MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"The conceptual and epistemological effects of the Newton's Third Law Open Source Tutorial in Austrian high schools"

verfasst von / submitted by

Iva Nunes Sampaio-Kronister, BEd

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

Master of Education (MEd)

Wien, 2020 / Vienna 2020

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) UF Mathematik, UF Physik
Betreut von / Supervisor:	Univ.-Prof. Dr. Martin Hopf
Mitbetreut von / Co-Supervisor:	-

Acknowledgments

Throughout the research and writing of this thesis I have received a great deal of support and assistance. I would like to express **my very profound gratitude** to:

Dr. Michael Hull, my supervisor, for this wonderful opportunity to work with him and with something I truly became passionate about. He provided me with encouragement and guidance, insights and constructive comments; made himself available every single moment that I needed his expertise and advice; and in the end spared no effort to help me finish in the limited time frame.

Dr. Martin Hopf, my supervisor, who not only gave his support to this study, but also helped me immensely with his professionalism and sensibility as Head of the Physics Teacher Education Degree program.

School directors, for the support and kind hospitality. Teachers, for the time, assistance and for being so considerate adapting their schedule to help this study. Students, who are the central piece of this research and showed a kind willingness to participate.

My dear friend, Judith, who helped me in so many ways and in so many levels, not only in this thesis but throughout the whole studies.

My father, Sampaio, for being my mentor and not only teaching me a great deal of physics but to have discipline and perseverance. My mother, Assunção, for always showing me the light when it was too hard to see it and for her unconditional love. Together, you showed me a world of physics and arts, music and love, and you never gave me the idea that I couldn't do whatever I wanted to do, or be whoever I wanted to be. Well, you are my role models, my ultimate inspiration.

My daughter, Larinha, for making me want to set the best example I could and for having to endure with an absentee mother during the last years. For being my best friend and a part of me.

And finally, my husband Gerhard, for the unending support, patience and understanding. For always being my safe port and my love. And mostly, for making his world mine. *"Weus'd a Wahnsinn bist für mi"*.

Contents

"The whole of science is nothing more than a refinement of everyday thinking.
It is for this reason that the critical thinking of the physicist cannot possibly be restricted to the examination of concepts from his own specific field.
He cannot proceed without considering critically a much more difficult problem, the problem of analyzing the nature of everyday thinking."

Albert Einstein

1 Introduction

It is known that teaching physics with traditional methods has a low efficiency and that physics becomes a very disliked subject for most students. However, science education in schools is not only important for each individual student, with impacts in personal life. Science education in schools is very important for the citizen each student is, part of a democratic society and it impacts social, political, economic and cultural worlds. Therefore, when looking at Austrian students' results in the PISA (Programme for International Student Assessment) [2] test, it is not only worrisome but also quite sad. In OECD's iLibrary chapter [3] that focuses on student engagement with science and attitudes towards science, it can be seen that the Austrian students' "self-efficacy in science", as well as "enjoyment of learning science" had a negative shift, showing that it decreased between 2006 and 2015. However, students' beliefs about their abilities in science influence their performance in science classes. In addition, *motivation can be regarded as a driving force behind engagement, learning and choice of occupation in all fields. To nurture students' engagement with science,* ***school systems need to ensure*** *that students have not only the basic knowledge that is necessary to engage with complex scientific issues, but also the* ***interest and motivation*** *that will make them want to do so.* [3]

It has been well established that we, teachers, have to do a lot more than just stand in front of the classroom and "spread" knowledge, most times carried out like a monologue where there is almost no student interaction. First, if we want students to truly learn, they have to **"actively participate"** in the learning process. Second, the responsibility of a science teacher goes far beyond teaching the scientific concepts to students. Successfully finding the possibilities to make students gain a good conceptual understanding is for sure a great start, but that is not sufficient in helping students achieve everything that they "can" (and with our help they really could) achieve. For that to happen students **attitudes and views about the nature of science and about what it means to acquire scientific knowledge** have to be taken in consideration. Only then it is possible to ensure that students have a higher motivation and self-efficacy in science.

For these reasons, in this master thesis I decided to research a different "teaching method" in Austrian high schools. The method, called *Open Source Tutorials* was developed by the University of Maryland and presented to me in the University of Vienna by my supervisor Dr. Michael M. Hull. There are many different tutorials, for the different topics in physics, which were designed aiming both conceptual improvement and development in students' beliefs and views about the nature of physics knowledge and learning physics.

Newton's third law is a very difficult topic in school physics because it goes "against common sense". The tutorial for this topic encourages students not only to know the correct answers, but most importantly, to understand them as both plausible and intuitive. Thus, this topic was the one chosen for the research of the Open Source Tutorials' effects in Austrian high school students.

The research question in this thesis is:

What are the (conceptual and epistemological) effects of using the Newton's Third Law Open Source Tutorial in Austrian high schools?

This question will be answered in the Discussion (see chapter 4.2).

2 Theoretical framework

2.1 Physics' education

Since antiquity, human beings have shown curiosity about the nature that surrounds them. In those remote times, there was already a desire to understand phenomena, to seek answers, to solve problems that interfered in one way or another in the life of society. Physics seeks to understand and study the fundamental phenomena of nature that surrounds us, from the most elementary to the most complex. It is one of the oldest sciences and has helped societies understand simple things like the act of walking, or more complex ones like the movement of galaxies. Furthermore, the many discoveries in physics, over the last centuries, have enabled the development and growth of other sciences and of technologies that changed our whole world and the way we live. Physics is present in most everything in human life and, as such, understanding of physics is part of the minimum general knowledge that everyone should have. The teaching of physics allows students not only to better understand the phenomena of nature and the technological world in which they live in, but also help them develop the necessary skills to do so.

Physics is not just a collection of laws, rules or acquired knowledge but – much more – a cultural activity and its study in school play a very important role not only in one's life but also in society. Physics classes have no longer the sole task of imparting physics' expertise to students. Rather, it is required that students develop an understanding of how knowledge is gained in physics and what are the central elements of physics' thinking and research – which is known as *scientific literacy* – through physics lessons. The teaching of scientific literacy fulfills the general educational entitlement of school instruction and goes far beyond a justification for physics teaching. [4]

According to Walter Jung [5], the arguments that justify physics lessons can be developed in four areas: economic and political requirements (or needs) as well as anthropological and cultural desires (or needs). Social-needs arguments are formulated out of a socio-political interest. There is often the indication that science education creates a central "raw material" in highly industrialized but resource-poor countries. In order to make sufficient use of this resource, from an economic point of view, as many students as possible should develop appropriate skills. It is considered necessary to develop – or maintain – a positive attitude towards the natural sciences through school education. Based on this background, initiatives from the economic world, which should increase the interest in science subjects, are to be interpreted. In the political education discussion, reference is made to the fact that in order to be actively involved in political decision-making processes, the competence to judge (supposedly) scientific statements is often required. Scientific literacy is understood as part of political maturity and is therefore part of a democratically necessary education in a high-tech society. Arguments of desire are aimed at the individual situation of students. In the understanding of physics as a cultural activity, basic science education leads to a cultural embedding of students that is not only socially but also individually desirable. This gives them the opportunity to develop personal interests and skills by working with physics and to deepen them in their free time or in team-work. Science education contributes to the development of well-founded personal positions on technical and scientific issues. Appropriate physics classes enable students to make rational decisions and to recognize and evaluate pseudo-scientific explanations and arguments. [4]

Scientific knowledge is therefore a human construction, whose interests and actions are guided not only by personal and individual elements, but by instances of society – such as economics, politics, historical-social contexts, environmental elements, etc. Concerning this matter, the didactic of physics becomes one of the means to relate students to the whole complex involving science. So, the question is: how can teachers and instructors succeed in promoting a learning environment for students in which they acquire or develop the necessary skills and competences?

We can recall the normative statements and concepts conceived in the seventeenth century by Comenius [6] – often called the father of modern education [7] – and see how it converges to the current school system and teaching methods:
"The teacher is certain to have attentive pupils."... "The teacher will be able to know with certainty if his pupils have thoroughly grasped everything that he has taught them."... "If the same thing be frequently repeated, the dullest intelligences will grasp it at last, and will thus be able to keep pace with the others; while the brighter ones will be pleased at obtaining such a thorough grip of the subject."... "never give individual instruction, either privately out of school or publicly in school, but teach all the pupils at one and the same time. He should,

therefore, never step up to any one scholar or allow any one of them to come to him separately, but should remain in his seat, where he can be seen and heard by all, just as the sun sends forth its rays over all things. The scholars, on the other hand, must direct their ears, eyes, and thoughts towards him and attend to everything that he tells them by word of mouth or explains by means of his hand or of diagrams. Thus, with a single blow, not one but many flies are killed."... "to imbue them with the notion that (as really is the case) the mouth of the teacher is a spring from which streams of knowledge issue and flow over them, and that, whenever they see this spring open, they should place their attention, like a cistern, beneath it, and thus allow nothing that flows forth to escape." [6]

With some adjustments, these concepts do seem very similar to the "direct instruction" practiced in most classrooms still today. The questioning here is if this traditional method of instruction – or its slightly modified versions, which are so common – is really effective. An analogy to this kind of instruction is based on the idea of knowledge as a substance, where many people spontaneously imagine the process as the transport of this substance – also known as the "pipeline model", in which students can learn content with almost no effort and that teachers can convey everything to everyone. It does little in the understanding of the process meaningfully and it certainly does not support it practically. What is presented here as a substance cannot be collected in a container, nor can it be filled into someone's head or body at any point. [8]

"For the things we have to learn before we can do them, we learn by doing them" stated Aristotle [9].

According to Bernd Hackl [8] knowledge is just a form of skill (to *know* is to *be able to do* something). He describes knowledge as a mental ability, or a skill in the activity of our mind. This skill has the ability to execute a mental process that has to be practiced (mental exercise); and the preparation for this practice (of a certain knowledge, i.e. ability) describes every learning process. Furthermore, the real subject of every learning is the *appropriate handling* and its result is the ability to handle (to do) it. Thus, "do, what you would like to learn" builds the core of learning. Since you first have to learn what you cannot do yet, it is of course not to be expected that such an execution will be possible without any problems. Certain bridging processes are first required, which can lead to the correct execution actually taking place, and that is where the question of “learning process” becomes important. If there was no difference at all between executing and learning, we could take Aristotle’s quotation literally and simply do what we cannot do without the question of learning. So how can the learning process occur, or what is the appropriate handling?

Since the learning process must necessarily be an unsafe, hypothetical, searching activity, it can also be called *trying*. Its basic function is that it leads to the incorporation of structures of the practices carried out. The embodied practices then form an unexpressed (implicit) available skill, to which “all skills – including knowledge” – turn to (or make use of). Such a skill cannot be developed without trying it out. [8]

Likewise, many physics didactics’ researchers express similar conclusions about the “pipeline model” or the traditional (direct) instruction. Wiesner, Schecker and Hopf [4] emphasize that teachers cannot give (convey) information directly to students. Nevertheless, instructors often teach unknowingly, acting as if knowledge could be transported directly into the students’ heads, as if what’s important is that the teacher chooses the right representation and that the students pay attention. According to them, what teachers can do is offer suitable data which students can establish a connection to their previous knowledge and thus make information that will be effective for the learning process. Good explanations and thorough arrangements of the learning environment are still very important tasks for teachers. However, the role of students in the teaching-learning process has at least the same importance.

Furthermore, physics education is developed in the concrete professional examination of physics’ facts and their explicit reflection. To be able to address the special character of physics approaches, students have to *“actively practice physics”* to an appropriate extent in learning situations. Thus, students should develop solutions and find answers, which are not easily or directly provided by the teacher. In order to develop solutions, it is necessary to have both available knowledge − instrumental knowledge of the methods and on physics (definitions, laws, device knowledge, experimental procedures) − and knowledge about which areas of life physics’ knowledge supports judging processes or making decisions; and how to infer or yield physics’ knowledge in such cases. These competences are developed in practical applied situations. [4]

Hestenes, Wells and Swackhamer [10] wrote that regardless of the fact that traditional instruction works for some students, it is not efficient. For them, even though problem-solving skill is regarded as the *sine qua non* (an essential condition) of physics understanding in traditional instruction, certain concepts and modes of reasoning should be developed before problem-solving instruction can be effective.

Mazur [11] reported that many students in his class were concentrating on learning "recipes", or "problem-solving" strategies, without considering the underlying concepts. For him what aggravates the problem is that traditional instruction is almost always carried out as a monologue in front of a passive audience.

Also according to Redish [12], lectures have limited success in helping students make sense of the physics they are learning. Even though there are some things you can do to get your students more engaged during a lecture, models that involve more structured interactions with the students have been shown to produce dramatic improvements in student learning. He says that if learning is carried out through social interactions, then it is more effectively for most people.

The list of researchers or arguments could go on, but the bottom line is simply that traditional instruction is maybe not the best option trying to create a successful setting in physics classrooms in order to achieve the educational goals that we, teachers, should try to achieve. That is a big motivation − at the very least − to explore or investigate alternative methods, interactive instruction curricula and different ways to accomplish this.

2.1.1 Students' misconceptions

"If I had to reduce all of educational psychology to just one principle, I would say this: The most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly." [13]

Students have years of personal experiences and different encounters in life − including previous lessons − which establish a system of commonsense beliefs, that are brought to the classroom and influence strongly how they interpret the physics presented or taught to them.

These beliefs, which students have, illustrate dispositions (or tendencies) to interpret the concepts of physics or describe phenomena in a certain way that differ or contradict the actual physics concept (or scientific conception) − therefore, they will be referred to as students' *misconceptions*. In students' reasoning, different misconceptions can coexist on the same subject, side by side and parallel to physics concepts and ideas. [14]

These misconceptions, which are brought to the classroom, can be originated or influenced by the scope and quality of previous physics classes, although they are mainly developed outside of the school context (everyday language, media and experiences). An important source of students' misconceptions is the informal or **colloquial language** used in everyday situations or in the media. In everyday language, words have meanings that often differ from their meanings in physics. [14] For instance, the word *impulse* is used in everyday language as "instant action", "spontaneity", "something you do without thinking", i.e., associated with a short time duration (while in physics the longer the duration, the greater the impulse) [15]. This means that different conceptual meanings are related to the same word. In everyday conversation an understanding is achieved because the context in which the word is used makes it clear. Therefore, learning physics also means learning a new language – not only in the sense of a new technical vocabulary, but also in terms of understanding the advantages of conceptual precision in the professional discourse. Some students' misconceptions are based on **graphic illustrations** (or models) that are used in the media or even in the classroom (where too little attention is paid emphasizing the model character of illustrations and students end up taking the model for reality). For example, the lines of force that *represent* a (magnetic, electrical,...) field. Students think that these lines *are* the field and that there is nothing between the lines. And finally, it is not in the classroom – but rather **in everyday life** – where students encounter many of physics topics for the first time (including the interpretation of sensorial stimulations and experiences). [14]

Originated in the 1970's, the research on students' beliefs and ideas about physics concepts and phenomena is probably the most successful and profitable topic in physics didactics and has influenced, with a time lag, the research on lesson design [16]. Physics education research has established that these beliefs play a dominant role in physics' classrooms and instruction that does not take them into account is almost totally ineffective, at least for the majority of students. *"We now have strong evidence that misconceptions must be taken into account to improve the efficiency of physics instruction."* [10]

Students' misconceptions are of great importance for learning the fundamental physics concepts and phenomena, and not only do they influence the processing of new learned content, but they are also often very resistant to changes through teaching [14]. Furthermore, since they represent the basis for the construction of signification, a completely different meaning will be produced within students, than the one intended by the teacher [4]. So, if the knowledge of students' misconceptions is so important for teaching; then, how can teachers elicit and deal with them when teaching?

In order for students' misconceptions to be evident in the classroom, students must have opportunities and dare to express their thoughts and reasonings on their own initiative, even if these may not be correct. They should be able to trust that the teacher clearly differentiates between learning phases and performance assessments. [14]

The physics research community has been developing a wide range of data collection using different methods, some of which will be briefly described. [14]

Written proceedings

• Developed *written tests* are the most common method used to elicit students' misconceptions. Most of them consist of questions with more options of answers (*multiple choice*), which present solutions based on previously known misconceptions and one correct solution.

• If the misconceptions of the subject in question are still not well known, tasks with *open-answers* format are more suitable.

Oral proceedings

• *Interviews*, which are also a good start to research unknown misconceptions.

• There is also an alternative called *acceptance survey* (or teaching experiments), where the interviewer presents the correct description of a physics' concept during the conversation and explores to what extent the student understands it and assesses it as useful, or whether there is resistance.

• Oral surveys, which are often supported by pictures or sketches.

• The creation and analysis of physics classes' *transcripts*, based on recordings of classroom discussions. For a profitable analysis, it is good if the teacher is open minded about the situation, i.e., hold him/herself back not giving judging or guiding statements and encourages the students to express themselves freely.

The question as to how a teacher should deal with current students' misconceptions often arises in the classroom: should teachers ignore them or bring them up? There are two fundamentally different models for dealing with this knowledge when planning and implementing classes. One possibility is to suddenly change students' thinking by creating a **discontinuous learning path**, where a cognitive conflict is created and then resolved by a conceptual change. For doing so, the students should recognize in experiments, for example, a discrepancy between their predictions − based on their misconceptions − and the actual result. The other possibility is to develop the physics' concepts gradually, trying to connect with the existing, expandable misconceptions, and either to delimit them and from this point on to lead them step by step to the desired concepts and ideas, or to create a new interpretation of them. This would be a **continuous, seamless learning path**. [14]

Both models to deal with misconceptions are based on a moderate constructivist view of learning, i.e. the idea that learning and general human knowledge is only possible on the basis of existing knowledge. Learning is an active process of the students, who have to construct their own knowledge, that is, they either have to build it up or rebuild it themselves. Which one of the models a teacher chooses for planning a lesson depends also on other goals. If there is the wish to address scientific theory aspects in physics lessons and make them consciously aware or to make metacognitive considerations, i.e. explicitly discussing the path of one's own knowledge acquisition with the students, then cognitive conflicts are appropriate. However, if these conflicts seem too challenging for students, the continuous path should be chosen, with the prerequisite that there are enough connection points to expandable existing misconceptions. [14]

The knowledge of students' misconceptions should enable teachers not only to better understand their students, but also to better promote conversation with them. Students' statements that appear superficially physically wrong can certainly have been based on careful, independent considerations. A teacher has to be aware that physical knowledge systems – to be developed by students – represent complex and abstract models. Children are able to learn complex and abstract relationships, but it takes considerable effort to restructure such a system. It is recommended to explicitly address these processes in class. [14]

It should be particularly emphasized that students' misconceptions are not mistakes made by children and teenagers, but often meaningful reconstructions of the world. Most of the time it will be the case that everyday ideas and physics' concepts coexist and can be meaningfully activated depending on the context. It has been proved in didactics research that these ideas should be seen as resources for learning physics, of which its use is a great challenge but also opens up great opportunities. One should therefore carefully consider whether a discontinuous strategy (aimed at an abrupt change of concepts) can really be used effectively or if the continuous path should be chosen. [14]

Regardless of the path, teachers may wish that all their students understand and adopt the physical perspective and then only reason in a way that is physically correct. This would mean that the physically incorrect misconceptions are eliminated in students' thinking or exchanged for the physically correct concepts. However, as psychological and didactic research shows, this is not realistic. Every learning path from students' misconceptions towards scientific concepts is referred to as *conceptual change*. Conceptual change – or the development of physics' concepts – is not about replacing physically incorrect concepts with correct ones. The assumption that this could be achieved persistently through physics lessons (approaching a broad range of topics) has proven to be hopeless. Conceptual change is about providing students with a physical perspective and elucidating it as a special perspective that differs in many aspects from everyday thinking. [14]

"Student misconceptions should be elicited and treated when they are prone to conflict with the physics concepts.[...] This requires planning, preparation, and practice. It is not easy to do well, but it can be very rewarding for teacher and students alike." [10]

2.1.2 Epistemology

As discussed in the previous chapter, it has been already well established the fact that engaging in research-based curricula, which take students' misconceptions in consideration, and that promote "active learning", has positive results in students' conceptual understanding of physics and phenomena. Still, one question remains unanswered: is that enough? Is focusing in students conceptual understanding of the physics they are learning enough to fulfill the responsibility we have as instructors – as science educators; or is it enough for students to achieve the learning goals or develop the competences we wish they do?

