, a person who has been exposed to radiation is also contaminated by it, which can lead to anxiety. Also in this idea, radiation is deposited and stored in the body (cf. p. 232).

Henriksen (1996) also finds in his study that insufficient distinction is made between radioactivity and radioactive material. Likewise, they find the idea of the deposition of radioactivity in the body and goes into more detail on the concept of absorption. They use the analogy of a sponge, which absorbs water and then releases it again. Thus, in this conception, an object can also release absorbed radiation again (cf. p. 193). Furthermore, Henriksen (1996) also states that radioactivity is sometimes confused with other environmental hazards. Also, radioactivity has been more or less equated with other terms such as electric or magnetic fields. Furthermore, only 62% of the respondents stated that they thought radioactive materials were present in their homes (cf. p. 194). Finally, Henriksen (1996) also talks about the media, which according to them, also played a role in shaping the result of their study. Certain false statements made by the people studied could be attributed directly to statements made by the press (cf. p. 195).

All these ideas were also found by Prather and Harrington (2001) in a study, in which college students took part. Specifically, they asked a question about whether a strawberry itself becomes radioactive when irradiated by a radioactive source. Most participants stated directly or indirectly that the strawberry does become radioactive in this case. In addition, some also stated that objects can become radioactive by absorbing radiation (cf. p. 90). Prather and Harrington (2001) also point out that although students would sometimes give correct answers to certain questions, they might give incorrect reasons for their answer. They believe that a good understanding of what happens in the atom during the decay process is needed. (cf. p .92).

Beyond that, Prather (2005) also cites some student perceptions specifically related to half-life. For example, some students would say that the mass or volume of a radioactive substance is halved after a half-life has elapsed (cf. p. 353). In addition, some students would state that after a half-life, only half of a substance would be radioactive (cf. Prather, 2005, p. 352).

Hull and Hopf (2020) share a similar opinion stating that: "... educators can best help students make sense of nuclear radioactivity despite its stochastic nature by focusing on the properties of the individual nucleus, ..." (p. 30). In their study, Hull and Hopf (2020) looked more closely at how students deal with those aspects of radioactivity that only occur when several particles are observed over a longer period of time. This includes, above all, the half-life. The aim of the study was to clarify whether student ideas about radioactivity or emergence are stable or fluid. In this context, fluid student ideas refer to conceptions that are very context-dependent and can therefore lead to the wrong solution in some situations, but to the right solution in others (cf. p. 22). Although the results of this study cannot be generalised, Hull and Hopf (2020) show initial data that student ideas about radioactivity tend to be fluid (cf. p. 30). Students' justifications would thus depend very much on the context and can differ from situation to situation.

Providing another research perspective, Janksy (2019) conducted a study on student ideas that revolved around radioactivity and chance with regard to radioactive decay. In this study, the author was able to discern several different ideas. Firstly, students believed that an atomic nucleus decays until there is nothing left of it. According to the students, the nucleus breaks either into individual parts or decays continuously. Secondly, the idea that radioactivity is some kind of substance was identified in pupils' answers. Finally, several ideas about half-life were found in the study. The above-mentioned idea that half of a substance has disappeared after a half-life has elapsed is also found in this study. In addition, some pupils stated that the half-life is the point in time with the greatest probability of an atomic nucleus to decay. Other ideas about the half-life were also found in isolated cases (cf. p. 184f). In this study, people with an average age of 17 to 18 years were interviewed. It should be emphasised that the respondents included pupils who had taken an advanced physics course. Nevertheless, some of the same ideas as just described emerged (cf. Jansky, 2019, p. 180).

Finally, I would like to mention a study by Eijkelhof et al. (1990). They asked experts what perceptions they regularly encounter and to what extent these represent an obstacle when it comes to correctly assessing the risks of radioactivity (cf. p. 186f). Some of the ideas that the experts stated coincide with those that have already been presented. The most common five were “Radioactive substances and radiation are not distinguished”, “activity and radiation dose are not distinguished”, “a radioactive substance is always dangerous”, “irradiation of food results in radioactive food”, and “a nuclear power station can explode like a nuclear bomb” (Eijkelhof et al., 1990, p. 188). These ideas can be found in one form or another in studies conducted by pupils or students (see above).

It seems as if there are many ideas about what radioactivity is, what effects it has and whether it is dangerous or not. When teaching the topic of radioactivity in school or designing a teaching concept, it is certainly an advantage to be aware of these ideas and keep them in mind.

1.3 The importance of teaching about radioactivity

Addressing ideas about radioactivity is not only relevant for the school context. As shown above, among other things, studies have found that the public has strong beliefs about the risk of radioactivity that are often negative and differ from experts’ views (cf. Slovic, 1996, p. 1). Eijkelhof comes to the same conclusion in saying that “the public has a lot of knowledge and beliefs about radiation risks which in the expert’s view may be wrong and inconsistent, but which are not so perceived by the audience” (1996, p. 277). As a consequence, Slovic suggests better information in general and also better education (1996, p.165).

Extending on this notion, Henriksen (1996) also mentions three reasons why the public should know something about radioactivity: 1. the pragmatic reason. This means that people should be able to protect themselves from radiation and at the same time not be particularly afraid of it. 2. the democratic reason. People should be able to form an informed opinion in political discussions that revolve around radioactivity. 3. the educational reason. This reason is about the individual learning more about the world around him or her (cf. p. 191). This implies that learning about radioactivity has a certain claim of generality and is thus relevant to a society and the individuals in it.

Eijkelhof also raises the concern that “… most students in Europe are not in school to prepare themselves for future studies in the physical sciences, but to be educated for life in modern society” (1996, p. 273). There is clearly a need for better education regarding radioactivity that is suitable for any student, no matter their career choices afterwards. As mentioned above, knowledge about radioactivity is not only relevant for people working in specific professions, but for every individual in a society.

1.4 What is HEC?

Hypothesis-Experiment Class is a special method of teaching physics developed by Kiyonobu Itakura. The name already indicates that hypotheses and experiments, as well as the relationship between the two, play a central role. The starting point for this teaching concept is the examination of "mental recognition", which is described as "investigated perception" (cf. Itakura, 2019, p. 4) and "requires the perceiver to conduct some action with intent to confirm something about the object" (cf. Itakura, 2019, p. 3). From this, “experiment” is defined as a special form of observation that involves a specific intention to investigate (cf. Itakura, 2019, p. 4). Any action is therefore only an experiment if there is also an intention on the part of the person carrying out the action to find something out. A “hypothesis”, on the other hand, is described as a special form of prediction that refers to a "more-or-less general theory or principle". It differs from a simple expectation, which refers only to a single event. Also, progress in science would takes place through a cyclical process in which hypotheses and experiments alternate, which describes the relationship of the two aspects with each other (cf. Itakura, 2019, p. 9).

According to Itakura (2019), there are some contrasts to this in educational practice. Experimental facts are presented first and then a theory is derived from an inductive process. Theories are literally imposed on students, which leads to their rejection. Additionally, experiments are also frequently used incorrectly. Replicating a natural phenomenon is not the same as conducting an experiment (cf. p. 10f).

Consistent with these definitions, when students conduct an “experiment” in HEC lessons, they should maintain an intention to investigate. They should also make a concrete prediction beforehand about what the hypothesis should reflect. Therefore, students should go through a process of prediction and experiment. At the end, students should be able to make correct predictions by applying a certain concept or rule (cf. Itakura, 2019, p. 12f).

Moreover, according to Itakura (2019), science is a social process. A theory becomes meaningful only when it has been accepted by society, and when there is sufficient evidence for it to be accepted by each hypothetical individual. In this sense, a theory should also be presented to students in a way that they can accept it (cf. p. 13f). Ideally, students should build their own understanding of a theory. Any naïve ideas students may have should be considered. When presenting new theories, especially at school, these should be considered beforehand (cf. Itakura, 2019, p. 16f).

HEC aims to realise the above aspects in the school context. This is implemented by running through a certain cycle of expectations and experiments. In this way, the social aspects of science are also included, as well as making predictions and doing experiments. The concrete procedure is as follows:

Learners are asked a question. They choose one possible answer from several options, thus showing their expectations. All answers are collected. Then individual learners are invited to explain why they have chosen a particular answer. At the end of this discussion, the learners are given the opportunity to change their answer. Finally, an experiment decides objectively which answer option was the correct one. The result is recorded by the students and the next question is asked. Everything is summarised in a so-called Classbook. This also contains explanations that are read together with the learners. The role of the teacher is a rather passive one. They are seen more as facilitators of the learning process (cf. Itakura, 2019, p. 25ff).

Through this approach, the above aspects are implemented. Theories are presented in a way that allows learners to build their own understanding and access. The lively discussion in the group turns it into a social process in which the opinion of the majority is given weight. Hypotheses are developed together. Experiments are only carried out if there is a real intention to find out something concrete.

HEC sets itself three specific goals:

Goal 1: "Make sure each and every student gains the ability to use the central theory or concept" (Itakura, 2019, p. 20). This means that more than 90% of the students should be able to predict the correct answer to the last problem in a Classbook. The last problem is designed accordingly. In addition, it is advised to give a test on the material covered one to two weeks after completing a Classbook (cf. Itakura, 2019, p. 20).

Goal 2: "Structure the class so that most students report that they like both science and these science classes" (Itakura, 2019, p. 21). The explicit goal is that over 50% of learners report that they "liked" or "really liked" science and HEC and that, with a few exceptions, no one reports that they "disliked" HEC. This is asked directly after the HEC lesson (cf. Itakura, 2019, p. 21).

Goal 3: "Make all necessary preparations so that any teacher sufficiently passionate about education, not just special veteran teachers, will be able to teach this type of class (Itakura, 2019, p. 22)". In short, this means that HEC lessons should be easy to carry out with the help of the Classbook by any teacher, regardless of their personal experience.

Itakura claims that all three goals are realistic. Although he describes a study in which 150 school classes and over 20 teachers participated no citation to a published work is provided (cf. Itakura, 2019, p. 23). Furthermore, it is claimed that "almost all classes run with the Classbooks developed ... have demonstrated good reproducibility of class performance, independent of teachers and students" (Itakura, 2019, p. 25). Again, no citation could be found.

1.5 The traditional way of teaching radioactivity

The present study looks at a Classbook that deals with the topic of radioactivity and is entitled "The Radiation Around Us". It is interesting to look at how the Classbook differs from the usual or traditional way of teaching radioactivity. Although individual teachers may have their own way of teaching this topic, there are probably some similar tendencies, at least in Europe. "The usual approach is to start with the structure of the atom and the nucleus, followed by concepts such as half-life, activity, nuclear fission and fusion. Towards the end of the series of lessons some applications are usually mentioned, such as carbon dating, irradiation, and the nuclear reactor. Safety issues are dealt with only superficially." (cf. Eijkelhof, 1996, p. 273).

This assessment by Eijkelhof essentially coincides with those of Hopf and Schecker (cf. 2021, p. 373f). According to these authors, too, the types of radiation, their properties, the decay law, biological effects and technical applications play the central role in traditional physics teaching. The atomic structure discussed at the beginning is framed with historical aspects, whereby the “Rutherford experiment”, for example, is examined in more detail. Certain technical terms, such as the nucleon number, are also introduced. As far as the actual radiation is concerned, it mainly deals with how and with which devices it can be detected. Hopf and Schecker (2021) mention that radioactive samples and measuring instruments have increasingly disappeared from schools, which is mainly due to various regulations. Radiation is therefore mainly analysed qualitatively, focusing on aspects such as absorption, range and deflection. Differences between alpha, beta and gamma radiation are often summarised in tables. In addition to the types of radiation and their properties, the half-life also plays a major role in traditional physics teaching. This is taught with demonstration experiments or analogy experiments. In connection with this, the C14-method is also mentioned as a noteworthy application. Radiation protection is another major topic. Various quantities are introduced, and both the usefulness and the negative, biological effects of radiation are discussed. Finally, nuclear fission and especially its applications are discussed (cf. p. 373f).

If one looks at various Austrian textbooks intended for the 8th grade, this largely coincides with the statements of Eijkelhof and Hopf and Schecker. Here, historical contexts are also often found at the beginning of the chapter (e.g., Mašin & Grois, 2020; Fürnstahl et al., 2020; Apolin, 2018) This usual approach is also particularly suitable in Austria, because it objectively covers the Austrian curriculum (for Gymnasium and Mittelschule as well), which includes the following aspects (see Curriculum for AHS-Unterstufe):

- gain insights into changes in the atomic nucleus as the cause of "radioactivity" (properties of alpha, beta and gamma rays);

- recognise radioactive decay as a constantly occurring process;

- understand the basic processes of energy conversion in the sun, stars, and nuclear reactions (nuclear fusion, nuclear fission).

According to Hopf and Schecker (2021), traditional radioactivity teaching has not been studied a lot, at least in German-speaking countries (p. 374).

The Classbook "The Radiation Around Us" is very different from the usual path. Nuclear decay, which is usually found at the beginning of the chapter on radiation in traditional teaching materials plays a subordinate role in the Classbook. In general, technical terms are used rather sparingly in the Classbook (hence the choice of the word “Radiation” in the title instead of “Radioactivity”) and are often only roughly explained or an overview is given. The focus is on the fact that radioactivity is a natural, constantly occurring phenomenon. It thus completely covers the second point in the Austrian curriculum, the others only marginally. If the Classbook is to be used in an Austrian teaching context, certain additional topics not covered by the Classbook would have to be added to the instructional sequence in accordance with the Austrian curriculum.

1.6 Existing curricula for radioactivity

Besides the traditional way of teaching radioactivity, it is also interesting to take a look at other teaching concepts for this topic. According to Hopf and Schecker (2021), for some topics of modern physics, including radioactivity, there are not many teaching concepts, and the ones which exist are only sparingly researched. For some topics, there are no teaching concepts at all (cf. p. 370f). Specifically on radioactivity, Hopf and Schecker (2021) present a teaching concept, which primarily addresses the half-life and was developed and evaluated by Jansky (2019). In this method, the concept of half-life is approached by using probability theory. Using everyday objects, the aim is to achieve an understanding of the fact that individual events can occur by chance but can still be described statistically. To begin with, several dice or drawing pins are thrown, and later rubber poppers are examined in order to introduce a probability per time. The behaviour of nuclei is then compared with that of the poppers. Both show a random change of state, whose probability depends on the time. Finally, the half-life is introduced, again using the everyday objects mentioned above (cf. Hopf and Schecker, 2021, p.376ff).

According to Hopf and Schecker (2021), other aspects of radioactivity, such as the different types of radiation or technical applications, are not addressed by this teaching concept. At the same time, the authors believe that the concept can be complemented by traditional teaching (cf. p. 376). So, like "The Radiation Around Us", Jansky’s teaching concept does not cover the entire curriculum. It is interesting that the two concepts would even complement each other, as the half-life is not mentioned in "The Radiation Around Us" and is not given any special attention, whereas it is the main focus in Jansky's concept. Other concepts on radioactivity are not mentioned by Hopf and Schecker (2021).

Millar et al. (1990) talk about an alternative approach to teaching radioactivity. They outline a long teaching sequence that spans ten 70-minute units and is divided into four blocks: 1. Phenomenological orientation. 2. Qualitative macroscopic treatment. 3. Quantitative macroscopic treatment. 4. Microscopic treatment. The individual steps are roughly outlined. In the first step, one should establish references to everyday life. In addition, different types of radiation, not only ionising radiation, can be examined. In the second step, student ideas should be taken into account. Differences between open and closed sources or between irradiation and contamination should be underlined. In addition, the different types of ionising radiation and their properties should be addressed in this step. Again, comparisons can be made with other types of radiation, such as light. At the end of this step, students should be able to use the words "radioactive matter" and "radiation" correctly in scientific terms. In the third step, terms such as activity, dose equivalent and half-life should be established. This is to be achieved by asking how to define and measure different quantities that have something to do with radioactivity. Here, one tries to carry out concrete measurements for the first time and tries to put various quantities into numbers. In the fourth and last step, the atom itself is examined more closely. One tries to clarify what happens in the atom when it emits ionising radiation. In addition, effects of ionising radiation are discussed in this step, such as its connection with cancer (cf. p. 340f). After the four steps have been presented, Miller et al. (1990) specifically address the relevance to everyday life. They list various examples of how a reference to reality which has a certain degree of relevance for the pupils can be established in each of the four steps, which has a certain degree of relevance for the pupils. For example, in the first step one could talk about protection against radiation and in the second step about sterilisation of medical products. In the third step, the problems with the disposal of radioactive waste could be discussed, and in the fourth step, the problem of radon in buildings (cf. p. 341).

In contrast to Jansky's teaching concept on half-life or "The Radiation Around Us" described above, this teaching proposal by Miller et al. (1990) is not dedicated to one particular aspect of radioactivity. Instead, they propose a concept that tries to cover all topics and at the same time differs from the traditional way. It is based on research into student ideas on radioactivity (cf. p. 342). The descriptions of the actual concept or the individual steps are very short and rather rough. Therefore, it is not clear to me how the individual aspects of radioactivity are to be taught in concrete terms. Only the order in which they should be taught is apparent to me. Miller et al. (1990) also state that the concept they present should be regarded as a kind of basic framework. The individual steps would still need to be fleshed out and then researched in more detail (cf. p. 342).

Another concept that deals with radioactivity in the classroom is the "Inquiry into Radioacitivity (IiR)" project or "Radiation By Inquiry (RBI)" project. The aim of this project is to develop different tools and techniques to help students to become radiation literate. The target groups are high school students and non-science college students. A person who is radiation literate should know three things: 1) that ionising radiation consists of small particles emanating from a specific source and moving at high speed, 2) that the sources of ionising radiation are unstable atomic nuclei and special machines, such as X-ray machines, 3) what effect ionising radiation has on molecules, cells and the human body, how to protect against it and how it is used in medicine (cf. Johnson and Maidl, 2014, p. 1). DeKay and Maidl (2012) also speak of the scientific view in this context (cf. p.1). The project tries to oppose the so-called "undifferentiated view". This in turn includes the following three aspects: "radiation is matter-like, there is no difference between radiation and radioactive, and radiation can be transferred to other objects making them radioactive" (cf. Johnson and Maidl, 2014, p. 2). Students work in small groups in this project. Their approach is guided by the materials. They discuss key questions and carry out different investigations. They also work in practical laboratories or use different simulations. Concrete activities include, for example, irradiating objects and then examining them to see that contamination has not occurred, counter to common student expectations (cf. Dekay and Maidl, 2012, p. 3f). The materials follow a certain cycle. In each cycle, the pupils' ideas are first elicited. Then students are guided to develop new ideas. At the end of a cycle, a certain resolution takes place. There are four such cycles in total: 1) Basic Characteristics of Radiation, 2) Atoms and Radioactivity, 3) Interaction of Radiation with Matter, 4) Nuclear Power and Nuclear Waste (cf. Hafele, 2012, p. 3).

DeKay and Maidl (2012) examined 35 students who learned with the IiR project materials (cf. p. 3). The use of the materials helped students to move from the "undifferentiated view" to the scientific view. 91% of the participants had the undifferentiated view before the intervention, afterwards only less than 5%. In contrast, less than 5% had the scientific view before and 67% after. A certain percentage of participants could not differentiate all three aspects of the undifferentiated view, which were therefore classified as partly differentiated. The research also found that the transition to the scientific view was difficult for the students and some occasionally reverted to old ideas in the process. The idea that objects themselves become radioactive when irradiated by a radioactive source proved to be particularly persistent. In addition, understanding how ionising radiation interacts with matter was challenging (cf. DeKay and Maidl, 2012, p. 6f).

Inquiry into Radiation differs from The Radiation Around Us, or from HEC in general, mainly in the way it is conducted. In the former approach, the students work in groups and carry out experiments themselves. In the latter, the pupils are always in the class and discussions are also held with the whole plenary. At the same time, both concepts are similar in one respect: DeKay and Maidl (2012) write that in Inquiry into Radiation "the students and the class as a whole are the primary source of ideas and explanations - which places a higher burden on sense-making by students". (cf. p. 3). In HEC something similar is the case. There too, as mentioned above, the teacher has a passive role. The sense-making in HEC is also done by the students themselves. The fact that both concepts work in certain cycles is also similar. However, they differ in length. In HEC, a cycle consists of only one question, one experiment and possibly a little information for the pupils. Several cycles are embedded in one topic. In IiR, on the other hand, one cycle already covers one topic.

Another teaching approach (which heavily inspired IiR) that puts the understanding process in the hands of the students was developed by Klaasen (1995) and called the "Problem Posing Approach to Teaching the Topic of Radioactivity". Klaasen (1995) refers to the rationality of the students as the driving force. They describe the didactic structure of his approach not as a series of tasks, but as a scenario in which, among other things, the connections between individual tasks are described. In addition, a textbook is a central component of their approach. However, this is not to be understood as a conventional textbook. Rather, it contains a special arrangement of tasks and worksheets (cf. p. 141). Likewise, the textbook hardly contains explanations of theory or instructions for experiments. Instead, the tasks in it are designed to encourage students to think for themselves. The textbook is thus designed to give students more responsibility and control over their own learning process. The role of the teacher in this approach, as in IiR and HEC, is more passive. The teacher in the Problem Posing Approach is responsible for managing the whole process. Also in relation to the students, the teacher should not actively try to get something out of the students, but rather passively receive their comments and react appropriately to them (cf. Klaasen, 1995, p. 143f).

The Problem Posing Approach and HEC are particularly similar in the use of a book that contains tasks for students to think for themselves. The textbook presented by Klassen (1995) covers more topics than the Classbook “The Radiation Around Us” (cf. p. 142f). In addition, the Problem Posing Approach also involves group work, which is never the case in HEC (e.g., Klaasen, 1995, p. 146).

2 Methodology

2.1 Research interest

Studies regarding the successfulness in terms of the three self-imposed goals of HEC Classbooks for The Radiation Around Us could not be found. Therefore, this master's thesis is a study aiming to investigate the extent to which these goals can be achieved with the Classbook "The Radiation Around Us" in Austrian schools, focusing on goal 1 and goal 2. The aim is to find out whether there is a significant increase in understanding when pupils are taught with the Classbook, in comparison with pupils who are not taught with the Classbook. In addition, I want to find out whether students like being taught with the Classbook and whether there is also a difference here compared to students who are not taught with the Classbook. I predict that pupils have a better understanding of radioactivity in relation to the fact that it is a natural, constantly occurring phenomenon, in comparison to those who are not taught with the Classbook. I also predict that pupils will enjoy learning with the Classbook.

2.1.1 Research questions

Therefore, the research questions are as follows:

- Do students who have used the Classbook have a significantly better understanding of radioactivity and specifically that radioactivity is a naturally occurring phenomenon than students who have not been taught with the Classbook?

- Do the students like the lessons with the Classbook and is there a difference to classes where the Classbook is not used (traditional physics lessons)?

2.2 Procedure of the study

The present study is a summative evaluation of the presented teaching concept, in particular of the Classbook "The Radiation Around Us". Reinders et al. (2011) distinguish between a summative and a formative evaluation. While a formative evaluation should serve to further develop something, a summative evaluation has a summarising and concluding character (cf. p. 134).

In this sense, the aim of this study was to find out what effect the Classbook has on the aspects mentioned above and to summarise any results. For this purpose, quantitative data was collected with a questionnaire. In order to make learning outcomes visible, the study follows the pre-test-post-test approach. Thus, primary data were collected before and after students were taught about radioactivity. This data was also compared with a control group who were taught about radioactivity in a traditional way but did not use the Classbook.

The study took place in four schools with 12 classes of 13- to 14-year-old students. The questionnaire was administered in both groups once before the lesson on radioactivity and once within the next two weeks after the last lesson on this topic. The questionnaire was given to the learners online via a link from the teacher. Before the questionnaire could be handed out to the students, a consent form had to be signed by the parents. In this document, the parents were informed about the study and in particular about what happens with the data. The quantitative data was collected between March 2022 and June 2022.

In order to better compare the experimental and control groups, interviews were also conducted with all teachers participating in the study, if they agreed to do so. In addition, a short questionnaire was given to the teachers to determine how long the subject of radioactivity or different chapters of it had been taught. This took place between July 2022 and August 2022.

2.3 Research design

As mentioned above, this study is an evaluation study. As far as the study design is concerned, I have chosen a field study. In the selection, considerations of internal and external validity of a study were taken into account. Hussy et al. (2013) define high internal validity as "when changes in a dependent variable can be attributed exclusively to variation in the independent variable". External validity, on the other hand, captures "the extent to which the results can be generalised (cf. p. 137). In other words, it refers to how easily any results found can be transferred to other, comparable situations. In my study, a high internal validity would therefore guarantee that, for example, any greater learning gains in the experimental group could actually be attributed to the use of the Classbook. A high external validity would guarantee that such an effect could also be reproduced in other classes. An overarching goal of this study is to make the use of HEC in schools more attractive, provided that truly learning-enhancing effects can be demonstrated. For this reason, I would like to keep external validity high, albeit at the expense of internal validity. Wilhelm and Hopf (2014) also believe that the results of science didactic research are often only reflected in classroom practice to a limited extent. According to school practitioners, proposals made by education researchers would often be impractical, unrealistic, or not relevant to the curriculum (cf. p. 32). In their design-based research approach, Wilhelm and Hopf (2014) deliberately do not rely on laboratory studies because they assume that when considering a learning environment, one cannot examine individual influencing factors in isolation. Instead, they speak of a "complex interplay of different influencing factors". According to the authors, if one examines a teaching approach and suspects effectiveness, they should "not be limited by the strict control of boundary conditions" (cf. p. 33). To me, this is another argument in favour of a field study.

In order to keep internal validity as high as possible, results from the experimental group are compared with a control group (cf. Reinders et al., 2011, p. 139). According to them, the inclusion of a control group is important in order to determine any effect of a treatment or intervention (cf. p. 139). The groups differed as follows:

The Classbook "The Radiation Around Us" is used in experimental groups and not in control groups. Because the Classbook does not cover the entire Austrian curriculum, the following procedure was decided on for the experimental groups: teachers start the topic of radioactivity with the Classbook. Once they are done with it, they finish teaching the other items listed in the curriculum at their own discretion. In control groups, teachers teach radioactivity without the Classbook, drawing on the same materials as they usually do.

The teachers of the experimental group have received a teacher's guide to guide them in the implementation of the Classbook. The teacher's guide is not specific to TRAU and is provided by ASHEC to HEC practitioners [https://www.kasetsu.org/4_english/aboutus.html]. In addition to translating this guide into German, only minor modifications were made, since the Classbook was not printed out, but rather the teachers received a digital copy. The Classbook can be viewed under this link: https://u.kyoto-u.jp/hec-trau-d. The official version was slightly modified for this study. For logistical reasons, it was not possible to build a cloud chamber in each class. This part was therefore omitted. The questions related to the observations in the cloud chamber were answered with the help of videos. These videos were given to the teachers with the Classbook. Apart from that, the teachers of the experimental group were not instructed in detail. It should be mentioned here that previous experience with the teaching concept varied between the teachers. A more detailed analysis of this can be found in the master's thesis of Wintersteller (in press).

Reinders et al. (2011) write that experimental and control groups should ideally only differ in terms of the treatment or intervention applied. Any differences measured could otherwise be due to other effects. At the same time, it is also acknowledged that evaluation studies often take place in settings that are difficult to control and that this condition is therefore difficult to fulfil (cf. p. 139).

Also, it was not entirely possible to control the two groups in the course of this study. Therefore, efforts were made to better compare the groups in order to increase internal validity. As mentioned above, interviews were conducted with the teachers of the control group involved in the study. The aim was to draw a rough picture of what their teaching looked like. A questionnaire was created to gain insight into how much time individual teachers spent on the topic as a whole and on individual parts of it. This was given out to all teachers who participated in the study. Interview guidelines and questionnaires for teachers can be found in the appendix. In a study by Burde (2018), a teaching concept was evaluated in a similar way as here. There, too, the learning progress of an experimental group using a new teaching concept was compared with that of a control group taught in the traditional way (cf. p. 180). Burde (2018) discusses some ways of getting a picture of what the teaching looked like in each group or class. The teachers involved in the study were asked to keep a teaching diary throughout the investigation. Due to the large number of participants, other options such as observation or videography of lessons were not used (cf. p. 183f). All of these possibilities place different demands on a teacher participating in the study. Wilhelm and Hopf (2014) write that there are several factors to consider when working with teachers. One of them is the fact that everyday school life is already very time-consuming for teachers (cf. p. 36). I therefore wanted to keep the effort for teachers participating in the study as low as possible. That is why I decided to use a questionnaire and interviews and conduct the interviews during the summer holidays.