"The educational goals of physics teaching include not only learning physics, but also learning *about* physics." [4]

In addition to students' beliefs about concepts and phenomena, their ideas about the nature of sciences and their beliefs about the nature of knowledge and knowing are also important for physics classes. Here, as well, considerable teaching efforts are required to support a conceptual change. [14] These beliefs or views about the *nature of knowledge and knowing* will be referred as *epistemology* in this thesis. For the physics education research community, epistemology has become a growing area of interest. [17]

Students' epistemology in physics can have many aspects (related for example not only to their ideas of physics knowledge, but also to their beliefs about how physics knowledge is acquired). For instance, some students may believe that learning physics involves applying and modifying their own understandings [18]; that physics is a coherent system of ideas and that the formalism is a way to express and work with those ideas [19]; that physics is present in their everyday life and it helps them understand the world; or that the physics concepts they learn make sense and correspond to their intuition. These are all examples of very developed epistemology and every teacher could only wish all his/her students would have these beliefs or views. Unfortunately, most of them don't. Among others, students often can:

- underestimate the importance of theoretical assumptions and believe that physics' knowledge is derived directly from experiments. They assume that you can always make an objective and clear decision whether a physical statement is true. [4]

- get disappointed when working with models in physics classes. They want to know "how it really is", or they view the models as "reality". [4]
- believe that learning physics is memorizing facts and formulas given by the teacher. [18]
- have the idea that physics knowledge is just a bunch of facts, formulas, and methods to solve problems, mainly without any connection to everyday thinking. [19]
- think of physics as being something so complicated that only physicists can really understand or make sense of it.

In the same way that conceptual change researchers have described students' prior knowledge as consisting of stable, robust misconceptions that cannot contribute to expert understanding, epistemology researchers have mostly described naive epistemologies in terms of stable, counter-productive beliefs that must be replaced in order to achieve sophistication. [19] Students' epistemology is important, because it can influence the way they prepare for classes and their self-evaluation of how well they "know" the material being taught. Research shows that students' epistemology in physics correlate with their interest in physics, with the courses they chose to take, with conceptual gain in the class, and with the decision to become a physicist. [17] There is also suggestion for positive correlations between students' epistemology and their performance motivation (highly motivated adolescents show a stronger belief in the development of knowledge in the natural sciences) [20]. Furthermore, epistemology is not only important at the level of the classroom, but also at the level of a democratic society, since people's perception of science affect both the financial support given to scientific research [17], and also the different degree of trust government officials place on scientists − as one can see now, in this pandemic moment (COVID-19).

Instructors assume students will develop a positive attitude automatically as their concepts get better. There are signs, though, that this is a negative assumption, even in courses that use a reform curricula and "active learning", which focus on improving student conceptual understanding. Research shows that it is common for students' epistemological beliefs to become worse after physics instruction that does not take epistemology in consideration. [17]

There have been instructional practices and curricular elements designed to improve students' epistemology in both college and high school. They have been successful in helping students develop considerably more sophisticated beliefs about knowledge and knowing. [21] For a course to succeed it must engage students *in activities that encourage them to approach their learning in epistemologically sophisticated ways.* Many of these courses contain discussions and activities that **explicitly** address the nature of scientific knowledge. However, curriculum which contains only an **implicit** focus on epistemology has also reported to achieve positive results improving students' epistemology. Independently of dealing with it explicitly or implicitly, what all these effective courses did is maintain the epistemological focus during the whole duration of the course and in all its elements. [17]

According to Elby [21], *so many excellent physics courses fail to foster significant epistemological change*, even courses that include some elements with focus in epistemology. Research suggests that *isolated pieces of epistemologically focused curriculum aren't enough. Instead, the epistemological focus must suffuse every aspect of the course. Therefore, the instructor's commitment to an epistemological agenda must go beyond a willingness to implement certain curricular elements.* Finally, he states that there is *no reason to think that partial adoption of* [a suite of curricular elements demonstrated to improve epistemology] *will lead to epistemological change.*

The reports of improvement in students' epistemology have only recently begun to appear in the literature. Because there are still not so many successful cases, it is not clear what are the conditions required and sufficient in order to achieve a positive shift in students' epistemology. [17]

Since these conditions are still not very clear, and there have been reported cases of some epistemology improvement with five to six interventions in high school [22], in this thesis' research it was decided to also evaluate students' beliefs and views about the nature of physics knowledge and knowing, despite Elby's warnings in the above statements.

2.2 Newton's third law

Among the many different topics in physics, "Newton's third law" (N3) − which involves many students' misconceptions − was the one chosen for this research. Therefore its original description as well as a brief explanation will be given as follow. In his 1687 major work *"Philosophiae Naturalis Principia Mathematica"* [23] (with the fundamentals of mechanics), Newton describes the third law of motion in Latin as:

"LEX III.
Actioni contrariam semper & æqualem esse reactionem: sive corporum duorum actiones in se mutuo semper esse æquales & in partes contrarias dirigi."

which translated "to the letter" would be:

"LAW III.
To every action there is always opposed an equal reaction: or the mutual actions of two bodies upon each other are always equal, and directed to contrary parts." [24]

Paul Hewitt [25] explains it in a simple way stating that this motion law *"describes the relationship between two forces in an interaction"* and says that *"To every action there is always an opposed equal reaction"*, or *"Whenever one object exerts a force on a second object, the second object exerts an equal and opposite force on the first"*.

2.2.1 Students' N3 misconceptions and intuitions

Newtonian mechanics is still one of the most difficult physics' topics in school. Among other reasons (like lessons' structure and instructors' inappropriate presentation of coherence) is students' misconceptions. [26] In this chapter the misconceptions involved in the physics of Newton's third law will be presented, starting with the concept of *force*, which is a broad-meaning colloquial term and is the central concept of Newtonian mechanics.

In physics, **force** is denoted by the strength and direction of an external influence on a body. If a body has a constant mass, a force acting on this body changes its speed (makes it faster or slower) and/or changes its direction of movement, i.e., changes its velocity. If several individual forces act on the same body, then the acceleration is proportional to the vector sum of all these forces (resulting force). [14]

Notwithstanding its physical meaning, students use the word *force* in very different ways, which can be related to a physics' concept. When students talk about a rolling ball **storing** *force*, which becomes visible when it hits an obstacle, physicists often talk about *kinetic energy*; when students say that the earth exerts a *force* on a falling ball, physicists speak about *Newton's force*; and when students say "In a collision the rolling ball **transfers** part of its *force* to the resting one", physicits usually talk about *momentum.* With such reinterpretations, one should not assume that students simply use the wrong word for a physical quantity that was actually correctly understood. Rather, it is a collective or cluster term of *force* as *energy, strength, power, swing.* [14] Students' *force* has also been described as a "vague wholeness", which is associated with "effort" [27]. In students' understanding, *force* is an extensive but vague term that stands for a whole bunch of different meanings. What students mean by it depends on the context of use.

Students' cluster term *force* is still present as a misconception even after mechanics' classes. Students acquire the ability to correctly calculate physical quantities regardless of whether they also understand that *force* has a specific meaning in physics that differs fundamentally from the everyday concept of *force.* This may be one of the reasons why teachers systematically overestimate the success of their mechanics' classes. [14]

Noteworthy students' misconceptions regarding the physics' concept *force* include [4]:

- *"Force is, like a property, something a body can possess or store."* It is not seen as a magnitude (quantity) of interaction. The essential aspect of the concept of force that a body acts on a second body (force as a relation, force as a measure of the intensity of the interaction) does not come into play. (Additionally, the heavier, the stronger. The force stored in a body is greater, the heavier the body is. *The earth attracts the moon more than the moon attracts the earth.* [14])
- *"Force is the potential ability to effectuate something (to do something). A body can therefore have force without exerting it."* This concept of force roughly corresponds to the physics' concept of energy.
- *"Force is something transferable: the force that one body has can be transferred to another body."* This concept of force is very similar to the hitting force, in which momentum is transmitted.

- When you start pulling on an object, you feel a force on your hand in the opposite direction. *"This force arises from the change in movement in the body and counteracts the acting force"*. For students it should be clear that the force you feel on the hand is not the one changing the state of motion, but its 3^{rd} law reaction.

- *"Forces are only exerted by "active" bodies, for example living beings or moving bodies. A ball hitting a wall exerts a force on it, but the wall does not hit the ball. The ball pushes itself off the wall."* (*"Passive bodies just offer resistance. Tables have no force on a book lying on it; the table resists it falling down."* [14]).

In addition, students' make a connection between force and motion resulting in, among others, the following misconceptions [4]:

- *"A moving body has force; and the faster it moves, the bigger the force."* ($F \sim v$).

- *"Bodies move in the direction of the acting force."*

- *"If the sum of the acting forces is zero, the body does not move."*

- *"The motion force is "used" up during the movement"* and *"a force in the direction of movement is required to maintain the movement."*

Many of these misconceptions cause serious implications in students' conceptual understanding of Newton's third law. *"Students often interpret the term 'interaction' by a conflict metaphor. They see an interaction as a 'struggle between opposing forces.' It follows from the metaphor that 'victory belongs to the stronger.' Hence, students find Newton's third law unreasonable, and they prefer some version of the dominance principle: in a conflict, the 'more forceful' exerts the greater force. Here 'more forceful' can mean 'bigger', 'greater mass', or 'more active'. Because of its strong metaphorical base, the dominance principle (though it is seldom clearly articulated) is so natural to students that it is one of the last misconceptions to be overcome in the transition to Newtonian thinking."* [10]

However, the previously mentioned misconceptions are not only reason which leads to students' misinterpretations when it comes to the interaction principle. Poor teaching has also its share of damage. It starts with the literal use of Newton's short form

"Actio equals Reactio" in the classroom, which suggests a slightly delayed response to an active cause (since a "re-action" to something is usually after an "action" has happened). Both forces act equally and simultaneously and neither is "first there" (disregarding situations where the speed of propagation of the interaction is relevant [28]). Instead, teachers could say *"Forces always occur in pairs"*. The phrase *"force equals counterforce"* is even more problematic. This leads to confusion with the balance of forces among students and causes a technically inappropriate concept. On the one hand, students associate this with an inner resistance of the body against the external force (a kind of passive resistance). On the other hand, a pattern of *force contest* or *dominance principle* is (re-)activated, so that students think: *"Force and counterforce act on the same body. As long as they are the same or the force is less than the counterforce, nothing happens. Only when the active force becomes somewhat larger than the counterforce, then there will be an effect."* An example of this is when pushing a heavy box. Students assume that the force exerted on the box is initially less than the static friction. Only after reaching a certain force magnitude, the "active thrust" overcomes the "resistance" and the box starts to move. According to students, if the force and counterforce are always the same, the box can never move. The interaction force of the box, acting on the person pushing it, is not considered by the students as counterforce in the sense of the 3^{rd} law. [14]

Another example of great amount of confusion, which involves more than one misconception, can be seen in the explanation given from a student when asked about the clothes in a washing machine: *"The centripetal force is a reaction of the centrifugal force... which is the action. So, both forces act on the clothes and cancel themselves, since force and counterforce are the same but in opposite directions."*

Smith and Wittmann [29] analyzed student reasoning about N3 in various cases separating them in two categories: pushing situations and collision situations. These two situations can be distinguished by the period of time which the two bodies are in contact. While collision situations happen when the bodies interact for a very short period of time, with pushing situations they are in contact for a longer period of time. Though, what causes problems for students and affect how they reason about the forces exerted is not much the period of time the bodies interact, but the various possible combinations of relative velocities and masses (they can be speeding up, moving at constant speed, or slowing down; and they can have different masses) – despite the fact that these differences lead to the same result. Thereby, in addition to the two different categories, Smith and Wittmann describe students' responses in terms of facets of reasoning. From the cluster facet *"The student indicates that the forces in a force pair do not have equal magnitude because the objects are dissimilar in some property e.g., bigger, stronger, faster."*, they found three variations to be important to their study and analysis, grouping all student responses into them: **action** dependence facet (*the more active or energetic object exerts more force*); **mass** dependence facet (*the bigger or heavier object exerts more force*); **velocity** dependence facet (*the moving object or a faster-moving object exerts a greater force*). [29]

*"The **action** dependence facet embodies the notion that one object causes a force, and the other object feels that force. This facet is most likely influenced by the common (mis)statement of N3, "for every action there is an equal, but opposite, reaction." Typically, students are more likely to focus on the action-reaction aspect of this statement rather than the equal-opposite portion. The action dependence facet manifests itself slightly differently in pushing and collision situations. In pushing situations, a student might state that the object doing the pushing is exerting a greater force than the object being pushed. In collision situations, a student might state that the object that initially has a greater speed exerts more force than the object initially at a slower speed."* [29]

*"The **mass** dependence facet expresses the notion that more massive bodies always exert more force than less massive bodies. Students will often cite Newton's second law ($F = m \cdot a$) as evidence of this; however, students often forget that Newton's second law deals with the net force on an object, not each individual force. The is utilized similarly in pushing and collision situations."* [29]

*"The **velocity** dependence facet arises from a confusion between velocity and acceleration. Students often think of force as an intrinsic property of a body in motion similar to momentum rather than a product of the interaction of two bodies. In pushing situations, a student might express the thought that the forces the bodies exert on each other are equal only if the two bodies are moving at a constant velocity. They would lead to a correct answer for incorrect reasons. In collision situations, a student might discuss the force of a moving car being transferred or imparted to a stationary*

one as it starts to move." [29]

Students may also use two or more of these facets simultaneously to solve a single problem. This can cause students to either have a "more consistent" result (for example a more massive object smashing into a smaller-mass, stationary object − might elicit all three facets); or to give the correct answer because the different facets "compensate" each other (for example a crash between two objects moving towards each other: a smaller-mass object with higher velocity and a more massive object with lower velocity). Smith and Wittmann [29] found that the **action** dependence facet was the most common incorrect reasoning used by students in their study.

All of these reasonings students have about N3 make sense for them and they can answer a question about it completely intuitively without even taking time to think or reflect on it. So, when deciding on which approach to use in order to deal with these misconceptions (discontinuous or continuous learning path (see chapter 2.1.1)), teachers should be very careful so that students do not get a stronger feeling that physics doesn't make sense and the logic behind it is only understandable by physicists, i.e., worsening students' epistemology (as known to happen with traditional instruction).

To conclude, the knowledge about students' misconceptions − and their epistemological implications − related to Newton's third law is definitely very important for instructors to consider when planning and teaching the lessons. Since these beliefs can have massive consequences in students' understanding and interpretation of N3, only by knowing how students reason about it can teachers provide their students with conceptual change and the feeling that the physics they are learning make sense.

"...it has been established that commonsense beliefs about motion and force are incompatible with Newtonian concepts in most respects; and conventional physics instruction produces little change in these beliefs. The implications could not be more serious. Since the students have evidently not learned the most basic Newtonian concepts, they must have failed to comprehend most of the material in the course. They have been forced to cope with the subject by rote memorization of isolated fragments and by carrying out meaningless tasks. No wonder so many are repelled!" [10]

3 Methodology

In this chapter, a description of all the necessary steps and processes used in the research of this thesis – the teaching method employed in the intervention (which is the central point of the research question); the tests applied before and after intervention to control conceptual growth and epistemology development; the tools used to analyze the results; and the entire research course of action – will be presented and explained.

3.1 Open Source Tutorials

Even with the knowledge of students' misconceptions and intuitions, it is not always easy for teachers to prepare lessons and come up with solutions, to successfully deal with them. Therefore, exploring methods that have been developed or that the physics research community around the world has shown to have positive results, could come as an important help to many instructors. In this section the teaching method chosen for this thesis' research will be presented and explained thoroughly.

Created by the University of Maryland Physics Education Research Group, the *Open Source Tutorials* (OST) are a collection of active-learning worksheets intended for use in the classroom. They follow the model (or take place within the structure) of a wide range of physics topics tutorials developed by the Physics Education Group at the University of Washington. These tutorials are research-based (ergo, their topics were chosen as a result of extensive physics education research on what topics are particularly difficult for students; and they have been developed through extensive testing, with many groups of students being observed in class to see how they interact with the lesson) and have been shown to be substantially more effective than traditional instruction in helping students build a good conceptual understanding of physics. [30]

The differential of OSTs is that, in addition to conceptual understanding, it focuses on helping students make sense of physics, having an explicit epistemological emphasis – *their physical intuition, the understanding of scientific reasoning, and relating physics to their everyday experience* [30]. In this context, students are expected to understand and deal with their prior knowledge and intuition, establish to what extent these are acceptable, and recognize how it matches the concept being studied. *"Rather than just* *helping students resolve their difficulties, the OST help students understand when their intuitive ideas are applicable and when not."* It has not only the important steps of *"eliciting and confronting misconceptions, but the process of resolution is explicitly one of refining existing (and misapplied) intuitions*, which helps improving students' views about physics knowledge and learning." [31]

Therefore, the OSTs, which make frequent use of both (discontinuous and continuous) [12] cognitive learning paths (see chapter 2.1.1), have many purposes, as follow: helping students improve their conceptual understanding of physics; getting students to learn to reason qualitatively about physics; getting students to build their physics intuition by making connections to their everyday experience; and helping students reconcile their misconceptions of how things work physically **without undermining their trust in their intuitions**. According to OST's developers this last goal is tricky: *"Students often have misinterpretations of their intuitions that lead to "wrong" answers. It's often easy to convince them that they are wrong, but it is not so easy to get them to transform their intuitions. There is a serious danger that they will just say "physics makes no sense to me – I just have to memorize the results (and I will then ignore everything I have learned when I get out of this silly class)." It is therefore very important never to put down a student's intuition, but to try to find why they have that intuition. There is usually a grain of truthful experience in even their incorrect assumptions. Often they have generalized an experience they know well incorrectly."* [32]

The tutorials were developed to be given in a 50-minute class with approximately 15 to 30 students, which are divided in small 3 to 5 student-groups. The students are guided by the tutorials' (4 to 6 pages) carefully designed worksheets, and each group should work and reason together, while instructor(s) float around interacting with them. They were created to be functional also in settings where computer tools are not available for each student, making modest use of digital resources. In order to stimulate students to say what they really think, promote good students' interaction and discussions (instead of giving the answers they think the teacher wants to hear), OSTs are usually not graded (rather, they get feedbacks from instructors during the tutorial's session). Furthermore, OSTs were designed so that each teacher can "make adjustments" or change them in order to fit their individual instructional frame or needs. Notwithstanding, their implementation is not trivial, i.e., the teacher or instructor can easily undermine the intent behind them by inappropriate interactions. [30]

"Many curriculum developers strive for "high-fidelity" implementation of their materials. This project takes another tack for two reasons. First, research shows that teachers modify curricula no matter how hard developers try to prevent it. Second, modifications can be productive if they adhere to the spirit of the original lesson but make adjustments based on a nuanced interpretation of students' needs and progress." [33]

To avoid mistakes when administering the tutorials, the developers of OST call attention to the following guidelines [32]:

"What to do in tutorials

1. *Listen – getting them to explain their thinking or confusion is the best starting point for delivering effective hints or suggestions;*
2. *Encourage collaboration – Get them to confirm every item with other members of their group. I have often found that each student thinks the answer is "obvious" and as a result assumes that the other students agree with them – even when they don't;*
3. *Let them work – If they are having a good discussion (even if some of them are wrong) let them go and try to work it out themselves. Especially for students new to tutorials it's good to smile, nod, and say something like "good work, keep it up!";*
4. *Check them – Listen every once in a while and ask about some of the critical items. If they have all agreed on a wrong answer that will not be addressed later in the tutorial (often these tutorials set them up to say something wrong to work it out later) ask them to reconsider what they have concluded. You might mention some factor they have ignored that has led them astray. (I will sometimes do this even if they are right – but then be sure to come back and make sure they have stuck to their answer!)*
5. *Keep them on task – Encourage them to stay on task and keep on moving. Sometimes they don't appreciate that getting through these tutorials will often take their serious attention for the entire period.*

What not to do in tutorials

1. *Do not talk too much. If you find yourself going on for more than a minute without an exchange and a substantial response (much more than a nod or a "yeah, I see") from a student, you are probably distracting them from productive work.*

2. *Do not pick up a pen or pencil to show them how to do something. Often* teachers *will "miss their blackboard" where they can show how smart they are and how well they understand the topic. That is not your job here! You are trying to guide them to make sense of the material in their own heads, not to appreciate how well you have made sense of it in yours!*

3. *Do not interrupt what appears to be a productive discussion unless it seems to be getting out of hand with people getting angry or refusing to consider others' points of view. In the latter case it may be worth while intervening and asking people to recap their arguments. A bit of guidance might help get them back on track, but listen first so you know what the argument is about.*

4. *Do not confirm the correctness of their reasoning too quickly. Part of what they are trying to learn here is to evaluate their own thinking. If they come to a conclusion and you immediately confirm their correctness, you steal the opportunity from them to learn how to do that. A brief "Why are you concerned?" or "Why do you think that's the right answer?" before a confirmation sends the message that you don't only expect them to figure out the right answer but that you expect them to know that it's the right answer.*

5. *Give <u>some</u> confirmation. While you shouldn't be giving to many answers or confirmations, do not avoid giving them. Sometimes the students need a confirmation and support – but finding the right time to give them is a skill that you want to develop.*" [32]

In addition to these important guidelines, OSTs have an "instructor's guide" for each specific tutorial with tips and explanations of the intentions behind each particular question or step of the worksheet in question. Annotated video clips were also prepared as a teacher-development tool (helps *teachers become better at listening and responding to their students' thinking – a key element of making active learning work* [33]).

All the materials (including homework activities for each tutorial) are available for teachers online. [34]

Research studies have shown that reformed curricula (with elements that have focus on epistemology), with the prolonged use of the tutorials, have a very positive result both in students conceptual understanding and in promoting epistemological improvements (not only in colleges, but also in high schools, as Elby's findings showed [21]).

Also internationally (outside the United States), positive results have been obtained with translated versions of OSTs. After implementing them in a university in Japan, Michael Hull [35] stated: *"Students at Gakugei entered with the typical view that learning physics means listening to and memorizing knowledge from the teacher's lecture and solving lots of problems by rote. However, they quickly changed their view to instead think that physics can and should be learned with reference to intuition and one's own experiences. They came to see the importance of conceptual reasoning, and of making sense of material for themselves."* Furthermore, Hull and Elby [36] revealed that many students *"somehow learned to integrate mathematics into their "constructivist" epistemologies of physics, even though OSTs do not emphasize this integration."*

When fulfilling the necessary subjects for the teacher education masters program in the University of Vienna, I had the opportunity to participate in Dr. Hull's OST seminar [37], where we would, not only, "be the students" doing the tutorial's worksheets and homeworks but also had the chance to administer them both to our colleagues and in school − discussing the instructor's guide and understanding the intentions behind each step. Having achieved quite a critical view of traditional instruction − not only due to classes I had taken, professors I had had and previous teaching experiences, but also by seeing the reality in schools during the various practical trainings − I have been keen about trying out "active-learning" and mixing up different teaching methods. As I began to do the tutorials and to understand OSTs' approach − "refining intuitions" − which helps students form new physics perspectives considering (not abandoning!) their beliefs about the world that surrounds them, and which pushes them toward Einstein's perspective that science is *"a refinement of everyday thinking"* [1], I knew exactly that this was what I wanted to research on my master thesis.

3.1.1 N3 OST

Newton's third law Open Source Tutorial (N3 OST) [38] considers students' intuitions in a collision situation between a heavy truck that rams into a parked car. This is a situation in which all three facets of reasoning (action, mass, and velocity dependence facets – see chapter 2.2.1) can be triggered.

In the N3 OST instructor's guide [39] the developers say: *"In our experience, this tutorial is perhaps the most jarring and memorable one of the set. Newton's Third Law is a very counterintuitive piece of physics when placed in the context of common collisions. We've set this up so that students start off making incorrect predictions about the forces between objects in a collision."*

Before administering the N3 OST in the researched classes, all the steps in the tutorial and how each one of them should be conducted, were thoroughly discussed with Dr. Hull. Since students had already had mechanics traditional instruction, in the beginning of each class special emphasis was given that whenever they were asked about *common sense* or *intuitions*, they should give their intuitive answer. Furthermore, since students (from all the tested classes) rarely work in groups, it was really pointed out that they should work together, discussing with each other everyone's ideas.

So, the tutorial begins with the following instructions:

The main point of this tutorial is helping you learn more strategies for learning physics concepts that seem to defy common sense.

I. Newton's third law and common sense

According to Newton's third law, when two objects interact,

The force exerted by object A on object B is equal in strength (but opposite in direction) to the force exerted by object B on object A.

Often, this law makes perfect sense. But in some cases, it seems not to.

Consider a heavy truck ramming into a parked, unoccupied car.

Figure 1: N3 OST - explanation of the situation in part I

Then, followed by three questions to be discussed and answered with the group:

> A. (*Work together*) According to *common sense*, which force (if either) is larger during the collision: the force exerted by the truck on the car, or the force exerted by the car on the truck? Explain the intuitive reasoning.
>
> B. (*Work together*) We've asked this question of many students, and a typical response goes like this:
>
> Intuitively, the car reacts more during the collision. (You'd rather be riding in the truck!) So the car feels the bigger force.
>
> Is your group's explanation in part A similar to or different from this? Explain.
>
> C. (*Work together*) According to Newton's third law, which of those forces (if either) is bigger?

Figure 2: N3 OST - part I questions

In question A just about every student answered that the truck's force is larger (as predicted). After discussing it, as groups, all of them agreed and gave the same intuitive answer. In B it was an immediate general consensus, and the question's intention (to cement the response hoped from A) was achieved (the very few who had followed Newton's law strictly to answer A were also in complete agreement in B). This question establishes the common way of discussing the collision as *"car reacts more"*. In question C they confront their intuitions for the first time in the tutorial.

The next step of the N3 OST is an experiment where they confront their intuitions for the second time. The groups were called almost always individually to come the teacher's desk, so that the students could really be close to the experiment and be able to express themselves freely. Very few times this was not possible (when more groups had finished at the same time), and two groups would be called to come together.

> D. *Experiment.* Is this a case where Newton's third law doesn't apply? At the front of the room, the TA has set up an experiment that simulates a truck ramming a car. Go do the experiment and record the results here. You can also test whether Newton's third law holds for other collisions.

Figure 3: N3 OST - part I experiment

At the teachers' desk, after showing and explaining the objects involved in the demonstration, and before actually carrying out the experiment, students were asked to give their opinions to what was going to happen or, more precisely, which force (if any) would be larger. Most students predicted that the truck's force would be larger, but some very few (usually the same ones who began saying that the forces were the same in question A) did say that it "had" to be the same because of Newton's law. Sometimes, after hearing this from a colleague, a student would quickly say they agreed, but if asked again *"is that what **you** really think?"*, they would change back their opinion. This kind of interaction with students is an example of the instructions given from the OST developers, and that teachers improve with practice (see item 4. on *"What not to do in tutorials"*, chapter 3.1).

- **The experiment and iOLab**

The experiment, which reproduces the situation of a parked car (of mass m) being hit by a truck (of mass $2m$) was performed with two iOLab [40] remotes.

The iOLab remotes are devices (little boxes) that have wheels on one side (which is great to be used as carts, simulating cars) and many different built in sensors and electronic devices (force probe, magnetometer, accelerometer, gyroscope, etc.), which can measure numerous physical quantities (force, acceleration, velocity, atmospheric pressure, light intensity, sound intensity, etc.). Before using them, they should be carefully calibrated (it was done once a day, before the intervention classes). They come with a USB powered radio transmitter (dongle), which is plugged in the computer and connects to the radio transmitter inside the remote. A software can be installed in the computer and after synchronizing (one or more remotes) to the dongle, everything that happens with the "carts" is analyzed immediately (no time delay) and can be seen at the computer screen. There is the choice of showing sensors from one cart or both carts on the software, as well as the different physical quantities' graphs.

For the experiment, a small box with weights inside (of exactly same mass as the iOLab remote) was attached to one of the remotes and this was the "truck". Before the demonstration, the masses of both "car" and "truck" were shown to students with a kitchen scale, so they could be sure that the truck's mass was double the car's mass. Afterwards students' predictions were asked and most of the them wouldn't just be that the force exerted by the truck would be *"larger"* but *"two times larger"*.

On the screen both remote sensors could be seen with graphs of *force versus time*, and still before the experiment was carried out, students often asked which one of the graphs (left or right) belonged to the car or truck. Since the truck was always placed on the left side and the car parked on the right side, the plots were respectively placed on the screen. The positive axis for the carts were defined in opposite directions (both truck and car had positive force when hitting their front), so it would be easier to compare both spikes in the same direction. For the collision both remotes had a flat disk screwed to the force probe, so apart from the weight box, they looked exactly the same. Furthermore, a one-meter-long "track" or "road lane" was built to make sure the iOLabs would hit each other exactly front-to-front and to avoid that they fell from the desk. Thus, the car would always hit its back on the end of the track, generating a second spike on the car's force plot (see graph on the right of Figure 4).

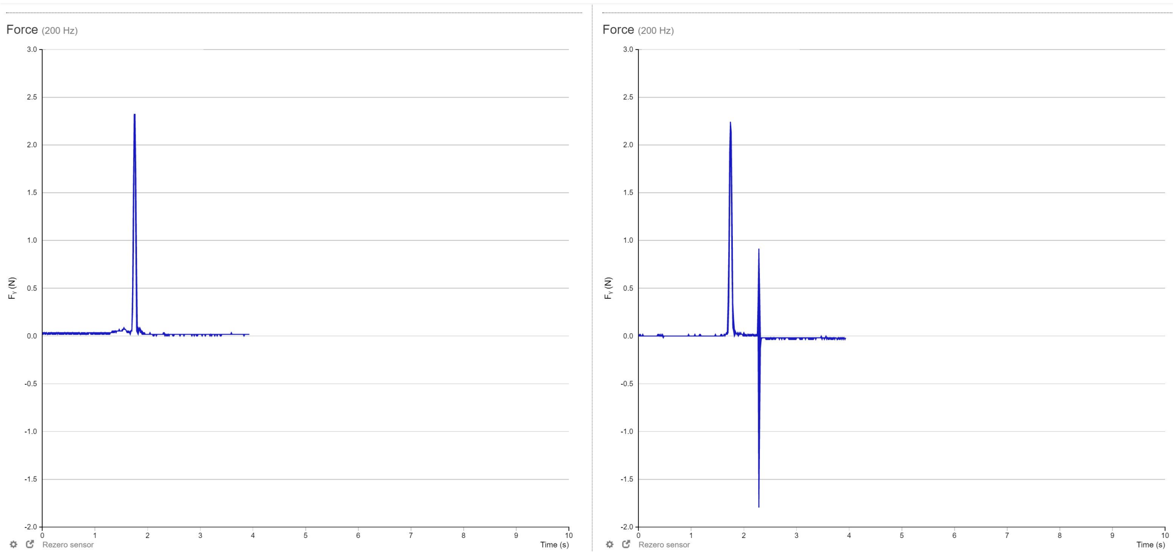


Figure 4: *Force-time* plot obtained in one of the classes (iOLab software)

The N3 OST instructor's guide [39] expresses the concern that, even with calibration, the forces may not show up as equal. It suggests that *"students need to see what it means for a force to be equal. We set up our force probes to take data over a few seconds. You can get one of your students to come up and take two force carts (equal mass). Tell him or her to push them together equally. Your students should expect* ***those*** *forces to be equal. Check out what the computer gives, and indicate that this is "what equal looks like" on the computer"*. There are a couple of reasons why this "pre-experiment" was not done when administering the tutorial for this research. First, testing the experiment several times before classes showed that the difference between the two spikes on the plots were very small (i.e., insignificant; see Figure 4). Then, there was the time issue, since the truck had been previously prepared with the weight box attached, it would take longer to remove it, make the demonstration, attach it back, weight both carts, take their predictions and then carry out the actual experiment. Instead, some time right after the collision was given for students to "make their wishes" (sometimes they would ask to switch the positions from car and truck, others to make the car hit the truck); and to point out "small measuring errors" and the downward spike due to the car hitting the track. However, since their predictions were never *"the force will be a little bit larger"*, this concern was never an issue (absolutely every student agreed or said that the forces on both graphs were the same).

Furthermore, a *zoom in* on the time axis was easily done to show students what was happening on both car and truck at the same time (see Figure 5 below).

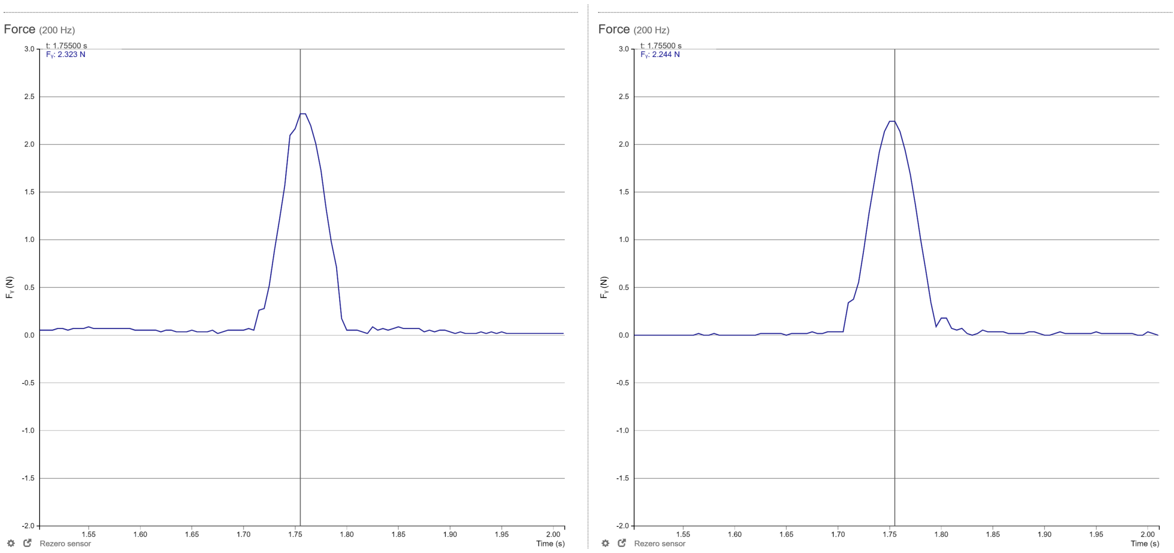


Figure 5: *Force-time* plot (*zoom in* on the *time* axis)

Subsequently on the tutorial, students go back to their places and work out the next question individually. This is the only question in the whole N3 OST which should not be discussed with the group (at first). Right after the experiment, while students were getting back to their seats, this was emphasized.

This second part of the tutorial (inserted below) drives students to indicate their beliefs and attitudes about physics (focus on epistemology). Therefore, students were told that this was a personal opinion question and that the "correct" answer was the one they believed in. Furthermore, if they could not agree to any of the three options (i., ii. or iii.), they had the choice of making up a fourth one (iv.), as long as they explained it. When discussing with the group, in question B., students would sometimes "change their mind" and agree with other members of the group. Special attention was also given telling students that, even though this question was not about the concepts of physics itself, it was very important that they did discuss it after answering A.

II. What to do with the contradiction between common sense and Newton's third law?

Before moving on to the next part of our Newton's third law lesson, let's consider the contradiction we just found between physics and common sense.

A. (*Work individually*) In summary, for most people, Newton's third law contradicts the common-sense intuition that the car reacts more during the collision. Which one of the following best expresses your attitude toward this contradiction?

   i. We shouldn't dwell on these kinds of contradictions and should instead focus on learning exactly when Newton's third law does and doesn't apply.

   ii. There's probably some way to reconcile common sense with Newton's third law, though I don't see how.

   iii. Although physics usually can be reconciled with common sense, here the contradiction between physics and common sense is so blatant that we have to accept it.

   Briefly explain why you chose the answer you chose.

B. Discuss your answer with your group. Is there a consensus or do people disagree?

Figure 6: N3 OST - part II questions

The recommendations in the instructor's guide for this part are: *"A. Although this should be asked about at the checkpoint, wander around and see what students say in response to this question. This will give you a feel for where your groups are when it is time to check them out. B. The students can say whatever they want here, but again, if there's even a slight difference in reasoning between even one of your students and the rest, make sure to talk about it at the checkpoint."* [39]

The goal of the next (third) section of the tutorial (Refining intuition) is to get students to see that Newton's law does hold, by going through some simple calculations which **take their intuition** (that the car *reacts more*) **in consideration**. So, they don't have to discard their common sense prediction, they just have to really understand it and refine it. On question A. it was important that every student expressed their ideas of how much speed the car would gain and that all of them were discussed in the group.

III. A new strategy: Refining intuition

Before accepting that there's an irreconcilable contradiction between Newton's third law and the intuition that the car reacts more during the collision, let's try a reconciliation strategy called *refining your intuitions*.

A. (*Work together*) We'll start with a new question. Suppose the truck's mass is 2000 kg while the car's mass is 1000 kg. And suppose the truck slows down by 5 m/s during the collision. Intuitively, how much speed does the car gain during the collision? (Apply the intuition that the car reacts more during the collision, keeping in mind that the truck is twice as heavy.) Explain your intuitive reasoning.

★ *Consult an instructor before you proceed.*

Figure 7: N3 OST - part III question A

The majority of groups got it really fast and easy, and would just say something like: *"well, intuitively it should be two times more, since the mass of the truck is two times larger"*. Some students, though, would say that, since they saw on the experiment that the forces are the same, the car also had to change its speed by $5\frac{m}{s}$. Luckily, there was usually someone on the group that would contradict the idea with the argument that "intuitively" it would change its speed by $10\frac{m}{s}$. On the rare cases, in which this did not happen, and all other students in a group would just agree, one or another comment or reflection would do the trick. For example, asking them what was their first intuition (on part I of the tutorial) and telling them to use that intuition; or as in two groups (as they were stuck with the idea of the same speed change) making the comparison to a bowling ball hitting a pin and how they compare the change in speed of the bowling ball to the change in speed of the pin; and a last resource, which was used once (since the group was convinced that *"it had to be more, but it had to be* $\frac{1}{2}$ *more, so it would be* $\frac{3}{2}$ *of what the truck changed and that would be* $\frac{15}{2} = 7,5\frac{m}{s}$*"*, and started to loose too much time with it), was simply saying: *hey, I heard one of the groups talking about* $10\frac{m}{s}$ *... what do you think about that? Talk about it! Does it make sense considering your initial intuitions?* Then one of the group members convinced the others that "according to their previous intuitions" $10\frac{m}{s}$ made sense.

⋆ Checkpoint 1: *There are two main objectives for this checkpoint. First, find out what your students said to the question in II.A. and talk out any disagreements. Early in the semester, you need to keep stressing that these questions aren't disposable and worthless of attention. Next, check their answers to III.A. Make sure that the* $10\frac{m}{s}$ *idea is out there on the table, and see that all the students are at least sympathetic to it. That's the number you need to use in the next part to get the desired "final conceptual payoff."* [39]

Following, is one the most exciting parts of N3 OST, when students realize that the forces really **are** the same and that their intuitive guess actually **does** agree with Newton's third law. They get this big *"A-ha"* moment, which is quite nice to see. Most groups had great discussions, (at question 3.) when – only then – many realized that the car's acceleration is two times larger than the truck's acceleration. For the calculations in B. the equation for acceleration $\vec{a} = \frac{\Delta\vec{v}}{\Delta t}$ and Newton's second law $F = m \cdot a$ were written on the black board.

B. Does your answer to part A agree with Newton's third law? To find out, we'll lead you through some quick calculations.

1. Suppose the car and truck remain in contact for 0.50 seconds before bouncing off each other. Calculate:
   i. the truck's acceleration during the collision.
   ii. the car's acceleration during the collision (assuming your guess about its change in speed is correct).
2. To good approximation, the forces that the car and truck exert on each other are the *net* horizontal forces they feel during the collision. Starting with the accelerations you just calculated, use Newton's second law (the one relating net force to acceleration) to find:
   i. the force felt by the truck during the collision.
   ii. the force felt by the car during the collision.
3. The accelerations and forces you just calculated were all based on your guess about the car's gain in speed – a guess based on the intuition that the car reacts more during the collision. Does that intuitive guess agree or disagree with Newton's third law? How do you know?

☆ *Consult an instructor before you proceed.*

Figure 8: N3 OST - part III question B

In a few groups some discussion came up due to disagreements about the idea that the force *"felt by the truck"* was *"produced by the car"* and vice-versa, which lead students to conclude that $F_{truck} = m_{car} \cdot a_{car}$ and $F_{car} = m_{truck} \cdot a_{truck}$ respectively. In fact, this would also lead them to the correct answer (since the forces **are** the same). However, since it could have lead to misconceptions in Newton's second law, a short reminder of it was provided to those students (*"if you take (and multiply) the mass and acceleration from a certain body, you find the force that was exerted on this* ***same*** *body (or felt by this body)"*).

⋆ Checkpoint 2: *This checkpoint is to ensure that the calculations are correct and that the students see agreement between N3 and their intuition.* [39]

IV. What just happened?

A. (*Work together*) We need to sort out what to do with the *car-reacts-more* intuition, *i.e.,* the idea that the car reacts more than the truck during the collision. At the beginning of this tutorial, when you answered a question about the forces acting on the car vs. the truck, that intuition led to a wrong answer that disagreed with Newton's third law. But in section III above, when you answered a question about changes in velocity, that same intuition led to a right answer that agrees with Newton's third law. So, what's up with the *car-reacts-more* intuition? Is it wrong? Is it right? Is it something else? Explain.

B. (*Work together*) We now see that the car "reacts" more during the collision in the sense that it undergoes a greater change in velocity, *i.e.,* it experiences a larger acceleration. Give a common-sense explanation for why the car reacts (accelerates) more during the collision even though it feels a force no bigger than the truck feels. Ask your TA for a hint, if needed. This is the second most important question in the tutorial.

C. (*Work together*) We intended this tutorial as a lesson not just about Newton's third law, but also about strategies for learning physics concepts that seem to contradict common sense. What general strategies are suggested by this tutorial—strategies you might be able to use with counterintuitive concepts appearing later in the course? This is the most important question in the tutorial.