2.4 Student-questionnaire

2.4.1 General structure

For data collection, a questionnaire was used that consists of several parts and changes slightly from pre-test to intermediate test to post-test. The questionnaire starts with an explanation at the beginning. Although consent was obtained from the parents beforehand, it is explained here again what happens to the collected data and reference is made to the voluntary nature of the data. This happens in the pre-, intermediate- and post-test, although the introductory words change somewhat. After that, there is a section investigating the students' interests, followed by a part aiming to discern the students' self-concept. The next part is about students' conceptual understanding, as well as a little bit about their personal attitude or emotion towards radioactivity. Finally, a personal code and a class code are entered by the students, and a short acknowledgement follows at the end. In the following section, the individual parts of the questionnaire are presented in more detail.

2.4.2 Conceptual understanding

To determine any increase in conceptual understanding, a questionnaire was prepared containing multiple choice items from the questionnaire developed by Holzinger (2022). This questionnaire was aimed at people who have already graduated from school. Therefore, I removed the items of the questionnaire that were irrelevant to my study. Mainly, all items that refer to the conceptual understanding of the respondents were used. These questions cover several aspects of radioactivity. Since not all questions were relevant to the contents of the Classbook and so that the questionnaire would not become too long for my study, not all items were used here either. In particular, all questions relating to the half-life were excluded. The reason for this is that the half-life is not mentioned in the Classbook "The Radiation Around Us". Any increase in understanding in this area in the experimental group could therefore not, or at least not directly, be attributed to the use of the Classbook, but would rather be due to the subsequent lessons. Other items, such as the one that refers to the types of radiation, are not discussed in more detail in the Classbook, but these terms are at least mentioned in the Classbook. This part of the questionnaire can be seen as a test in which the students receive points for correct answers. They result in a test score that can be compared between pre- and post-test as well as between the two groups. Here are two examples from the conceptual understanding or knowledge test. For these two items, I expected a large increase in the experimental groups, because they are directly addressed by the Classbook.

Q1	Du siehst nun eine Reihe von Begriffen. Kreuze alle Begriffe an, bei welchen ein Messgerät für Radioaktivität klicken würde. Mehrfachantwort möglich. Smartphone; Umgebungsluft; Ein Schulkind; Betonwände in der Rezeption eines Krankenhauses; Betonwände eines Krankenhausraumes, in dem Röntgenaufnahmen gemacht wurden; Betonwände eines Schulgebäudes; Zigarettenrauch; Sonnenlicht; Popcorn aus der Microwelle; Banane
Q10	Wie verändert sich die Strahlung in einem Flugzeug im Vergleich zum Boden? A) Sie nimmt mit der Höhe zu B) Sie nimmt mit der Höhe ab C) Sie bleibt gleich D) Keine der Antwortmöglichkeiten ist richtig

For any given question, either one or more answers can be correct. As a rule, one point is awarded if all correct answer options and no incorrect answer options are ticked. Partial points are only awarded for question 3, which refers to the types of radiation. The scoring rubric is thus consistent with the one used by the author of the questionnaire (see Holzinger, 2022). The question referring to the emotion regarding radioactivity was included but not scored.

2.4.3 Interest and self-concept

I measured the variable of student enjoyment of learning with the Classbook in three ways. Two of these metrics were 1) a set of items probing respondent interest and 2) a set of items probing self-concept. For this purpose, the corresponding items from the IPN interest study are used (Hoffmann et al., 1998). These items are administered at the same time as the above conceptual questionnaire. In the experimental groups, an additional questionnaire containing only the items on interest and self-concept is administered immediately after the Classbook lessons have ended, the intermediate test. This should make it possible to determine more precisely what effect the Classbook has on these two aspects, as opposed to the rest of the lessons on radioactivity. All items on interest and self-concept follow a 5-rank scale. The third metric I used is a question that is typically administered at the end of HEC Classbooks, referred to as the "tanodo" (literally, "degree of enjoyment") question. In this question, students are asked directly whether they liked the lessons in the last few days.

2.4.4 Class- and personal code

In the course of investigation, students enter a class code at the end of the survey to take class effects into account later. In addition, they enter a personal code which enables me to analyse changes of individuals. The code is created by providing 3 pieces of information from the participant. The selection of the corresponding questions was based on Pöge (2008). They emphasise that the education researchers should strive to keep these codes as simple as possible (cf. p. 69). The questions were therefore essentially taken from the code sheet presented by Pöge (2008, p. 5) and only slightly modified to ensure this simplicity. As far as the length of the codes is concerned, Pöge (2008) describes a procedure in which more and more errors are allowed when matching the codes of two questionnaires. If two questionnaires with supposedly identical codes are found, the handwriting of both sheets is checked to determine whether they really belong together. If obvious differences in the handwriting are found during these so-called plausibility checks, the two sheets are classified as not belonging together and discarded. The longer the code, the more errors could be successively ignored during the check in order to find matches. At the same time, it is also mentioned that if two questionnaires have codes that match in all places, i.e. if zero errors are allowed, the questionnaire pair was actually filled out by the same person. As soon as you allow a mistake, you find significantly more pairs that do not match. (cf. p. 64ff). Since the questionnaires for my study are handed out virtually and thus a possible check of the handwriting afterwards is not possible no advantage can be gained from giving a long code. Every additional code digit has the potential to introduce an error. Therefore, I have decided on a 5-digit code, which, however, is only formed from three questions. The first question asked for the first two letters of one's mother's first name. The second question asked for the day of the participant's birthday, which had to be entered in two digits. For example, if one's birthday was on the fifth of the month, one filled in "05" for the code. The third question asked for the last letter of one's eye colour. As far as the class code is concerned, an arbitrary code was chosen, which should serve to be able to distinguish the classes afterwards.

2.5 Interviews and teacher-questionnaire

The aim of the interviews, as explained above, was to draw a rough picture of the teaching of the individual teachers. The purpose of this is to better compare the two groups and to increase the validity of the study. For the interview, the "narrative interview" described by Reinders et al. (2011) was chosen. In this type of interview, the aim is to understand a person's actions and their interpretation of them. The procedure is also described as being "maximally open" (cf. p. 90), which means that the course of the conversation is little or not at all predetermined by the interviewer, but rather determined by the interviewee.

The interview questions relate to different aspects of teaching. On the one hand, media used in the classroom are discussed. On the other hand, it is asked whether and how many experiments on the topic of radioactivity were shown in class. Furthermore, it is asked what proportion of the teaching time was allocated to the teacher or the students, in the sense of various activities and/or speaking time. Finally, personal emphasis is also investigated.

To triangulate with these interviews, I also administered a survey to teachers that asked whether and how long they have taught individual contents or concepts on the topic of radioactivity. In this section, the teachers of the experimental group also indicate which contents or concepts they have taught beyond the Classbook.

2.6 Participants

Several schools and teachers were contacted by email and invited to participate in the study. In the end, 6 classes were found for each group, whereby one class from the experimental group was taught by myself.

The Classbook is taught in six school classes of the 8th grade. Four of them in the same Gymnasium, the other two in the same Mittelschule. These form the experimental group. The control group is composed of six classes of 8th grade, spread across 2 different Gymnasiums.

Experimental Group: 6 classes total		Control Group: 6 classes total	
Gym. 1: 4 classes	MS 1: 2 classes	Gym. 2: 4 classes	Gym. 3: 2 classes

Table 1: composition of the two groups

The number of participants varied across the different tests in both groups. In the experimental group, 106 people completed the pre-test, 90 people completed the intermediate test and 97 people completed the post-test. In the control group, 113 people completed the pre-test and 113 people completed the post-test. The fluctuations can be attributed to the fact that, although the questionnaires were to be completed online and given out via a link, the students usually completed the questionnaires in the classroom during class time. There can be many different reasons for a child not being present at a given lesson. In addition, mistakes have occasionally occurred in the distribution of the questionnaires. For example, 1 person of the experimental group filled in the intermediate test a second time, instead of the post-test. That means, regarding the conceptual understanding section, only 96 observations can be analysed for the post test, since the intermediate test does not include conceptual questions. Also, 12 people from the control group filled out the pre-test twice for unknown reasons. Pre-test and post-test only differ in the question of how much one liked the lesson. Thus, the analysis must take into account that the number of participants for this aspect is lower than for the conceptual understanding aspect, for example. It is possible that the links from the questionnaires were mixed up by teachers or students. It should also be noted that the 113 people who completed the pre-test in the control group are not necessarily the same 113 people who completed the post-test in the same group.

2.7 Analysis

The data was analysed within the framework of classical test theory. Microsoft® Excel® for Microsoft 365 MSO and RStudio (2023.06.0+421 "Mountain Hydrangea" Release) were used for this purpose. At the beginning, the collected data was processed a little. From the outset, only fully completed questionnaires were included in the data list. As mentioned above, it was noticeable that, for example, in one class, about half of the students had filled in the pre-test a second time instead of the post-test. This was evident from the times at which the questionnaires were completed and the class code. The identifier of these data sets was changed accordingly. Apart from this, there were isolated errors in the identifier of some data sets, which were corrected in this first step. Afterwards, mean values and standard deviations of individual items were considered in order to get an overview.

Special attention was paid to the understanding section of the questionnaire for the students. In principle, the overall test performance was compared within the groups between pre- and post-test and between the two groups.

With the help of the personal codes, the individual gain of students could also be calculated and compared between the groups. A pair of codes was only used if there were exactly two matching codes. If a code could only be found once or more than twice in the data set, these variables were not included in the analysis of individual gains. In the experimental group, it was possible to calculate the gain for 65 people in this way. This corresponds to about 67% of the people who completed the post-test in this group. In the control group, gains could be determined for 80 people, which corresponds to about 71% of the people who completed the post-test in this group. The rate is only a little below that reported by Pöge (2008) after applying their improvement procedures (cf. p. 76). It is therefore satisfactory from my point of view.

As mentioned above, only one point could be scored per question in the knowledge test. Especially for question 1, this point was only awarded if all 10 answer options were ticked. Even if students ticked significantly more answer options in this question after the lesson, this increase would be invisible if we only consider whether someone got a point for the question or not. This question will therefore be examined in more detail to see how many, and which answer options were specifically ticked before and after. In addition, the two groups will also be compared specifically with regard to question 1, since, as mentioned above, this question is directly addressed by the Classbook. It is therefore an indicator of the effect the Classbook has had in the experimental group.

2.7.1 Significance tests

In order to test different results for significance, the Mann-Whitney U test is used (Mann and Whitney, 1947). According to Rasch et al. (2009), this test is used in the evaluation of experiments in which there are two groups whose conditions differ in an independent variable. The test then checks whether the two groups differ significantly with regard to a dependent variable. It is also suitable for ordinal scaled data, i.e. data that follow a certain marginal order (cf. p. 144f). In this case, this independent variable is the Classbook that was used in the experimental group and not in the control group. The dependent variables are the students' test results. According to Rasch et al. (2009), the U-test is thus similar to the t-test. The t-test measures values directly, whereas the U-test assigns rankings to the measured values and compares them. It is preferable to the t-test, among other things, when the characteristic under investigation does not follow a normal distribution (cf. p. 144f). As I will show in the results, a normal distribution cannot be assumed for the data collected, which is why I will use the U-test. The null hypothesis of the U-test is "that there is no difference between the two groups studied with regard to the characteristic surveyed" (Rasch et al., 2009, p. 145). A so-called U-value is determined for this purpose. If this is sufficiently improbable, assuming that the null hypothesis is valid, it is rejected. Whether a significant result exists is decided after comparison with a previously determined significance level α (cf. Rasch et al., 2009, p. 153). For this study, $\alpha=0.05$ is set, which means that the null hypothesis of the U-test is rejected if the U value (also called p-value later on) is less than 0.05. Wilcox and Lewandowski (2017) conducted a comparative study with three groups that also followed a pre-post approach. Their aim was to find out how students' responses to the E-CLASS Survey (see Zwickl et al., 2014) differed depending on the focus of the physics course they took. A distinction was made between courses that focused on physics concepts, laboratory work, or both. The courses were categorised according to the teachers who taught the course (cf. p. 2f). Because their data did not follow a normal distribution, Wilcox and Lewandowski (2017) also used the Mann-Whitney U-test to test their data for significance (p. 4). In another study, Hoyer and Girwidz (2020) investigated which assistance improved students' ability to transfer data from computer-based experiments to graphs. In addition, eye-tracking software was used in this study to analyse where students looked during experiments, how long their gaze stayed there and how often their gaze changed (cf. p. 2f). Students completed a pre-test and post-test as part of the study and one of three different trainings in between. In the first training, the assistance consisted of animations, in the second of interactive, immediate feedback and in the third of static images (cf. Hoyer and Girwidz, 2020, p. 6). How much time the students in the different groups spent with the trainings was not normally distributed. Similarly, data on where and how long students lingered with their gaze and how many jumps there were in between, were not normally distributed. For this reason, the authors also used the Mann-Whitney U test to test the three groups each for significant differences in this regard (cf. p. 9). Eid et al. (2013) also state that the Mann-Whitney U test is used when a normal distribution cannot be assumed but the shape of the distributions is comparable (cf. p. 322).

Data samples are tested for normality using the Shapiro-Wilk test (Shapiro and Wilk, 1965). In order to determine whether the distribution shapes of two data samples are similar, histograms of the corresponding samples are examined. In this context, it is also interesting to look at the variance of the two samples. Hayes (2000) reports that randomisation tests become more conservative when the variance of the larger sample is greater. In contrast, type-1 errors increase when the variance of the smaller sample is larger (cf. p. 654). A type-1 error, or alpha error, is committed when one chooses to reject the null hypothesis when it is true in reality (cf. Eid et al, 2013, p. 196). In this study, the null hypothesis would be roughly described as TRAU had no impact on students' understanding or interest. According to Hayes (2000), if the variances of two samples are equal, the result of a randomisation test would be valid. If the samples are of equal size, type-1 errors can be controlled, regardless of whether the variances are comparable (cf. p. 654f). The facts mentioned above are taken into account when indicating significance. Whether the variances of two samples are comparable is checked with the Levene test.

This study also compares the performance of the two groups with regard to individual items of the questionnaire. As mentioned above, you can get either 0 or 1 point for most items. This means that a participant can either answer a question incorrectly or correctly with nothing in between. The resulting data is no longer ordinal but nominal data, which is also referred to as "categorical" or "discrete". The

Chi Square Test is used for such data (cf. Rasch et al., 2009, p. 171). If the examined characteristics are divided into different levels, the two-dimensional Chi Square Test is used (cf. Rasch et al., 2009, p.185). The null hypothesis of this test states that the observed characteristics are stochastically independent of each other. Accordingly, the alternative hypothesis states that there is a correlation between the levels of the individual characteristics (cf. Rasch et al, 2009, 186). The significance level is again given as α=0.05. If the calculated p-value of the test falls below this value, the null hypothesis can be rejected, and a correlation of the characteristics can be assumed. In this study, two characteristics are examined. One characteristic is "wrong" or "right" for answering a question in the questionnaire. The other characteristic is "experimental group" or "control group", depending on which group a student was in. Because both characteristics have exactly two levels, i.e. are dichotomous, a special case of the two-dimensional Chi Square Test can be used, the so-called four-field Chi Square Test. In this case, the data can be plotted in a 2x2 contingency table (cf. Rasch et al., 2009, 196). In this study, a table is formed in this way for each item, except for item 3, since here, as already mentioned, partial points were possible. For item 3, the first characteristic has the levels $0; \frac{1}{3}; \frac{2}{3}, 1$. An example for all other items is given in Table 2.

	Wrong (0)	Right (1)
Experimental group		
Control group		

Table 2: Example of a 2x2 contingency table for the four-field Chi Square test for one item of the questionnaire (except item 3)

Rasch et al. (2009) also list the prerequisites that must be met in order to conduct a Chi Square test. These include 1. the individual observations must be independent of each other, 2. each person examined must be clearly assignable to a combination of characteristics, 3. the expected frequencies must be greater than five in 80% of the cells of the experimental design (cf. p. 198). The first prerequisite is fulfilled in this study because the answers of one person to the questionnaire do not influence the answers of another person. The second prerequisite is also met because a student is either from one group or the other, and can answer all items except item 3, either correctly or incorrectly. This procedure is also suitable for item 3, because here too the score achieved can be clearly assigned to a person. The third condition is fulfilled in most cases. As I will show in the results, there are cases in which this condition is not fulfilled. In these cases, Fisher's exact test is used, more on this later.

The Chi Squared Test was also used in a comparable way by Burde (2018) in their study. Here, too, the data on individual items from an experimental and a control group were checked for significant differences using the chi-square test (e.g., p. 200).

As mentioned above, in some cases the Fisher's exact test is used instead of the Chi-square test because values smaller than 5 occur in the corresponding 2x2 contingency table. Hess and Hess (2017) describe the Fisher's exact test as an alternative to the Chi-square test, which is often applied to 2x2 tables with small cell numbers (cf. p. 878f).

2.7.2 Effect size and average normalized gain

Because a non-parametric test is used to test the results for significance, non-parametric procedures must also be used to determine the effect size. Various non-parametric effect sizes can be found in the literature, such as Cliff's Delta (Cliff, 1993) or the Vargha-Delaney A (Vargha and Delaney, 2000).

In addition, the so-called average normalized gain according to Hake (1998) is used in educational research. In the literature, this value is compared with the Cohens' d effect size, although opinions differ. Nissen et al. (2018) advise against using the average normalized gain because it would be biased in favour of high pre-test results (cf. p. 10). Coletta and Steinert (2020), on the other hand, provide evidence to argue against this claim. Additionally, they argue that normalized gain has provided valuable information to instructors for the last 20 years and thus it should be used (cf. p. 16). In the study by Burde (2018), the so-called relative learning gain of individual students is also calculated and compared between the experimental and control groups (cf. p. 225f). This relative gain is calculated similarly to the average normalised gain in Hake (1998). In Burde (2018), the relative gain of individual persons is calculated and then the mean value is calculated and compared between groups (e.g., p. 255f). According to Burde (2018), this method is another frequently used method to calculate the average normalised gain (cf. p. 3f).

I consider several non-parametric effect sizes, as well as the average normalised gain, in order to assess how large the effect of the Classbook was. Here, too, the comparison with the control group is interesting. In addition to the absolute learning gain, the relative learning gain of the groups is also considered, analogous to the study by Burde (cf. 2018, p. 255f).

2.7.3 Analysis of the interviews

Since the interviews are not central to the study, but only serve to better compare the two groups, a complete transcription was not made. Instead, a “summary protocol” was prepared. In such a protocol, the material from the tape is already selected and bundled so that only what is relevant for the study is written down (cf. Reinders et al, 2011, p. 112).

3 Results

In the following section, the results of the study are presented. First, the results of the student questionnaire will be presented, focusing first on the conceptual understanding part and then on self-concept, interest and the tanodo question. Finally, the results of the teacher interviews and the teacher questionnaire are briefly presented.

3.1 Conceptual understanding

As explained above, this study was a pre-post study with a control group. Therefore, comparisons within the groups as well as between the two groups are of interest. I would like to start by comparing the results of the pre- and post-test within each group. I will start with the experimental group and then continue with the control group. After that, the pre-test and post-test results of the two groups will be compared. To examine the overall test results for significance, the Mann-Whitney U test is used. Because the individual items are dichotomous or categorical response formats, the Chi Square Test or Fisher's Exact Test is used here. The significance level for all tests was set $\alpha = 0.05$. Finally, the individual classes are compared with each other.

3.1.1.1 Analysis of conceptual understanding within the groups

As already mentioned, the first comparison is within the two groups. I will start by analysing the individual items. I will then compare the average test results before and after. Question 1 will be given special attention because the Classbook refers to it specifically. The average number of items chosen for question 1 will be analysed.

3.1.1.2 Experimental group

<u>Individual items of the experimental group</u>

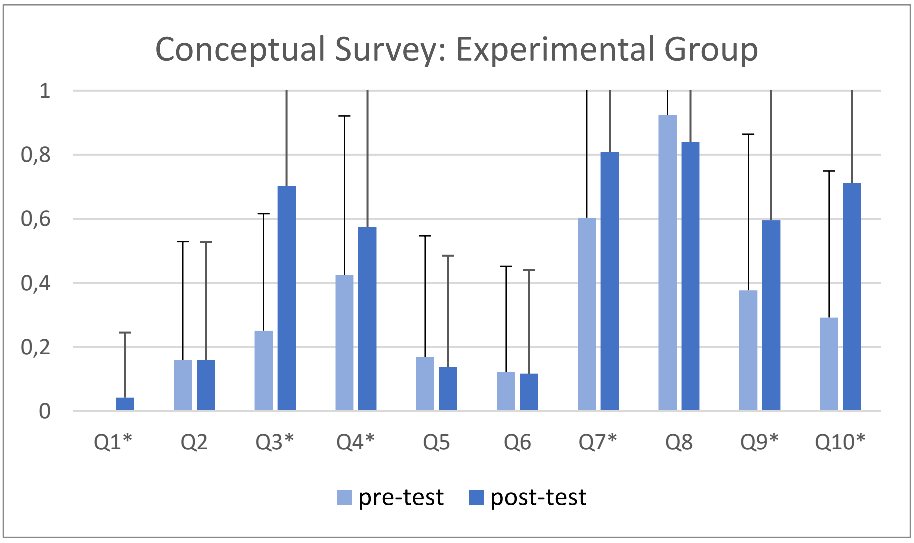

Figure 1: Pre-test and post-test results of individual questions from the conceptual survey for the experimental group. Error bars show standard deviation.

Fig. 1 shows the mean values of the points achieved per question. Error bars show standard deviation. A maximum of 1 point could be achieved per question. Since all questions except question 3, on which partial credit was given, could be scored either 0 or 1, this graph also shows the percentage of students who solved these questions correctly. For example, these results show that question 7 was solved correctly by about 60% of the students in the pre-test and by about 80% of the students in the post-test. This conclusion cannot be drawn for question 3, as partial points were possible here.

There was a particularly large increase in questions 3 and 10. In general, there was an increase in all questions except questions 2, 5, 6 and 8. Whether any of the differences were significant was checked using the Chi-squared test. For Item 1 Fisher's exact test was used instead, because the corresponding contingency table showed values below 5 (see section 2.7.1). The mean score and score difference between the two groups is listed in Table 3. In addition, the p-value is given, which, as just mentioned, was determined with the Chi-square test or with the Fisher's exact test.

Item #	Pre	Post	Difference	p-Value
1	0.00	0.04	0.04	**<0,05**
2	0.16	0.16	0.00	1
3	0.25	0.70	0.45	**<0.001**
4	0.42	0.57	0.15	**<0.05**
5	0.17	0.14	-0.03	0.063
6	0.12	0.12	-0.01	1
7	0.60	0.81	0.20	**<0.01**
8	0.92	0.84	-0.08	0.11
9	0.38	0.60	0.22	**<0.01**
10	0.29	0.71	0.42	**<0.001**

Table 3: Overview of the average number of points per question in the experimental group, difference between pre- and post-test and p-values. Significant values are printed in bold.

Question 1 is particularly interesting. A large increase was expected here because the Classbook deals directly with the content of this question. At first glance, this increase did not occur, although the change was statistically significant. It should be noted that only one point was awarded for this question if all 10 answer options were ticked. For this reason, a closer look was taken at how many choices were ticked on average and how often individual answer options were selected. More on this later.

<u>Total results of the experimental group</u>

The mean value of the points achieved by all participants from the experimental group in the pre-test ($N = 106$) was 3.3 ($SD = 1.3$). In the post-test ($N = 96$), the mean value was 4.6 ($SD = 1.4$). The maximum number of points was 10. These results are to be tested for significance using the Mann-Whitney U test. To do this, the distribution form of the data sets or their variance must be comparable. Fig. 2 shows the corresponding histograms. Based on these, the distributions appear to be comparable, because both are unimodal and peak around 3. The Shapiro-Wilk test shows for both the pre-test ($W = 0.97$, p-value = 0.023) and for the post-test ($W = 0.97$, p-value = 0.03) of the experimental group, that the data are not normally distributed. Levene's test shows that the variances for these two data sets are comparable ($F = 1.83$ p-value = 0.18). The Mann-Whitney U test shows a significant difference for these two data sets ($W = 7513$, p-value < 0.001). The average test score of the experimental group was thus significantly better after the intervention. As for effect sizes see section 3.1.2.

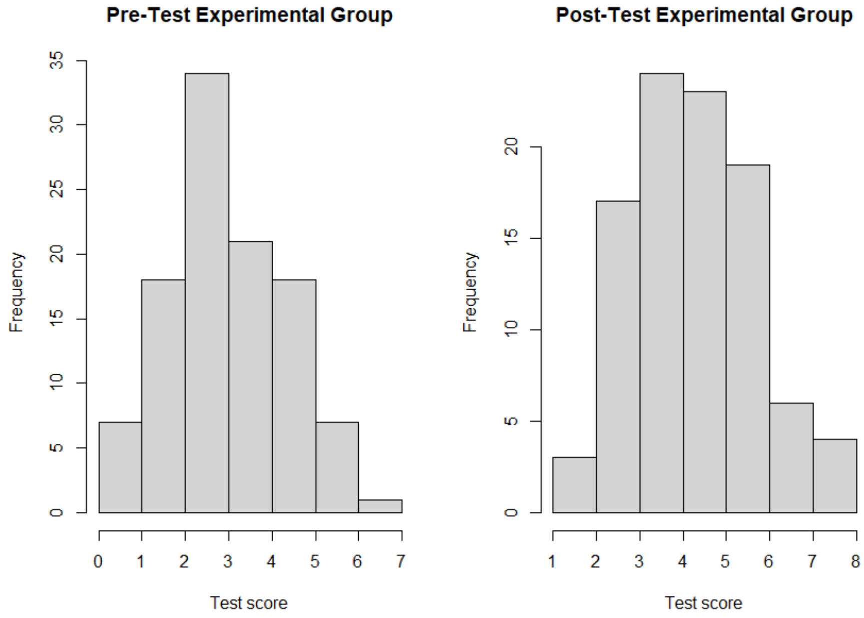

Figure 2: Histograms of the total number of points achieved by individual students in the experimental group; left: pre-test, right: post-test

<u>Question 1 in the experimental group</u>

As mentioned above, question 1 will be examined more closely. In the pre-test, an average of 3.3 (SD = 1.5) answer options of question 1 were selected, in the post-test 5.2 (SD = 2.3). A Mann-Whitney U test is used to determine whether this change is significant. Fig. 3 shows the histograms in which is shown how many choices were selected by students in the experimental group for question 1. The Shapiro-Wilk test shows that the data concerning question 1 in the experimental group are neither normally distributed in the pre-test (W = 0.94, p-value < 0.001) nor in the post-test (W = 0.96, p-value < 0.01). Levene's test does not guarantee variance homogeneity for these two data sets (F = 4.93, p-value < 0.001). Moreover, the variance of the larger sample from the pre-test (N = 106) is 2.1, which is smaller than the variance of the smaller sample from the post-test (N = 96), which is 5.1. This means that the Mann-Whitney U test becomes more liberal (see section 2.7.1). It would give a significant result for these samples (W = 7636.5, p-value = 4.959e-10). I have decided to give the exact p-value in this case, as it is extraordinarily low. Since I cannot estimate by how much the actual p-value of the test is underestimated due to variance inhomogeneity, I will not make a statement about possible significance. In my view, significance can neither be confirmed nor ruled out in this case.