Figure 9: N3 OST - part IV questions

A 50-minute lesson is very tight for the tutorial but all groups could get this far (which was good, at least so they could see that N3 does hold and that their intuitions were not "completely" wrong). But the fourth section of the tutorial (see Figure 9) is important, not only to give closure to the main ideas, but also to make them think about the nature of physics and of learning physics. Some groups could discuss A., very few B. and no group got to C. In general, a few minutes before the end of class, questions A. and B. would be brought up and discussed for the whole class together. This was an attempt not to let them leave without having at least heard the ideas of A. and B. At this point, in some classes, Hewitt's [25] technique of exaggerating the letters was used on the board to help them visualize the intuitive relation: ${\Large m} \cdot \Delta\vec{v} \sim m \cdot {\Large \Delta\vec{v}}$.

The results reported here reflect what one might expect if the OST is used in Austrian high schools, and not the ideal of the OST developers (as shown below in the instructions for this part). So, in my opinion, maybe the comments of *"being the second and the most important questions"* in B. and C. could be removed from the tutorial when used, for example in Austrian high schools, which could avoid possible frustration from not being able to discuss them.

The instructions for this part are:
A. The goal is for them to see that car-reacts-more isn't wrong or right in itself, but depends on context. If you say the car feels more force, that's wrong. If you associate "acceleration" with "reaction," you're now correct. Some students will recognize this as a bit of trickiness arising from our use of language.
B. The answer to this question is that a car's smaller mass will accelerate more than a heavier mass given equal forces. It's not a complicated answer, but it's important that all of your students get this.
C. We call this the "most important question," but the students often see it as the least important question since it doesn't contain any new physics. NOTE: Do not let your students leave tutorial until you've seen their answers to B and C here. If they are allowed to run off without discussing these questions, you will have great difficulty focusing their attention on questions like this in future weeks. In preparing to teach this lesson, it might help to think about what responses would be good to C. [39]

Studies have shown that the N3 OST helps students have a better conceptual understanding, not only compared to traditional instruction, but to other active-learning tutorials. Smith and Wittmann [29] have researched and compared three methods for teaching Newton's third law and stated: *"we find that students using the OST version of the tutorial perform better than students using either of the other two."*

However, as previously stated, *isolated pieces of epistemologically focused curriculum aren't enough. Instead, the epistemological focus must suffuse every aspect of the course. Therefore, the instructor's commitment to an epistemological agenda must go beyond a willingness to implement certain curricular elements.* There is *no reason to think that partial adoption of the* [epistemologically focused curriculum described in the paper] *will lead to epistemological change.* [21]

Personaly, I found working with the N3 OST in high schools during this time very enjoyable! I could see that students had a great time going through the tutorial's activities and that even those who are usually less active in class (which I realized when I later returned to the classroom to observe the teaching of the instructor) had participated with enthusiasm. In my opinion, that alone was a great gain. Very seldom there would be a group which needed a little extra time "to get it running", but as soon as they started, there would be no other topics on the table. At the end of all N3 OST classes, independently of the results I would get from this research, I knew this would be a tutorial to be "inserted" in my future lessons' plan. The feedbacks from the classes' teachers were very positive, as well as from Dr. Hull after listening to the recordings from the first N3 OST lesson: *" It is also clear that the students had a good time discussing with each other and seemed to reach the important points."*

Apart from the first N3 OST intervention, where students spoke Standard German because it was being recorded, students spoke mainly Austrian dialect during classes. The German version of the N3 OST, used to do this research was translated by two native German speakers and underwent several cycles of revision as a result of consultation with Dr. Hull for authenticity of the translation. The German version of N3 OST is in the Attachments (see Figure 42)

3.2 Test instruments

A very hard but also important task in education research is assessing what students know and learn. In both education research and in teaching practices, pre-post testing are useful and helpful methods, which are commonly used. There are many test instruments, already well researched and known by the physics' education research community, to assess not only students' conceptual understanding but also epistemological beliefs in physics.

"When using these methods in practice, researchers and teachers are often concerned about the length of the assessment and memorization from pre- to post-test. If a test is too long, instructors may be reluctant to administer the test because it might take up too much time in an already tightly scheduled classroom. If the same test is used in both pre- and post-testing, ***at least 5 weeks*** *of time must elapse between the two tests in order to reduce the influence from memorization."* [41]

In the following chapters the test instruments, which are important and which were used in this thesis' research − the *Force Concept Inventory* (FCI) [10] for conceptual understanding, the *Maryland Physics Expectations Survey* (MPEX) [42] for epistemological beliefs, the *split task* [43] "way" of answering multiple-choice questions, as well as the *Force-Test* (built for this research) − will be presented and discussed.

3.2.1 FCI

The *Force Concept Inventory* (FCI) is a test instrument created by Hestenes, Wells and Swackhamer [10] to evaluate student understanding of the fundamental concepts of Newtonian physics. The FCI is, since the 1990s, the most internationally used test for students' misconceptions in kinematics and dynamics, and it has established an extensive collection of tests results. The FCI has two versions − the original from 1992 and the revised version from 1995, which is the one used for this thesis' research.

The FCI is composed of 30 multiple-choice questions (each with five possible answers − from A to E), which have been categorized into six major conceptual dimensions (Kinematics, First Law, Second Law, **Third Law**, Superposition Principle and Kinds of Force). There are no questions where calculations are necessary (the questions are purely conceptual). By answering the FCI multiple-choice questions, the student is compelled to make a decision between Newtonian concepts (represented by the one correct answer in each question) and commonsense alternatives – or misconceptions (represented by the other four answer possibilities). At first glance the FCI questions appear to be quite banal (therefore not very revealing) to the majority of physics teachers, but in the moment that they find how badly their own students did on the test, this becomes an astonishment. However, rather than an intelligence test, the FCI as a whole is a good indicator of Newtonian thinking. [10]

It has been argued that the FCI can be used for several different purposes – as a *diagnostic tool*, to identify and classify misconceptions, to raise the awareness of misconceptions among teachers' students; for *evaluating instruction*, being a very accurate and reliable instrument; and as a *placement exam* in colleges and universities (not in high school, since it is not a test of ability) to help determine if student understanding of introductory physics is sufficient for a more advanced course. [10]

Even with the considerable length of which the authors have gone through, in order to validate the FCI, conducting follow-up interviews with students, the meaningfulness of the test results has been controversial and discussed. Huffman and Heller [44] analyzed the correlations between all items in order to select groups of items that all appear to measure the same idea. By examining the items that group together (factors), they determine if the test actually measures the concepts it appears to measure. Among their conclusions, here we should consider these two:

- *"Caution is also advised in analyzing any of the six conceptual dimensions separately. The fact that the items did not group together on the six conceptual dimensions of the force concept indicates that the FCI should not be decomposed into the six dimensions originally proposed by its authors."*

- …*"it appears that the inventory can still be used as a diagnostic tool and as a means for evaluating instruction [...] However, the results do not warrant the use of the inventory as a placement exam."*

Even though they explicitly advised not to use the conceptual dimensions separately, Huffman and Heller did find that **all four questions from the inventory that addressed Newton's third law grouped together on the significant factor for**

high school. [44] This finding was considered important for this thesis, since N3 is the conceptual dimension researched and analyzed.

Hestenes and Halloun stated that the FCI can be used for several different purposes, but that **evaluating the effectiveness of instruction is the most important one** [45]. The FCI test remains a standard tool (and is the most used one internationally) for evaluating the success of mechanics' classes [14].

Another research done on the test is Wang and Bao's analysis of the FCI with *item response theory.* [46] For each item of the FCI, it was established – among two other parameters – values for the *chance of guessing correctly* parameter (ranging from 0 to 30%). This parameter describes the probability of giving a correct answer by an examinee with very low proficiency (instead of the often estimated 20%, which implies the probability of choosing each answer to be equal). Sometimes students give the correct answer on a question without understanding the physics content being tested in that question, so it is said that these are *false positives.* Even though, when a student randomly guesses the correct answer it is also a false positive, it was found that additional false positives arise on some of the FCI items due to **incorrect reasoning**. [47]

The four questions in the *conceptual dimension Third Law* of the FCI will be presented with some of their aspects (misconceptions [10] and guessing parameters [46]).

4. A large truck collides head-on with a small compact car. During the collision:
   (A) the truck exerts a greater amount of force on the car than the car exerts on the truck.
   (B) the car exerts a greater amount of force on the truck than the truck exerts on the car.
   (C) neither exerts a force on the other, the car gets smashed simply because it gets in the way of the truck.
   (D) the truck exerts a force on the car but the car does not exert a force on the truck.
   (E) the truck exerts the same amount of force on the car as the car exerts on the truck.

Figure 10: FCI N3 question number 4

Question number 4 (see Figure 10) is the **only N3 collision** item on the FCI. The alternative answers address the misconceptions of *"greater mass implies greater force"* (answers A and D) and *"obstacles exert no force"* (answer C). Here, since both bodies are in movement (and their velocities are not specified), the action and velocity facets are not addressed. This question's guessing parameter is 15%. Most students answer A (55% on the pre-test). The correct answer to this question is E.

USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT TWO QUESTIONS (15 and 16).

A large truck breaks down out on the road and receives a push back into town by a small compact car as shown in the figure below.

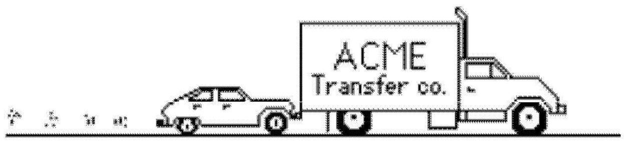


Figure 11: FCI N3 instruction for questions 15 and 16

The next two N3 questions (15 and 16) are about the previous pushing situation (see Figure 11). In both questions the alternative answers address the misconceptions of *"greater mass implies greater force"* (answer B), *"most active agent produces greatest force"* (answer C), *"only active agents exert forces"* (answer D) and *"obstacles exert no force"* (answer E). Since both car and truck have the same velocity, the velocity facet is not addressed. The correct answer to both is A.

15. While the car, still pushing the truck, is speeding up to get up to cruising speed:
    (A) the amount of force with which the car pushes on the truck is equal to that with which the truck pushes back on the car.
    (B) the amount of force with which the car pushes on the truck is smaller than that with which the truck pushes back on the car.
    (C) the amount of force with which the car pushes on the truck is greater than that with which the truck pushes back on the car.
    (D) the car's engine is running so the car pushes against the truck, but the truck's engine is not running so the truck cannot push back against the car. The truck is pushed forward simply because it is in the way of the car.
    (E) neither the car nor the truck exert any force on the other. The truck is pushed forward simply because it is in the way of the car.

Figure 12: FCI N3 question number 15

Question 15 (see Figure 12) has a guessing parameter of 13%. Most students answer

C (53% on the pre-test), since the car is speeding up. Students confuse the balance of forces on a system (Newton's second law) with the equal and opposite interaction forces between two objects (Newton's third law).

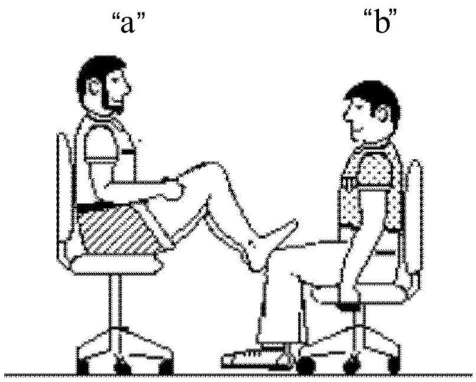
16. After the car reaches the constant cruising speed at which its driver wishes to push the truck:
    (A) the amount of force with which the car pushes on the truck is equal to that with which the truck pushes back on the car.
    (B) the amount of force with which the car pushes on the truck is smaller than that with which the truck pushes back on the car.
    (C) the amount of force with which the car pushes on the truck is greater than that with which the truck pushes back on the car.
    (D) the car's engine is running so the car pushes against the truck, but the truck's engine is not running so the truck cannot push back against the car. The truck is pushed forward simply because it is in the way of the car.
    (E) neither the car nor the truck exert any force on the other. The truck is pushed forward simply because it is in the way of the car.

Figure 13: FCI N3 question number 16

Question 16 (see Figure 13) is the most problematic from the N3 conceptual dimension, or perhaps from the whole FCI. It has a guessing parameter of 34%. Again, students confuse the balance of forces on a system (either Newton's second law – no acceleration means no force; or Newton's first law – constant speed means no force) with the equal and opposite interaction forces between two objects (Newton's third law). However, this time, most students answer the correct option A (43% on the pre-test), but for the wrong reasons (false positives).

28. In the figure at right, student "a" has a mass of 95 kg and student "b" has a mass of 77 kg. They sit in identical office chairs facing each other.

Student "a" places his bare feet on the knees of student "b", as shown. Student "a" then suddenly pushes outward with his feet, causing both chairs to move.



During the push and while the students are still touching each other:

(A) neither student exerts a force on the other.
(B) student "a" exerts a force on student "b", but "b" does not exert any force on "a".
(C) each student exerts a force on the other, but "b" exerts the larger force.
(D) each student exerts a force on the other, but "a" exerts the larger force.
(E) each student exerts the same amount of force on the other.

Figure 14: FCI N3 question number 28

The last question of the N3 conceptual dimension in the FCI is question 28 (see Figure 14). The alternative answers address the misconceptions of *"only active agents exert forces"* (answer B) and *"greater mass implies greater force" and "most active agent produces greatest force"* (answer D). Here, since both bodies are not moving at the beginning, the velocity facet is not addressed. This question's guessing parameter is 13%. Most students answer D (46% on the pre-test), but since the active body is also the more massive one, it is unclear which of the two facets (if not both) arised. The correct answer to this question is E.

None of the four questions addressed the misconception *"motion implies active force"*, that is, the velocity facet.

Following are three other questions from the FCI, which were part of the test administered in this thesis. Therefore, they will also be presented (topics, misconceptions [10] and guessing parameters [46]). The importance of these questions will be explained later in the Split task (see chapter 3.2.3).

12. A ball is fired by a cannon from the top of a cliff as shown in the figure below. Which of the paths would the cannon ball most closely follow?

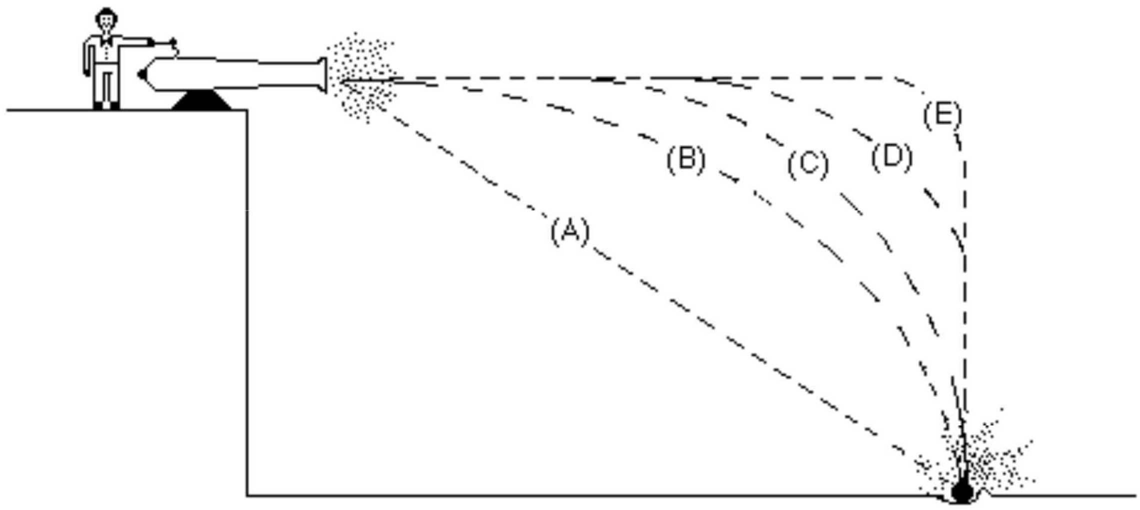


Figure 15: FCI question number 12

The first one is question 12 of the FCI (see Figure 15). This question's topics are *Kinematics – Constant acceleration entails parabolic orbit* and *Kinds of force – Gravitation.* Its alternative answers address the misconceptions of *"impetus dissipation"* (answers C, D and E), *"force compromise determines motion"* (answer A), and *"gravity acts after impetus wears down"* (answer E). This question's guessing parameter is 13%. The correct answer to this question is B.

25. A woman exerts a constant horizontal force on a large box. As a result, the box moves across a horizontal floor at a constant speed "v_0".

The constant horizontal force applied by the woman:

(A) has the same magnitude as the weight of the box.

(B) is greater than the weight of the box.

(C) has the same magnitude as the total force which resists the motion of the box.

(D) is greater than the total force which resists the motion of the box.

(E) is greater than either the weight of the box or the total force which resists its motion.

Figure 16: FCI question number 25

The next one is question 25 of the FCI (see Figure 16). This question's topics are *First Law – with cancelling forces* and *Superposition Principle – Canceling forces.* Its alternative answers address the misconceptions of *"velocity proportional to applied force"* (answer A), *"motion when force overcomes resistance"* (answers B and D), and *"resistance opposes force/impetus"* (answer E). This question's guessing parameter is 10%. The correct answer to this question is C.

30. Despite a very strong wind, a tennis player manages to hit a tennis ball with her racquet so that the ball passes over the net and lands in her opponent's court.
Consider the following forces:
    1. A downward force of gravity.
    2. A force by the "hit".
    3. A force exerted by the air.

Which of the above forces is (are) acting on the tennis ball after it has left contact with the racquet and before it touches the ground?
(A) 1 only.
(B) 1 and 2.
(C) 1 and 3.
(D) 2 and 3.
(E) 1, 2, and 3.

Figure 17: FCI question number 30

The last one is question 30 of the FCI (see Figure 17). This question's topics are *Kinds of force – Fluid contact – air resistance* and *Kinds of force – Gravitation.* Its alternative answers address the misconceptions of *"impetus supplied by hit"* (answers B, D and E) and *"only active agents exert forces"* (answer A). This question's guessing parameter is 9%. The correct answer to this question is C.

3.2.2 MPEX

The Maryland Physics Expectations (MPEX) Survey is a test instrument created by Redish, Saul and Steinberg Redish [42] to probe some aspects of student expectations. The developers explore student attitudes and beliefs about physics and how they change as a result of physics instruction. They developed the MPEX survey to meet the need of studying larger populations of students with a test instrument, which students can complete in less than thirty minutes and which can be analyzed by a computer. The authors started to develop the MPEX survey in 1992 at the University of Washington. They validated the items in a number of ways: by discussing with other faculty and experts in physics education, carrying out student interviews, probing a variety of "experts" with the survey, and administering the survey repeatedly to the same group of students. [42]

The final version of the survey has 34 items about the nature of physics, the study of physics, and students' relation to it. Each item can be rated by the student on a five point scale from strongly disagree (1) to strongly agree (5), i.e., 5-option Likert-scale. When creating the MPEX survey, focus was given on issues that have an effect on the way students interpret and process the physics in the class. Student's feelings about physics, its value or its importance were not considered. Thus, six issues or dimensions to categorize student attitudes towards the appropriate way to do physics and to classify student beliefs about the nature of learning physics were proposed. The survey's items, which were designed to probe a particular dimension is referred to as *cluster*. For each cluster there is a "favorable" view (which agrees with that of most mature scientists) and a "unfavorable" view (which agrees with that of most beginning students). [42]

The description of each dimension, of its favorable and unfavorable views, and an example of item from its cluster (with indication of the favorable view) will be presented as follow [42]:

- **Independence** – beliefs about learning physics – whether it means receiving information or involves an active process of reconstructing one's own understanding; (cluster: 1, 8, 13, 14, 17, 27)
  Favorable: takes responsibility for constructing own understanding;

Unfavorable: takes what is given by authorities (teacher, text) without evaluation.
Example: #1: *All I need to do to understand most of the basic ideas in this course is just read the text, work most of the problems, and/or pay close attention in class.* Favorable view: (strongly) disagree.

- **Coherence** – beliefs about the structure of physics knowledge – as a collection of isolated pieces or as a single coherent system; (cluster: 12, 15, 16, 21, 29)
  Favorable: believes physics needs to be considered as a connected, consistent framework;
  Unfavorable: believes physics can be treated as unrelated facts or "pieces".
  Example: #21: *If I came up with two different approaches to a problem and they gave different answers, I would not worry about it; I would just choose the answer that seemed most reasonable. (Assume the answer is not in the back of the book.)* Favorable view: (strongly) disagree.

- **Concepts** – beliefs about the content of physics knowledge – as formulas or as concepts that underlie the formulas; (cluster: 4, 19, 26, 27, 32)
  Favorable: stresses understanding of the underlying ideas and concepts;
  Unfavorable: focuses on memorizing and using formulas.
  Example: #26: *When I solve most exam or homework problems, I explicitly think about the concepts that underlie the problem.* Favorable view: (strongly) agree.

- **Reality Link** – beliefs about the connection between physics and reality – whether physics is unrelated to experiences outside the classroom or whether it is useful to think about them together; (cluster: 10, 18, 22, 25)
  Favorable: believes ideas learned in physics are relevant and useful in a wide variety of real contexts;
  Unfavorable: believes ideas learned in physics has little relation to experiences outside the classroom.
  Example: #10: *Physical laws have little relation to what I experience in the real world.* Favorable view: (strongly) disagree.

- **Math Link** – beliefs about the role of mathematics in learning physics – whether the mathematical formalism is used as a way of representing information about physical phenomena or mathematics is just used to calculate numbers;

(cluster: 2, 6, 8, 16, 20)
Favorable: considers mathematics as a convenient way of representing physical phenomena;
Unfavorable: views the physics and the math as independent with little relationship between them.
Example: #2: *All I learn from a derivation of a formula is that the formula obtained is valid and that it is OK to use it in problems.* Favorable view: (strongly) disagree.

- **Effort** – beliefs about the kind of activities and work necessary to make sense out of physics – whether they expect to think carefully and evaluate what they are doing based on available materials and feedback or not. (cluster: 3, 6, 7, 24, 31)
  Favorable: makes the effort to use information available and tries to make sense of it;
  Unfavorable: does not attempt to use available in formation effectively.
  Example: #6: *I spend a lot of time figuring out and understanding at least some of the derivations or proofs given either in class or in the text.* Favorable view: (strongly) agree.

By giving a modified version of the MPEX, in which students were asked to explain (in writing) their Likert choice, Omasits and Wagner [48] investigated the validity of the test (the possibility of false negative or positive responses). They found that over 95% of the explanations were consistent with the corresponding choice. Only few questions generated multiple student responses, which were later reviewed. They concluded that most items on the MPEX are valid in a calculus-based introductory course and that the number of evident inconsistent explanations is, mostly, insignificant.

The MPEX survey is one of the first and best known student expectation survey [49] and it has been given to thousands of students at many institutions across the United States [50]. It has also been administered in other countries, for example in Germany [22] and in Thailand [49]. For the study done by Wutchana and Emarat [49] in Thailand, the MPEX survey was translated to the native language of Thai people and validated by academic staff and graduate students in the physics department. They were asked to do the Thai test first and only afterwards the English test (thus, they could not know what exactly the original English questions meant when doing the Thai version). All of the staff and students obtained the same answers for each question in both languages, with minor adjustment of the translation.

The German version of the MPEX was translated by Wilhelm [22]. However, a validation of the translated test was not done. Thus, it does not constitute a standardized test instrument yet. One way of investigating the translation would be a back translation into English and a subsequent comparison with the original. Another possibility would be to ask students about their understanding of the items by conducting interviews. Wilhelm also says that in addition to the problem of possibly changing the items' intension with the translation, it should be noted that the learning and teaching situations in American high schools and universities differ from the ones in German high schools.

Despite these problems, the German version of the MPEX translated by Wilhelm was the one used for this thesis' research. Unfortunately, to investigate the validity of the test was not possible within the scope of this thesis.

An important point, which was observed when administering the test in Austrian high schools, is that the Likert scale from 1 to 5 is very confusing for students here. This problem arises due to Austria's note range. In schools students get a note on exams and subjects that range exactly from 1 to 5. However, 5 is the worse note ("failed") but for the MPEX Likert scale it is "strongly agree". Likewise 1 is the best note for them ("very well") and for the test scale it is "strongly disagree". This issue was not only addressed when giving the instructions before they began doing the test, but many times emphasized throughout the duration of the test. Still, when inserting the test answers to the computer, it was clear that some students hat temporarily forgotten and as they remembered (or were remembered), they scratched a set of answers and circled or crossed the corresponding opposite answer (for example, 2 scratched and 4 circled, 5 scratched and 1 circled). The question of how many students did not realized that and went on answering the "opposite" answers was not possible to know.

3.2.3 Split task

The split task is nothing more than instructions, or "a way" of asking students to answer multiple choice tests – but an amazing way to do so. Before explaining precisely what the split task is, a very interesting statement from Mazur [11] (from when he decided to test his students with the FCI – or an older version of it) leads exactly to the point: *"Intrigued, I decided to test my own students' conceptual understanding [...] The first warning came when I gave the Halloun and Hestenes test to my class and a student asked, "Professor Mazur, how should I answer these questions?* ***According to what you taught us, or by the way I think about these things?""***

This student's question puts any teacher into reflexion, and it touches the core of one aspect of students' epistemology: to what extent do students "believe" what they have been learning? Do these physics concepts make sense to students? And even further: do students think that these physics concepts should make sense? McCaskey and Elby [43] have researched to explore the question: *"Do students really believe the physical principles they learn in class?"* and have very interesting and useful findings.

They used the FCI as the basis for a new probe called a **split task**. The instruction they gave students was to give **two answers** to each question. One answer (where the student would make a circle) represented the answer which the student "really believed". The other (where the student would make a square) represented the answer which the student thought a scientist would give. Additionally, it was made clear that students could circle and square the same answer, if appropriate. [43]

On their study they found that the average split rate was 24% but that, individually, students' split rate ranged from 0% to 90%. Furthermore, on the four FCI items from the dimension Third Law, they found that even among the students who indicated the correct scientists' answers to all of those items, 80% of students split on at least one item. This means that, even if students give the "accepted scientist answers", it is not necessarily the case that they "believe" all of those answers. [43]

Afterwards, searching for the meaning of these splits, they conducted validation interviews. The data obtained with the interviews gave them another perspective for the reason why students split. *"A split never indicated that the student disbelieves the scientists' answer."* Sometimes students split indicated a lack of self confidence in their "belief" thinking and often expressed clear epistemological views, for example this student's statement: *"It is true, if you listen to a scientist talk, a lot of times, you don't understand everything, and so if I didn't understand everything in the answer, it seemed like a plausible scientist explanation."* Thus, the reasons for students' splitting were more subtle and varied from student to student, but often they said that their circled answer corresponded to their **intuition**. [43]

Mainly because of these results, the "circle" instruction was changed to *"circle the answer that makes the most intuitive sense to you"* (see Figure 18), helping to find students' epistemological views about the role of common-sense ideas in learning physics. The results showed that "intuition splits" *really do indicate a discrepancy between a student's common-sense ideas and the answer he thinks a scientist would give.* Furthermore, the validation interviews verified that focusing on the intuition split actually does provide insight into the tendency and capacity students have to reconcile their intuitive ideas with physics' concepts. The interviews also showed students' wish to reconcile the squared and circled answers. [43]

For each question:
– Please **circle** the answer that makes the most intuitive sense to you;
– Please draw a **square** around the answer you think scientists would give.

Example:
A) Answer, which makes most sense for me.
B) Some random answer.
C) Answer I think a scientist would give.
D) Some random answer.
E) Some random answer.

If the answer, which makes most sense for you, **agrees** with the answer you think a scientist would give, draw a **circle** and **square** around that same answer.

Example:
A) Some random answer.
B) Answer, which makes most sense for me, and which I think a scientist would give.
C) Some random answer.

Figure 18: Split task instructions

In cases where the students had received a tutorial (reform students) on the relevant topic between the initial survey and the interview, they often nullified their split. Students had, in fact, **refined their intuitive ideas to be consistent with the physics concepts**. These students are more inclined and/or able to achieve reconciliation when compared with the traditional students who answered the questions correctly. *The reform students split a far lower percentage of the time*, even though a *much smaller — and therefore more selected — percentage of the traditionally taught students got the scientists' answers right.* Furthermore, most of the reformed students expressed an expectation that they would maybe some day obtain a similar reconciliation on the topics they had not studied yet. [43]

To conclude, students can (and often do) easily memorize and repeat information they have learned. Most test instruments, though, do not assess whether a student gives the answers that make sense to them. Students may answer questions in a way that contradicts their beliefs or intuitions, as they try to tell what their teachers want to hear. [43] Differently then the MPEX, which covers many epistemological domains, the final version of the FCI split task considers a **new epistemological instruction, dealing with one aspect of epistemology** and epistemological development − the reconciliation of the physics' concepts with students' intuitions. [51]

In this thesis the split task was used on the Force-Test to find out students splits in Austrian high schools both after the N3 traditional instruction and after the N3 OST. The instructions to the test were not only written on the test, but were also verbally emphasized while giving examples on the black board. At the end, as students returned their answered tests, each question was inspected to see if any circle or square was missing. Those students, who eventually had missing answers, received their tests again to complete it. The test was not hard to administer, but not so easy to score (as also previously reported by Mc-Caskey [51]).

The German version of the split task instruction is in the Attachments (see Figure 43).

3.2.4 Force-Test

In order to measure the conceptual effects after N3 traditional instruction and after the N3 OST intervention, it was important to test students with the four FCI questions that relate to Newton's third law. To apply the whole FCI seemed senseless and a little exaggerated (since the focus is only N3), and both tests (FCI and MPEX) would turn the pre- and post-tests unnecessarily long. However, a test containing only four questions could lead to some problems like, for example, floor effect and lack of a control measure. Floor effect happens when the test exceeds in difficulty for examinees and if many of them score zero, it limits the reliable data values. A control measure was important in order to compare students' conceptual changes in other topics with those of N3. If these changes are similar, then the change cannot be attributed to the intervention.

Taking these points into consideration, a *"Force-Test"* with ten questions about forces was designed containing the following items:

- the **four FCI N3 questions** (4, 15, 16 and 28) – questions 6, 3, 4 and 9 on the Force-Test, respectively;
- three other FCI questions (12, 25 and 30) – questions 8, 2 and 5 on the Force-Test, respectively;
- three N3 "very easy" made up questions – questions 1, 7 and 10 on the Force-Test.

The four FCI N3 questions are the important ones for the study and are described (addressed misconceptions, false positives, guessing parameter values, etc.) in the chapter FCI (see chapter 3.2.1).

The three other FCI questions were included to be the control measure. Therefore they were carefully chosen, taking not only their topics (each one in a different topic), but also their guessing parameter value (which should be low, such that they give a more accurate measure of student ability) in consideration. A description of these questions are also in the chapter FCI (see chapter 3.2.1).

The three N3 made up questions were included to avoid floor effect. This means that they had to be as simple and easy as they could, making sure that all (or most) students would answer as many as possible correctly. These questions are also multiple choice with five possible answers, following the FCI's design. However, to avoid more conceptual over-thinking and too much reasoning from students, these three questions do not follow the answer pattern from the FCI. The answers were formulated to be very simple: containing only one equation with one force (for example F_1) on the left side of the equation and either a multiple or fraction of the other force (for example $3 \cdot F_2$, $2 \cdot F_2$, $\frac{1}{2} \cdot F_2$ or F_2) on the right side of the equation, where the correct answer is always $F_1 = F_2$ (Newton's third law). In all these four answers to all three questions only the magnitude of the forces were considered to make it easier (focusing only on the "equal" part and not on the "opposite directions" of the N3 law). The last answer possibility was: *"none of the above"*. Two of the questions showed a collision situation and one a pulling situation.

Following are the three questions (translated to English) designed for the Force-Test used in this thesis' research.

1. In this game, two balls, which are exactly the same, hit each other again and again in the middle. Let F_{LR} be the magnitude of the force that the left ball exerts on the right one, and F_{RL} the magnitude of the force that the right ball exerts on the left one.

It is correct to say that:

A) $F_{LR} = \frac{1}{2} F_{RL}$

B) $F_{LR} = F_{RL}$

C) $F_{LR} = 2 \cdot F_{RL}$

D) $F_{LR} = 3 \cdot F_{RL}$

E) None of the above.

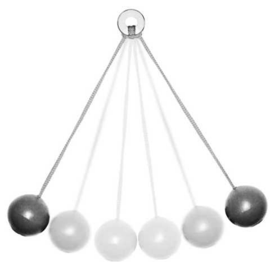


Figure 19: N3 created question 1

7. A girl and a boy hold each other and turn in circles. Let F_{MB} be the magnitude of the force that the girl exerts on the boy, and F_{BM} the magnitude of the force that the boy exerts on the girl. Correct is:

A) $F_{MB} = 3 \cdot F_{BM}$

B) $F_{MB} = 2 \cdot F_{BM}$

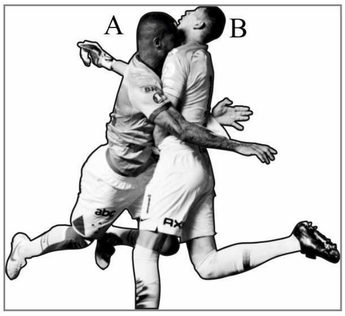
C) $F_{MB} = \frac{1}{2} F_{BM}$

D) $F_{MB} = F_{BM}$

E) None of the above.

Figure 20: N3 created question 7

10. In a soccer game, player A and player B chase the ball that is in the air and collide. Exactly at the moment of the collision, the magnitude of the force that player A exerts on B (F_{AB}), compared to the size of the force that player B exerts on A (F_{BA}), is:



A) $F_{AB} = 3 \cdot F_{BA}$

B) $F_{AB} = 2 \cdot F_{BA}$

C) $F_{AB} = \frac{1}{2} F_{BA}$

D) $F_{AB} = F_{BA}$

E) None of the above.

Figure 21: N3 created question 10

One noticed "flaw" on the design of the test was the position of the FCI question 12 (question 8 of the Force-Test), which was on the upper part of the last sheet (on the back) and students often did not "see" it. Since every question was checked for double responses (split task), whenever students missed this question, they would get it back to complete it.

The German version of the whole Force-Test is in the Attachements (see Figure 44).

3.3 Analysis' tools

Since pre- and post-tests have become a very used method of assessment in education, researchers developed many tools to make its analysis. Each scientific research establishes which of these tools should be used, which parameters must be calculated or obtained for a better quantification or certainty of the results, or for better understanding them. The tools to quantify the improvement (g-factor) and to critically analyze these quantifications statistically (p-value), both used to analyze this thesis' data, will be described and discussed below.

3.3.1 g-factor

For pre-post testing, the *normalized gain* (g) – or g-factor – is often used by the physics education community. It is the number of points gained compared to the number of points that could have been gained, i.e., it is the realized improvement to the maximum possible improvement:

$$g = \frac{y - x}{1 - x} \tag{1}$$

where x is the pre-test score and y the post-test score (both fractional, that is: the ratio of the number of correct answers to the total number of questions).

Example: If there were 10 questions in a test and the student's pre-test score was 2 and post-test score was 4, the g-factor would be as follow: $g = \frac{0,4-0,2}{1-0,2} = \frac{0,2}{0,8} = 0,25$. This means that 25% of the maximum possible improvement was achieved (2 out of 8 questions). Now if the student's pre-test score was 8 and post-test score was 9, the g-factor would be: $g = \frac{0,9-0,8}{1-0,8} = \frac{0,1}{0,2} = 0,5$. In this case the student achieved 50% of the maximum possible improvement (1 out of 2).

As we can see, a student that had one point more on the post-test could have a higher normalized gain than a student who had two points more on the post-test (if the second had more "room" for improvement). So, the normalized gain has big advantages over simply calculating the absolute gain ($y - x$). In theory, it allows the comparison of increases in scores independently of pre-tests scores for larger student populations. [52]

To compare whole populations of students – and not just singular cases – the normalized gain g from all the students can be added and divide by the total number of students, i.e., calculate the *average normalized gain* ($\bar{g}$) of the population:

$$\bar{g} = \frac{1}{N}\sum_{k=1}^{N} g_k = \frac{1}{N}\sum_{k=1}^{N} \frac{y_k - x_k}{1 - x_k} \tag{2}$$

where N is the total number of students being tested, and x_k and y_k are the pre- and post-test scores, respectively, of the k^{th} student in the population.

Another possibility is to calculate the populations' average pre- and post-test scores and thus to calculate the *normalized gain of the arithmetic means* (g_m):

$$g_m = \frac{\bar{y} - \bar{x}}{1 - \bar{x}} = \frac{\sum_{k=1}^{N} y_k - \sum_{k=1}^{N} x_k}{N - \sum_{k=1}^{N} x_k} = \frac{\sum_{k=1}^{N} (y_k - x_k)}{\sum_{k=1}^{N} (1 - x_k)} \tag{3}$$

where $\bar{x}$ and $\bar{y}$ are the fractional pre- and post-test average score of the population, respectively.

Marx and Cummings [52] argue that the calculations of normalized gain have limitations, especially if students have a lower post-test score than the pre-test score. Their arguments are that: it has a low test-score bias; it generates a non-symmetric range of scores (making interpretation difficult in some cases); and that if a student achieves a perfect pre-test score it makes it impossible to calculate $\bar{g}$ (Equation 2), forcing the calculation of g_m (Equation 3) instead. They propose a normalized **change** (c) instead, which takes the different possible results of pre- and post-tests.

In order to calculate c, three different situations should be considered: if the pre-test score is equal or lower than the post-test score (same or gain), then Equation 1 is used; if the pre-test score is higher than the post-test score (loss), then an analogous expression is used, which is the ratio of the actual loss to the maximum possible loss; and if the student's performance is either beyond or below the scope of the measurement instrument (that is if pre- and post-tests scores are both either perfect or zero),

then these students' scores should be removed from the data sets.
Thus, the equation for c is:

$$c = \begin{cases} \dfrac{y-x}{1-x} & \text{if} \quad y \geqslant x \\ \\ \dfrac{y-x}{x} & \text{if} \quad y < x \\ \\ \text{drop} & \text{if} \quad y = x = (1 \text{ or } 0) \end{cases} \tag{4}$$

having the substantial advantage that students who have post-test scores that increase or decrease by the same percentage relative to the maximum possible gain or loss, respectively, obtain the same magnitude of c. [52]

For example, in a test with 10 questions, a student that has a pre-test score of 4 and post test score of 2 has $c = -0,5$ (decreased by half of the maximum possible amount) and a student that has a pre-test score of 8 and post-test score of 9 has $c = +0,5$ (increased by half of the maximum possible amount). In contrast, Equation 1 would give $g = -0,33$ and $g = +0,5$, respectively.

Avoiding the problem with "exploding" g when a student achieves a perfect pre-test score, Equation 4 allows the calculation of normalized changes for every student under all circumstances, providing the option of calculating the *average normalized change* ($\bar{c}$) for populations of students by averaging their individual normalized changes.

$$\bar{c} = \frac{1}{N}\left(\sum_{y_k \geqslant x_k} \frac{y_k - x_k}{1 - x_k} + \sum_{y_k < x_k} \frac{y_k - x_k}{x_k} \right), k = [1, N] \quad \text{except} \quad y_k = x_k = (1 \text{ or } 0) \tag{5}$$

To alleviate the concern of averaging numbers generated by two different relations (one for gains and one for losses) Marx and Cummings [52] also presented c in a single relation, which will not be addressed here.

In this thesis the idea of handling each individual's score appropriately was considered important. Therefore, in the analysis the results of the *average normalized change* ($\bar{c}$) of the Force-Test will be presented. For purposes of comparing the findings with ones of existing research, the results of $\bar{g}$ and g_m for this test will be presented as well.

3.3.2 p-value

In order to critically analyze the quantified improvements ($\bar{c}$, $\bar{g}$, %, etc...) a statistical test is often required. In this thesis the t-Test paired (two samples for means) will be used to determine if there is a significant difference between the arithmetic means of pre- and post-tests.

"If, given a pre-specified uncertainty level, a statistical test based on the data supports the hypothesis about the population, we say that this hypothesis is statistically proven. Note that the research question has to be operationalized before it can be tested by a statistical test." "An important point is that the research hypothesis which we want to prove has to be formulated as the statistical alternative hypothesis, often denoted by H_1. The opposite of the research hypothesis has to be formulated as the statistical null hypothesis, denoted by H_0". [53]

Considering the arithmetic means of pre-test scores $\bar{X}$ and of post-test scores $\bar{Y}$, the alternative hypothesis can be: $H_1 : \bar{Y} \neq \bar{X}$ (the OST instruction **did** make a difference). In contrast, the null hypothesis assumes that there was no change in the means (the OST instruction did **not** make a difference): $H_0 : \bar{Y} = \bar{X}$.

When a hypothesis is tested, there are four possible situations, in which two types of error can occur [53]:

	H_0 is true	H_0 is not true
H_0 is not rejected	Correct decision	Type II error
H_0 is rejected	Type I error	Correct decision

It is defined the *significance level* $\alpha = P(H_1|H_0)$ as the probability of rejecting H_0 (accepting H_1) considering H_0 to be true (which is the probability of type I error occurring). Its value is pre-specified in a test and $\alpha = 0,05$ is very usually used. On the other hand, $\beta = P(H_0|H_1)$ which is the probability of not rejecting H_0 considering H_0 to be not true (probability of type II error occurring) is not controlled by the construction of the test and can become very high, therefore **a test not rejecting H_0 is not a (statistical) proof of H_0!** [53] This means that, if the comparison between post-tests' results and pre-tests' results show (statistically) that there is no evidence that they are "different", that does not necessarily mean that they are the "same".