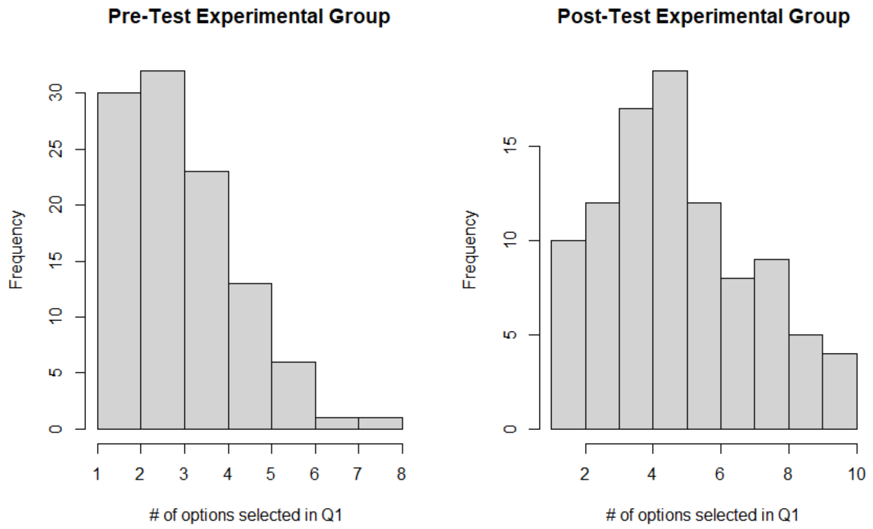

Figure 3: Histogram of response options for question 1 of individual students of the experimental group; left: pre-test, right: post-test

<u>Individual answer choices of question 1 in the experimental group</u>

Therefore, we will take a closer look at the individual items of question 1. The overview of the percentage of participants who selected individual answer options can be found in Fig. 4. There were increases in all response options except the one related to cigarette smoke, where there was a decrease.

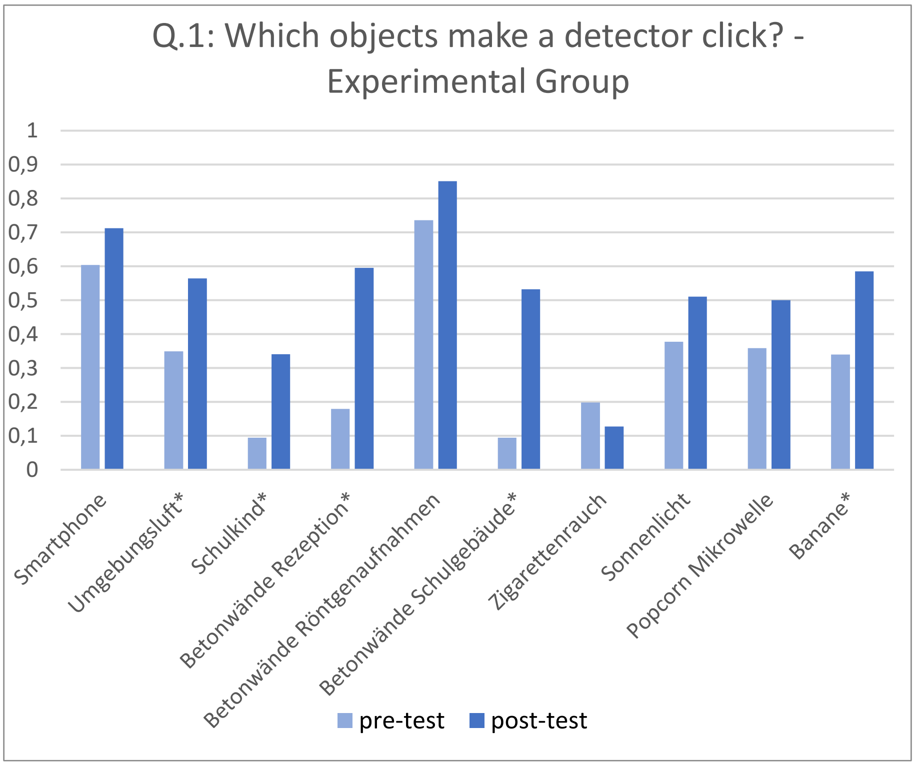

Figure 4: Pre-test and post-test results for individual answer options of question 1 for the experimental group. Bars show the percentage of students who chose that answer option.

Again, a Chi-squared test should clarify whether the differences are significant. Table 4 shows an overview.

Answer option	Pre	Post	Difference	p-Value
Smartphone	60.4	71.3	10.90	0.21
Umgebungsluft	34.9	56.4	21.48	**<0.01**
Schulkind	9.4	34.0	24.61	**<0.001**
Betonwände Rezeption	17.9	59.6	41.65	**<0.001**
Betonwände Röntgenaufnahmen	73.6	85.1	11.52	0.13
Betonwände Schulgebäude	9.4	53.2	43.76	**<0.001**
Zigarettenrauch	19.8	12.8	-7.05	0.23
Sonnenlicht	37.7	51.1	13.33	0.11
Popcorn Mikrowelle	35.8	50.0	14.15	0.08
Banane	34.0	58.5	24.55	**<0.01**

Table 4: Overview of selection frequencies for individual choices in question 1 in the experimental group of the pre- and post-test, difference, and p-values (significant values are printed in bold).

3.1.1.3 Control group

<u>Individual Items</u>

Now the same results are presented for the control group. The average number of points achieved per question can be seen in Fig. 5. Error bars again show standard deviation.

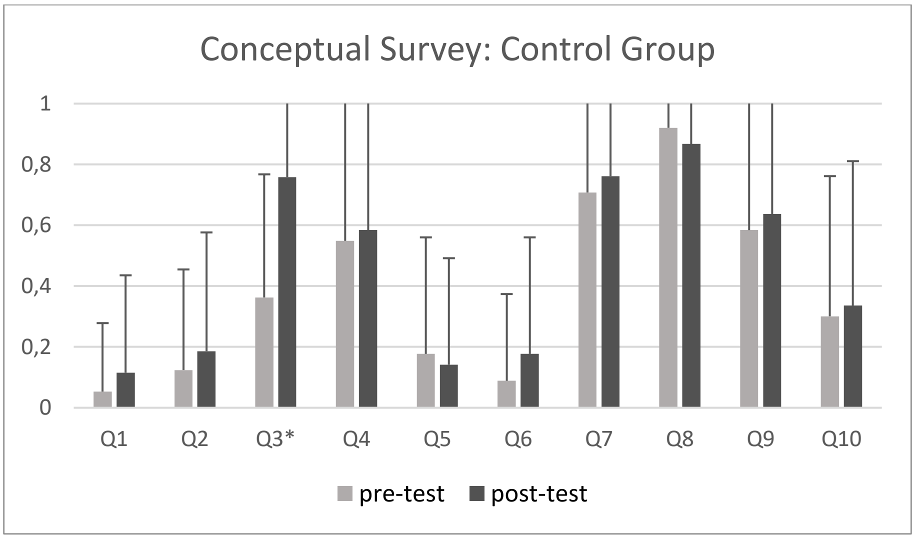

Figure 5: Pre-test and post-test results of individual questions from the conceptual survey for the control group. Error bars show standard deviation.

In the control group there was also an increase in all questions except 5 and 8. There is also a large increase in question 3 in this group, but the increase in question 10 is much smaller compared to the experimental group. Overall, the increases are much smaller and there are decreases as well. A Chi-squared test shall show, which differences are significant. The information can be seen in table 5.

Item #	Pre	Post	Difference	p-Value
1	0.05	0.12	0.06	0.15
2	0.12	0.19	0.06	0.27
3	0.36	0.76	0.40	**<0.001**
4	0.55	0.58	0.04	0.69
5	0.18	0.14	-0.04	0.59
6	0.09	0.18	0.09	0.08
7	0.71	0.76	0.05	0.45
8	0.92	0.87	-0.05	0.28
9	0.58	0.64	0.05	0.50
10	0.30	0.34	0.04	0.67

Table 5: Overview of the average number of points per question in the control group, difference between pre- and post-test and p-values. Significant values are printed in bold.

Total result of the control group

In the pre-test (N = 113), students in the control group achieved an average of 3.9 (SD = 1.5) points. In the post-test (N = 113) an average of 4.6 (SD = 1.7) points were achieved. Fig. 6 shows the histograms for the test results from the pre- and post-test for the control group. The data from the pre-test (W = 0.96, p-value < 0.001) and from the post-test (W = 0.97, p-value = 0.03) of the control group are not normally distributed, as shown by the Shapiro-Wilk test. However, their distributions appear to be comparable based on the histograms, as they are both unimodal and peak around 3. Variance homogeneity was confirmed for these data with Levene's test (F = 1.47 p-value = 0.23). Apart from that, the sample size in this case is identical between pre- and post-test. The Mann-Whitney U test also shows that the test results for the control group differ significantly between pre- and post-test (W = 8054.5, p-value < 0.001).

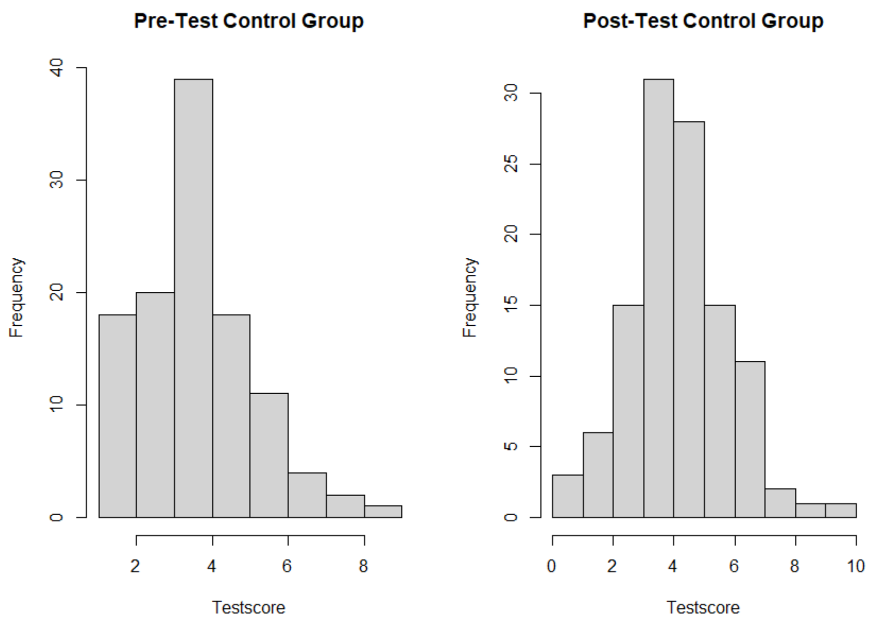

Figure 6: Histograms of the total number of points achieved by individual students in the control group; left: pre-test, right: post-test

Question 1 in the control group

Question 1 was also examined more closely for this group. In the control group, an average of 4.3 answer options (SD = 2.2) were ticked in the pre-test, more than in the experimental group. In the post-test, an average of 4.8 answer options (SD = 2.7) were ticked, less than in the experimental group. This difference is also tested for significance. Fig. 7 again shows the corresponding histograms. The distributions seem to be similar. In any case, they are not normally distributed, which a Shapiro-Wilk test shows on the one hand for the pre-test ($W = 0.89$, p-value < 0.001), and on the other hand for the post-test ($W = 0.90$, p-value < 0.001). Since the sample sizes for the control group in the pre-test and post-test are identical, type 1 errors can be controlled when applying a Mann-Whitney U test here (see section 2.7.1). According to the Mann-Whitney U test, the number of selected answer options for question 1 does not differ significantly for the control group ($W = 6848$, p-value $= 0.34$).

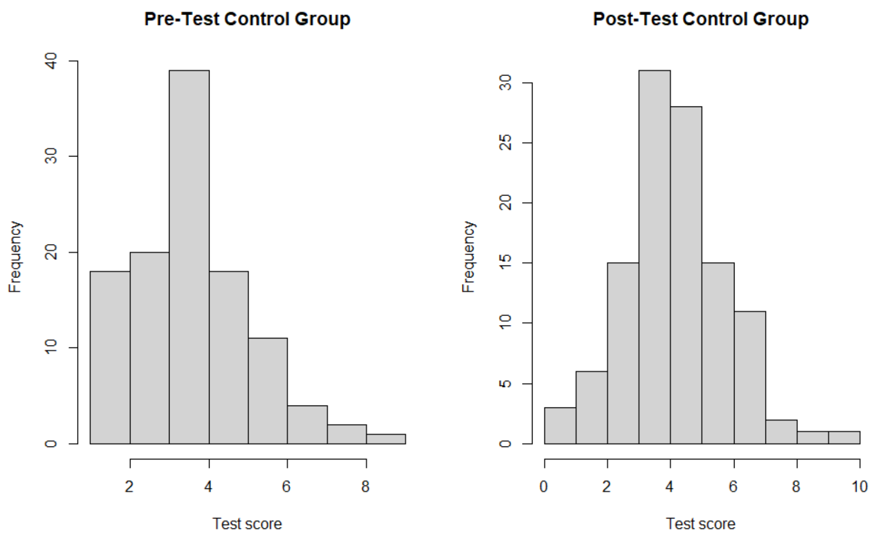

Figure 7: Histogram of response options for question 1 of individual students of the control group; left: pre-test, right: post-test

The analysis of the individual answer options is nevertheless interesting. Figure 8 shows the percentage of students in the control group who selected a particular option. There is a notable increase in answer options 3, 4 and 6. There are decreases as well.

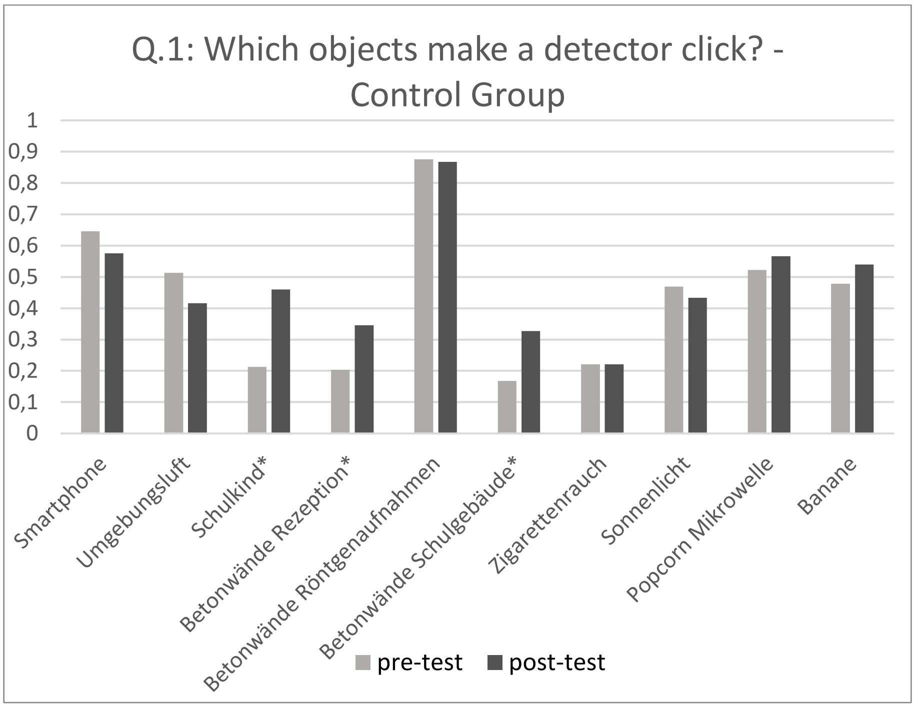

Figure 8: Pre-test and post-test results for individual answer options of question 1 for the control group. Bars show the percentage of students who chose that answer option.

Again, the Chi-squared test shows, which of the differences are significant. Consider table 6 for the corresponding values.

Answer option	Pre	Post	Difference	p-Value
Smartphone	64.6	57.5	-7.08	0.34
Umgebungsluft	51.3	41.6	-9.73	0.18
Schulkind	21.2	46.0	24.78	**<0.001**
Betonwände Rezeption	20.4	34.5	14.16	**<0.05**
Betonwände Röntgenaufnahmen	87.6	86.7	-0.88	1
Betonwände Schulgebäude	16.8	32.7	15.93	**<0.01**
Zigarettenrauch	22.1	22.1	0.00	1
Sonnenlicht	46.9	43.4	-3.54	0.69
Popcorn Mikrowelle	52.2	56.6	4.42	0.59
Banane	47.8	54.0	6.19	0.42

Table 6: Overview of selection frequencies for individual choices in question 1 in the control group of the pre- and post-test, difference, and p-values (significant values are printed in bold).

Analysis of conceptual understanding between the groups

In the following section, the two groups are compared with each other. In the first section, the results of the pre-tests are compared. This should serve to give insight into the prior knowledge bases of the groups and compare them with each other. After that, the post-test results of the two groups are compared. The corresponding data compares the groups' conceptual understanding in the final section. The procedure is the same as above.

3.1.1.4 Pre-Test

<u>Individual items</u>

First, the average points achieved for each question are considered individually. As mentioned above, for all questions except question 3, the number obtained can be considered as the percentage of students who got this question correct. Figure 9 compares the points achieved per item in the pre-test.

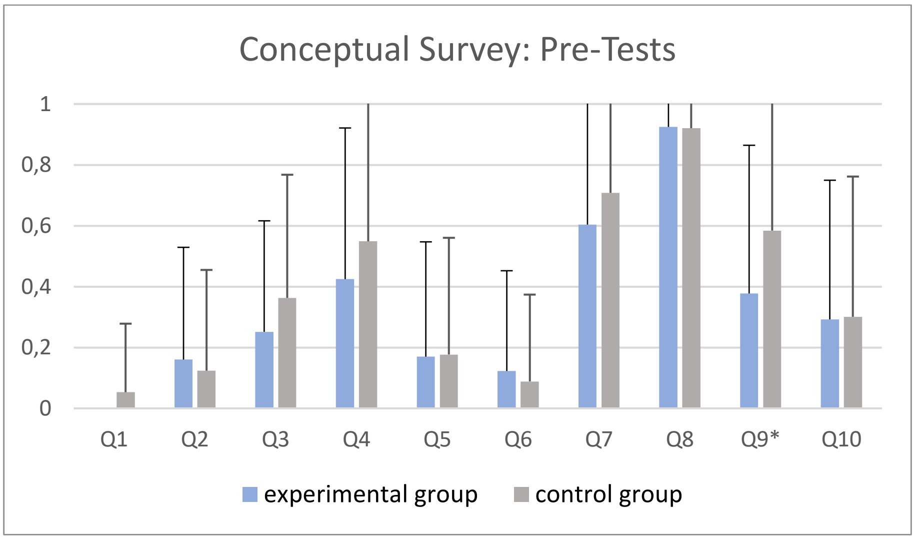

Figure 9: Pre-test results for individual questions of the conceptual survey for both groups

The Chi-square test was used to check whether the differences between the two groups with regard to the individual questions were significant. Similarly, to before, in the case of question 1 Fisher's Exact Test was used because values below 5 occurred in the corresponding cross-tabulation (see section 2.7.1). For relevant values see Table 7.

Item #	EG	CG	Difference	p-Value
1	0.00	0.05	0.05	0.15
2	0.16	0.12	-0.04	0.56
3	0.25	0.36	0.11	0.16
4	0.42	0.55	0.12	0.09
5	0.17	0.18	0.01	1
6	0.12	0.09	-0.03	0.55
7	0.60	0.71	0.10	0.14
8	0.92	0.92	0.00	1
9	0.38	0.58	0.21	**<0.01**
10	0.29	0.30	0.01	1

Table 7: Overview of the average number of points per question in the pre-test, difference between the experimental group (EG) and the control group (CG) and p-values. Significant values are printed in bold.

The points achieved were therefore significantly different for 1 of the total of 10 questions, with the control group showing an advantage over the experimental group. Question 1 was answered very rarely correctly by both groups. In the control group, only 5% of the students answered it correctly and in the experimental group nobody answered it correctly, so we are seeing a floor effect with this item. Question 9 refers to the danger of radioactivity. Here, the students in the control group seem to have had more prior knowledge than the students in the experimental group.

Total result of the pre-test

Subsequently, the average test results of the pre-test of the respective groups will be compared and tested for significance. Figure 10 shows the histograms for the test results of both groups. The distributions look comparable because both have one peak at around 3. With the Shapiro Wilk test, I have already shown above that the data sets are not normally distributed. The comparability of the variances is given after checking with Levene's test (F = 0.51, p-value = 0.47).

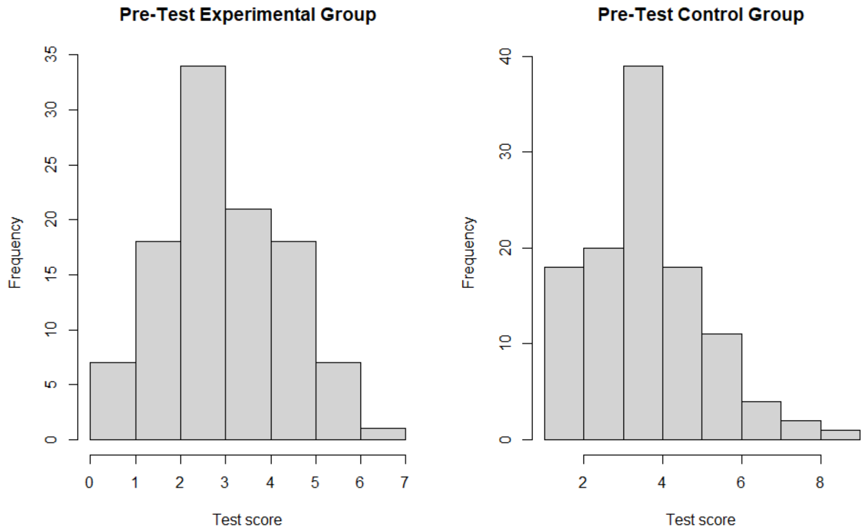

Figure 10: Histograms of the total number of points achieved by individual students in the pre-test; left: experimental group, right: control group

As mentioned above, the experimental group (N = 106) scored an average of 3.3 points (SD = 1.3). In contrast, the control group (N = 113) scored an average of 3.9 points (SD = 1.5). According to the Mann-Whitney U test, these results differed significantly from each other (W = 7226, p-value < 0.01). For this reason, among others, the absolute and relative growth of individual students will be analysed in more detail later.

Question 1 in the pre-test

Question 1 is also inspected more closely here in order to be able to better assess what the prior knowledge of the two groups was with regard to this question. As already mentioned, the experimental group selected an average of 3.3 (SD = 1.5) answer options. In the control group, the average was 4.3 (SD = 2.2). Fig. 11 shows the corresponding histograms of these two data sets. The distributions appear comparable because of the peaks at the same point, even though there is a second small peak in the control group. Both data sets have already been shown to be non-normally distributed using the Shapiro-Wilk test. The variance of the two data sets is not comparable according to Levene's test (F = 7.98, p < 0.01). However, the variance of the smaller sample (N = 106) with 2.1 is lower than that of the larger sample (N = 113) with 4.86. Moreover, the samples are similar in size. Type-1 errors should therefore be kept under control (see section 2.7.1). The Mann-Whitney U test shows that the data sets differ significantly from each other (W = 7447, p-value < 0.01).

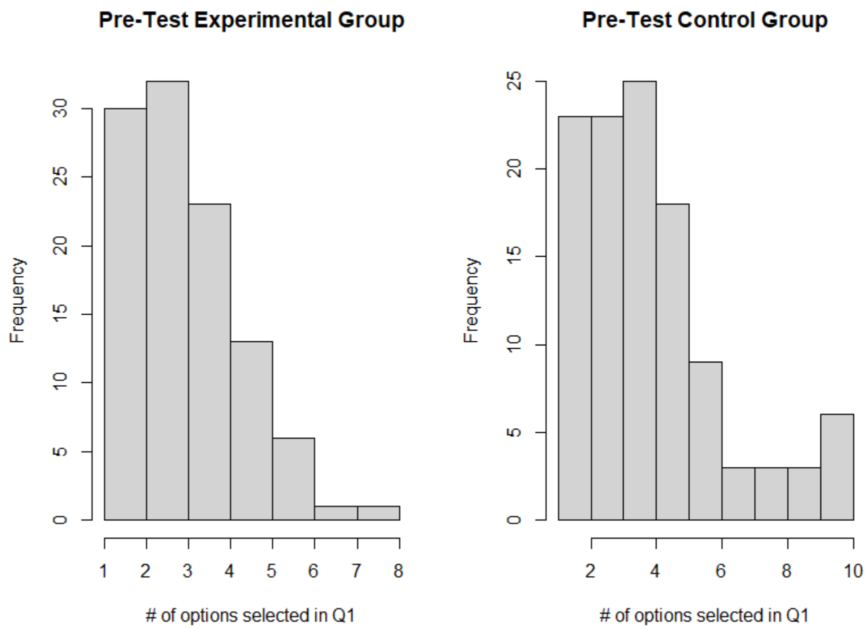

Figure 11: Histogram of response options for question 1 of individual students of the pre-test; left: experimental Group, right: control Group

<u>Individual answer choices of question 1 in the pre-test</u>

Also, the individual items of question 1 have been analysed in more detail for the pre-test. Fig. 12 shows the relative proportions of those students who chose the respective answer option in the pre-test.

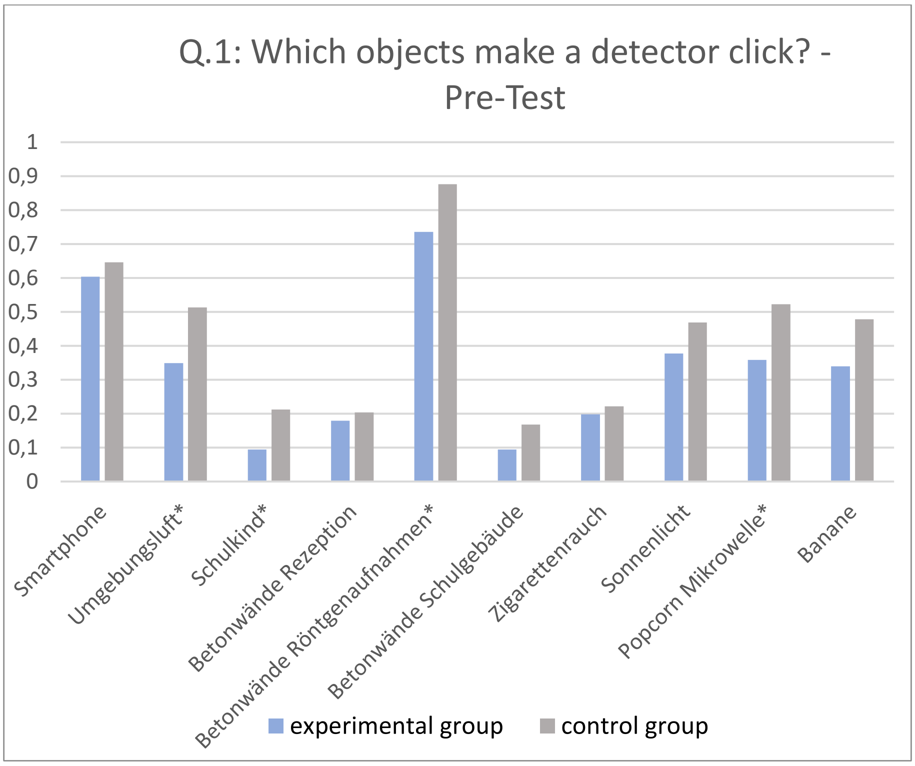

Figure 12: Pre-test results for individual answer options of question 1 for the experimental and control group. Bars show the percentage of students who chose that answer option.

These results were again tested for significance using the Chi-square test. An overview can be found in Table 8. Out of the 10 possible answer options, four were chosen significantly more often by students in the control group than by students in the experimental group. For one answer option the p-value was very close to 0.05. The control group seems to have had more prior knowledge about this question.

Answer option	EG	CG	Difference	p-Value
Smartphone	60.4	64.6	4.2	0.61
Umgebungsluft	34.9	51.3	16.4	**0.02**
Schulkind	9.4	21.2	11.8	**0.03**
Betonwände Rezeption	17.9	20.4	2.4	0.78
Betonwände Röntgenaufnahmen	73.6	87.6	14.0	**0.01**
Betonwände Schulgebäude	9.4	16.8	7.4	0.16
Zigarettenrauch	19.8	22.1	2.3	0.80
Sonnenlicht	37.7	46.9	9.2	0.22
Popcorn Mikrowelle	35.8	52.2	16.4	**0.02**
Banane	34.0	47.8	13.8	0.052

Table 8: Overview of selection frequencies for individual choices in question 1 in the pre-test of the experimental group (EG) and the control group (CG), difference, and p-values (significant values are printed in bold).

3.1.1.5 Post-Test

<u>Individual items</u>

Again, the average points for each question are analysed individually first. Fig. 13 again shows the points achieved per question and group, this time for the post-test.

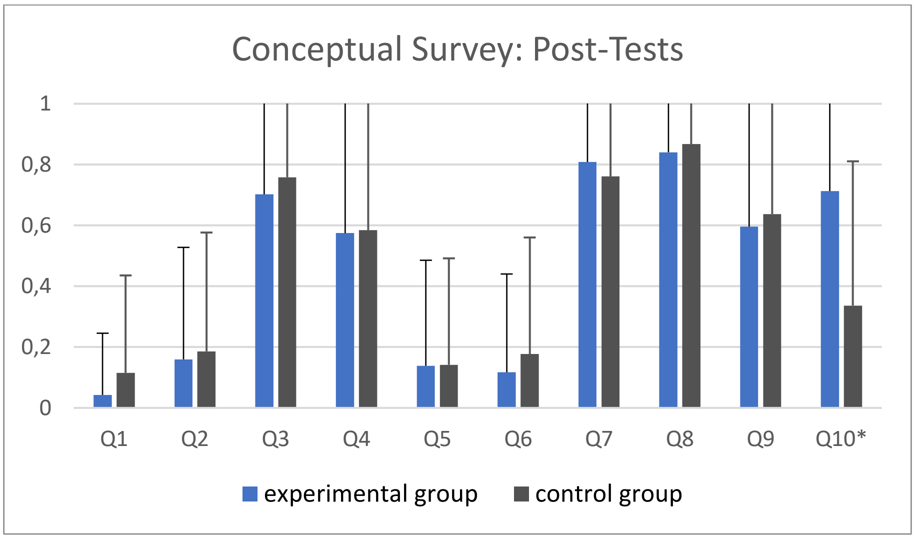

Figure 13: Post-test results for individual questions of the conceptual survey for both groups

The differences between the two groups are again clearly presented in Table 9. The Chi-square test serves to show which differences are significant. This time, only one significant difference can be found. Question 10, which is also directly addressed by the Classbook, was significantly more often answered correctly by the experimental group than by the control group. The difference is considerable at 38%. With all these data, it must be remembered that the pre-test results were significantly different.

Item #	EG	CG	Difference	p-Value
1	0.04	0.12	0.07	0.09
2	0.16	0.19	0.03	0.70
3	0.70	0.76	0.06	0.49
4	0.57	0.58	0.01	0.98
5	0.14	0.14	0.00	1
6	0.12	0.18	0.06	0.28
7	0.81	0.76	0.05	0.59
8	0.84	0.87	0.03	0.78
9	0.60	0.64	0.04	0.51
10	0.71	0.34	0.38	**<0.001**

Table 9: Overview of selection frequencies for individual choices in question 1 in the post-test of the experimental group (EG) and the control group (CG), difference, and p-values (significant values are printed in bold).

<u>Total result of the post-test</u>

Now the total test result of the post-tests is compared. Significance is again to be tested with the Mann-Whitney U test. Fig. 14 shows the histograms for both groups. The distributions seem comparable based on these. Although there is a greater difference in peak width on the post-test between groups, with the experimental group being wider, both graphs continue to be unimodal. Both cases are not normally distributed. This has already been shown by the Shapiro-Wilk test above. According to Levene's test, the variances for these two samples are comparable (F = 0.39, p-value = 0.54).

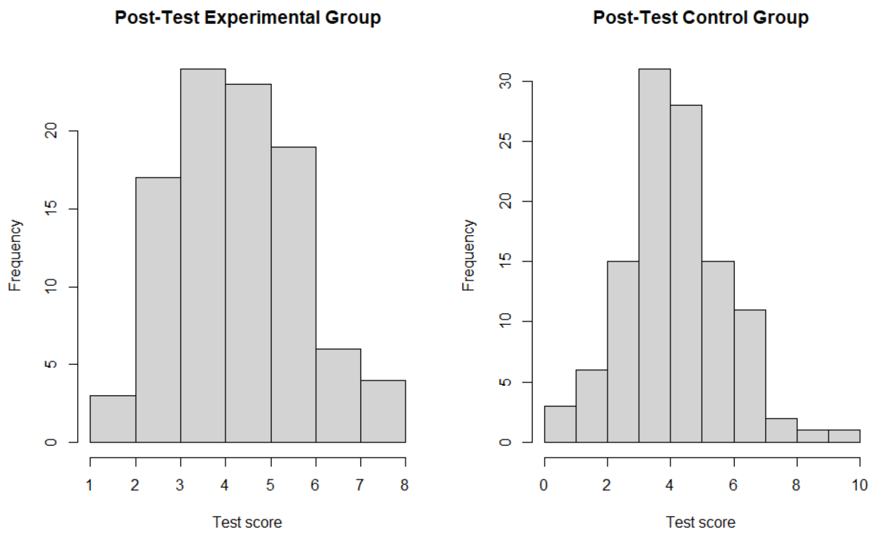

Figure 14: Histograms of the total number of points achieved by individual students in the post-test; left: experimental group, right: control group

The average test score of the experimental group in the post-test (N = 96) was 4.6 (SD = 1.4). In the control group (N = 113), the average test result was also 4.6 (SD = 1.7). Not surprisingly, these results do not differ significantly from each other according to the Mann-Whitney U test (W = 5244.5, p-value = 0.68).

Question 1 in the post-test

In question 1, the experimental group achieved a better result with 5.2 answer choices ticked on average, compared with the control group, which ticked 4.8 answer choices on average. This question is also examined in more detail here. Fig. 15 shows the two histograms of the data sets for question 1 for the post-test. At first glance, the two distributions do not seem comparable. In the control group, two peaks can be seen at the edge of the distribution. In the experimental group, on the other hand, one peak can be seen in the middle. Normality has already been excluded for both data sets. Levene's test would predict variance homogeneity for both data sets (F = 2.07, p-value = 0.15). According to the Mann-Whitney U test, there would be no significant differences between the two data sets, although the p-value is very small (W = 4616.5, p-value = 0.06). Since the distributions of the data sets are not strictly comparable, I would not speak of a significant difference here.

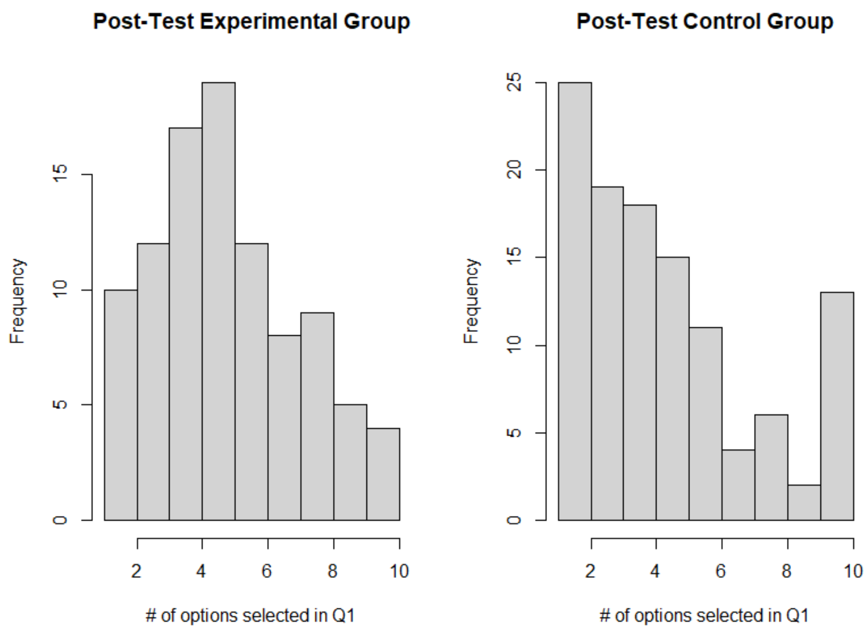

Figure 15: Histogram of response options for question 1 of individual students of the post-test; left: experimental group, right: control group

Nevertheless, it is interesting to take a closer look at the individual answer options. Fig. 16 again shows the proportion of students who selected the corresponding answer option.

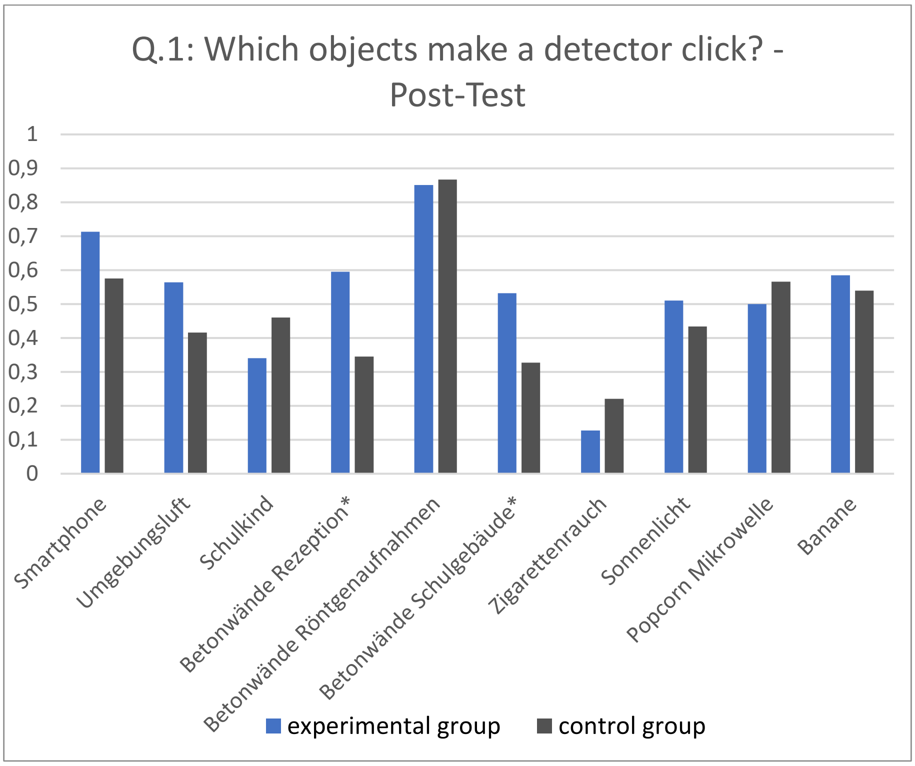

Figure 16: Pre-test results for Question 1 for both groups.

Of the 10 response options, 6 were chosen more often by the experimental group. In 2 of these cases the difference between the groups is significant, in another case the p-value is small but not significant. Conversely, 4 answer options were chosen more often by the control group. The difference is never significant, but the p-value is low in 2 cases. Especially the answer option "a school child" was chosen more often by the control group. Exact values for differences and p-values can be found in Table 10.

Answer option	EG	CG	Difference	p-Value
Smartphone	71.3	57.5	13.8	0.09
Umgebungsluft	56.4	41.6	14.8	0.07
Schulkind	34.0	46.0	-12.0	0.09
Betonwände Rezeption	59.6	34.5	25.1	**<0.001**
Betonwände Röntgenaufnahmen	85.1	86.7	-1.6	0.62
Betonwände Schulgebäude	53.2	32.7	20.4	**<0.01**
Zigarettenrauch	12.8	22.1	-9.4	0.10
Sonnenlicht	51.1	43.4	7.7	0.41
Popcorn Mikrowelle	50.0	56.6	-6.6	0.33
Banane	58.5	54.0	4.5	0.73

Table 10: Overview of selection frequencies for individual choices in question 1 in the post-test of the experimental group (EG) and the control group (CG), difference, and p-values (significant values are printed in bold).

3.1.2 Learning gains and effect sizes

In this section, the learning gains of the two groups will be analysed and compared. This is interesting because the post-test results did not differ significantly between the groups, but the pre-test results did. The difference in points between the pre-test and the post-test is used as the absolute learning gain. The average test scores of the individual groups have already been examined above. The average test results in the experimental group differ by 1.3 points, those in the control group by 0.7 points. The datasets differ significantly (see section 3.1.1). Thus, both groups experienced significant learning gains in terms of average test scores.

3.1.2.1 Individual learning gains

In order to determine the learning gains of individual students, the personal codes were used. As mentioned above, 65 valid code pairs were found for the experimental group and 80 valid code pairs for the control group. The individual growth could therefore be determined for the respective number of pupils. For the 65 students in the experimental group, the average individual growth was 1.4 points, which is somewhat higher than the average unmatched growth (average of all experimental post-tests – average of all experimental pre-tests). For the 80 students in the control group, the average increase in points was 0.8, which is also slightly higher than the average unmatched growth of the control group.

A Mann-Whitney U test will again be used to test for significance. Fig. 17 shows the histograms for the respective increases from both groups. The distributions appear to be roughly comparable, as both peak around 1. In this case, the Shapiro-Wilk test demonstrates normality of the data for the control group (W = 0.98, p-value = 0.25), but not for the experimental group (W = 0.95, p-value = 0.01). Levene's test, however, just detects variance homogeneity (F = 3.77, p-value = 0.054). In this case, the variance of the smaller sample with 3.11 would be greater than that of the larger sample with 1.86. The result of the Mann-Whitney U test should therefore be viewed with caution. It would indicate a significant difference between the gains of the two groups in terms of the test score (W = 1933.5, p-value = 0.008). However, the actual p-value may again be underestimated. In this case, I would still speak of a significant difference.

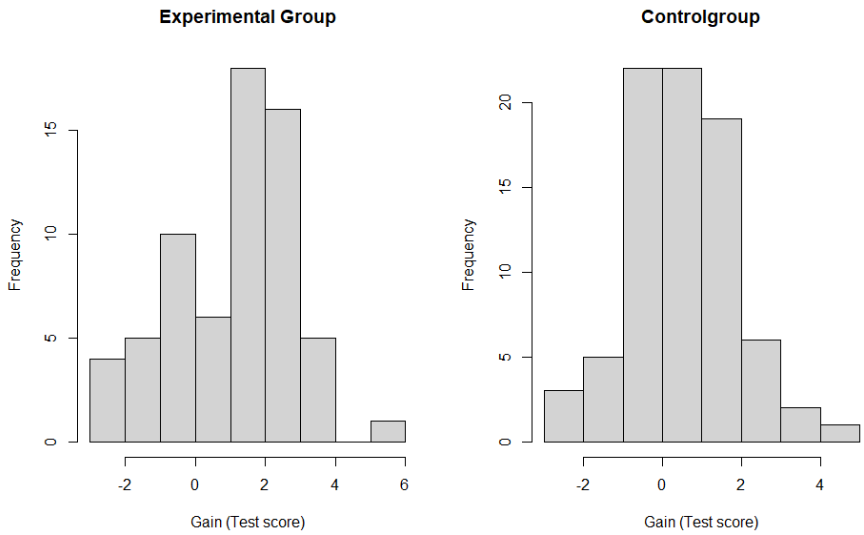

Figure 17: Histograms of the absolute learning gains of individual pupils; left: experimental group, right: control group (learning gain = difference between pre-test score and post-test score)

Average and individual gains regarding Question 1

As far as question 1 is concerned, on average 1.9 more answer options were selected by the experimental group when comparing between pre- and post-test. However, it is unclear whether this increase is significant. In the control group, an average of 0.4 more answer options were selected for question 1. This difference is not significant, as shown above.

Of the 65 students in the experimental group whose codes were comparable, 1.8 more answer options were selected on average, which is slightly below the unmatched average for them. For the 80 students in the control group, 0.5 more answer options were selected on average, slightly more than the unmatched average for the control group.

To determine any significant difference between these two increases, a Mann-Whitney U test was again performed. Fig. 18 again shows the histograms of the corresponding distributions, which are judged to be roughly comparable, because of the peak at the same value. Neither the data of the experimental group (W = 0.95, p-value = 0.01) nor the data of the control group (W = 0.93, p-value < 0.001) are normally distributed, which was checked with a Shapiro-Wilk test. Variance homogeneity can be assumed thanks to Levene's test (F = 0.17, p-value = 0.68). The Mann-Whitney U test shows a significant difference between the two gains (W = 1594.5, p-value < 0.001).

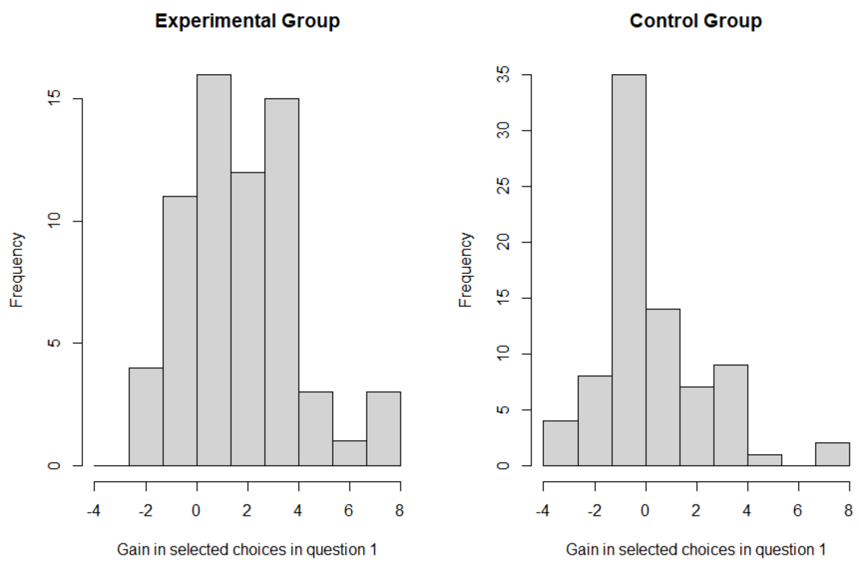

Figure 18: Histograms of how many more answer choices were selected by students in question 1 between pre-test and post-test; left: experimental group, right: control group

Effect size of the absolute and individual learning gain

In order to assess how large the above learning gains are within the individual groups, Cliff's Delta, Vargha-Delaney A, and the average normalised gain were calculated for each group. First, the average gain of 1.3 points on the test score in the experimental group was considered. Cliff's delta of 0.48, 95% CI [0.33, 0.60], and Vargha-Delaney A of 0.74 would indicate a large effect. The average normalised gain, on the other hand, is only 0.20, which would indicate a small effect.

As for the average test score gain of 0.7 points in the control group, both Cliff's delta of 0.26, 95% CI [0.11, 0.40], Vargha-Delaney A of 0.63, and the average normalised gain of 0.11 suggest a small effect.

Next, the individual gains of the two groups were compared. Cliff's Delta and Vargha-Delaney A were calculated to access the size of the effect. These individual gains were, as a reminder, 1.4 for the 65 students of the experimental group and 0.8 for the 80 students of the control group. Cliff's delta of -0.26, 95% CI [-0.43, -0.06] and Vargha-Delaney A of 0.37 suggest a small effect.

Also, the normalised gain for individual students was calculated, again using the person codes. Calculating the normalised gain for individual students and taking the mean value, we obtain 0.18 for the experimental group and 0.13 for the control group, which indicates a small effect in both cases. According to the Mann-Whitney U test, the difference between the two values is significant (W = 2046.5, p-value = 0.03). The effect size according to Cliff's Delta is -0.21, 95% CI [-0.39, -0.02] and Vargha-Delaney A is 0.39, which indicates a small effect in both cases.

Effect size regarding question 1

As far as question 1 is concerned, a similar picture emerges. According to Cliff's Delta of 0.50, 95% CI [0.35, 0.62] and Vargha-Delaney A of 0.75, the average gain of 1.9 selected answer options in the experimental group can be described as large. However, the normalised average gain for this aspect is only 0.28, which can still be described as small.

With regard to the control group, the gain of 0.4 selected response options with a Cliff's Delta of 0.07, 95% CI [-0.08, 0.22] and Vargha-Delaney A of 0.54 can be described as negligible. According to the normalised average gain of 0.08, the effect can be described as small.

Effect sizes were also calculated for the comparison of the individual gains on selected answer options to question 1. They were 1.8 for the individual gains of the experimental group and 0.5 for the individual gains of the control group. According to Cliff's Delta of -0.39, 95% CI [-0.54, -0.20] and Vargha-Delaney A of 0.30, the difference is of medium magnitude.

3.1.3 Class comparison regarding conceptual understanding

In the following section, the average test scores, the absolute, and the relative learning gains of the individual classes are compared. For test scores, I will look at pre-test results and then post-test results. For the absolute learning gains, the difference in scores between the pre-test and post-test of the individual classes is again used. In addition, the normalised gain at class level is also calculated and compared.

3.1.3.1 Test scores in class comparison

It is interesting to see whether certain classes performed visibly better than others in the pre- or post-test. For this purpose, the test scores of the individual students were calculated and a mean value was calculated for each class. In Figure 19, these data are sorted by group in ascending order. Fig. 20 shows these results again for the post-test. The class numbers are identical in both figures.

As shown above, the prior knowledge in the control group was clearly better than in the experimental group. This is also evident in the comparison of the individual classes. If we look at Fig. 19, it is particularly noticeable that one class in the experimental group performed considerably worse in the pre-test than all the others. In addition, it is noticeable that there are two classes in the control group that performed considerably better than all the others. Furthermore, all classes in the experimental group are below the mean of the control group, while the control group classes are all above the mean of the experimental group.

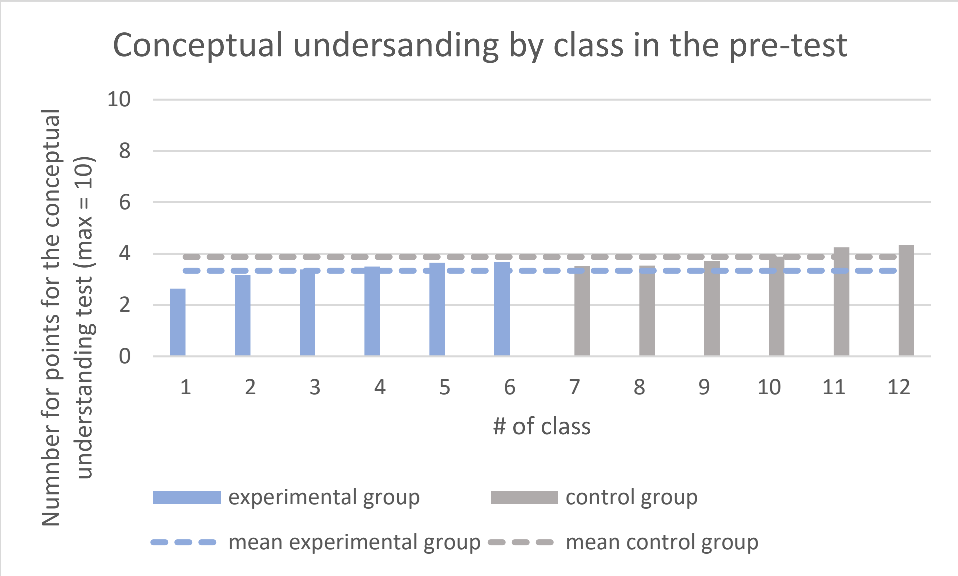

Figure 19: Mean values of the test scores of the 12 individual classes for the pre-test. Mean of the experimental group = 3.34; mean of the control group = 3.87

The results from the post-test discussed above are also reflected in the class comparison. Fig. 20 shows that the mean values have converged. At the same time, the inhomogeneous picture in the experimental group is striking. Two classes have improved enormously and are above all other classes. At the same time two other classes in the experimental group have hardly improved at all. In the control group, the picture looks more homogeneous. Three classes have improved visibly, the other three hardly at all. In the end, only two classes in the experimental group are above the average of the control group, but they are clearly above. Conversely, 5 classes in the control group are above the average of the experimental group, but only slightly.

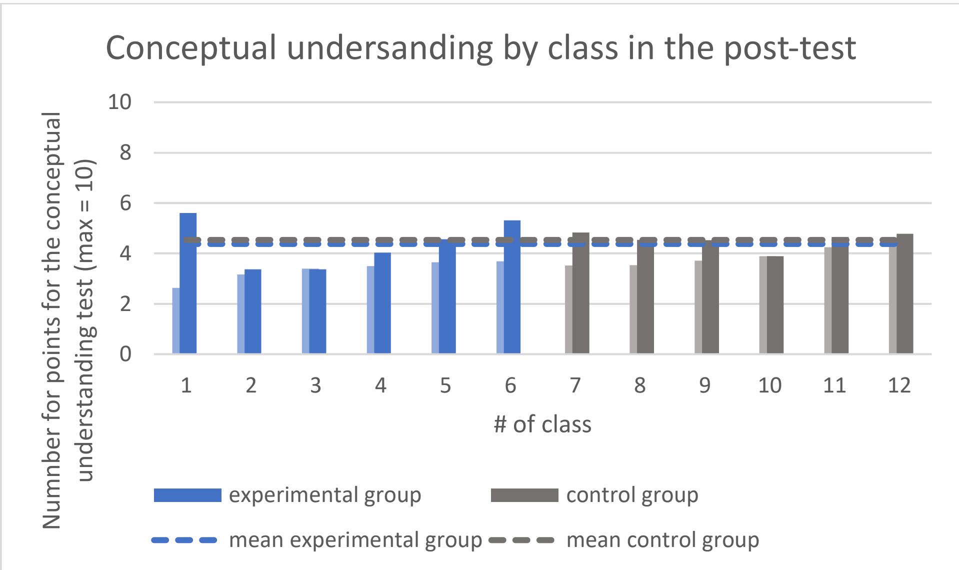

Figure 20: Mean values of the test scores of the 12 individual classes for the post-test. Mean of the experimental group = 4.38; mean of the control group = 4.54; bars in the background show pre-test results of these classes

3.1.3.2 Absolute gains in class comparison

In the following section, I will present the absolute learning gains of the 12 individual classes. An overview is given in Fig. 21. Note that the class numbers are not the same, as shown in Fig 19. and Fig. 20. Instead, the data was arranged once more in ascending order. Here again, we can clearly see the picture that has already emerged above. Two classes in the experimental group had a particularly large increase in their conceptual understanding. Two others had almost none. In the control group, the picture is more homogeneous. It is noticeable that the learning gains of four of the six classes from the experimental group were greater than the average of the control group. Conversely, only the learning gains of one class in the control group were better than the average of the experimental group. In addition, the learning gains of two classes in the experimental group were greater than those of all the control groups.

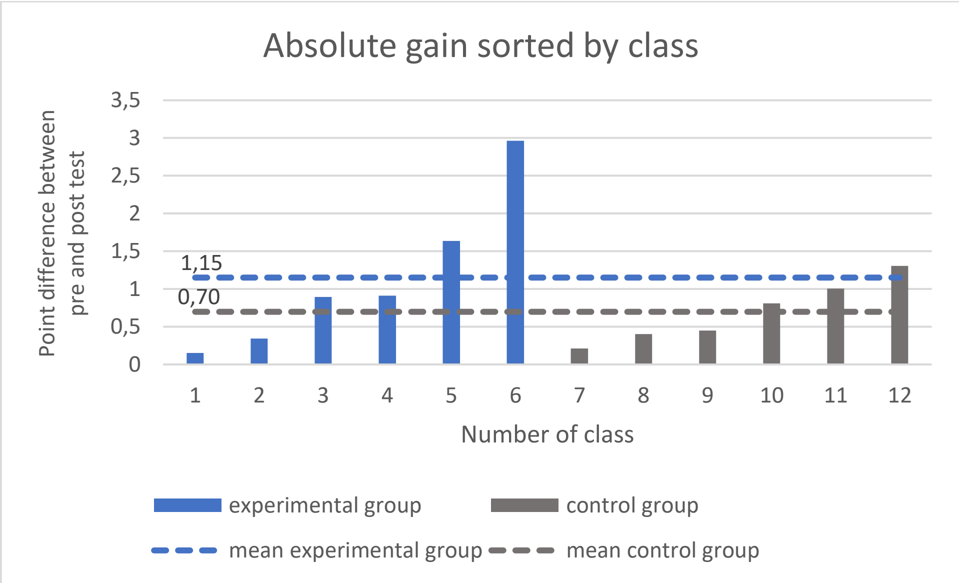

Figure 21: Absolute learning gains of the individual classes sorted by size and group

3.1.3.3 Normalised gains in class comparison

Looking at the normalised gains in Fig. 22, a very similar picture can be observed. As far as the average values are concerned, we see the same situation as before. In addition, the normalised gain of two classes in the experimental group was also greater than that of all classes in the control group.