There are different ways to run a t-Test, but the one which compares means from the same group of students (or population) at different times – which is the case comparing pre- and post-tests – is called "two-*dependent*-samples" or a "paired data problem". [53]

The test calculates a parameter p, whose value is used to decide whether to accept or reject the null hypothesis. It "*quantifies the probability of observing results at least as extreme as the ones observed given that the null hypothesis is true. It is then compared against the pre-determined significance level (α).*" [54] Informally, the p-value is a measure of the evidence against H_0: the smaller the p-value, the stronger the evidence against H_0. [55] If the p-value is smaller than α, then H_0 is rejected, but if the p-value is larger than α, then H_0 cannot be rejected.

It is typical for researchers to consider the scale below:

p-value	evidence
$< 0,01$	very strong evidence against H_0
$0,01 - 0,05$	strong evidence against H_0
$0,05 - 0,10$	week evidence against H_0
$> 0,1$	little or no evidence against H_0

The p-value is sometimes also called the *significance*, but it is better to use this term only in the context of a test result: a test is (statistically) significant if (and only if) H_0 can be rejected. [53]

When dealing with the p-value some warnings should be considered [55]:
– A large p-value is not a strong evidence in favor of H_0. A large p-value can occur for two reasons: (i) H_0 is true or (ii) H_0 is false but the test has low power.
– The p-value is not the probability that H_0 is true.
– When we reject H_0 we often say that the result is statistically significant. A result might be statistically significant and yet the size of the effect might be small. In such a case we have a result that is statistically significant but not scientifically or practically significant.

Even though the p-value has been widely used in both the social and natural sciences' researches, "*substantial literature has been produced critiquing how p-values are used and understood.*" Vidgen and Yasseri [54] reviewed this critical literature and provided a picture of what the main criticisms are.

Some of the arguments are that: "*it is relatively easy to achieve results that can be labeled significant when a "nil" hypothesis (where the effect size of* H_0 *is set at zero) is used rather than a true "null" hypothesis (where the direction of the effect, or even the effect size, is especified)." ... "indicating what the expected effect size is (thereby setting a nil rather than a null hypothesis) is something that most researchers rarely do." ... "it encourages the use of terminology such as significant/nonsignificant. This dichotomizes the p-value on an arbitrary basis, and converts a probability into a certainty." ... the real problem lies not with p-values, but with* α *and how this has led to p-values being interpreted dichotomously: too much importance is attached to the arbitrary cutoff* $\alpha \leq 0.05$*."* Furthermore, they "*recommend that far lower significance levels are used, such as 0.01 or 0.001.* Besides these criticisms, the fact that many researchers do "selective reporting", where only the positive results are reported or where there is a lack of transparency with the results, is also a criticism reason. [54]

All the t-Tests – of the obtained data in this research (see Figures 39,40 and 41 in the Attachments) – were performed in Excel and a summary of the pre-determined parameters and hypothesis used when analyzing the p-value in this thesis is presented as follow:
$H_0 : \bar{Y} = \bar{X}$
$H_1 : \bar{Y} \neq \bar{X}$
significance level $\alpha = 0,05$
$p > \alpha \Rightarrow H_0$ is not rejected
$p < \alpha \Rightarrow H_0$ is rejected $\Rightarrow H_1$ is statistically significant

Note that these are the usual used values, calculated in order to compare the findings with those of existing researches. Nevertheless, the criticism on the "null" or "nil" hypothesis and significance level α were also taken in consideration in this thesis and additional results will also be presented.

3.4 Subjects and setting

Focusing on the questions of this thesis' research, it was important to be able to **compare the difference between traditional instruction with the intervention using OST on the same sample of students**, having the same parameters (conditions, background, culture, i.e., in Austria) and not the comparison with other samples of students of existing findings from other countries. This comparison with existing results and findings was, ergo, in a second plane. Since the tests and tutorial were administered in the summer semester 2019, students had already had traditional mechanics' lessons from their teachers in the first semester of school year (winter 2018). Therefore, the pre-test gave a good measure of students' status (in N3 conceptual understanding and their epistemology) with traditional instruction. Hence, we could say that the "pre-test" was in fact a ***post-test*** *for the* ***traditional instruction***.

The students took the Force-Test and MPEX as pre-tests on the last week of April 2019 and the N3 OST instructions occurred about two weeks later in the middle of May. Around four weeks after that, in June, the Force-Test and MPEX were administered as post-tests. Between pre- and post-test there was a time elapse of approximately six weeks, which is safe considering the advised minimum of five weeks (see chapter 3.2). All teachers guaranteed that they would not give any mechanics' classes during this period of time. This means that, the N3 OST intervention was the only instruction about force or Newton's laws between the pre- and post-tests. Therefore, the post-test gave a good measure of students' status (in N3 conceptual understanding and their epistemology) with the OST intervention. Hence, we could say that (comparing to the above sentence) the "post-test" was in fact a ***post-test*** *for the* ***N3 OST instruction***.

When deciding whether to administer the post-test right after the N3 OST intervention or in the end of the school year, various arguments for both came up (beside the advised 5-week interval between pre- and post-tests):

- On the one hand, any number of things could occur in between the intervention and the post-test to increase student understanding, that are not related to the intervention itself (like teachers "secretely" giving instruction about forces, which would increase the gains as a whole). However, in this case, it seems reasonable to think that the questions on the Force-Test, which are not related to N3, would also have learning gains. And if these happenings are isolated, like for instance a student watching a video about forces (maybe the tutorial aroused some extra curiosity or interest on that student), then they are "normal" happenings and its data should be considered. The influence of these isolated cases in the whole data should be small, considering a larger sample of students.

- On the other hand, if learning gains exist immediately after the intervention, then one can think that they are not as relevant as if they are long-lasting. Furthermore, since most classes had finished their mechanics' instruction in over a month before the pre-test, in order to make a "fair" *traditional instruction* $\times$ *OST* comparison (trying not to put traditional instruction in disadvantage), it seemed reasonable to wait at least four weeks after the N3 OST instruction occurred to administer the post-test.

In order to have a wide sample of students from different courses and schools, i.e., a generalized and heterogeneous representation of Austrian high school students, the data obtained for the analysis of this thesis' research were collected in three different schools, nine different classes. Since not only the general high school (AHS − *Allgemein bildende höhere Schule*, including *Gymnasium* and *Realgymnasium*) but also the vocational school (BHS − *Berufsbildende höhere Schule*) accounts for a great percentage of youth's education in Austria [56], both school types were tested (half of the students from AHS and half from BHS). All together, there were seven different course focuses (language, informatics, music, arts, sports, agriculture and agricultural machinery). This also enabled a good balance between gender (in some classes there were mainly young women, but in others mainly young men).

The decision as to which classes (in sense of school year) to administer the tests and tutorial was made based on the official high schools' curricula (*Lehrplan*), provided by the Austrian Federal Ministry of Education, Science and Research (bmbwf) [57]. Both the *Lehrplan* for AHS [58] and the *Lehrplan* for BHS [59] prescribe the study of Newton's laws in the first **physics' year in** high school (which is not necessarily the first high school year; depending on school and course types, it can be either on the first, second or third high school year). Therefore, with the variety of courses in the (nine) classes, the age range was quite large, as seen below.

# of classes	age	high school year	school year	% of students
3	$14 \sim 15$	1^{st}	9^{th}	$\sim 37,5$
3	$15 \sim 16$	2^{nd}	10^{th}	$\sim 33,3$
2	$16 \sim 17$	3^{rd}	11^{th}	$\sim 16,7$
1	$17 \sim 19$	1^{st}	12^{th}	$\sim 12,5$

The total number of students, which took part in either the pre-test, the tutorial lesson, or the post-test (or in two of these) was 240. The Force-Test results, though, were only analyzed from the **181 students, who participated in all three phases of the process**. For the MPEX, seven students had to be disconsidered, since they left one whole side of the test blank (front or back, either on the pre- or on the post-test). The results considering these "blanks" had only a very small difference to the results disconsidering them. Anyhow, for this thesis' analisys all seven were removed from the data, leaving 174 analyzed students for the MPEX.

The decision to test this amount of students, instead of using power analysis (with similar effect sizes from the ones of existing researches' results) to determine the minimum sample size necessary, was made based not only on the wish to have this wide and generalized representation of students − of different course focuses and school type, but also:

- to have different teachers giving mechanics' traditional instruction (obtaining a broader example of the "normal lessons" in schools - avoiding the argument, in case of learning gains, that OST could be better than "the one" teacher who had taught the classes);
- to have less interference due to possible influences on individual students (as mentioned before);
- to have a certain safety regarding absences (to be sure there would be a good sample even removing the students who missed one of the three phases); and
- a larger sample expectedly improves statistical results, i.e., the research's reliability is fore sure greater.

The contact to teachers and school directors was made almost two months before the pre-tests (beginning of March). First, the teachers were asked if they had first-year physics classes, interest, and available lessons to participate in the research. Then, the directors were asked for permission to carry out the research in their schools. They were guarantied anonymity, i.e., that neither the schools and teachers' names, nor the students' names would appear in the thesis).

In all three phases, which took place in schools, everything worked out without any problems. For the tests, most classes had the "desk divider-shield" which, together with two instructors in class, made it really difficult for students to "cheat". Furthermore, explicit explanation about not being graded (that their teachers would not have any access to their tests and this would not at all affect, in any way, their official grades in school), besides informing that their names would be removed from the tests, might have given students a stronger feeling of safety, increasing the reliability of their answers on the tests. An interesting fact was that, expressing the value of being part of a research, that is trying to improve physics lessons in schools − and this was, amazingly, emphasized a couple of times by the teachers − might have made students not only feel more special, I think, but also helped them understand the importance of their honesty on the tests. Of course, there were always a few students who didn't seem to have much patience, specially with the MPEX (it was clear as they turned the page and saw nearly twenty questions still left, they would groan something like *"Aaarrhh"*).

In three classes there were too many students (over thirty) to make the N3 OST tutorial successful, so the teacher was, not only supportive, but kind enough to split the class in two and use a double-hour lesson, where he would follow on with his teaching plan with half of the class in one classroom, while I administered the tutorial to the other half in the physics lab. Then, at half-time, students would switch classrooms and we would do the same with the other half. Also, in one of the post-tests' class, many students were absent, so the test was administered (again) a couple of days later for those who had missed class. In total there were 9 pre-testing sessions, 12 N3 OST instructions, and 10 post-testing sessions, hence, a sum of 31 lessons in schools.

For each one of the $\sim$ 240 pre-tests a code was made up and written on the first page of the test. Then, after the OST instruction − and later the post-tests − each student's name was searched on the classes' pile, and the same code would be written on both the worksheets and on the post-tests. In some few cases there were students who wrote their first names in one and their nicknames, or last name, on the other (though not only would it be clear which student it was from exclusion, but usually their handwriting made it possible to identify). Subsequently, all the results were manually inserted in an excel table and marked whether they had or not participated on the OST lesson. In case the student had not, or in case the student had only one of the tests, this student would be removed from the data.

The processes of “labeling” students’ (nearly 1.200) tests and worksheets and of manually inserting each one of the (nearly 25.000) answers in the computer, was very time consuming. If it had to be done again, I would think on a different way to do so. For the labeling, in case the calculation of each students’ individual g-factor or c (see 3.3.1) is not important (i.e., not calculating $\bar{g}$ or $\bar{c}$ but only g_m), then students would not have to write their names at all, and a simple presence list from the teachers would make it simple to eliminate the data from students who were absent in one or more of the three occasions. For the tests’ results, I believe that the MPEX could be scored electronically, but with the split task Force-Test, electronic scoring would probably be difficult, since each student give two answers for each one of the questions (see 3.2.3).

The next step was to meticulously figure out the correct equations in Excel to correctly count every possibility of the split task (both correct, scientist correct and intuition wrong, intuition correct and scientist wrong, both wrong). Afterwards, the analysis was done with the percentage of students who answered each of the four possibilities, their graphs, the calculation of normalized gain and change (which equations also had to be adapted for excel) and the p-value. For the MPEX, all the results were inserted into an available template [60], which made all the analysis. The more difficult part of the MPEX results was separating all the clusters and all the students favorable and unfavorable answers to calculate their p-values (see chapter 3.3.2).

The whole process from “creating” the Force-Test, printing and folding tests and tutorials, finding schools and teachers willing to “give up” their lessons, organizing dates, administering the tests and N3 OST, sorting and labeling all the material, inserting the data into the computer, to analyzing the results was like a roller coaster with incredibly amazing moments to really difficult ones; and it gives me great satisfaction to have done it all, when I see the results (see next chapter 4.1).

4 Analysis

The aim of this master's thesis is to make a critical analysis of the conceptual and epistemological effects of the N3 OST in Austrian high school students (compared to traditional instruction), in order to answer the research question. The analyses of students' pre- and post-test scores will be presented in the Results chapter (see 4.1) and these results will be discussed in the Discussion chapter (see 4.2).

The analyses of all the collected data from both the Force-Test and the MPEX (template [60]) were done in Excel.

Two tables with students score on the Force-Test for "all seven N3 questions" (items 1, 3, 4, 6, 7, 9 and 10 of the Force-Test), "FCI N3 questions" (items 3, 4, 6 and 9 of the Force-Test), "FCI N3 collision question" (item 6 of the Force-Test) and "Non N3 questions" (items 2, 5 and 8 of the Force-Test) are in the Attachments (see Figure 37 and Figure 38). The tables containing students' score to each individual question (for both Force-Test and MPEX) will not be included, seeing that their size and amount of values would be unsuitable here.

4.1 Results

4.1.1 Force-Test

The results are given for four different groupings of the Force-Test questions:

- all seven N3 questions (items 1, 3, 4, 6, 7, 9 and 10);
- **four FCI N3 questions** (items 3, 4, 6 and 9);
- FCI N3 collision question (item 6); and
- Non-N3 FCI questions – control questions (items 2, 5 and 8).

The split task has four distinct possibilities:

- **"Right reconciled"** means the correct answer was given for both the scientist's and the intuition's answer (not split).

- **"Right scientist"** means the correct answer was given as the scientist's answer but a different answer was given as the intuitive one (split).
- **"Right intuition"** means the correct answer was given as the intuitive answer but a different answer was given as the scientist's one (split).
- **"Wrong"** means both scientist's and intuitive answers were wrong.

Provided that the splits which occur when students get the scientist's answer correct show a lack of reconciliation between the student's intuitions and what they successfully learned, these splits are most important. Therefore, two graphs for each of the four different grouping of questions will be presented.

- In *black and white* bar graphs both "Right reconciled" and "Right scientist" will be presented as "Right", and both "Right intuition" and "Wrong" will be presented as "Wrong".
  The calculations of normalized gain of the arithmetic means (g_m), the average normalized gain ($\bar{g}$), the average normalized change ($\bar{c}$), and the p-values, are based on these results of "Right" and "Wrong". Here, the p_0 refers to the usual p-value calculated with a null hypothesis and significance level $\alpha_0 = 0,05$. The p_1 refers to the p-value calculated with a "nil" hypothesis and significance level $\alpha_1 = 0,01$. The "nil" hypothesis presumes that each student who did not achieve the maximum possible score on the pre-test will improve his or her score with one point more, i.e., the "nil" hypothesis has a *hypothesized mean difference* of the fraction of students who can improve. In the group "all N3 questions", 7 students (out of 181) achieved total score on the pre-test. Thus, the mean difference presumed on the "nil" hypothesis was $1 - \frac{7}{181} \approx 0,96$. For the group "FCI N3 questions", it was $1 - \frac{10}{181} \approx 0,94$ (10 students out of 181 achieved total score on the pre-test). (For further explanations on g-factor and p-value, see chapter 3.3; for the p-values' calculations presented here, see Figure 39 in the Attachements, and for further explanation on the split task see chapter 3.2.3.)
- In *colored* bar graphs, the split task will be presented with all four distinct possibilities. The calculation of the split percentage is the percentage of "Right scientist" answers in relation to (divided by) the percentage of "Right" answers.

The comparisons are always between the "pre-tests" (**traditional instruction post-test**) and "post-tests" (**N3 OST post-test**).

All seven N3 questions

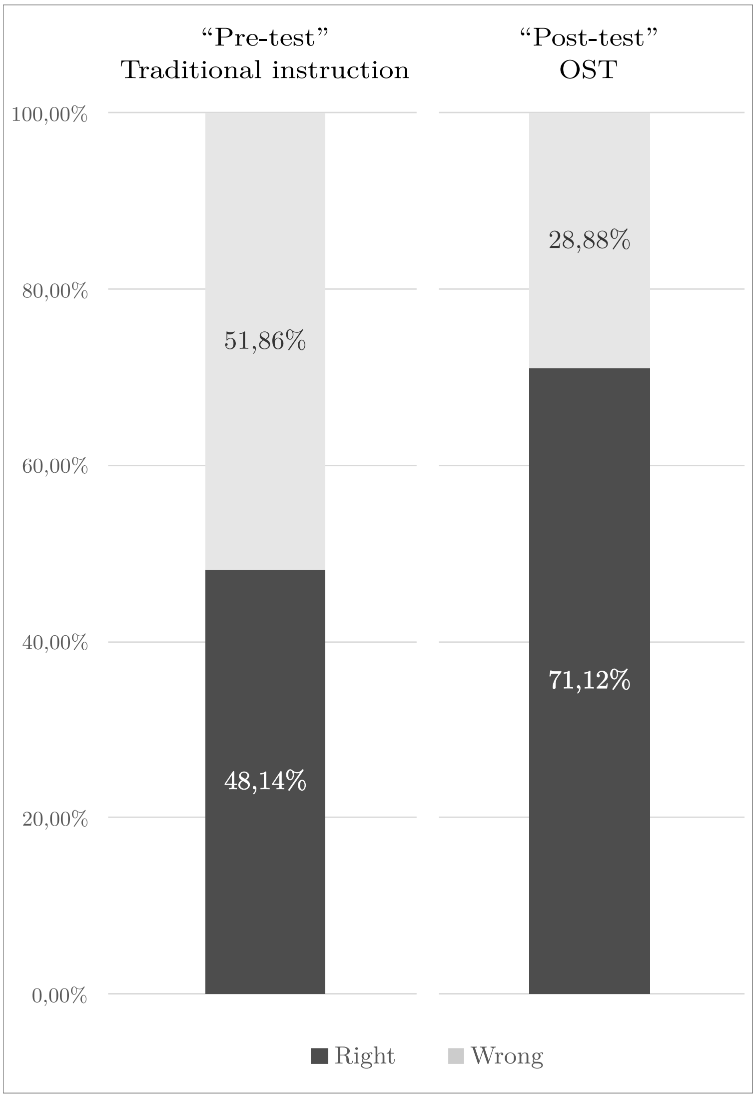


Figure 22: Graph - all seven N3 questions - Right and Wrong percentage graph

• The amount of right answers given after **traditional instruction** was **48,14%** and after the **N3 OST** was **71,12%**.
• The difference between the post-test mean and the pre-test mean was $+1,61$.
• The g-factors and p-values were:

$$g_m = 0,4429 \approx 0,44 \qquad \bar{g} = 0,4329 \approx 0,43 \qquad \bar{c} = 0,4369 \approx 0,44$$

$$p_0 = 6,33 \cdot 10^{-25} \ll \alpha_0 \qquad p_1 = 2,56 \cdot 10^{-6} \ll \alpha_1$$

All seven N3 questions

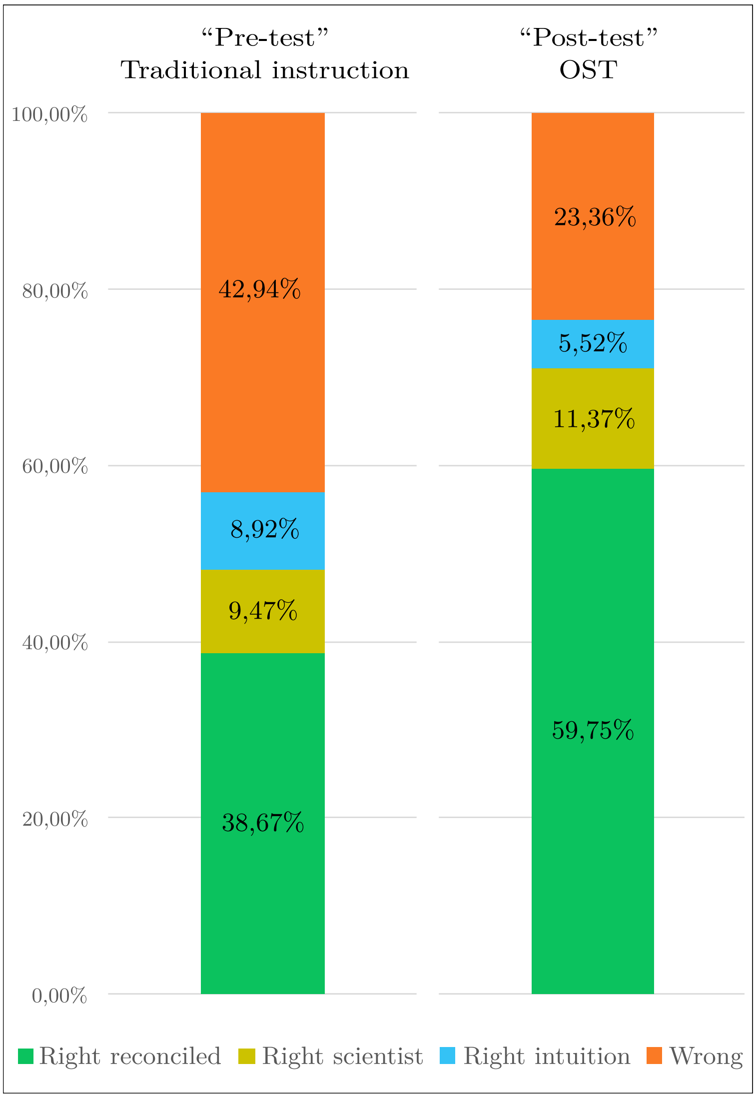


Figure 23: Graph - all seven N3 questions - Split task percentage graph

- The **split percentage** in the right answers was:

$\frac{9{,}47}{9{,}47+38{,}67} = \frac{9{,}47}{48{,}14} = 19,67\%$ after **traditional instruction**

$\frac{11{,}37}{11{,}37+59{,}75} = \frac{11{,}37}{71{,}12} = 15,99\%$ after the **N3 OST intervention**

Four FCI N3 questions

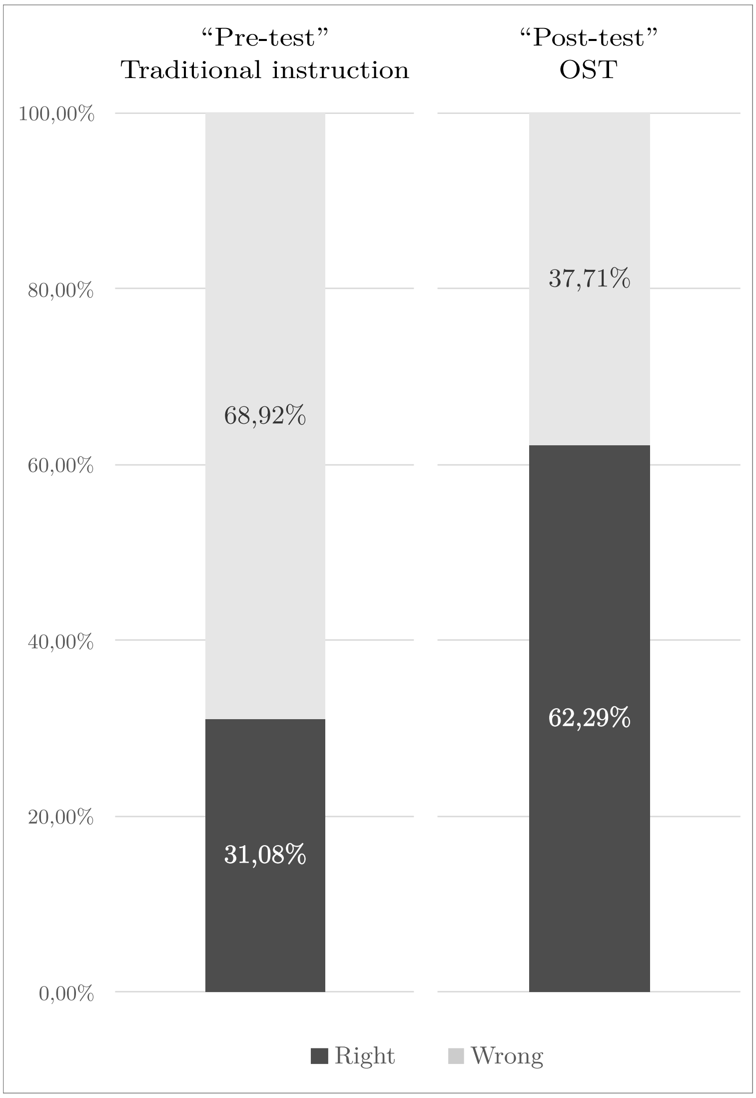


Figure 24: Graph - four FCI N3 questions - Right and Wrong percentage graph

• The amount of right answers given after **traditional instruction** was **31,08%** and after the **N3 OST** was **62,29%**.
• The difference between the post-test mean and the pre-test mean was $+1,25$.
• The g-factors and p-values were:

$$g_m = 0,4529 \approx 0,45 \qquad \bar{g} = 0,4254 \approx 0,43 \qquad \bar{c} = 0,4487 \approx 0,45$$

$$p_0 = 6,71 \cdot 10^{-25} \ll \alpha_0 \qquad p_1 = 0,0033 < \alpha_1$$

Four FCI N3 questions

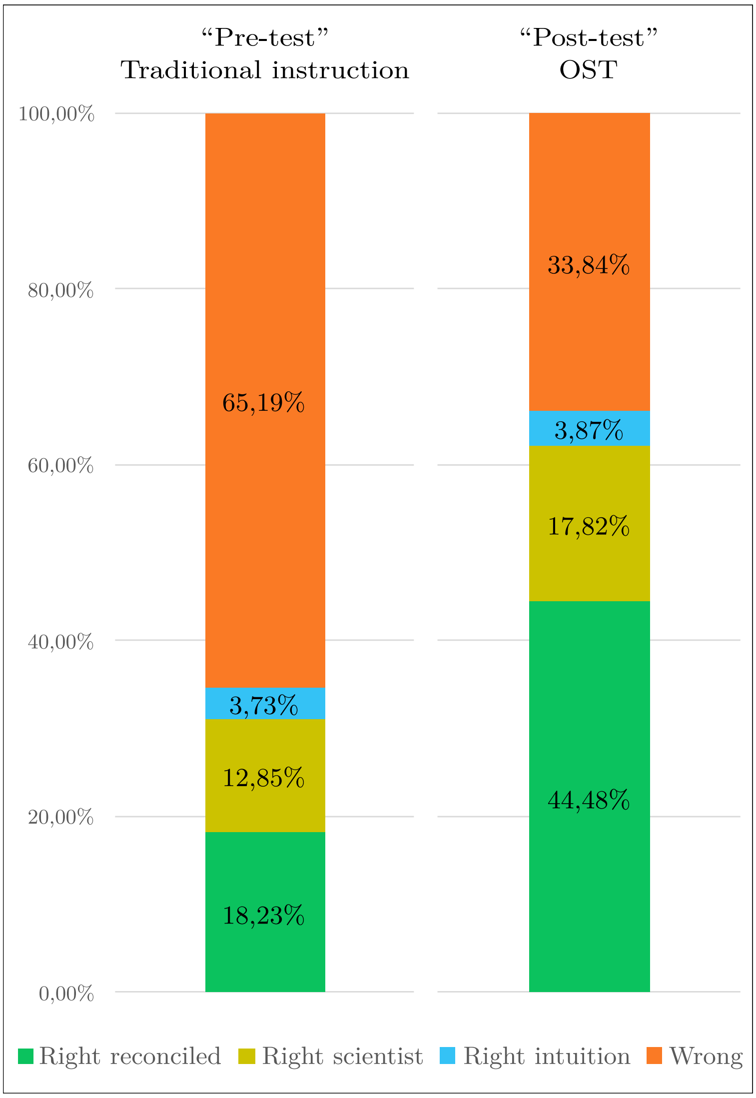


Figure 25: Graph - four FCI N3 questions - Split task percentage graph

- The **split percentage** in the right answers was:

$\frac{12{,}85}{12{,}85+18{,}23} = \frac{12{,}85}{31{,}08} = 41,34\%$ after **traditional instruction**

$\frac{17{,}82}{17{,}82+44{,}48} = \frac{17{,}82}{62{,}29} = 28,61\%$ after the **N3 OST intervention**

FCI N3 collision question

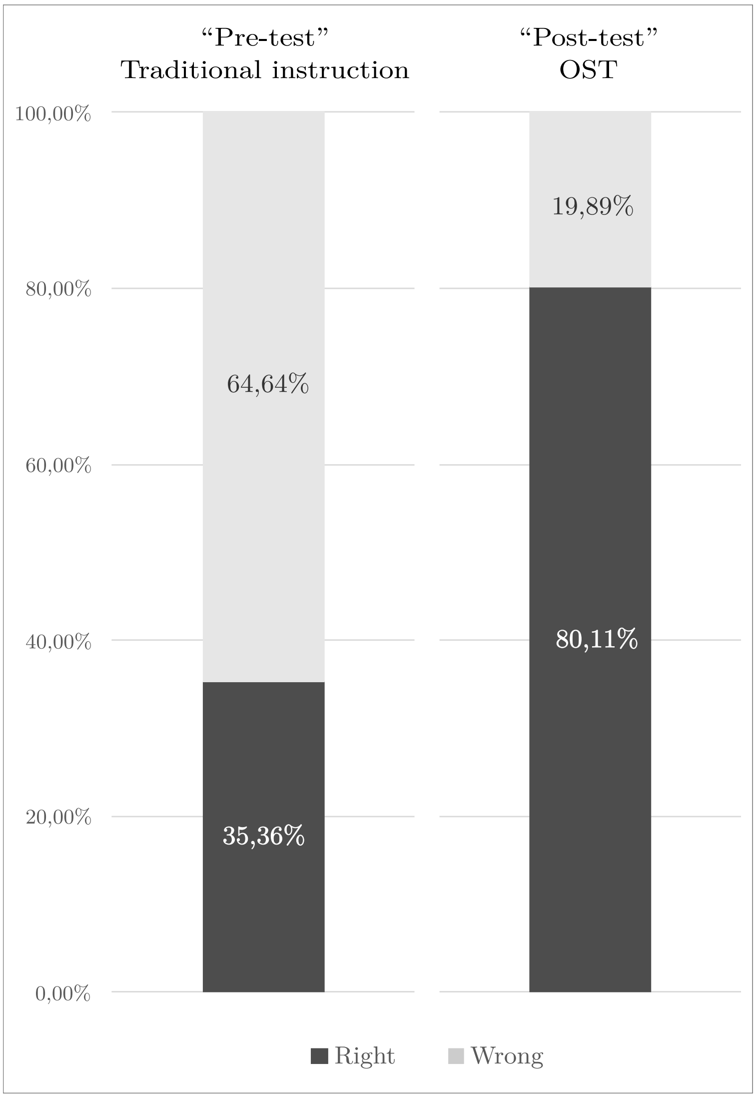


Figure 26: Graph - FCI N3 collision question - Right and Wrong percentage graph

• The amount of right answers given after **traditional instruction** was **35,36%** and after the **N3 OST** was **80,11%**.
• The difference between the post-test mean and the pre-test mean was $+0,45$.
• The g-factor and p-value were:

$$g_m = 0,6923 \approx 0,69 \qquad p_0 = 1,47 \cdot 10^{-19} \ll \alpha_0$$

($\bar{g}$ and $\bar{c}$ cannot be calculated for one single question - scores 0% or 100% - see 3.3.1.)

FCI N3 collision question

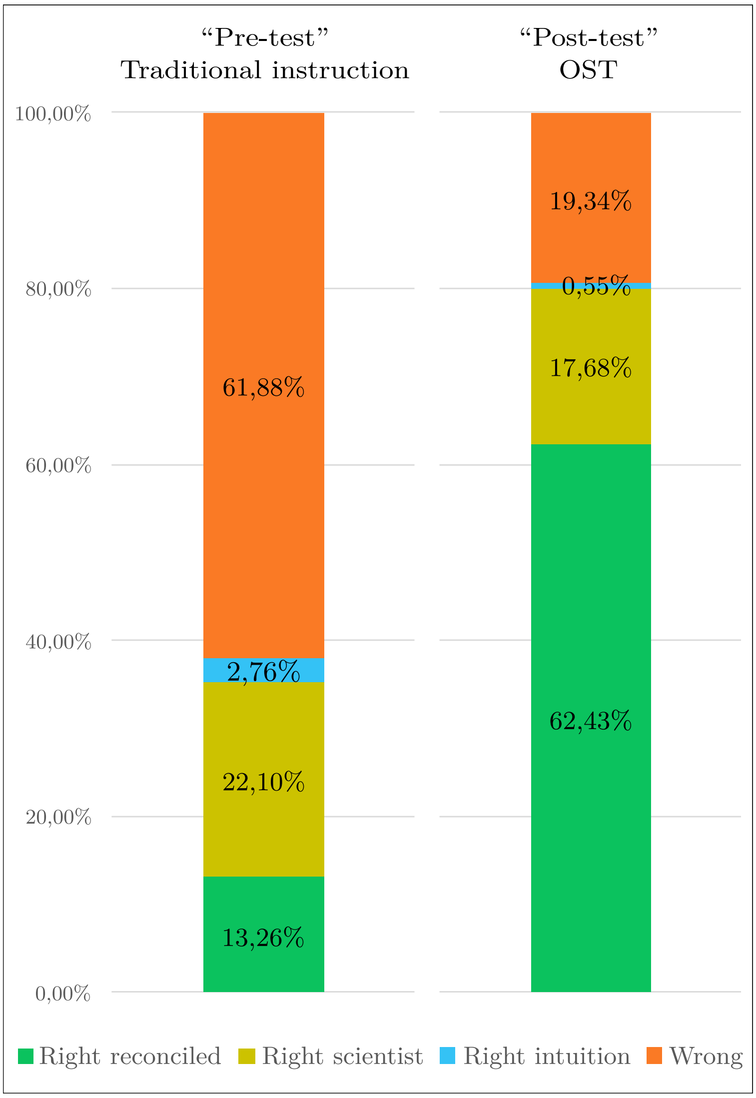


Figure 27: Graph - FCI N3 collision question - Split task percentage graph

- The **split percentage** in the right answers was:

$\frac{22,10}{22,10+13,26} = \frac{22,10}{35,36} = 62,50\%$ after **traditional instruction**

$\frac{17,68}{17,68+62,43} = \frac{17,68}{80,11} = 22,07\%$ after the **N3 OST intervention**

Non N3 questions

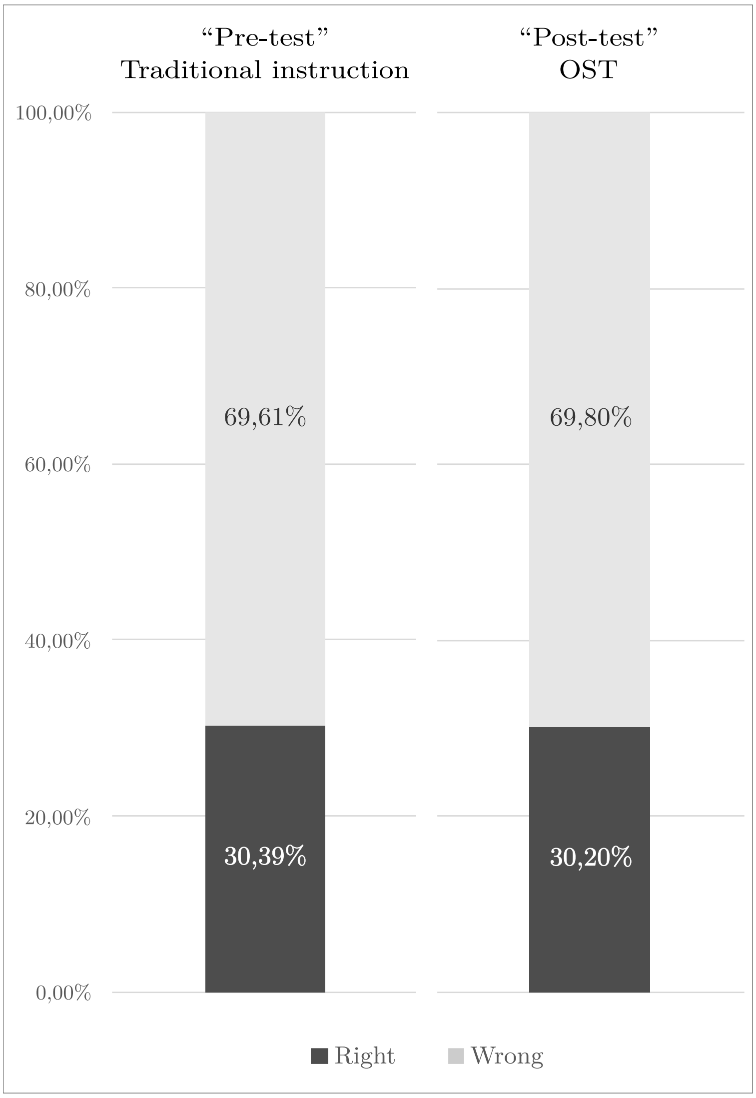


Figure 28: Graph - non N3 FCI questions - Right and Wrong percentage graph

• The amount of right answers given after **traditional instruction** was **30,39%** and after the **N3 OST** was **30,20%**.
• The difference between the post-test mean and the pre-test mean was $-0,006$.
• The g-factors and p-value were:

$$g_m = -0,0026 \approx 0 \qquad \bar{g} = -0,1004 \approx -0,10 \qquad \bar{c} = -0,1108 \approx -0,11$$

$$p_0 = 0,93 > \alpha_0$$

Non N3 questions

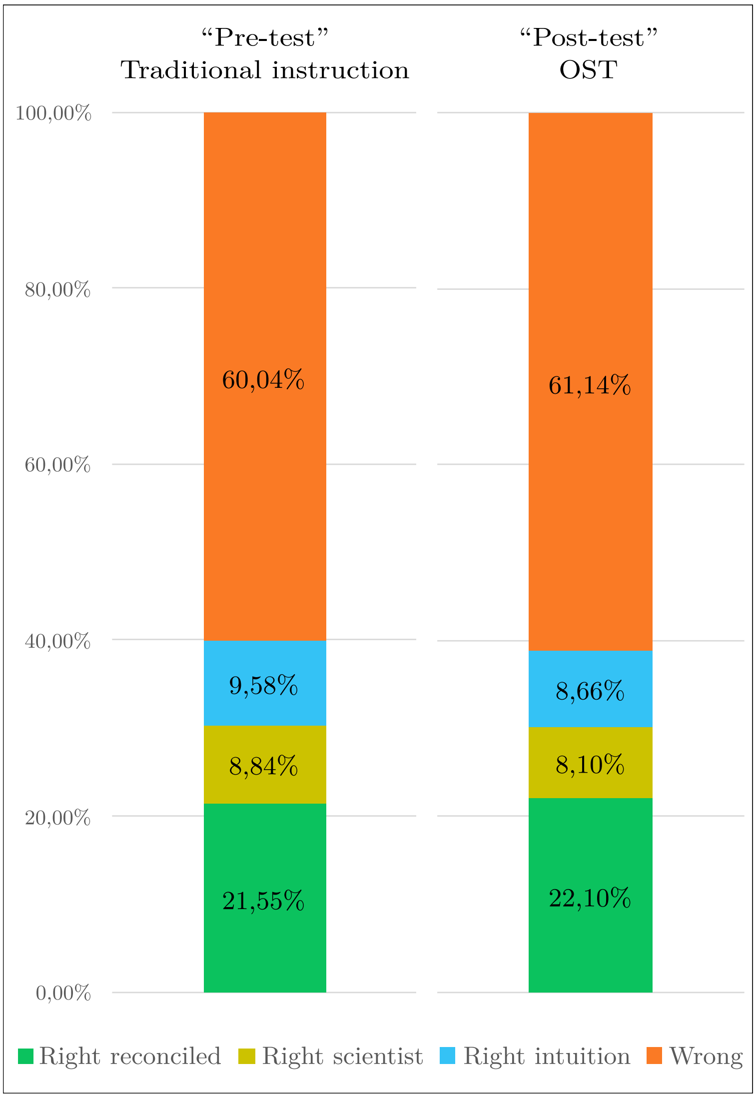


Figure 29: Graph - non N3 FCI questions - Split task percentage graph

- The **split percentage** in the right answers was:

$\frac{8{,}84}{8{,}84+21{,}55} = \frac{8{,}84}{30{,}39} = 29,09\%$ after **traditional instruction**

$\frac{8{,}10}{8{,}10+22{,}10} = \frac{8{,}10}{30{,}20} = 26,82\%$ after the **N3 OST intervention**

4.1.2 MPEX

The MPEX results are given separately for the favorable answers and unfavorable answers. The percentage of both will be presented for each cluster with their respective *p*-value (in percent with a significance level of 5%). All the *p*-values' calculations are in the Attachments (see 40 and 41). The comparisons are always between the "pre-tests" (**traditional instruction post-test**) and "post-tests" (**N3 OST post-test**).