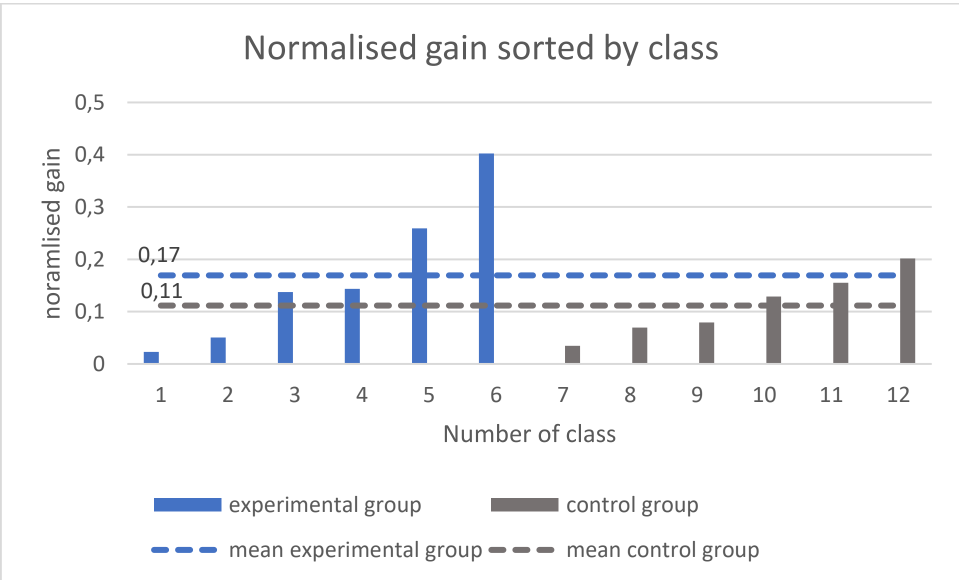

Figure 22: Normalised gains of the individual classes sorted by size and group

3.1.4 Emotion towards radioactivity

The part of the questionnaire that tested conceptual understanding also contained a question about personal emotion towards radioactivity. This question was not scored and therefore did not count towards a person's total test score. It is analysed separately here.

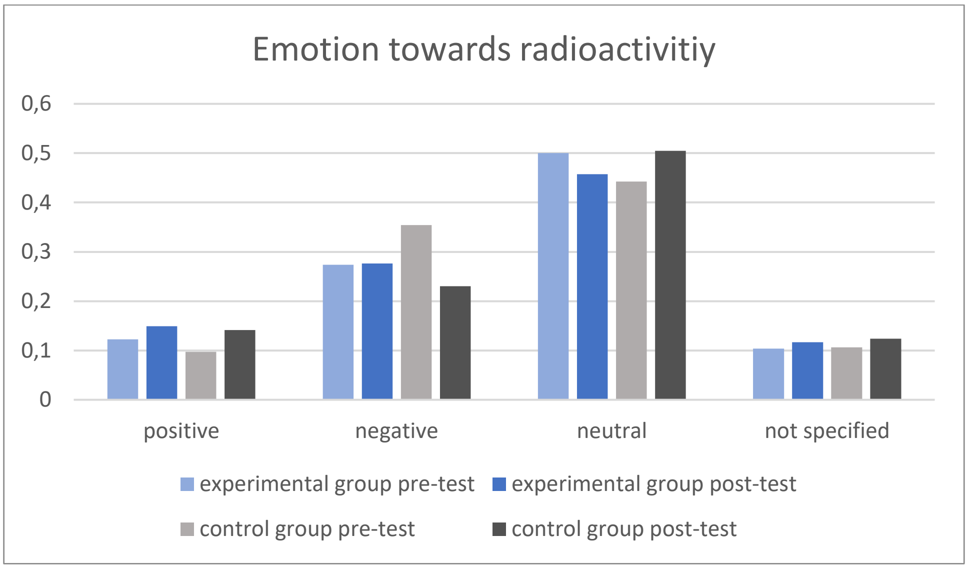

Figure 23: relative frequency of the individual answer options sorted by groups and tests

As can be seen in Fig. 23, the difference between the groups and the individual tests was not very large. The negative emotions outweighed the positive ones in every case. In addition, the answer option "neutral" was chosen most often in each case. In order to check whether there were significant differences in the choice of answer options in any case, either within the groups or between the groups, a Fisher's exact test was carried out for each case. Within the groups, the test yielded a p-value of 0.83 for the experimental group and 0.22 for the control group. Between the groups, the test yielded a p-value of 0.61 for the pre-tests and 0.83 for the post-tests. There were therefore no significant differences anywhere.

3.2 Self-concept

In this section, the results regarding the self-concept of the students are presented. First, comparisons are made within the groups again. As already mentioned, an intermediate test was given in the experimental group, which, among other things, checked the self-concept again. This should help to assess the influence of the Classbook on this aspect more precisely. Afterwards, the groups are compared with each other to see if there is a significant difference between the pre-tests and the post-tests of the two groups. At the end, the individual classes will again be compared with each other. The intermediate test was completed by 90 people in the experimental group. As a reminder, the pre-test was completed by 106 people in the experimental group and the post-test by 97 people. In the control group, both the pre-test and the post-test were completed by 113 people.

The person codes were not included in this part. As I will show in the next section, there were few significant differences between the tests within the groups. Comparing the gains of individuals between groups is unlikely to yield additional significant information.

3.2.1 Analysis of self-concept within the groups

First, the results within the two groups are compared. I will start with the experimental group, presenting an overview of the data first. Then I will compare the results of the pre-test with those of the intermediate test. Following that, I will contrast the results of the intermediate test with those of the post-test. Finally, I will also compare the pre-test with the post-test directly in order to be able to compare better with the corresponding results of the control group. For the control group, the results of the pre-test are then compared with those of the post-test.

3.2.1.1 Self-concept in the experimental group

Fig. 24 shows an overview of the mean values of the results regarding the self-concept of the experimental group. In general, the students assessed themselves very positively. Therefore, it is possible that a ceiling effect occurs here. The figure shows that the data did not change much at first glance. For each of the variables, a Mann-Whitney U test was therefore carried out between every two tests in order to determine significant changes.

Figure 24: Mean values of the data for individual items of the self-concept test of the experimental group for pre-test, intermediate test, and post-test. Error bars show the standard deviation.

Pre- and intermediate test

From pre- to intermediate test, the mean values for all seven items were larger than before. The differences are small. The biggest difference was for the first item with 0.32, the second biggest for the second item with 0.30, both significant. The first item refers to whether a student thinks they understand the physics material well. The second refers to whether the material is retained, i.e. whether one still knows the material after a longer period of time. Otherwise, the differences were either 0.2 or smaller or even 0.05 or smaller. Table 11 provides an overview of the differences.

Item #	Pre	Int	Difference	p-Value
1	4.06	4.38	0.32	**0.01**
2	3.90	4.20	0.30	**<0.01**
3	3.95	4.16	0.20	0.1
4	3.95	4.00	0.05	0.79
5	3.75	3.80	0.05	0.53
6	3.93	4.10	0.17	0.32
7	4.52	4.54	0.03	0.82

*Table 11: Overview of the means for individual items of **experimental** group's self-concept for **pre- and intermediate** test, difference, and p-values (significant values are printed in bold).*

Intermediate and post-test

From intermediate test to post-test, an opposing picture emerged. The mean values decreased for all seven items. The difference was greatest for item 6, which was the only item where the difference was significant. For all other items, the difference was 0.2 or less and not significant. Item 6 was about the assessment of a pupil in terms of how the teacher assesses the pupil. An overview of means, differences and the corresponding p-values can be found in Table 12.

Item #	Int	Pos	Difference	p-Value
1	4.38	4.18	-0.20	0.18
2	4.20	4.01	-0.19	0.13
3	4.16	4.04	-0.11	0.44
4	4.00	3.85	-0.15	0.29
5	3.80	3.70	-0.10	0.5
6	4.10	3.76	-0.34	**0.04**
7	4.54	4.43	-0.12	0.57

*Table 12: Overview of the means for individual items of **experimental** group's self-concept for **intermediate and post**-test, difference, and p-values (significant values are printed in bold).*

Pre- and post-test

If we compare the pre-test with the post-test of the experimental group directly with regard to the self-concept, there are no major differences. The biggest difference is found in item 6 with 0.17. The mean value increased in items 1, 2 and 3, but decreased in all other items. None of the differences were significant. The corresponding details, as well as p-values, can be found in Table 13.

Item #	Pre	Pos	Difference	p-Value
1	4.06	4.18	0.12	0.33
2	3.90	4.01	0.11	0.29
3	3.95	4.04	0.09	0.43
4	3.95	3.85	0.10	0.45
5	3.75	3.70	0.06	0.89
6	3.93	3.76	0.17	0.28
7	4.52	4.43	0.09	0.71

*Table 13: Overview of the means for individual items of **experimental** group's self-concept for **pre- and post**-test, difference, and p-values (significant values are printed in bold).*

3.2.1.2 Self-concept in the control group

In the control group, the picture is similar to that in the experimental group if only the pre-test and the post-test are considered. The mean values for the individual items are shown in Fig. 25. Mean values were rather high overall, which is why a ceiling effect may also be noticeable here. The results were again tested for significance using the Mann-Whitney U test.

Figure 25: Mean values of the data for individual items of the self-concept test of the control group for pre-test with the post-test. Error bars show the standard deviation.

The differences are not very large here either, none of them is greater than 0.15. The mean value has only decreased in item 7. However, it has increased for all other items. No difference was statistically significant. The exact values, with p-values, can be found in Table 14.

Item #	Pre	Pos	Difference	p-Value
1	3.76	3.83	0.07	0.57
2	3.39	3.47	0.08	0.58
3	3.75	3.90	0.15	0.23
4	3.62	3.71	0.09	0.35
5	3.69	3.74	0.05	0.66
6	3.75	3.80	0.04	0.73
7	4.34	4.19	0.15	0.18

*Table 14: Overview of the means for individual items of **control** group's self-concept for **pre- and post**-test, difference, and p-values (significant values are printed in bold).*

3.2.2 Analysis of self-concept between the groups

In this section, the results regarding self-concept from the corresponding pre-tests with the post-tests of the two groups are compared. Since the intermediate test only took place in the experimental group, a comparison is not possible here. First the pre-tests and then the post-tests are compared. Again, differences are tested for significance with the Mann-Whitney U test.

3.2.2.1 Self-concept in the pre-tests

Fig. 26 compares the mean values of the individual items of the self-concept part of the pre-tests from both groups. At first glance, the values are similar everywhere. In all cases, the mean values from the experimental group are greater than those from the control group.

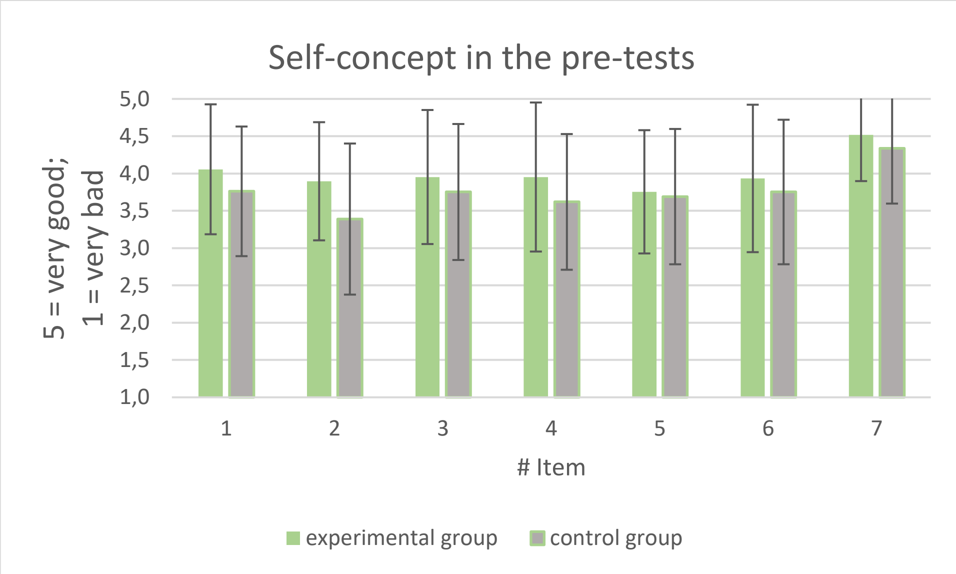

Figure 26: Mean values of the data for individual items of the self-concept test in the pre-test of the experimental and control group. Error bars show the standard deviation.

The difference is greatest in item 2. There, the mean value of the experimental group is greater by 0.51 than that of the control group. The differences were relatively large in item 1 (0.30) and item 4 (0.33). All these differences were statistically significant. Item 1, as mentioned above, relates to how well one understands the material, while item 2 relates to how well one remembers it, and item 4 to how well one participates in class. For all other items, the difference was 0.20 or less and not significant. Details can be found in Table 15.

Item #	Experimental	Control	Difference	p-Value
1	4.06	3.76	0.30	**<0.01**
2	3.90	3.39	0.51	**<0.001**
3	3.95	3.75	0.20	0.1
4	3.95	3.62	0.33	**<0.01**
5	3.75	3.69	0.06	0.59
6	3.93	3.75	0.18	0.12
7	4.52	4.34	0.18	0.08

*Table 15: Overview of the means for individual items of both group's self-concept for the **pre**-test, difference, and p-values (significant values are printed in bold).*

3.2.2.2 Self-concept in the post-tests

The post-tests also show a comparable picture. The differences in the mean values are not large at first glance, as can be seen in Figure 27. The mean values of the experimental group for items 4 and 5 were smaller than those of the control group, in all other cases they were larger.

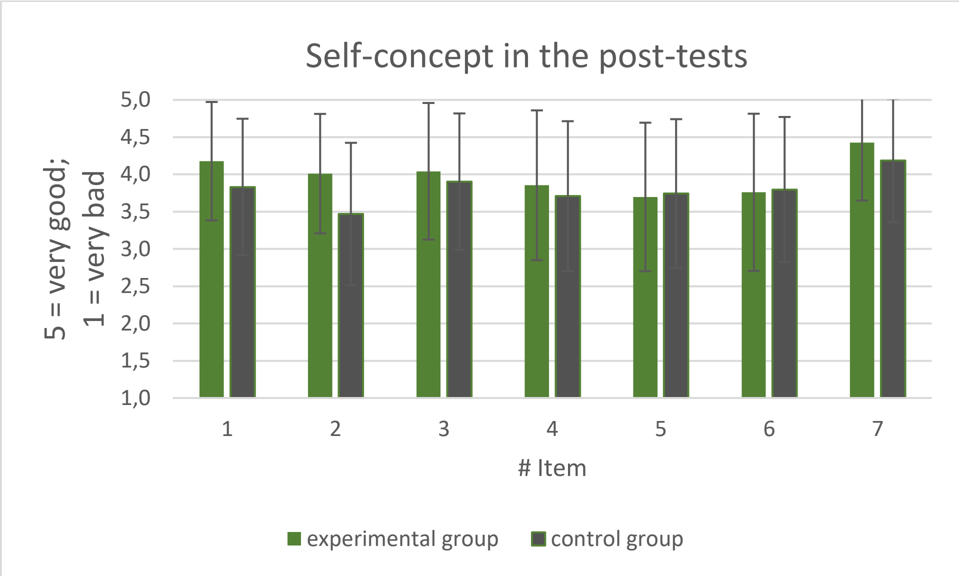

Figure 27: Mean values of the data for individual items of the self-concept test in the post-test of the experimental and control group. Error bars show the standard deviation.

The Mann-Whitney U test shows that the differences are significant in three cases. Namely, for item 1 with a difference of 0.35, for item 2 with a difference of 0.54 and for item 7 with a difference of 0.24. The latter refers to how a student assesses their own future performance in physics. In all these three cases, the mean value of the experimental group was greater than that of the control group. The exact values can be found in Table 16.

Item #	Experimental	Control	Difference	p-Value
1	4.18	3.83	0.35	**<0.01**
2	4.01	3.47	0.54	**<0.001**
3	4.04	3.90	0.14	0.24
4	3.85	3.71	0.15	0.28
5	3.70	3.74	0.05	0.91
6	3.76	3.80	0.04	0.92
7	4.43	4.19	0.24	**0.01**

*Table 16: Overview of the means for individual items of both group's self-concept for the **post**-test, difference, and p-values (significant values are printed in bold).*

3.2.3 Self-concept in class comparison

It is interesting to compare the individual classes in terms of self-concept. This will happen in a rough way in this section. For this purpose, the mean value of all 7 self-concept items was formed for each of the total of 12 classes. These were then compared in Figure 28 for the pre-test and sorted by group. Figure 29 shows the same results for the post-test. Here, the order of the classes was not changed, so that class 1 in the first figure also corresponds to class 1 in the second figure.

For the pre-test, the picture is rather homogeneous. The range of all mean values for the pre-test is 0.96. The self-concept of one class was thus nowhere considerably greater than that of another. Again, it should be noted that a ceiling effect may have occurred here.

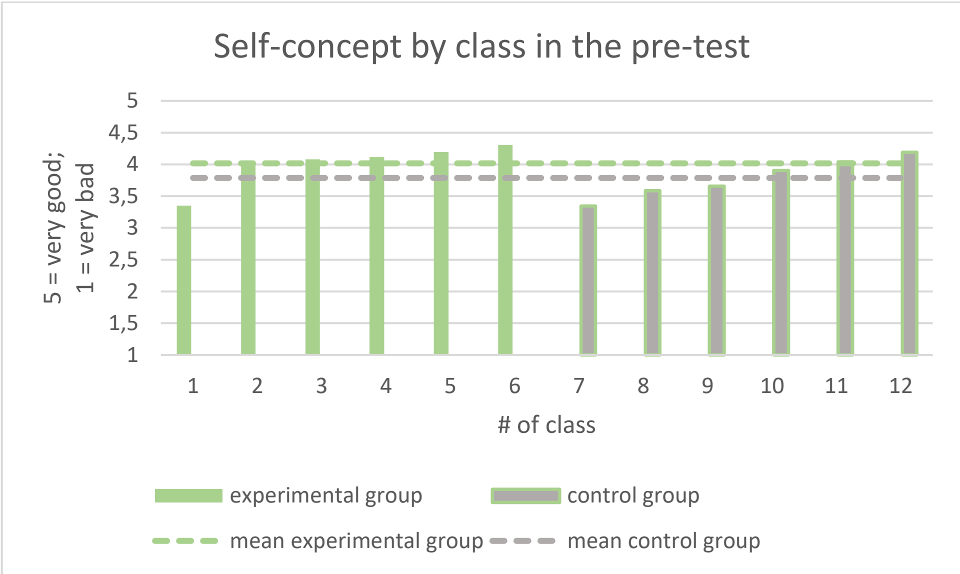

Figure 28: Mean values of all 7 self-concept items of the individual classes for the pre-test. Mean of the experimental group = 4.01; mean of the control group = 3.79

For the post-test, the comparison of classes shows a very similar picture to the pre-test with regard to the self-concept. The range has even decreased a little to 0.84. It should also be emphasised that the mean values have hardly changed from pre-test to post-test. This is in line with the observations above, where there were hardly any significant differences in the individual items between the pre- and post-test.

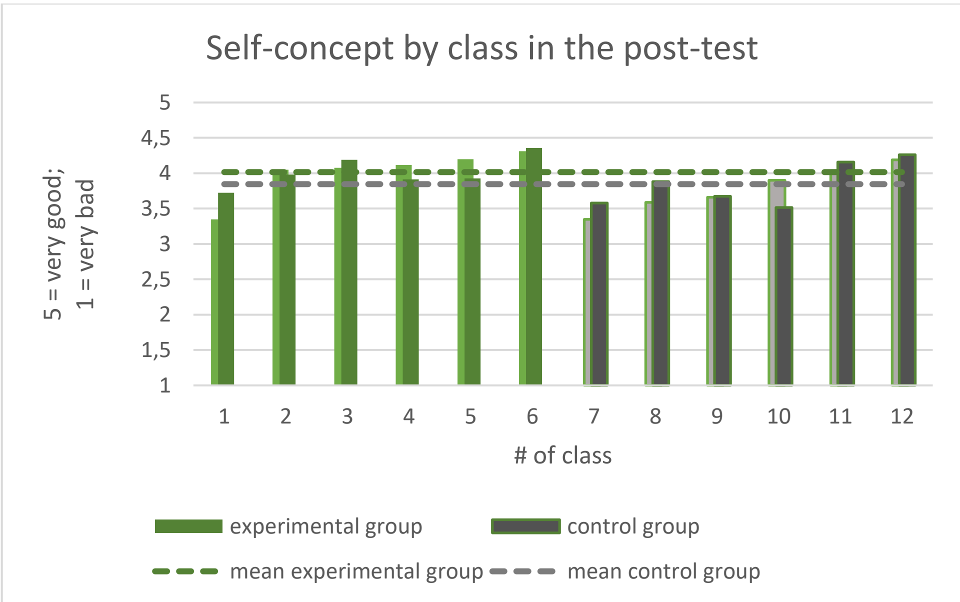

Figure 29: Mean values of all 7 self-concept items of the individual classes for the post-test. Mean of experimental group = 4.01; mean of control group = 3.84; bars in the background show pre-test results of these classes

3.3 Interest

In this section, the data is analysed in terms of interest. This is done in exactly the same way as in the section on self-concept. This means that comparisons are first made within the group. Again, the intermediate test must be taken into account. After that, comparisons are made between the groups. Finally, the individual classes are analysed. As before, the pre-test in the experimental group was completed by 106 participants, the intermediate test by 90 participants and the post-test by 97 participants. In the control group, 113 people completed the pre-test and post-test. In this section, too, the person codes were not included for the same reasons as in the section on the self-concept (there were few significant gains or losses in interest within the individual groups).

3.3.1 Analysis of interest within groups

I want to again start with the experimental group. First, I will give an overview of the data. After that, I will compare the pre- with the intermediate test and then the intermediate with the post-test. To be able to better compare the data with that of the control group, I also analyse the differences between the pre- and post-test of the experimental group. Finally, I will also analyse the pre- and post-test for the control group.

3.3.1.1 Interest in the experimental group

As can be seen in Figure 30, the interest of the students in the experimental group was above 3 (neutral). There was no average interest value below 3. As with the results for self-concept, the differences between the individual tests are not strikingly different at first glance. Again, a Mann-Whitney U test was conducted for each variable between every two tests to determine whether there were significant differences.

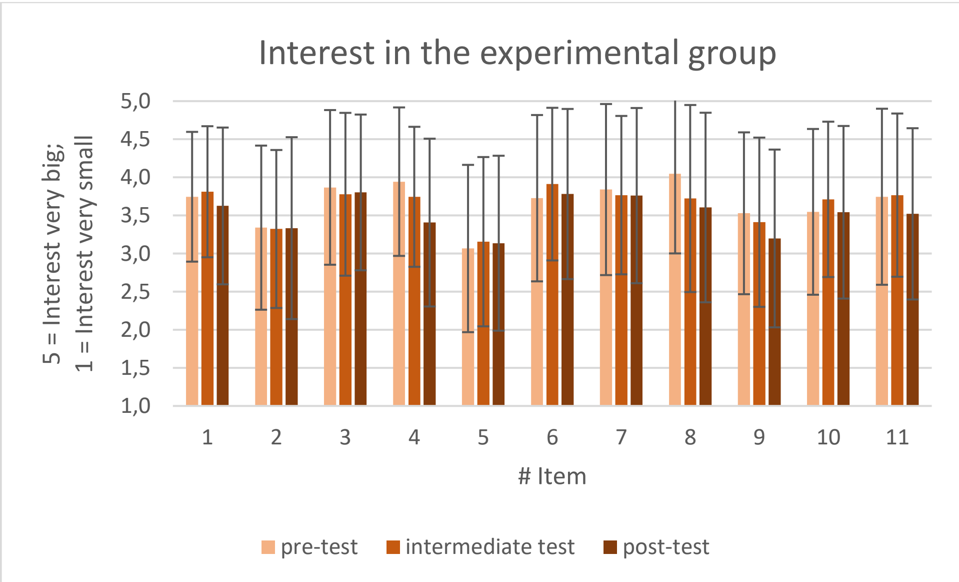

Figure 30: Mean values of the data for individual items of the interest test of the experimental group for pre-test, intermediate test, and post-test. Error bars show the standard deviation.

<u>Pre- and intermediate test</u>

The differences in the interest of the experimental group between the pre- and intermediate tests were very small on average. For 6 out of the 11 items the mean value decreased, for the others it increased. The difference was nowhere significant. The biggest difference was in item 4 with 0.20 and item 8 with 0.32, whereby in both cases the mean was lower in the intermediate test than in the pre-test. These differences were also not statistically significant. Item 4 was about interest in the different types of radiation and their effects, and item 8 was about interest in building a device to detect radiation. An overview of the mean values, their differences and the corresponding p-values can be found in Table 17.

Item #	Pre	Int	Difference	p-Value
1	3.75	3.81	0.07	0.46
2	3.34	3.32	-0.02	0.88
3	3.87	3.78	-0.09	0.59
4	3.94	3.74	-0.20	0.09
5	3.07	3.16	0.09	0.58
6	3.73	3.91	0.18	0.27
7	3.84	3.77	-0.07	0.55
8	4.05	3.72	-0.32	0.07
9	3.53	3.41	-0.12	0.54
10	3.55	3.71	0.16	0.32
11	3.75	3.77	0.02	0.95

*Table 17: Overview of the means for individual items of **experimental** group's interest for **pre- and intermediate** test, difference, and p-values (significant values are printed in bold).*

Intermediate and post-test

If we compare the intermediate test with the post-test in the experimental group, we see that interest has decreased for the most part. In only 2 items did interest increase very slightly on average, in all other items it decreased. The biggest difference was in item 4, where the average was 0.34 lower in the post-test than in the intermediate test. This difference was also statistically significant. Table 18 shows an overview of the relevant values.

Item #	Int	Post	Difference	p-Value
1	3.81	3.63	-0.19	0.20
2	3.32	3.33	0.01	0.92
3	3.78	3.80	0.02	0.91
4	3.74	3.41	-0.34	**0.04**
5	3.16	3.14	-0.02	0.92
6	3.91	3.78	-0.13	0.50
7	3.77	3.76	-0.01	0.86
8	3.72	3.60	-0.12	0.56
9	3.41	3.20	-0.21	0.24
10	3.71	3.54	-0.17	0.33
11	3.77	3.52	-0.25	0.12

*Table 18: Overview of the means for individual items of **experimental** group's interest for **intermediate and post**-test, difference, and p-values (significant values are printed in bold).*

Pre- and post-test

The analysis of the differences between pre-test and post-test confirms what was already foreseeable in the analysis with the intermediate test. The mean values decreased in all items except for items 5 and 6. The biggest difference is in item 4 with 0.54 and in item 8 with 0.44. Both values are statistically significant. Also worth noting is the comparatively large difference of 0.33 for item 9, which was just barely not statistically significant. Item 9 is about the interest in coming up with a method to detect radiation. An overview of relevant values can be found in Table 19.

Item #	Pre	Post	Difference	p-Value
1	3.75	3.63	-0.12	0.49
2	3.34	3.33	-0.01	0.81
3	3.87	3.80	-0.07	0.66
4	3.94	3.41	-0.54	**<0.001**
5	3.07	3.14	0.07	0.66
6	3.73	3.78	0.05	0.70
7	3.84	3.76	-0.08	0.64
8	4.05	3.60	-0.44	**0.01**
9	3.53	3.20	-0.33	0.06
10	3.55	3.54	-0.01	0.99
11	3.75	3.52	-0.22	0.10

*Table 19: Overview of the means for individual items of **experimental** group's interest for **pre- and post**-test, difference, and p-values (significant values are printed in bold).*

3.3.1.2 Interest in the control group

An overview of the mean values regarding interest in the control group can be found in Figure 31. A similar picture to the control group can be seen when comparing pre- and post-test. The mean values also decreased in all items, except for item 2 and item 5. In general, the interest in the control group was also comparatively high. Here only one mean value, in item 5, was below 3, otherwise they were always higher. A Mann-Whitney U test was again used to check which differences were significant.

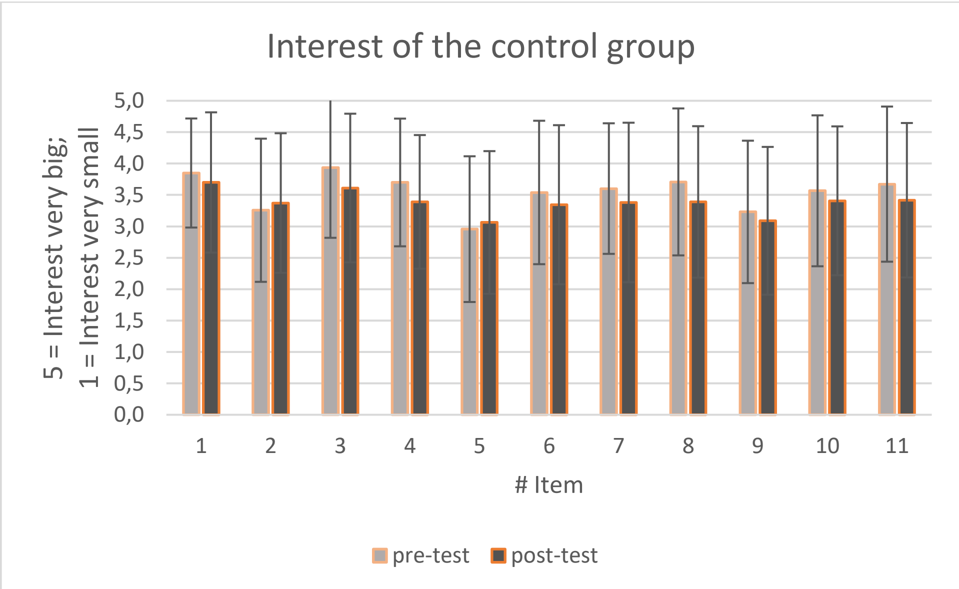

Figure 31: Mean values of the data for individual items of the interest test of the control group for pre-test with the post-test. Error bars show the standard deviation.

Again, the largest differences were for item 4 and item 8, with 0.31 and 0.32 respectively, and additionally for item 3, with 0.33. All these differences were statistically significant. Item 3 was about interest in the military and peaceful use of nuclear energy. An overview can be found in Table 20.

Item #	Pre	Post	Difference	p-Value
1	3.85	3.70	-0.15	0.64
2	3.26	3.37	0.12	0.43
3	3.94	3.61	-0.33	**0.03**
4	3.70	3.39	-0.31	**0.03**
5	2.96	3.06	0.11	0.48
6	3.54	3.35	-0.19	0.27
7	3.60	3.38	-0.22	0.28
8	3.71	3.39	-0.32	**0.04**
9	3.23	3.09	-0.14	0.47
10	3.57	3.41	-0.16	0.29
11	3.67	3.42	-0.26	0.10

*Table 20: Overview of the means for individual items of **control** group's interest for **pre- and post**-test, difference, and p-values (significant values are printed in bold).*

3.3.2 Analysis of interest between the groups

As with self-concept, the results regarding interest are now compared between the corresponding pre- and post-tests. Again, a comparison with the intermediate test of the experimental group is not possible, as there was no such test in the control group. Any differences were again tested for significance using the Mann-Whitney U test.

3.3.2.1 Interest in the pre-tests

First, the pre-tests of the two groups are compared. It is interesting to see if there was a different baseline in terms of interest. Taking a first look at Figure 32, this was not the case. As mentioned above, the interest of both groups was above 3 (neutral). The mean scores were greater for items 1, 3 and 10 from the control group than from the experimental group. In all other cases, the mean values of the experimental group were greater.

Figure 32: Mean values of the data for individual items of the interest test in the pre-test of the experimental and control group. Error bars show the standard deviation.

A closer analysis of the differences shows that some are statistically significant. This was the case for item 8 and item 9, where the mean value of the experimental group was 0.34 and 0.30 higher than that of the control group. In item 4 and item 7, the difference of 0.24 is noteworthy, but just barely not significant. An overview can be found in Table 21. As mentioned above, item 8 and item 9 were about developing a device or method to detect radiation. Item 4 relates to interest in different types of radiation and their effects. Item 7 relates to interest in the clinical or medical application of radiation.

Item #	Experimental	Control	Difference	p-Value
1	3.75	3.85	0.10	0.38
2	3.34	3.26	-0.08	0.58
3	3.87	3.94	0.07	0.43
4	3.94	3.70	-0.24	0.07
5	3.07	2.96	-0.11	0.45
6	3.73	3.54	-0.19	0.23
7	3.84	3.60	-0.24	0.07
8	4.05	3.71	-0.34	**0.03**
9	3.53	3.23	-0.30	**0.048**
10	3.55	3.57	0.02	0.71
11	3.75	3.67	-0.07	0.75

*Table 21: Overview of the means for individual items of both group's interest for the **pre**-test, difference, and p-values (significant values are printed in bold).*

3.3.2.2 Interest in the post-tests

The comparison of the post-tests again shows a similar picture to the comparison of the pre-tests. For items 1 and 2, the mean value of the control group was higher in the post-test, in all other cases that of the experimental group was higher. Figure 33 shows the corresponding mean values.

Figure 33: Mean values of the data for individual items of the interest test in the post-test of the experimental and control group. Error bars show the standard deviation.

For item 6 and item 7, the difference, at 0.44 and 0.38 respectively, was also statistically significant. Item 6 referred to the interest in working in a nuclear power plant, item 7, as already mentioned above, to the clinical or medical application of radiation. Otherwise, there were no significant differences. Table 22 provides an overview.

Item #	Experimental	Control	Difference	p-Value
1	3.63	3.70	0.07	0.35
2	3.33	3.37	0.04	0.64
3	3.80	3.61	-0.19	0.34
4	3.41	3.39	-0.02	0.89
5	3.14	3.06	-0.07	0.65
6	3.78	3.35	-0.44	**0.01**
7	3.76	3.38	-0.38	**0.03**
8	3.60	3.39	-0.21	0.15
9	3.20	3.09	-0.11	0.48
10	3.54	3.41	-0.13	0.50
11	3.52	3.42	-0.10	0.69

Table 22: Overview of the means for individual items of both group's interest for the ***post****-test, difference, and p-values (significant values are printed in bold).*

3.3.3 Interest in class comparison

In this section, as before, the interest of the individual classes is compared in an approximative way. Again, an average score was taken for all 11 items of interest for each class. These were plotted side by side in Figure 34.

All classes show the same interest with only minor differences. Accordingly, there is no notable difference between the two groups. The mean values for each group are very close to each other, 3.69 for the experimental group and 3.56 for the experimental group. The mean values are also close to the middle of the scale, so we do not need to be concerned about a ceiling effect.

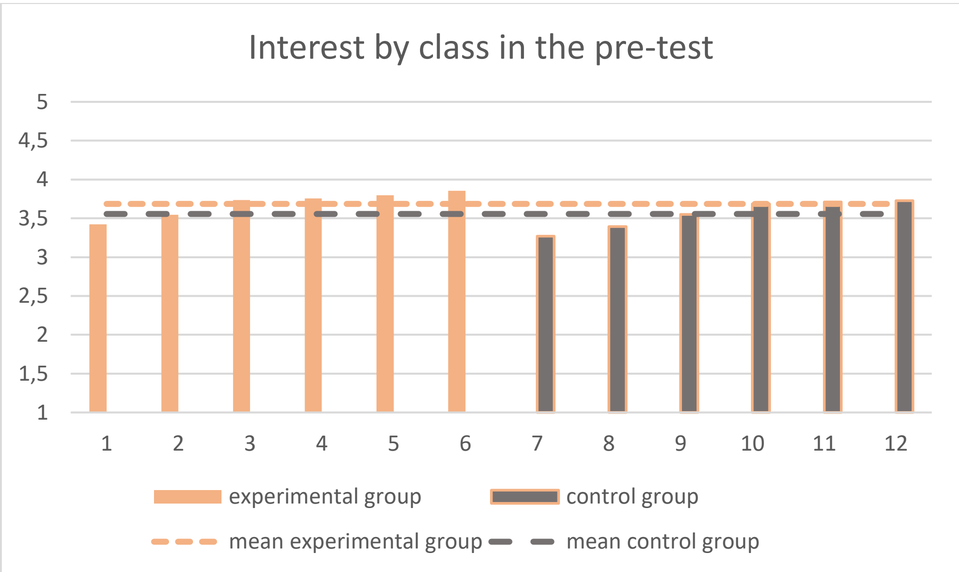

Figure 34: Mean values of all 11 interest items of the 12 individual classes for the pre-test. Mean of the experimental group = 3.69; mean of the control group = 3.56

If we look at Figure 35, which shows the interest of the individual classes for the post-test, we see the same situation. The interest has hardly changed in the individual classes. Neither have the mean values. Again, this fits with the observation from before, where there were some significant but not large differences between the groups.

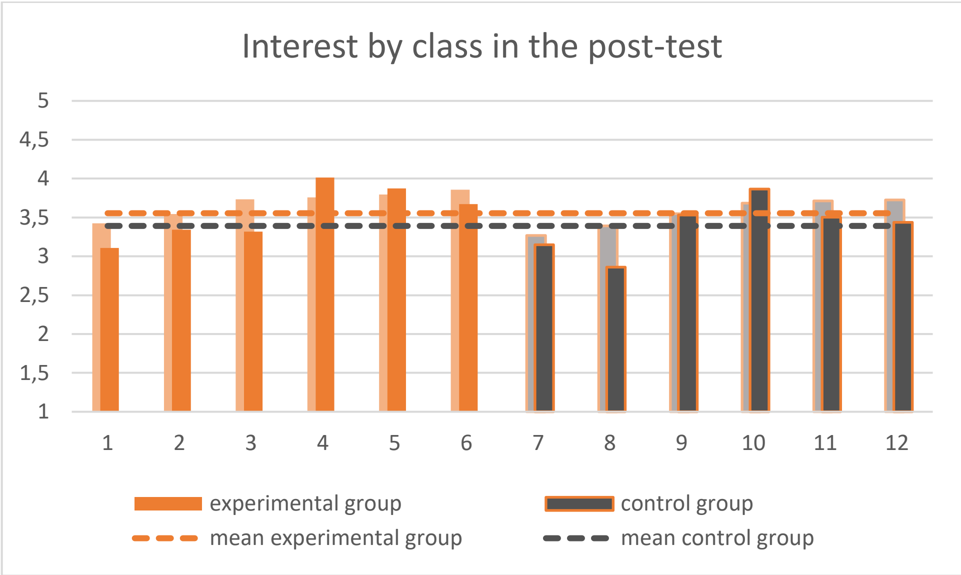

Figure 35: Mean values of all 11 interest items of the 12 individual classes for the post-test. Mean of the experimental group = 3.56; mean of the control group = 3.39; bars in the background show pre-test results of these classes

3.4 Tanodo question

As mentioned in section 2, a question was attached to the questionnaire asking the students directly how they liked the past lessons. The response options corresponded to a 5-point scale, ranging from 5 = "It was very enjoyable", to 3 = "No strong opinion either way", to 1 = "It was very boring". This question is also usually found at the end of a Classbook and as mentioned earlier, is called a "tanodo" question. It was asked both in the intermediate test of the experimental group and in the post-test of both groups, although the wording changed a little. The intermediate test, as already mentioned, was issued immediately after the use of the Classbook. The question thus referred to the lessons of the previous week. A varying amount of time could pass before the post-test was administered by different teachers, so the question was slightly reworded there. It then referred to all lessons on radioactivity of the previous weeks. This question was answered by 90 students in the intermediate test and 97 students in the post-test in the experimental group. In the control group 101 students answered this question in the post-test. The number was lower here, because some students of the control group filled out the pre-test twice instead of the post-test.

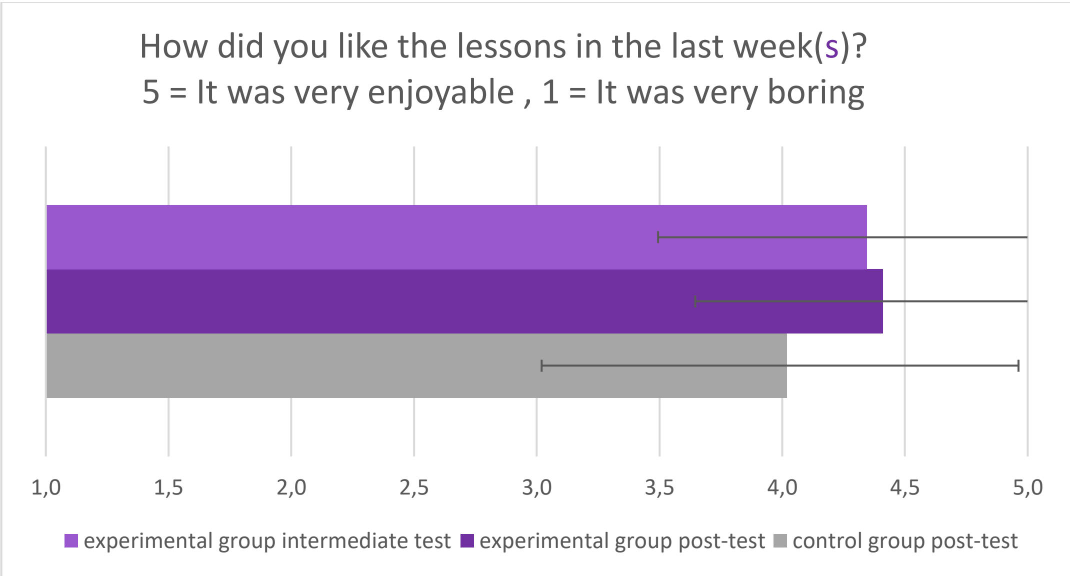

Figure 36: mean values of the answers to the tanodo question

Figure 36 shows the average result of the two groups. Both groups chose "enjoyable" more often than "boring". The mean values are 4.34 for the intermediate test of the experimental group, 4.41 for the post-test of the experimental group and 4.02 for the post-test of the control group. The large values again indicate a ceiling effect. At first glance, the individual mean values are comparable. A Mann-Whitney U test should now show whether there were any significant differences. First, I will compare the mean values within the groups, which means the intermediate test and the post-test of the experimental group in this case. Then we compare the results of the post-test of the respective groups.

3.4.1 Tanodo within groups

Within the groups, only the intermediate test of the experimental group can be compared with their post-test. Figure 37 shows the corresponding histograms. The distribution is comparable, as they show a peak at the same point and the values fall off comparably. Both data sets are not normally distributed, as a Shapiro-Wilk test shows both for the intermediate test ($W = 0.73$, p-value < 0.001) and for the post-test ($W = 0.74$, p-value < 0.001). Levene's test guarantees variance homogeneity for the two data sets ($F = 0.27$, p-value $= 0.60$). According to the Mann-Whitney U test, there is no significant difference between the two data sets ($W = 4447$, p-value $= 0.70$).

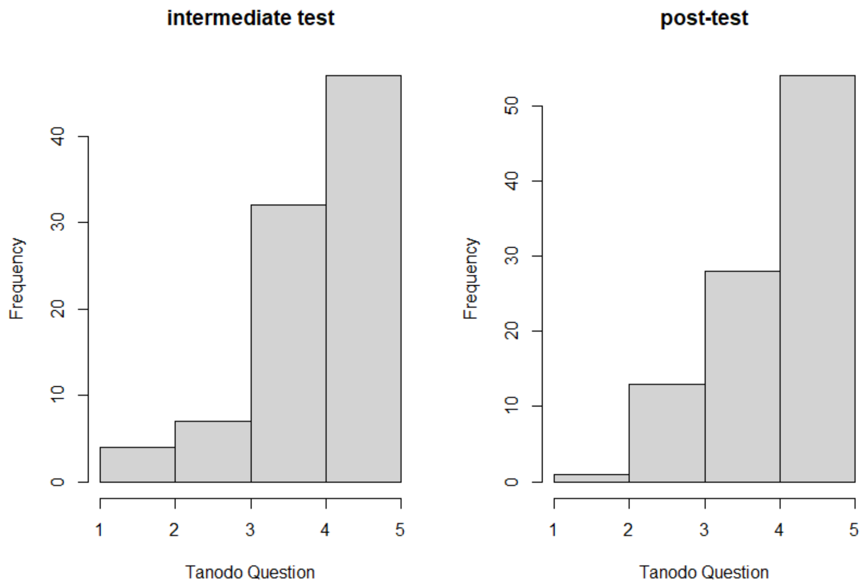

Figure 37: Histograms of the answers to the tanodo questions for the experimental group; left: intermediate test, right: post-test

3.4.2 Tanodo between groups

Between the groups, I will compare the two post-tests of experimental and control group respectively. Figure 38 shows the corresponding histograms. They seem comparable again because of the peak at the same place and the comparable decrease of the values. For the post-test of the experimental group, the Shapiro Wilk test has already shown that the data are not normally distributed ($W = 0.74$, $p < 0.001$). The data of the post-test of the control group are also not normally distributed according to this test ($W = 0.83$, $p < 0.001$). The variance homogeneity is given, as Levene's test shows ($F = 1.66$, $p = 0.20$). These two data sets differ significantly from each other according to the Mann-Whitney U-test ($W = 5906.5$, $p < 0.01$).

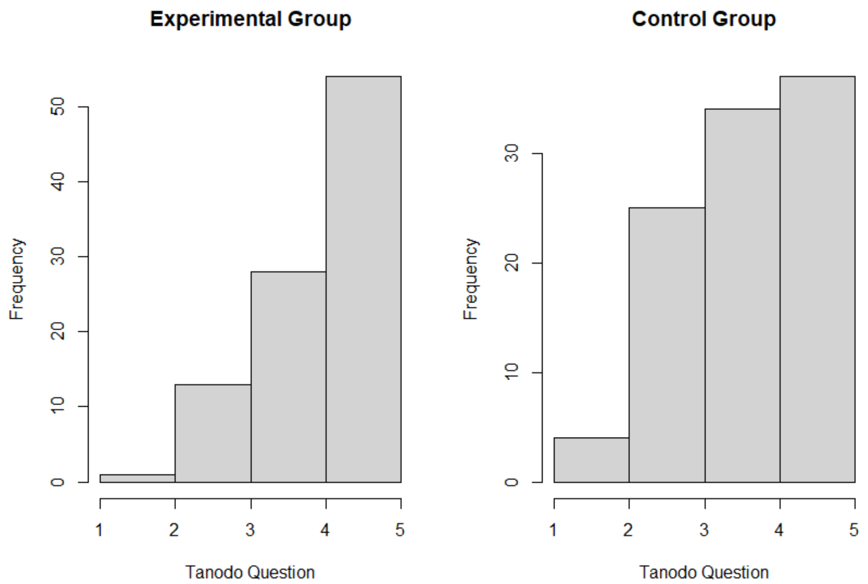

Figure 38: Histograms of the answers to the tanodo questions for the post-tests; left: Experimental Group, right: Control Group

3.4.2.1 *Effect size for the tanodo question*

Because the difference between the post-test results of the two classes on the tanodo question was significant, I will now give the relevant effect sizes. As a reminder, the mean of the experimental group was 4.41, that of the control group was 4.02. The difference is therefore 0.39. Cliff's delta for these data is -0.23, 95% CI [-0.37, -0.08], which indicates a small effect. Vargha-Delaney A of 0.38 also indicates a small effect.

3.4.3 Tanodo question in class comparison

Finally, I will roughly analyse how the tanodo question was answered in the individual classes. Figure 39 shows an overview. There are no obvious extreme differences in the individual classes. If we look specifically at the experimental group, it is noticeable that the mean value has visibly decreased in one class. On the other hand, it increased visibly in two other classes. In the other three classes of the experimental group, the mean values are very similar. This fits well with the findings above, where no significant difference was found in the experimental group between the intermediate test and the post-test.

The significant difference between the two post-tests of the two groups is also visible in Figure 39. Of the six classes in the experimental group, five were above the average of the control group. Conversely, not a single class in the control group was above the average of the experimental group.

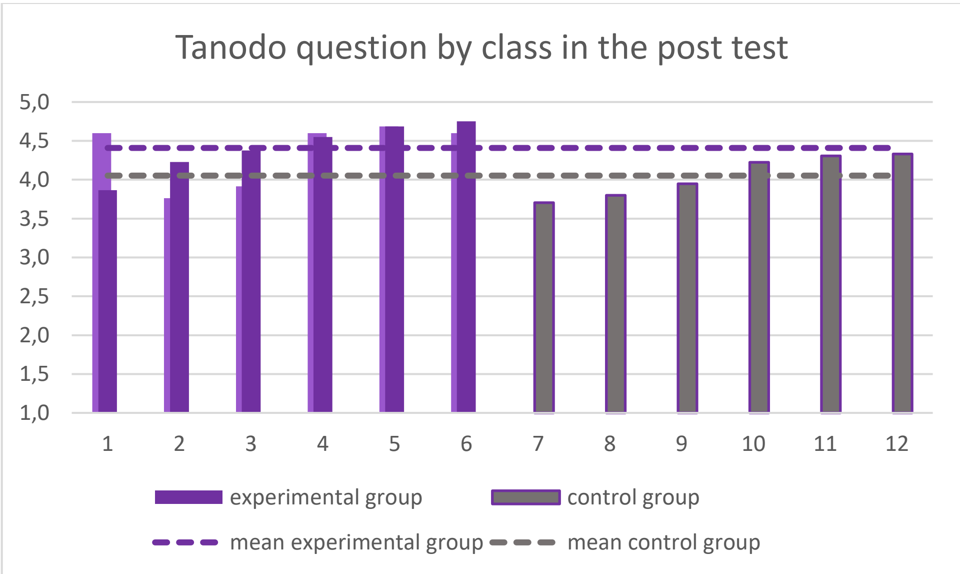

Figure 39: Mean values of the answers to the tanodo question of the 12 individual classes for the post-test. Mean of the experimental group = 4.41; mean of the control group = 4.05; bars in the background show intermediate test results of the experimental group

3.5 Teacher questionnaire and interviews

3.5.1 Interviews

As already mentioned, the interviews with the teachers in the control group were not transcribed. Instead, only a summary protocol was prepared. This revealed both similarities and differences between the teachers in the control group. As far as the general preparation for the lessons is concerned, two teachers stated that they mainly prepare the material for the lessons themselves, while the third stated that they adhere very closely to the textbook. In all three cases, the preparation for radioactivity lessons does not differ much from the preparation for physics lessons on other topics. One teacher uses the textbook frequently, the others less so. However, all teachers use exercises from either the textbook or workbooks, which are worked out by the students. Two teachers make more use of videos and films in radioactivity lessons. Among those, one teacher also uses simulations. In all three cases, experiments are conducted on the topic of radioactivity. In particular, the law of distance and also the shielding of ionising radiation are tested experimentally. In addition, everyday objects are also measured with a Geiger-Mühler counter. Two teachers stated that independent work on the part of the students would account for about one third of the teaching time on the topic of radioactivity. Frontal teaching would account for about half of the teaching time, while the rest would be spent on other activities. The third teacher stated that in their lessons, independent work on the part of the students and frontal teaching would be in a ratio of about 1 to 1. The focus of the three teachers was broadly similar with some differences in detail. All of them were concerned about protection from radioactivity and creating a good understanding of when radioactivity is dangerous and when it is harmless. The first teacher focused on the dangers of radioactivity in detail. For the second teacher, practical relevance was particularly important. This means protection against radioactivity and, in particular, the possible applications and opportunities of radioactivity. The third teacher also emphasised that the exact functioning of certain technical applications, such as nuclear power plants or atomic bombs, were deliberately not a focus.

In summary, it can be said that the teachers' preparation for teaching radioactivity is no different from that for other topics. Moreover, they all carried out the same experiments in class. The media used varied to some extent. As far as the proportion of independent work by students and frontal teaching is concerned, there are no noticeable differences here either. The focal points that the teachers have set for the topic of radioactivity in their lessons are in any case comparable and only differ in detail, if at all.

The Classbook "The Radiation Around Us" focuses on natural sources of radiation and specifically on the fact that the human body contains radioactive atoms. This content was also addressed in the control group. All three teachers in the control group stated that they had measured either everyday objects, the human body, or both with a Geiger-Mühler counter. The teacher questionnaire therefore also asked how many teaching hours were spent on these aspects (see below). If one compares the lessons in the control group with HEC in general, one finds differences. The activity of the teacher is very low when a Classbook is used. Frontal teaching is practically non-existent. All 3 teachers in the control group stated that frontal teaching made up about 50% of their lessons. In this respect, there is a huge difference between the groups as far as teaching with the Classbook is concerned.

<u>Teacher questionnaire</u>

The aim of the teacher questionnaire was to find out which topics related to radioactivity were taught by the teachers who participated in the study and for how long. Unfortunately, I did not receive data from two teachers, one from the experimental group and one from the control group.

One teacher in the control group stated that they taught all subject areas except nuclear fusion. All other teachers indicated that they taught all subject areas. The amount of time spent on individual topics was very similar on average. On average, the longest time was spent teaching the history of radioactivity. The second longest was natural sources of radiation. Then came, in descending order, the effects of and protection against ionising radiation, applications, nuclear fission, and types of radiation. On average, atomic structure, isotopes and half-life were taught the shortest. An overview can be found in Figure 40.

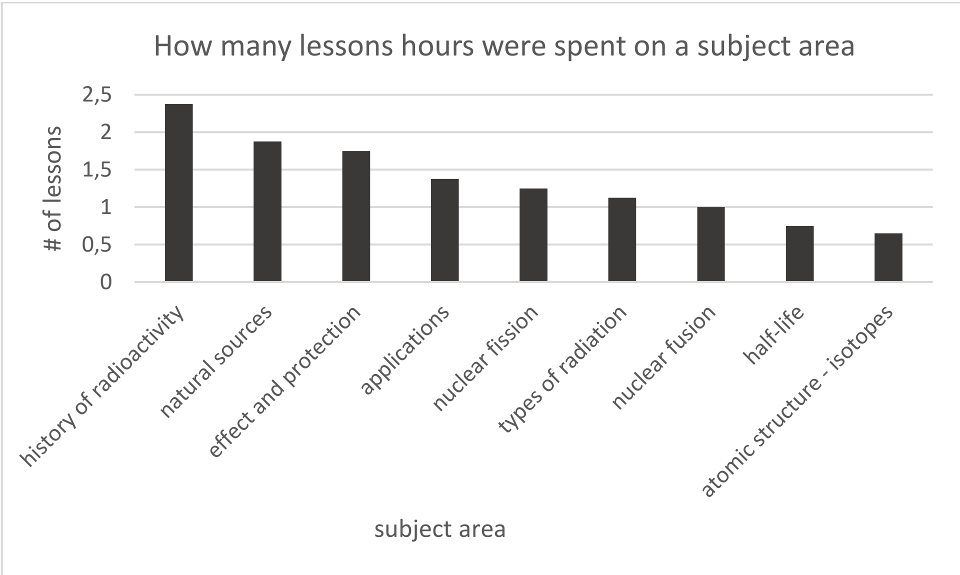

Figure 40: Overview of how many teaching hours on average were spent on individual subject areas.

Of particular interest for this study is the topic "natural sources of radiation". This aspect is fully explored in the Classbook "The Radiation Around Us". The Classbook also contains minor information on historical aspects, units, or types of radiation. However, the focus is clearly on natural sources of radiation and natural radiation exposure. In the teacher questionnaire, the teachers of the experimental group were asked to indicate for how many lessons they had taught the Classbook. These lessons were assigned to the topic area "natural sources of radiation". Figure 41 compares the average number of Classbook lessons in the experimental group with the average number of lessons on the topic "natural sources of radiation" in the control group. As can be seen, this subject area was taught noticeably longer in the experimental group due to the use of the Classbook. Both teachers in the experimental group stated that they had used the Classbook in three lessons. One teacher in the control group stated that they had taught this topic for half an hour, the other for a whole hour.

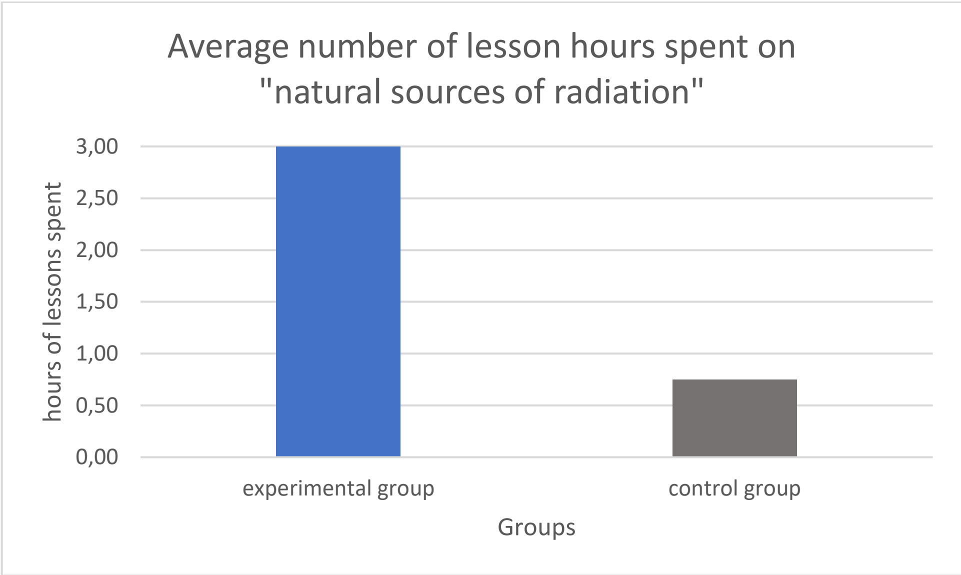

Figure 41: Average number on lessons spent on the topic of "natural sources of radiation", for the experimental group these equate to the lesson hours spent on the Classbook

4 Discussion

The aim of the present study was to find out whether students like the Classbook “The Radiation Around Us” and whether its use leads to a better understanding of radioactivity as a natural phenomenon. The results obtained will be discussed in this chapter with this in mind. First, I will give an overview of the data obtained and its relevance to the two research questions. I will then answer my research questions and make relevant arguments. This is followed by a discussion of what benefits can be drawn from the results of my research and point out some of its implications. Finally, I will state which aspects of my study were less successful and what could have been done better. I will also give an outlook on what future studies on HEC or Classbooks could look at.

4.1 Summary

In this chapter, I will summarise the most important results I obtained from the data. I will give an overview of the results regarding conceptual understanding, self-concept, interest and finally the tanodo question. I will highlight all those results that are useful to answer the research questions.

4.1.1 Conceptual understanding

4.1.1.1 Within groups

At the beginning I would like to summarise the results within the groups. It is interesting to see where there have been improvements or also worsening. The average test score of the experimental group increased from 3.3 to 4.6 points. This difference was significant, as a Mann-Whitney U test showed, with a large effect size. In contrast, the average score of the control group increased from 3.9 to 4.6. This difference was also significant, with a small effect size. So, both groups benefited from the teaching about radioactivity and were able to achieve a better test score. However, effect was greater overall in the experimental group. In detail, 6 questions were answered correctly significantly more often in the experimental group. The increase of 42 percentage points in question 10, which deals with ground and cosmic radiation, was gratifying. This is dealt with directly by the Classbook “The Radiation Around Us”. The difference of only 4 percentage points in question 1 was significant, but disappointingly small. The content of this question is aimed at the radioactivity of everyday objects, which is the central core of the Classbook. Question 1 was therefore examined more closely for the number of answer options selected, as already mentioned. Also interesting was the large increase in question 3, which has to do with decay types in terms of content. These were only mentioned incidentally in the Classbook. The large increase of 45 percentage points for this question is therefore more likely to be due to the rest of the lessons on radioactivity. In the control group, this was also the only question that was answered significantly more often in the post-test than in the pre-test, with a difference of 40 percentage points. All other questions had a maximum difference of 9 percentage points. In the experimental group, questions 4, 7 and 9 were also answered correctly with a significantly higher frequency. A closer look at the individual items also leads to the view that overall, the experimental group benefited more from the lessons on radioactivity.

As already mentioned, question 1 was analysed in more detail. There were 10 everyday objects to choose from and a person was supposed to tick all of them where a measuring device for radioactivity would click, which means display a measured value. The choices were 1. smartphone, 2. ambient air, 3. schoolchild, 4. concrete walls of a hospital reception, 5. concrete walls of a hospital room where X-rays were taken, 6. concrete walls of a school building, 7. cigarette smoke, 8. sunlight, 9. microwave popcorn and 10. banana. One point was given for this task only if all 10 choices were ticked. I am of the opinion that a person who, on average, has ticked significantly more answer options than another person also has a better understanding of the content of this question. Therefore, this question was also examined again in this respect. In the experimental group, an average of 3.3 answer options were initially ticked, then 5.2. It is unclear whether this difference is significant, as the corresponding variances were very heterogeneous here. In the control group, an average of 4.3 and then 4.8 answer options were ticked. This difference was not significant. At first glance, there was a greater increase in the experimental group.

Again, a closer look was taken at the individual answer options. The fact that ambient air, a school child, concrete walls in general and bananas are radioactive is mentioned directly in the Classbook. I therefore expected a large increase in these answer options in the experimental group. At the same time, it is also shown that the natural radiation exposure in the shade and in the sun does not differ noticeably. I therefore expected a possible drop for the corresponding answer choice. As for the other answer choices, I did not expect any significant changes. In fact, in the experimental group, every answer option was ticked more often in the post-test, except for cigarette smoke, which is not mentioned in the Classbook. There was a significant increase in the five answer options 2, 3, 4, 6 and 10, which is in line with expectations. Specifically, regarding the concrete walls of a hospital room where X-rays were taken, the difference was not significant because this answer option was already ticked very often in the pre-test. Nevertheless, the value still increased. A similar picture emerges for the smartphone. There were no significant changes in the other answer options, which is also in line with expectations. It is pleasing that sunlight was ticked more often, although not significantly. In general, it is disappointing that schoolchild was only ticked by 34% of the participants in the post-test. The last question of the Classbook, which is very important in HEC, is about human radioactivity. I would have expected a much bigger result.

In the control group, the picture is weaker. Significant differences were found for the schoolchild and the concrete walls. The exception is again the concrete walls of a hospital room where X-rays were taken, because this option was also ticked very often in the pre-test. Thus, there were significant increases in 3 answer options. Otherwise, there were no significant changes. It is interesting that in the control group, the school child was ticked by 46% of the participants in the post-test, more than in the experimental group. This may be related to the experiments conducted by the teachers in the control group (see section 3.5.1). In any case, there were fewer significant increases in the individual answer options in the control group compared to the experimental group. This fits with the averages from above. Again, it appears that the experimental group benefited a little more.

4.1.1.2 Between groups

In this section, I would like to summarise the comparison of the results between the groups. It is interesting to see which group had more prior knowledge, and which group had, disregarding other factors, performed better in the end. The analysis showed that the test result of 3.3 points of the experimental group was significantly different from the test result of 3.9 points of the control group. Thus, prior knowledge seems to have been slightly greater in the control group. In particular, one question was answered significantly more often by the control group than by the experimental group. This was question 9, which was about the danger of radioactivity.

As far as question 1 is concerned, the experimental group first ticked 3.3, the control group 4.3 answer options on average. This difference was also significant. In detail, 4 answer options were selected more often by students in the control group. These included ambient air, school child, concrete walls of a hospital room where X-rays were taken, and microwaved popcorn. This supports the assumption that the control group's prior knowledge was greater in this area in particular as well. It should also be noted that neither group answered the question correctly in the pre-test significantly more often than the other. This is again due to the fact that only one point was awarded for this question if everything was ticked.

The results in the post-test were not significantly different with 4.6 points each. There was only one question that was answered significantly more often by the experimental group and that was question

10. Otherwise, there were no significant differences. So, you can say that the two groups achieved about the same outcome. I would have expected the experimental group to do better in questions 1 and 10, since these two questions are directly addressed by the Classbook. That they did better in the case for question 10 is pleasing. Furthermore, question 1 is again given a closer look. In the post-test, the experimental group ticked 5.2, the control group 4.8 answer options, although the difference was not significant. There were two response options that were ticked significantly more often by the experimental group, and these were concrete walls in the reception of a hospital and concrete walls in a school building. As mentioned earlier, I expected a large increase for these answer options, so this result is encouraging. Disappointingly, the control group selected "schoolchild" more often, although not significantly. The experimental group may have gained a small advantage through the Classbook as far as this question is concerned.

4.1.1.3 Learning gains

The results so far suggest that the prior knowledge of the control group was a little greater and that both groups performed comparably well in the end. The tendency is that the experimental group did better. Therefore, the gains were examined more closely and above all the individual learning gains of individual students were compared, which could be calculated with the help of the person codes. The individual average gain of individual students was 1.4 points in the experimental group and 0.8 points in the control group. The difference between the two scores was significant, the effect size small. These results fit well with those mentioned previously. The experimental group caught up with the control group despite having less prior knowledge.

Individual increases in the number of answer choices selected were also calculated for question 1. The matched values resulted in an average increase in the number of selected answer options of 1.9 for the experimental group and 0.4 for the control group. This difference was significant at medium effect size. Although question 1 was not answered correctly significantly more often in the experimental group than in the control group, these results show that there was a greater improvement in the experimental group. This circumstance is pleasing and corresponds with the expectations.

Overall, the improvement of students in the experimental group with regard to questions 1 and 10 was significantly greater than in the control group.

4.1.2 Self-concept

In order to better understand the influence of the Classbook on the students' self-concept, they were asked about this in an intermediate test directly after using the Classbook. Finally, they were asked again at the end of the radioactivity lesson along with the questions about conceptual understanding.

It was my expectation that the self-concept in the experimental group would increase immediately after using the Classbook. After the students returned to traditional teaching, I assumed that the self-concept would remain the same or decrease. In the control group, I assumed that the self-concept would remain the same.

It turned out that the self-concept of both groups was high in the pre-test. In the comparison of classes, it could be seen that the mean values of individual classes on the 5-part scale were consistently above 3 and, in some cases, even above 4. The self-concept was slightly better in the experimental group than in the control group, where there were significant differences in three of the seven items. A comparison of the post-tests showed a similar picture. Again, the self-concept of the experimental group was significantly better in three items than that of the control group. In both cases, there was a significant difference in item 1, which is about how well one understands the material, and in item 2, which is about how well one remembers the material. This was checked using Chi-Squared and Fisher's Exact test.

As far as the actual groups are concerned, the self-concept in the experimental group improved slightly. In the items 1 and 2 just described, the students' self-concept was significantly better in the intermediate test, so directly after using the Classbook. The differences were very small, 0.32 and 0.30 respectively. This could be due to the ceiling effect. From the intermediate to the post-test, the students' self-concept worsened slightly again. The difference was only significant in item 6, which is about how well one assesses one's own performance. This overall picture is confirmed by the comparison of pre- and post-test in the experimental group. Here it was shown that there were no significant differences. The same was the case in the control group, where there were no significant differences between pre- and post-test. So, my expectation was partly fulfilled, although the effect was much smaller than expected.

4.1.3 Interest

As far as interest is concerned, I had the same expectations as for self-concept. An increase in the experimental group in the intermediate test and constancy or a decrease in the post-test. I did not expect any significant change in the control group. I tended to expect a decrease in interest, as physics education research has documented that student interest in physics tends to decrease after traditional instruction.

Just like self-concept, the interest in both groups was already high in the pre-test. The comparison of classes showed that the mean value of each class on the 5-part scale was everywhere above 3, but not above 4. The interest of the experimental group was minimally higher than that of the control group. In two of the 11 items, the interest of the experimental group was significantly higher. The same was the case in the post-tests. In two items, the interest of the experimental group was significantly higher, otherwise there were no significant differences. Again, this was checked with Chi-Squared and Fisher's Exact test. Comparing classes, it could be seen that interest decreased in most classes.

In detail, the interest of the experimental group did not change significantly anywhere directly after the application of the Classbook. From the intermediate to post-test, the interest in one item decreased significantly. Overall, interest tended to decrease in both cases. In the comparison of pre- and post-test it could be seen that the interest in two items decreased significantly. It was similar in the control group, where interest in three items decreased significantly in the comparison between pre- and post-test. These results were surprising for me with regard to the experimental group. I did not expect that the Classbook would have so little influence on the students' interest. As for the control group and traditional teaching, the data was in line with expectations.

4.1.4 Tanodo question

The so-called "tanodo" question, which is usually found at the end of every Classbook, was also interesting for this study. In this question, students indicate whether they found the lessons boring or enjoyable. The question was added to the intermediate test of the experimental group and to the post-tests of both groups. The analysis of the question provides information on whether the students enjoyed the lessons with the Classbook and whether they liked it significantly better than the traditional lessons. I expected that the students would like the lessons with the Classbook in general and even more than the traditional lessons.

On a 5-part scale, where 1 stands for "very boring" and 5 for "very enjoyable", the mean value in the intermediate test of the experimental group was 4.34, in the post-test 4.41. This difference was not significant according to the Mann-Whitney U test. Students thus found both the lessons with the Classbook and the traditional lessons afterwards to be enjoyable. The high value after the intermediate test corresponds to expectations. The fact that the value was still high in the post-test is surprising. It is therefore unclear whether the students liked the lessons with the Classbook better than the traditional lessons. It is possible, for example, that the students enjoyed the whole topic, regardless of the teaching method.

The control group had a mean score of 4.02 on this question. The difference to the experimental group turned out to be significant, which was shown with a Mann-Whitney U test with a small effect size. This result is also surprisingly high. In principle, a ceiling effect could have occurred again in this case, masking a potentially larger difference. Overall, however, the experimental group liked the lessons significantly better. Whether this is because of the Classbook or something else, is unclear.

4.2 Interpretation

Now I will restate the research questions and answer them based on the above findings.

4.2.1 Research question 1

The first research question was: “Do students who have used the Classbook have a significantly better understanding of radioactivity and specifically that radioactivity is a naturally occurring phenomenon than students who have not been taught with the Classbook?”

As far as the overall understanding of radioactivity is concerned, the experimental group did not have a significantly better understanding after the lessons than the control group, because both groups had the same average final test result. It could therefore be argued that in both cases the teaching brought the students to the same certain level of understanding, regardless of prior knowledge. At the same time, answering the research question only makes sense in my view if one can assume that both groups had comparable prior knowledge. In the results, however, you can see that the prior knowledge of the control group was greater. Therefore, I looked specifically at the gains of the students and specifically at the individual gains of certain students. The individual gains were greater in the experimental group, which indicates that this group improved more noticeably.

In terms of radioactivity being a naturally occurring phenomenon, the experimental group did have a better understanding. This was tested directly by question 1 and indirectly by question 10 of the knowledge test. Both questions were answered significantly better by the experimental group. The latter was answered correctly by 71% of the students in the experimental group in the post-test, 38 percentage points more than the control group. This difference was significant. Question 1 was not answered correctly significantly more often by the experimental group, but students in this group selected more answer options. This also speaks for a better understanding in this matter. Because the prior knowledge in the control group was also greater regarding question 1, individual increases were also considered here. The analysis of the individual increases in particular showed that the increase in the number of selected answer options was significantly greater in the experimental group than in the control group. A medium effect size in this aspect is good and certainly expandable.

So, from my point of view, the research question can be answered with a yes and no overall. Considering the topic of radioactivity as a whole, the Classbook "The Radiation Around Us" did not lead to an understanding greater than that of students who did not use TRAU. However, in comparison to students who did not learn with TRAU, TRAU was able to bring out a greater degree of understanding about radioactivity as a natural and everyday occurrence. In principle, one can say that one of the goals of HEC is achievable with this Classbook. In this respect, it is only disappointing that only about one third of the students in the experimental group chose the answer option "schoolchild" for question 1. This is far from 90% of the students who are supposed to correctly predict the last question in the Classbook. This target was not met in this study. There are several possible reasons for this. As mentioned at the beginning, the Classbook was slightly changed. It is conceivable that this change contributed to the poorer result. It is also unclear how closely the teachers in the experimental group followed the HEC procedure. Finally, it is also possible that the Classbook simply did not have the desired effect. It is also unclear whether this is due to the Classbook itself or to the HEC method.

4.2.2 Reserach question 2

The second research question was: "Do the students like the lessons with the Classbook and is there a difference to classes where the Classbook is not used (traditional physics lessons)?"

The students in the experimental group definitely liked the lessons with the Classbook. This is shown on the one hand by the high self-concept, which was high immediately after using the Classbook. The same applies to interest. The answers to the tanodo question at the end also show that the students enjoyed the lessons.

At the same time, not many significant differences can be noticed with regard to the classes that did not use the Classbook and only enjoyed traditional lessons. Interest and self-concept were comparable in both groups at the end with no noticeable significant differences. Only the tanodo question was answered significantly better in the experimental group than in the control group. However, it is unclear whether this is due to the Classbook alone. Even within the experimental group, there were no noticeable differences in terms of interest and self-concept when comparing lessons with the Classbook (intermediate test) and the subsequent traditional lessons (post-test). In the analysis of the self-concept, there were signs that the Classbook may have an advantage over traditional physics teaching. At the same time, a ceiling effect could be assumed, which might have masked greater differences.

From my point of view, the first part of the research question can be answered with an unequivocal yes. The students obviously liked the lessons with the Classbook. As far as the difference to the control group and the traditional physics lessons is concerned, I would answer the question with a cautious "no". Indicators that the Classbook brings an advantage in this respect were sparse and, in my opinion, not sufficient to assume such an advantage.

4.3 Implications

This study is intended to help physics educators make decisions about their teaching methods. As mentioned above, the number of teaching concepts for teaching radioactivity is very limited. Moreover, I am not aware of any teaching concept that deals so directly with the naturalness and everydayness of radioactivity. At the same time, HEC and thus "The Radiation Around Us" is very different from traditional physics teaching in Austria. Whereas in the latter, history, the structure of the atom and types of radiation are in the foreground at the beginning, in the Classbook this is only a minor matter or is not even mentioned in detail. The way in which HEC is carried out, on the one hand frontally and at the same time with the involvement of the whole class, is also different. It is therefore conceivable that physics teachers feel a certain reluctance to use such a foreign teaching concept. In this case, empirical research can be a remedy, showing that the concept is useful and that the students enjoy it. Both could be largely demonstrated in this study. The Classbook did not lead to a better understanding of all the contents related to radioactivity. At the same time, the students have enjoyed the Classbook and through its use, they have gained a better understanding of radioactivity as an everyday phenomenon and that it occurs naturally. In my opinion, the Classbook is a suitable alternative or rather a possible supplement to traditional radioactivity teaching or other teaching concepts on radioactivity. At the same time, I believe that no teacher should feel pressured to use the Classbook. Ultimately, there are different teaching styles that suit one teacher more than another. HEC can be one such style and complement a teacher's personal repertoire. Finally, it should be mentioned that the Classbook or HEC may be beginner-friendly and therefore interesting for teachers who have just started teaching, because one can stick to a certain procedure and does not have to prepare anything oneself. Wintersteller (in press) has investigated how differently experienced teachers have coped with the Classbook.

4.4 Limitations and further research

As mentioned at the beginning, this study tried to keep external validity high, which would indicate that my results can also be reproduced in other classes. In my view, this was largely successful. Except for the use of the Classbook, the classes were taught as usual by their own teacher in their own classroom. The internal validity suffered as a result. From my point of view, this was visible in the analysis of self-concept, interest and the tanodo question. In some cases, these aspects were better in the experimental group than in the control group. However, it is unclear whether these improvements can be attributed to the Classbook.

At this point, the changes that were made to the Classbook should also be mentioned. With none of the classes a cloud chamber was built and used, as was actually intended in the Classbook. It is possible that greater effects could have been observed in the experimental group if this part had not been omitted.

As far as conceptual understanding is concerned, noticeable differences between the individual classes of the experimental group were visible in the class comparison. On the one hand, there were classes that did particularly badly or particularly well in the pre-test. On the other hand, there were individual classes that improved particularly strongly, while others hardly improved at all. It would have been interesting to find out how these large differences came about. This could have been done, for example, by specifically questioning the respective teachers about it. In retrospect, this was not done in this study due to time and logistical reasons.

Finally, it should be mentioned that "The Radiation Around Us" is only one of several Classbooks that have all been developed for HEC. Future studies could examine other Classbooks to better describe the effectiveness of HEC as a whole. In addition, "The Radiation Around Us" could also be tested more extensively and especially in its original version. A laboratory study that allows for high internal validity is also imaginable.

5 References

Apolin, M. (2018). *Big Bang 4.* (1st ed.). Österreichischer Bundesverlag Schulbuch GmbH & Co. KG

Burde, J. (2018). *Konzeption und Evaluation eines Unterrichtskonzepts zu einfachen Stromkreisen auf Basis des Elektronengasmodells*.

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, *114*(3), 494–509. https://doi.org/10.1037/0033-2909.114.3.494

Coletta, V. P. & Steinert, J. J. (2020). Why normalized gain should continue to be used in analyzing preinstruction and postinstruction scores on concept inventories. *Physical review*, *16*(1). https://doi.org/10.1103/physrevphyseducres.16.010108

Curriculum of the AHS-Untestufe. BGBl. No. 88/1985 as amended by BGBl. II No. 1/2023 of 02 June 2023.

DeKay, N., & Maidl, R. (2012). Identifying & Resolving Problematic Student Thinking About Ionizing Radiation. In *Proceedings of The National Conference On Undergraduate Research (NCUR) 2012*. Weber State College, Utah

Demtröder, W. (2017). Experimentalphysik 4. In *Springer-Lehrbuch*. Springer Science+Business Media. https://doi.org/10.1007/978-3-662-52884-6

Eid, M., Gollwitzer, M. & Schmitt, M. (2013). *Statistik und Forschungsmethoden: Lehrbuch. Mit Online-Materialien*. Beltz.

Eijkelhof, H. (1996). Radiation Risk and Science Education. *Radiation Protection Dosimetry*, *68*(3), 273–278. https://doi.org/10.1093/oxfordjournals.rpd.a031878

Eijkelhof, H., Klaassen, C. W. J. M., Lijnse, P. & Scholte, R. L. J. (1990). Perceived incidence and importance of lay-ideas on ionizing radiation: Results of a delphi-study among radiation-experts. *Science Education*, *74*(2), 183–195. https://doi.org/10.1002/sce.3730740205

Fürnstahl, H., Janisch, S., Wolfbauer, M. (2020). *Physik heute 4.* VERITAS.

Hafele, A. (2012). Exploring Learning Difficulties Associated with Understanding Ionizing By Radiation. In *Proceedings of The National Conference On Undergraduate Research (NCUR) 2012*. Weber State College, Utah

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. https://doi.org/10.1119/1.18809

Hayes, A. F. (2000). Randomization tests and the equality of variance assumption when comparing group means. *Animal Behaviour*, *59*(3), 653–656. https://doi.org/10.1006/anbe.1999.1366

Henriksen, E. K. (1996). Laypeople's Understanding of Radioactivity and Radiation. *Radiation Protection Dosimetry*, *68*(3), 191–196. https://doi.org/10.1093/oxfordjournals.rpd.a031863

Hess, A. T. & Hess, J. R. (2017). Understanding tests of the association of categorical variables: the Pearson chi-square test and Fisher's exact test. *Transfusion*, *57*(4), 877–879. https://doi.org/10.1111/trf.14057

Hoffmann, L., Häußler, P. & Lehrke, M. (1998). *Die IPN-Interessenstudie Physik*. Ipn.

Holzinger, E. S. (2022). *The public's knowledge on radioactivity in Austria affected by the time passed after graduation and the type of school attended* [master's thesis, University of Vienna]. https://doi.org/10.25365/thesis.71474

Hopf, M. & Schecker, H. (2018). Schülervorstellungen zu fortgeschrittenen Themen der Schulphysik. In *Schülervorstellungen und Physikunterricht*. https://doi.org/10.1007/978-3-662-57270-2_11

Hopf, M. & Schecker, H. (2021). Unterrichtskonzeptionen zu fortgeschrittenen Themen der Schulphysik. In *Unterrichtskonzeptionen für den Physikunterricht*. https://doi.org/10.1007/978-3-662-63053-2_12

Hoyer, C. & Girwidz, R. (2020). Animation and interactivity in computer-based physics experiments to support the documentation of measured vector quantities in diagrams: An eye tracking study. *Physical review*, *16*(2). https://doi.org/10.1103/physrevphyseducres.16.020124

Hull, M. G. & Hopf, M. (2020). Student understanding of emergent aspects of radioactivity. *International Journal of Physics and Chemistry Education*, *12*(2), 19–33. https://doi.org/10.12973/ijpce/20529

Hussy, W., Schreier, M. & Echterhoff, G. (2013). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor*. Springer-Verlag.

Itakura, K. (2019). *Hypothesis-Experiment Class (Kasetsu)*.

Jansky, A. J. (2019). *Die Rolle von Schülervorstellungen zu Wahrscheinlichkeit und Zufall im naturwissenschaftlichen Kontext* [doctoral thesis, University of Vienna].

Johnson, A. & Maidl, R. (2014). *Students Coming to Understand Ionizing Radiation - A Radiation Literacy Challenge.* South Dakota Center for the Advancement of Math and Science Education. Black Hills State University

Klaassen, C. W. J. M. (1995). *A problem-posing approach to teaching the topic of radioactivity*. https://dspace.library.uu.nl/handle/1874/637

Krieger, H. (2019). Grundlagen der Strahlungsphysik und des Strahlenschutzes. In *Springer eBooks*. https://doi.org/10.1007/978-3-662-60584-4

Lijnse, P., Eijkelhof, H., Klaassen, C. W. J. M. & Scholte, R. L. J. (1990). Pupils' and mass-media ideas about radioactivity. *International Journal of Science Education*, *12*(1), 67–78. https://doi.org/10.1080/0950069900120106

Mann, H. J. & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

Mašin, C. & Grois, G. (2020). *Physik verstehen 4.* Österreichischer Bundesverlag Schulbuch GmbH & Co. KG

Millar, R., Klaassen, K. & Eijkelhof, H. (1990). Teaching about radioactivity and ionising radiation: an alternative approach. *Physics Education*, *25*(6), 338–342. https://doi.org/10.1088/0031-9120/25/6/310

Nissen, J. M., Talbot, R. W., Thompson, A. N. & Van Dusen, B. (2018). Comparison of normalized gain and Cohen'sdfor analyzing gains on concept inventories. *Physical review*, *14*(1). https://doi.org/10.1103/physrevphyseducres.14.010115

Pöge, A. (2008). Persönliche Codes 'reloaded'. *Methoden, Daten, Analysen (mda)*, 2(1), 59-70. https://nbnresolving.org/urn:nbn:de:0168-ssoar-126543

Prather, E. E. (2005). Students' Beliefs About the Role of Atoms in Radioactive Decay and Half-life. *Journal of geoscience education*, *53*(4), 345–354. https://doi.org/10.5408/1089-9995-53.4.345

Prather, E. E. & Harrington, R. (2001). Student Understanding of Ionizing Radiation and Radioactivity. *The journal of college science teaching*, *31*(2), 89–93. https://eric.ed.gov/?id=EJ636023

Rasch, B., Friese, M., Hofmann, W. J. & Naumann, E. (2009). *Quantitative Methoden 2. Einführung in die Statistik für Psychologen und Sozialwissenschaftler* (Springer-Lehrbuch) (3., erw. Aufl. 2010). Springer.

Reinders, H., Ditton, H., Gräsel, C. & Gniewosz, B. (2011). *Empirische Bildungsforschung: Strukturen und Methoden*. Springer Science & Business Media.

Schecker, H. & Duit, R. (2018). Schülervorstellungen und Physiklernen. In *Schülervorstellungen und Physikunterricht*. https://doi.org/10.1007/978-3-662-57270-2_1

Shapiro, S. M. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Slovic, P. (1996). Perception of Risk from Radiation. *Radiation Protection Dosimetry*, *68*(3), 165–180. https://doi.org/10.1093/oxfordjournals.rpd.a031860

Vargha, A., Delaney, H. D. & Vargha, A. (2000). A Critique and Improvement of the „CL" Common Language Effect Size Statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, *25*(2), 101. https://doi.org/10.2307/1165329

Wilcox, B. R. & Lewandowski, H. (2017). Developing skills versus reinforcing concepts in physics labs: Insight from a survey of students' beliefs about experimental physics. *Physical review*, *13*(1). https://doi.org/10.1103/physrevphyseducres.13.010108

Wilhelm, T. & Hopf, M. (2014). Design-Forschung. In *Springer eBooks* (S. 31–42). https://doi.org/10.1007/978-3-642-37827-0_3

Wintersteller, M. (in print). *Hypothesis Experiment Class - teacher's influence on teaching HEC.* [master's thesis, University of Vienna]

Zwickl, B. M., Hirokawa, T., Finkelstein, N. D. & Lewandowski, H. (2014). Epistemology and expectations survey about experimental physics: Development and initial results. *Physical Review Special Topics-physics Education Research*, *10*(1). https://doi.org/10.1103/physrevstper.10.010120

6 Appendix

6.1 Questionnaire for the experimental group (German)

Deckblatt 1: (pre-Test)

Liebe Schülerin, lieber Schüler!

Im Physikunterricht lernt man über viele verschiedene Themen. Dazu gehört auch die Radioaktivität. Wir möchten in unseren Masterarbeiten einen genauen Blick darauf werfen, wie dieses Thema unterrichtet wird.

Wir würden von dir gerne erfahren, wie sehr dich dieses Thema interessiert und wie du dich selbst im Physikunterricht einschätzt. Darüber hinaus würden wir gerne erfahren, was du schon über Radioaktivität weißt.

Bekomme kein schlechtes Gefühl, wenn du dir bei einer Antwort nicht sicher bist. Jede Antwort kann dazu beitragen, den Physikunterricht für dich in Zukunft interessanter zu machen. Aus diesem Grund würden wir dich bitten, den folgenden Fragebogen sorgfältig auszufüllen.

Die Teilnahme ist **freiwillig.** Die Daten werden gespeichert und ausschließlich zur Bearbeitung unserer beiden Masterarbeiten verwendet. Die Daten werden nur in **anonymisierter Form** veröffentlicht. Alle Daten werden **strikt vertraulich** behandelt. Die Beantwortung dauert etwa **10- 15 Minuten.**

Wenn du Fragen hast, wende dich gerne an deine Lehrperson oder direkt an uns, unter folgender E-Mail-Adresse: [redacted]

Vielen Dank für deine Mitarbeit!

[redacted]

<u>Deckblatt 2: (Zwischentest)</u>

Liebe Schülerin, lieber Schüler!

Du hast nun schon ein wenig über Radioaktivität im Alltag gehört. Wir würden nun gerne erfahren, ob sich etwas an deinem Interesse zu diesem Thema geändert hat. Ebenso, ob du dich selbst jetzt anders einschätzt als vorher.

Aus diesem Grund würden wir dich wieder bitten, den folgenden Fragebogen sorgfältig auszufüllen. Deine Antworten können wieder dazu beitragen, den Physikunterricht in Zukunft interessanter zu gestalten.

Es gilt nach wie vor: Die Teilnahme ist **freiwillig.** Die Daten werden gespeichert und ausschließlich zur Bearbeitung unserer beiden Masterarbeiten verwendet. Die Daten werden nur in **anonymisierter Form** veröffentlicht. Alle Daten werden **strikt vertraulich** behandelt. Die Beantwortung dauert etwa **5-10 Minuten.**

Wenn du Fragen hast, wende dich gerne an deine Lehrperson oder direkt an uns, unter folgender E-Mail-Adresse: [redacted]

Vielen Dank für deine Mitarbeit!

[redacted]

<u>Deckblatt 3: (post-Test)</u>

Liebe Schülerin, lieber Schüler!

Du hast nun einiges im Unterricht zum Thema Radioaktivität gehört. Wir würden nun gerne erfahren, ob sich etwas an deinem Interesse zu diesem Thema geändert hat. Ebenso, ob du dich selbst jetzt anders einschätzt als vorher. Uns interessiert auch, was du jetzt über Radioaktivität weißt und was du dir gemerkt hast.

Aus diesem Grund würden wir dich wieder bitten, den folgenden Fragebogen sorgfältig auszufüllen. Bekomme auch dieses Mal kein schlechtes Gefühl, wenn du dir bei einer Antwort nicht sicher bist. Jede deiner Antworten kann wieder dazu beitragen, den Physikunterricht in Zukunft interessanter zu gestalten.

Es gilt nach wie vor: Die Teilnahme ist **freiwillig.** Die Daten werden gespeichert und ausschließlich zur Bearbeitung unserer Masterarbeiten verwendet. Die Daten werden nur in **anonymisierter Form** veröffentlicht. Alle Daten werden **strikt vertraulich** behandelt. Die Beantwortung dauert etwa **10-15 Minuten.**

Wenn du Fragen hast, wende dich gerne an deine Lehrperson oder direkt an uns, unter folgender E-Mail-Adresse: [redacted]

Vielen Dank für deine Mitarbeit!

[redacted]

Teil 1

Gib bitte an, ob dein Interesse an den folgenden Inhalten zu Radioaktivität für dich sehr groß, groß, mittel, gering oder sehr gering ist. Kreuze das entsprechende Kästchen an. Bitte kreuze in jeder Zeile nur ein Kästchen an. Lass bitte keine Zeile aus.

	Mein Interesse daran ist				
	Sehr groß	Groß	Mittel	Gering	Sehr gering
1. Mehr darüber erfahren, welche Stoffe in unserer Umgebung radioaktiv sind	O	O	O	O	O
2. Mehr darüber erfahren, warum in einer kleinen Menge Uran eine so große Energiemenge steckt	O	O	O	O	O
3. Mehr darüber erfahren, wie die Kernenergie in militärischen und friedlichen Anwendungen genutzt wird	O	O	O	O	O
4. Mehr darüber erfahren, welche Arten von radioaktiven Strahlen es gibt und welche Wirkungen sie haben	O	O	O	O	O
5. Mehr darüber erfahren, was im Einzelnen passiert, wenn Neutronen auf Uran-Atome aufprallen	O	O	O	O	O
6. Mehr Einblick erhalten, wie die Angestellten eines Kernkraftwerkes dort arbeiten	O	O	O	O	O
7. Mehr Einblick erhalten, wie in einer Klinik radioaktive Stoffe bei Untersuchungen und zur Bestrahlung eingesetzt werden	O	O	O	O	O
8. Ein Gerät bauen, mit dem man radioaktive Strahlung nachweisen kann	O	O	O	O	O
9. Sich eine Methode ausdenken, wie man radioaktive Strahlen nachweisen könnte	O	O	O	O	O
10. Abschätzen, unter welchen Umständen ein Experiment mit Uran außer Kontrolle geraten kann	O	O	O	O	O
11. Sich mit der militärischen und friedlichen Anwendung der Kernenergie beschäftigen und darüber diskutieren, wie Unheil abgewendet werden kann	O	O	O	O	O

Teil 2

Bitte vervollständige die folgenden Sätze durch Ankreuzen derjenigen Ergänzung, die auf dich zutrifft:

	Sehr gut	gut	mittel	schlecht	Sehr schlecht
1. Ich verstehe den Stoff in Physik …	O	O	O	O	O
2. Ich behalte den Stoff in Physik …	O	O	O	O	O
3. Meine Leistungen in Physik sind nach meiner eigenen Einschätzung …	O	O	O	O	O
4. Ich beteilige mich am Physikunterricht …	O	O	O	O	O
5. Ich glaube, dass mich meine Mitschüler in Physik für … halten.	O	O	O	O	O
6. Ich glaube, dass meine Physiklehrerin / mein Physiklehrer meine Leistungen in Physik als … einschätzt.	O	O	O	O	O
7. Ich erwarte, dass in Zukunft meine Leistungen in Physik … sein werden.	O	O	O	O	O

(nur im Zwischentest):

8. Wie fandest du den Unterricht diese Woche?

Er hat mir sehr gut gefallen.	O
Er hat mir gefallen.	O
Keine starke Meinung in irgendeine Richtung.	O
Es war langweilig.	O
Es war sehr langweilig.	O

(nur im post-Test):

9. Wie fandest du den Unterricht in den letzten Wochen über Radioaktivität?

Er hat mir sehr gut gefallen.	O
Er hat mir gefallen.	O
Keine starke Meinung in irgendeine Richtung.	O
Es war langweilig.	O
Es war sehr langweilig.	O

Teil 3

1) Du siehst nun eine Reihe von Begriffen. Kreuze alle Begriffe an, bei welchen ein Messgerät für Radioaktivität klicken würde.
   Mehrfachantwort möglich.
   A) Smartphone
   B) Umgebungsluft
   C) Ein Schulkind
   D) Betonwände in der Rezeption eines Krankenhauses
   E) Betonwände eines Krankenhausraumes, in dem Röntgenaufnahmen gemacht wurden
   F) Betonwände eines Schulgebäudes
   G) Zigarettenrauch
   H) Sonnenlicht
   I) Popcorn aus der Microwelle
   J) Banane

2) Gib die Definition von Radioaktivität aufgrund folgender Aussagen an. Mehrfachantwort möglich.
   A) Radioaktivität geht nie mit einem Atomkernzerfall einher. Es beschreibt lediglich die Energieabgabe in Form von ionisierender Strahlung.
   B) Radioaktivität ist die fluoreszierende Eigenschaft von Atomen. Radioaktive Stoffe leuchten immer in Dunkelheit.
   C) Radioaktivität ist die Freigabe von Energie in Form von ionisierender Strahlung, nachdem ein instabiler Atomkern zerfallen ist.
   D) Radioaktivität ist die von einem Atomkern ausgesendete Strahlung. Diese Strahlung ist für den menschlichen Organismus sehr gefährlich.

3) Nenne die Arten von Strahlung, die beim Zerfall radioaktiver Stoffe von Atomkernen ausgesendet werden.
   Mehrfachantwort möglich.
   A) Radiowellen
   B) Alphastrahlung
   C) Gammastrahlung
   D) Betastrahlung
   E) Thetastrahlung

4) Du findest im Schwarzwald (Deutschland) ein unbekanntes Gestein. Dein Messgerät für Radioaktivität schlägt bedenklich weit aus. Es könnte sich hier also um radioaktives Gestein handeln. Was machst du?
   Mehrfachantwort möglich.
   A) Ich entferne mich von dem Gestein, um Abstand zur Quelle zu gewinnen.
   B) Ich entferne mich von dem Gestein, um nicht selbst radioaktiv zu werden.
   C) Ich stülpe einen Behälter über das Gestein, um einen Teil der Strahlung abzuschirmen.
   D) Ich ziehe Handschuhe an und wasche das Gestein mit Wasser des nächsten Baches ab, um seine Radioaktivität abzuwaschen.

5) Vor dir liegt folgender Versuchsaufbau, welcher in der Abbildung zu sehen ist. Die Abbildung zeigt eine radioaktive Quelle. Von dieser wird Strahlung ausgesendet (welche in dieser Skizze nicht gezeigt wird). Diese Strahlung trifft auf eine Kiwi. Die radioaktive Quelle und die Kiwi sind 2 cm voneinander entfernt.

Versuchsaufbau:

Abbildung 1: Versuchsaufbau

Gib jenes Objekt/jene Objekte (radioaktive Quelle, Strahlung und Kiwi) an, welches/welche radioaktiv ist/sind? (eine Antwort ist möglich)

A) Die radioaktive Quelle, die Strahlung und die Kiwi sind radioaktiv.
B) Die radioaktive Quelle und die Strahlung sind radioaktiv.
C) Die radioaktive Quelle und die Kiwi sind radioaktiv.
D) Die radioaktive Quelle ist radioaktiv.

6) Die radioaktive Quelle wird nach 24 Stunden entfernt. Ist die Kiwi jetzt radioaktiv?
A) Ja, die Kiwi ist durch die Strahlung radioaktiv geworden - unabhängig von der Distanz.
B) Ja, die Kiwi ist durch die Strahlung radioaktiv geworden, weil sie nah genug an der Quelle war.
C) Ja, die Kiwi ist durch die Strahlung radioaktiv geworden, weil sie während des Experiments lange genug neben der Quelle war.
D) Nein, die Kiwi ist durch die Strahlung nicht radioaktiv geworden, weil sie, um radioaktiv zu werden, eine längere Dauer neben der Strahlungsquelle verweilen müsste.
E) Nein, die Kiwi ist durch die Strahlung nicht radioaktiv geworden, weil die Strahlung Objekte nicht radioaktiv macht.

7) Radioaktivität ist:
A) Natürlich
B) Von Menschen erzeugt
C) Kann natürlich auftreten und kann vom Menschen erzeugt werden
D) Keine der Antwortmöglichkeiten ist richtig

8) Der radioaktive Zerfall ist:
A) Kontrollierbar
B) Unkontrollierbar
C) Kann kontrollierbar und kann unkontrollierbar sein
D) Keine der Antwortmöglichkeiten ist richtig

9) Radioaktivität ist für den Menschen:
   A) Ungefährlich
   B) Gefährlich
   C) Kann gefährlich und kann ungefährlich sein
   D) Keine der Antwortmöglichkeiten ist richtig

10) Wie verändert sich die Strahlung in einem Flugzeug im Vergleich zum Boden?
   E) Sie nimmt mit der Höhe zu
   F) Sie nimmt mit der Höhe ab
   G) Sie bleibt gleich
   H) Keine der Antwortmöglichkeiten ist richtig

11) Sind deine Emotionen bezüglich Radioaktivität positiv, negativ oder neutral?
   A) Positiv
   B) Negativ
   C) Neutral
   D) Keine Angabe

12) Welche Schulform besuchst du?
   A) Mittelschule
   B) Gymnasium

Um festzustellen, ob und wie sehr sich deine Antworten verändert haben, würde ich dich um folgenden Code bitten. Damit kann ich deine Ergebnisse vergleichen. Die Daten werden nur in **anonymisierter Form** (also ohne Code) veröffentlicht.

1) Bitte gib die **ersten zwei Buchstaben** des Vornamens deiner Mutter an (z.B. Theresia -> **TH**)
2) Bitte gib den **Tag deines Geburtstages** an (z.B. 05.08.1995 -> **05**)
3) Bitte gib den **letzten Buchstaben** deiner Augenfarbe an (z.B. blau → **U**)

Bitte gib hier den Klassencode ein:

____

6.2 Questionnaire for the control group (German)

<u>Deckblatt 1: (pre-Test)</u>

Liebe Schülerin, lieber Schüler!

Im Physikunterricht lernt man über viele verschiedene Themen. Dazu gehört auch die Radioaktivität. Ich möchte in meiner Masterarbeit einen genauen Blick darauf werfen, wie dieses Thema unterrichtet wird.

Ich würde von dir gerne erfahren, wie sehr dich dieses Thema interessiert und wie du dich selbst im Physikunterricht einschätzt. Darüber hinaus würde ich gerne erfahren, was du schon über Radioaktivität weißt.

Bekomme kein schlechtes Gefühl, wenn du dir bei einer Antwort nicht sicher bist. Jede Antwort kann dazu beitragen, den Physikunterricht für dich in Zukunft interessanter zu machen. Aus diesem Grund würde ich dich bitten, den folgenden Fragebogen sorgfältig auszufüllen.

Die Teilnahme ist **freiwillig.** Die Daten werden gespeichert und ausschließlich zur Bearbeitung meiner Masterarbeit verwendet. Die Daten werden nur in **anonymisierter Form** veröffentlicht. Alle Daten werden **strikt vertraulich** behandelt. Die Beantwortung dauert etwa **10- 15 Minuten.**

Wenn du Fragen hast, wende dich gerne an deine Lehrperson oder direkt an mich, unter folgender E-Mail-Adresse: [redacted]

Vielen Dank für deine Mitarbeit!

[redacted]

<u>Deckblatt 2: (post-Test)</u>

Liebe Schülerin, lieber Schüler!

Du hast nun einiges im Unterricht zum Thema Radioaktivität gehört. Ich würde nun gerne erfahren, ob sich etwas an deinem Interesse zu diesem Thema geändert hat. Ebenso, ob du dich selbst jetzt anders einschätzt als vorher. Mich interessiert auch, was du jetzt über Radioaktivität weißt und was du dir gemerkt hast.

Aus diesem Grund würde ich dich wieder bitten, den folgenden Fragebogen sorgfältig auszufüllen. Bekomme auch dieses Mal kein schlechtes Gefühl, wenn du dir bei einer Antwort nicht sicher bist. Jede deiner Antworten kann wieder dazu beitragen, den Physikunterricht in Zukunft interessanter zu gestalten.

Es gilt nach wie vor: Die Teilnahme ist **freiwillig.** Die Daten werden gespeichert und ausschließlich zur Bearbeitung meiner Masterarbeit verwendet. Die Daten werden nur in **anonymisierter Form** veröffentlicht. Alle Daten werden **strikt vertraulich** behandelt. Die Beantwortung dauert etwa **10-15 Minuten.**

Wenn du Fragen hast, wende dich gerne an deine Lehrperson oder direkt an mich, unter folgender E-Mail-Adresse: [redacted]

Vielen Dank für deine Mitarbeit!

[redacted]

Teil 1

Gib bitte an, ob dein Interesse an den folgenden Inhalten zu Radioaktivität für dich sehr groß, groß, mittel, gering oder sehr gering ist. Kreuze das entsprechende Kästchen an. Bitte kreuze in jeder Zeile nur ein Kästchen an. Lass bitte keine Zeile aus.

	Mein Interesse daran ist				
	Sehr groß	Groß	Mittel	Gering	Sehr gering
1. Mehr darüber erfahren, welche Stoffe in unserer Umgebung radioaktiv sind	O	O	O	O	O
2. Mehr darüber erfahren, warum in einer kleinen Menge Uran eine so große Energiemenge steckt	O	O	O	O	O
3. Mehr darüber erfahren, wie die Kernenergie in militärischen und friedlichen Anwendungen genutzt wird	O	O	O	O	O
4. Mehr darüber erfahren, welche Arten von radioaktiven Strahlen es gibt und welche Wirkungen sie haben	O	O	O	O	O
5. Mehr darüber erfahren, was im Einzelnen passiert, wenn Neutronen auf Uran-Atome aufprallen	O	O	O	O	O
6. Mehr Einblick erhalten, wie die Angestellten eines Kernkraftwerkes dort arbeiten	O	O	O	O	O
7. Mehr Einblick erhalten, wie in einer Klinik radioaktive Stoffe bei Untersuchungen und zur Bestrahlung eingesetzt werden	O	O	O	O	O
8. Ein Gerät bauen, mit dem man radioaktive Strahlung nachweisen kann	O	O	O	O	O
9. Sich eine Methode ausdenken, wie man radioaktive Strahlen nachweisen könnte	O	O	O	O	O
10. Abschätzen, unter welchen Umständen ein Experiment mit Uran außer Kontrolle geraten kann	O	O	O	O	O
11. Sich mit der militärischen und friedlichen Anwendung der Kernenergie beschäftigen und darüber diskutieren, wie Unheil abgewendet werden kann	O	O	O	O	O

Teil 2

Bitte vervollständige die folgenden Sätze durch Ankreuzen derjenigen Ergänzung, die auf dich zutrifft:

	Sehr gut	gut	mittel	schlecht	Sehr schlecht
1. Ich verstehe den Stoff in Physik …	O	O	O	O	O
2. Ich behalte den Stoff in Physik …	O	O	O	O	O
3. Meine Leistungen in Physik sind nach meiner eigenen Einschätzung …	O	O	O	O	O
4. Ich beteilige mich am Physikunterricht …	O	O	O	O	O
5. Ich glaube, dass mich meine Mitschüler in Physik für … halten.	O	O	O	O	O
6. Ich glaube, dass meine Physiklehrerin / mein Physiklehrer meine Leistungen in Physik als … einschätzt.	O	O	O	O	O
7. Ich erwarte, dass in Zukunft meine Leistungen in Physik … sein werden.	O	O	O	O	O

(nur im post-Test):

8. Wie fandest du den Unterricht in den letzten Wochen über Radioaktivität?

Er hat mir sehr gut gefallen.	O
Er hat mir gefallen.	O
Keine starke Meinung in irgendeine Richtung.	O
Es war langweilig.	O
Es war sehr langweilig.	O

Teil 3

1) Du siehst nun eine Reihe von Begriffen. Kreuze alle Begriffe an, bei welchen ein Messgerät für Radioaktivität klicken würde.
   Mehrfachantwort möglich.
   K) Smartphone
   L) Umgebungsluft
   M) Ein Schulkind
   N) Betonwände in der Rezeption eines Krankenhauses
   O) Betonwände eines Krankenhausraumes, in dem Röntgenaufnahmen gemacht wurden
   P) Betonwände eines Schulgebäudes
   Q) Zigarettenrauch
   R) Sonnenlicht
   S) Popcorn aus der Microwelle
   T) Banane

2) Gib die Definition von Radioaktivität aufgrund folgender Aussagen an. Mehrfachantwort möglich.
   E) Radioaktivität geht nie mit einem Atomkernzerfall einher. Es beschreibt lediglich die Energieabgabe in Form von ionisierender Strahlung.
   F) Radioaktivität ist die fluoreszierende Eigenschaft von Atomen. Radioaktive Stoffe leuchten immer in Dunkelheit.
   G) Radioaktivität ist die Freigabe von Energie in Form von ionisierender Strahlung, nachdem ein instabiler Atomkern zerfallen ist.
   H) Radioaktivität ist die von einem Atomkern ausgesendete Strahlung. Diese Strahlung ist für den menschlichen Organismus sehr gefährlich.

3) Nenne die Arten von Strahlung, die beim Zerfall radioaktiver Stoffe von Atomkernen ausgesendet werden.
   Mehrfachantwort möglich.
   F) Radiowellen
   G) Alphastrahlung
   H) Gammastrahlung
   I) Betastrahlung
   J) Thetastrahlung

4) Du findest im Schwarzwald (Deutschland) ein unbekanntes Gestein. Dein Messgerät für Radioaktivität schlägt bedenklich weit aus. Es könnte sich hier also um radioaktives Gestein handeln. Was machst du?
   Mehrfachantwort möglich.
   E) Ich entferne mich von dem Gestein, um Abstand zur Quelle zu gewinnen.
   F) Ich entferne mich von dem Gestein, um nicht selbst radioaktiv zu werden.
   G) Ich stülpe einen Behälter über das Gestein, um einen Teil der Strahlung abzuschirmen.
   H) Ich ziehe Handschuhe an und wasche das Gestein mit Wasser des nächsten Baches ab, um seine Radioaktivität abzuwaschen.

5) Vor dir liegt folgender Versuchsaufbau, welcher in der Abbildung zu sehen ist. Die Abbildung zeigt eine radioaktive Quelle. Von dieser wird Strahlung ausgesendet (welche in dieser Skizze nicht gezeigt wird). Diese Strahlung trifft auf eine Kiwi. Die radioaktive Quelle und die Kiwi sind 2 cm voneinander entfernt.

Versuchsaufbau:

Abbildung 2: Versuchsaufbau

Gib jenes Objekt/jene Objekte (radioaktive Quelle, Strahlung und Kiwi) an, welches/welche radioaktiv ist/sind? (eine Antwort ist möglich)

E) Die radioaktive Quelle, die Strahlung und die Kiwi sind radioaktiv.
F) Die radioaktive Quelle und die Strahlung sind radioaktiv.
G) Die radioaktive Quelle und die Kiwi sind radioaktiv.
H) Die radioaktive Quelle ist radioaktiv.

6) Die radioaktive Quelle wird nach 24 Stunden entfernt. Ist die Kiwi jetzt radioaktiv?
F) Ja, die Kiwi ist durch die Strahlung radioaktiv geworden - unabhängig von der Distanz.
G) Ja, die Kiwi ist durch die Strahlung radioaktiv geworden, weil sie nah genug an der Quelle war.
H) Ja, die Kiwi ist durch die Strahlung radioaktiv geworden, weil sie während des Experiments lange genug neben der Quelle war.
I) Nein, die Kiwi ist durch die Strahlung nicht radioaktiv geworden, weil sie, um radioaktiv zu werden, eine längere Dauer neben der Strahlungsquelle verweilen müsste.
J) Nein, die Kiwi ist durch die Strahlung nicht radioaktiv geworden, weil die Strahlung Objekte nicht radioaktiv macht.

7) Radioaktivität ist:
E) Natürlich
F) Von Menschen erzeugt
G) Kann natürlich auftreten und kann vom Menschen erzeugt werden
H) Keine der Antwortmöglichkeiten ist richtig

8) Der radioaktive Zerfall ist:
E) Kontrollierbar
F) Unkontrollierbar
G) Kann kontrollierbar und kann unkontrollierbar sein
H) Keine der Antwortmöglichkeiten ist richtig

9) Radioaktivität ist für den Menschen:
   E) Ungefährlich
   F) Gefährlich
   G) Kann gefährlich und kann ungefährlich sein
   H) Keine der Antwortmöglichkeiten ist richtig

10) Wie verändert sich die Strahlung in einem Flugzeug im Vergleich zum Boden?
   I) Sie nimmt mit der Höhe zu
   J) Sie nimmt mit der Höhe ab
   K) Sie bleibt gleich
   L) Keine der Antwortmöglichkeiten ist richtig

11) Sind deine Emotionen bezüglich Radioaktivität positiv, negativ oder neutral?
   E) Positiv
   F) Negativ
   G) Neutral
   H) Keine Angabe

12) Welche Schulform besuchst du?
   C) Mittelschule
   D) Gymnasium

Um festzustellen, ob und wie sehr sich deine Antworten verändert haben, würde ich dich um folgenden Code bitten. Damit kann ich deine Ergebnisse vergleichen. Die Daten werden nur in **anonymisierter Form** (also ohne Code) veröffentlicht.

4) Bitte gib die **ersten zwei Buchstaben** des Vornamens deiner Mutter an (z.B. Theresia -> **TH**)
5) Bitte gib den **Tag deines Geburtstages** an (z.B. 05.08.1995 -> **05**)
6) Bitte gib den **letzten Buchstaben** deiner Augenfarbe an (z.B. blau → **U**)

Bitte gib hier den Klassencode ein:

____

6.3 Interview guide (German)

Warm-up

- Erzählen Sie, wie Sie sich prinzipiell auf den Unterricht vorbereiten. Was bedenken Sie bei Ihrer Planung?
- Wie haben Sie sich auf den Unterricht zum Thema Radioaktivität vorbereitet?

Einsatz von Medien

- Verwenden Sie ein Buch für den Unterricht, wenn ja welches?
    - Wie oft haben Sie das Buch beim Unterrichten von Radioaktivität verwendet?
    - Warum?
- Haben Sie den Schüler*innen beim Unterrichten von Radioaktivität Filme, Videos oder Simulationen gezeigt?
    - Wenn ja, welche und warum?

Durchführung von Experimenten

- Haben Sie zum Thema Radioaktivität Experimente gezeigt oder von Schüler*innen durchführen lassen?
    - Wenn ja, welche und warum?

Zeit

- Welchen Anteil einer Unterrichtsstunde zum Thema Radioaktivität erarbeiten die Schüler*innen selbstständig Inhalte.
    - Um welche Aktivitäten handelt es sich dabei? (z.B. Durchführung von Experimenten, Rechercheaufgaben, Rechenaufgaben, …)
- Welchen Anteil einer Stunde sprechen bzw. erklären Sie im Unterricht?
- Wie gehen Sie mit Schüler*innenfragen um?

Ausklang

- Was ist Ihnen beim Unterrichten von Radioaktivität besonders wichtig? Setzen Sie einen bestimmten Schwerpunkt?

6.4 Teacher questionnaire (German)

Deckblatt:

Sehr geehrte Damen und Herren!

Zu Beginn möchte ich mich nochmal sehr herzlich bei Ihnen für die Teilnahme an meiner Studie bedanken. Im folgenden Fragebogen möchte ich ein paar grundlegende Informationen zu dem von Ihnen gehaltenen Unterricht zum Thema Radioaktivität einholen. Speziell geht es darum, wie lange Ihr Unterricht zu Radioaktivität gedauert hat und wie viel Zeit Sie für einzelne Inhalte bzw. Konzepte verwendet haben

Das Ausfüllen des Fragebogens dauert **weniger als 5 Minuten** und es werden **keine** personenbezogenen Daten abgefragt. Zusätzlich werden alle Daten **strikt vertraulich** behandelt. Die Teilnahme ist zudem **freiwillig**.

Wenn Sie Fragen haben, melden Sie sich gerne bei mir unter: [redacted]

Vielen Dank für Ihre Mitarbeit!
[redacted]

Fragebogen:

1. Über wie viele Schulstunden wurde das Thema „Radioaktivität" insgesamt unterrichtet?
2. Bitte kreuzen Sie an, welche der unten angeführten Inhalte bzw. Konzepte Sie im Unterricht thematisiert haben. Bitte tragen Sie in das Feld daneben ein, wie viele Unterrichtsstunden Sie zur Bearbeitung des Themas verwendet haben.

	Inhalte/Konzepte	**Dauer des Unterrichts**
O	Geschichtliches zu Radioaktivität (Entdeckung, etc…)	
O	Atomaufbau - Isotope	
O	Strahlungsarten (Alpha-, Beta-, Gammastrahlung)	
O	Halbwertszeit	
O	Kernspaltung	
O	Kernfusion	
O	Wirkung und Schutz vor ionisierender Strahlung	
O	Anwendungen ionisierender Strahlung (z.B. Kernreaktoren)	
O	Natürliche Strahlungsquellen (z.B.: Beton, Lebensmittel, Höhenstrahlung, Menschen)	
O	Sonstiges:	

Danksagung:

Ich bedanke mich nochmal sehr herzlich für die Teilnahme!

Mit freundlichen Grüßen,
[redacted]