Answers in %	Favorable		
	Pre-test	Post-test	p-value
Overall	37,44	34,18	**0,03**
Independence	34,58	30,94	**4,07**
Coherence	38,16	31,95	**0,27**
Concepts	31,61	31,26	85,81
Reality	47,13	45,98	59,28
Math	33,45	29,77	6,81
Effort	34,02	28,05	**0,40**

Answers in %	Unfavorable		
	Pre-test	Post-test	p-value
Overall	45,62	45,01	50,54
Independence	49,04	48,08	62,11
Coherence	43,45	45,98	18,94
Concepts	48,62	46,09	21,24
Reality	37,93	34,34	9,87
Math	47,36	49,43	34,15
Effort	50,92	54,02	12,24

Figure 30: Table - MPEX results for favorable and unfavorable views.The values in **bold** indicate the statistically significant results ($p < 5\%$)

The number of students (absolute and percentage) who did the same, worse or better on both Overall favorable and Overall unfavorable views are presented as follow.
For favorable:
same = **same** amount of favorable answers on both post- and pre-test;
worse = **less** favorable answers on the post-test than on the pre-test;
better = **more** favorable answers on the post-test than on the pre-test.
For unfavorable:
same = **same** amount of unfavorable answers on both post- and pre-test;
worse = **more** unfavorable answers on the post-test than on the pre-test;
better = **less** unfavorable answers on the post-test than on the pre-test.

Favorable		
same	18	10,34%
worse	98	56,32%
better	58	33,33%

Unfavorable		
same	16	9,20%
worse	69	39,66%
better	89	51,15%

Figure 31: Table - number of students who did same, worse or better on the MPEX

On the following graph of the MPEX results (for both favorable and unfavorable views), the percentage range considering the extreme values is marked for pre- and post-tests. Note that for the unfavorable views to have a positive or "good" movement means that the values should go "left".

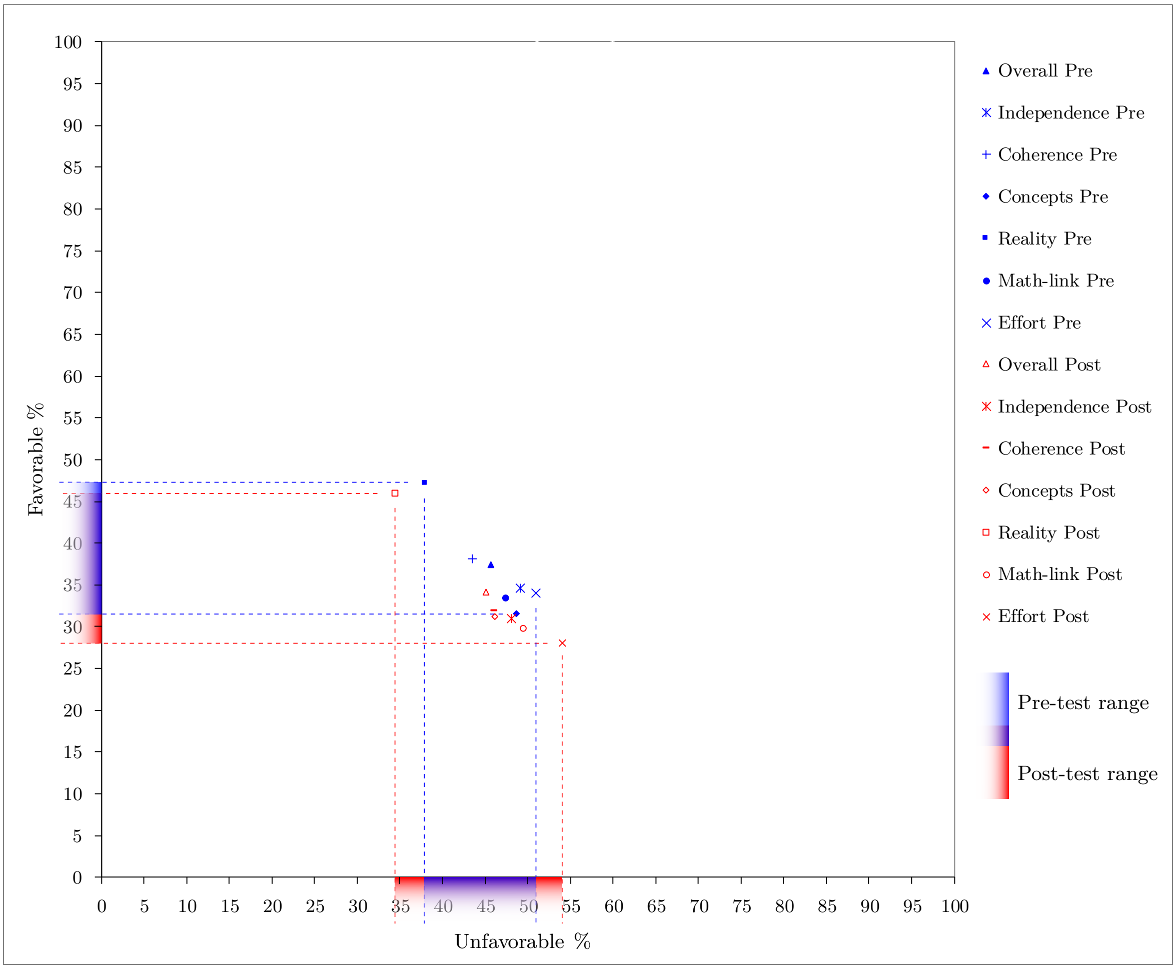


Figure 32: Graph - MPEX results: pre-tests and post-tests in percentage

The percentage range from both pre-tests and post-tests are marked with blue and red, respectively, for both favorable and unfavorable views.

The following graph shows the normalized movement between pre- and post-test for both favorable and unfavorable views. Note that for the unfavorable motion to have a positive or "good" movement means that the values should go "left" (to the negative side of the unfavorable axis).

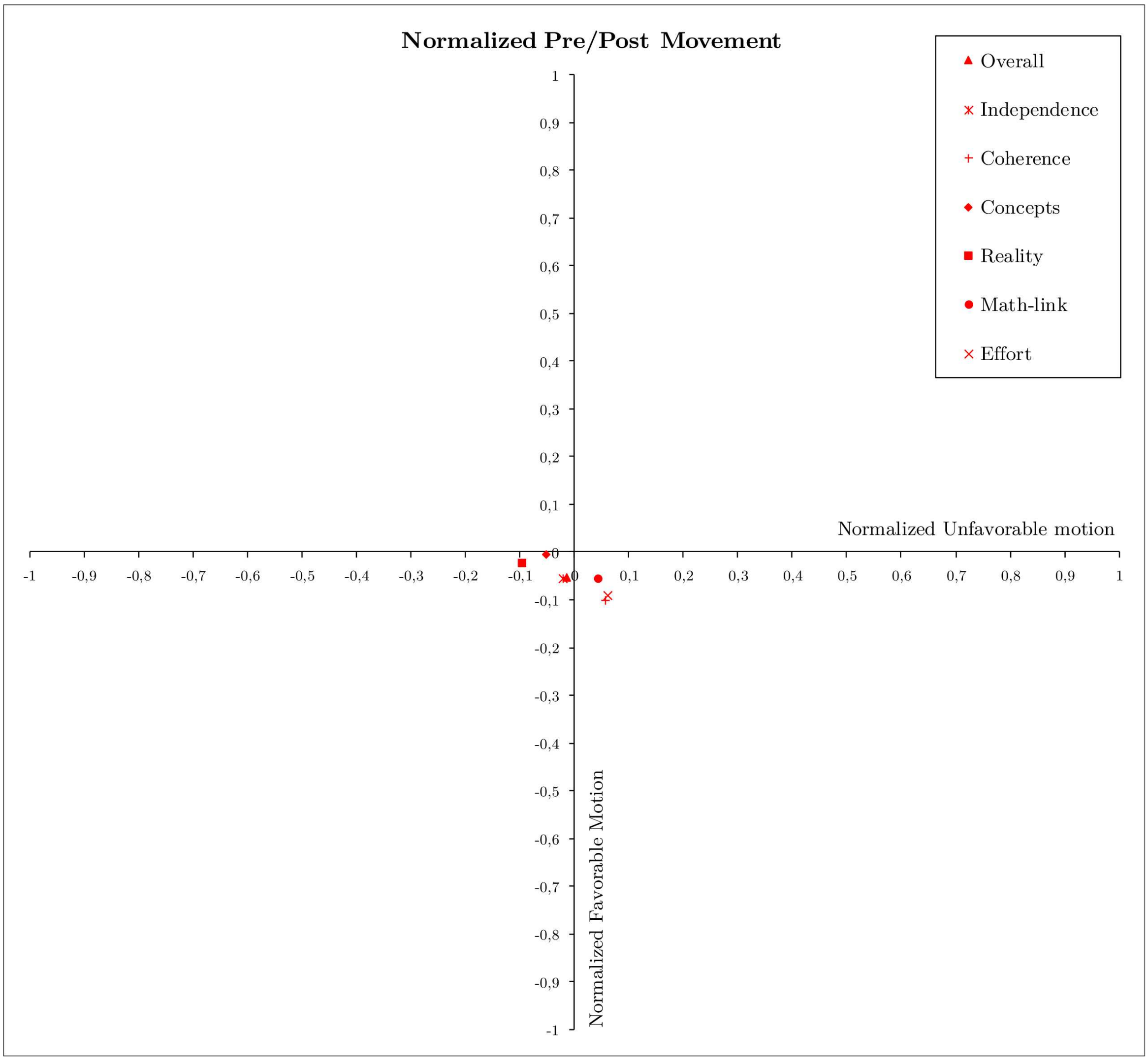


Figure 33: Graph - MPEX results - normalized pre/post movement

On the following graph a zoom from the previous percentage graph was done, marking those differences, which are statistically significant.These were only present in the favorable views, therefore the stronger lines indicate the change in favorable views (vertical component of the whole movement).

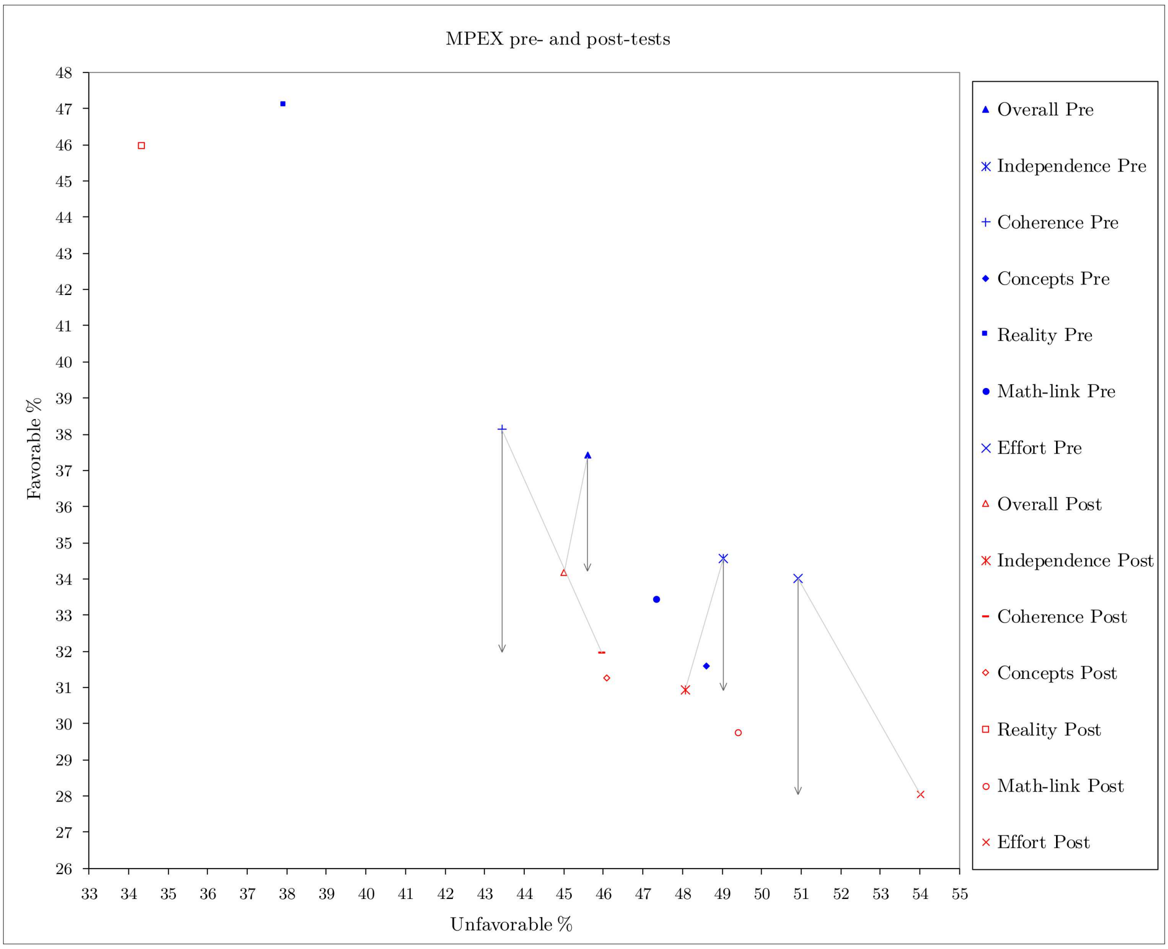


Figure 34: Graph - MPEX results: zoom in of the percentage graph

The following graph is a zoom from the previous normalized movement between pre- and post-test for both favorable and unfavorable views, with the statistically significant values marked. Again, since these were only present in the favorable views, the lines are all vertical, indicating the change in the favorable views.

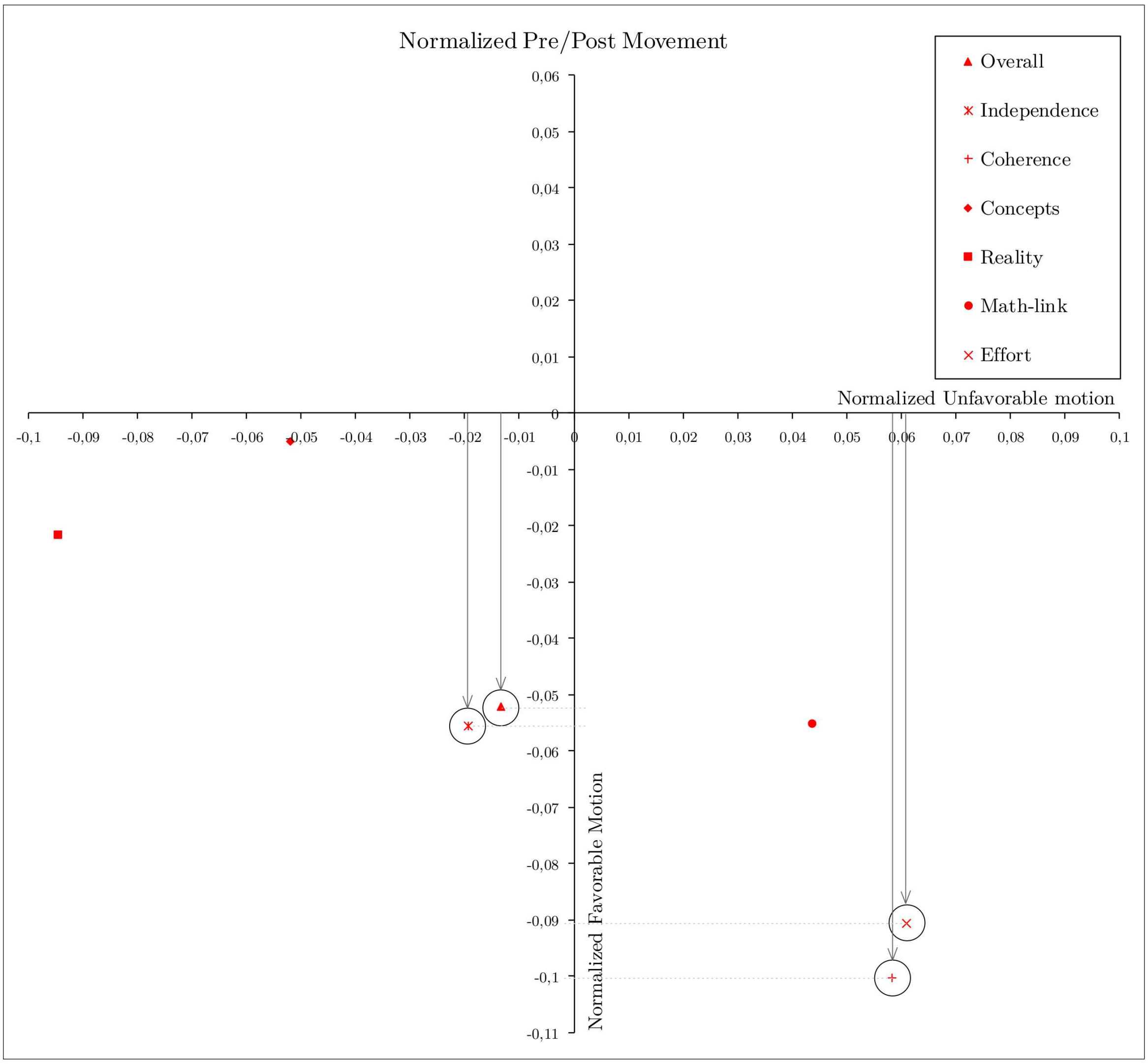


Figure 35: Graph - MPEX results - zoom in of the normalized pre/post movement

4.2 Discussion

When comparing the results presented on the previous chapter, what can be noticed is that the normalized gain of the arithmetic means (g_m), the average normalized gain ($\bar{g}$), and the average normalized change ($\bar{c}$) all have very similar values. This means that, for this sample, the individual analysis of students scores ($\bar{g}$ and $\bar{c}$) would not, in fact, be necessary. Having knowledge of that would have saved a great time in the analysis (labeling, matching tests and tutorials, calculating individually, etc.), as mentioned in the chapter Subjects and setting (see 3.4). Of course, this was not something one could be sure in advance.

The discussion about the conceptual and epistemological effects caused by the N3 OST intervention in three Austrian high schools (9 classes, 240 students) will be done as follow.

4.2.1 Conceptual effects

The results from the group "all N3 questions" (items 1, 3, 4, 6, 7, 9 and 10) from the Force-Test clearly reveal that students had a better understanding of the physics' concepts involving Newton's third law, i.e., it shows a conceptual improvement from what they had learned with traditional instruction to what they then learned with the N3 OST. This visible improvement is also statistically significant (not likely to have occurred randomly). However, since these results involve the three "created" questions (see chapter Force-Test 3.2.4), they are not as important as the ones from the group "FCI N3 questions".

Even though there may have criticism in analyzing the individual dimensions of the FCI, as discussed previously (see chapter 3.2.1), the Third Law dimension was found to be a legitimate cluster (it is not surprising that many researches have explored it [43][51][61][62]). The results from the group "FCI N3 questions" (items 3, 4, 6 and 9 which correspond to items 15, 16, 4, and 28 of the FCI, respectively) also give a clear statement that students had a conceptual improvement. This gain is also sustained statistically, with a great certainty that the collected data do not represent a random result, but rather a true difference in the effectiveness of the tutorial in comparison to traditional instruction.

Existing research establishes that for typical traditional instructions in high school and college one could expect a normalized gain of the arithmetic means on the order of $g_m \sim 0,2$, for modest reforms on the order of $g_m \sim 0,35$ and for more extensive reforms on the order of $g_m \sim 0,6$ [61]. Even though extensive research show that "active learning" methods and interactive engagement are much more effective than traditional instruction (see chapter 2.1.1), the fact that **the result achieved in Austrian high schools for one single 50-minute intervention was** $g_m \sim 0,45$, is above what one would expect. Although, it must be stated that this comparison is not very precise. These values of g-factors were calculated with scores from the whole FCI and not only the Third Law dimension, like the one calculated here. Furthermore, these values of g_m take in consideration the pre-test done **before any instruction**. In the case of this thesis, by the time of the pre-test students had already had mechanics class, so the gain g_m is between what **they had already learned** with traditional instruction to what they then learned with the N3 OST. Therefore, if the pre-test had been given before traditional instruction, the obtained gain would probably have been larger.

The decision to make the analysis for the one FCI N3 collision question (item 6, which corresponds to item 4 of the FCI), even though testing one single question can be problematic, came from the wish to see how these results compare to the ones from the whole FCI N3 cluster. Since N3 OST focuses only in a collision situation, it was a wish to see if the conclusions from Smith and Wittmann [29] made sense. They analyzed students' reasoning separated into collision and pushing situations (see chapter 2.2.1) and one of their suggestions was that curriculum developers should focus on pushing situations to help students learn N3 more effectively, indicating that this is the area with most room for improvement. In fact, the results do show that the number of students who answered this question correctly after the N3 OST intervention was dramatically more than after traditional instruction. On the "pre-test" 64 out of 181 students answered this question correctly but on the "post-test" this number escalated to 145 (2,27 times more). Then, by comparing the normalized gain of the arithmetic means (g_m) from the group "four FCI N3 question" (0,45), which include three pushing items and only one collision item, with the g_m from the "one FCI N3 collision question" (0,69), it is clear that there is a remarkable difference.

The analysis' result of the group "non N3 FCI" (items 2, 5 and 8 which correspond to items 25, 30 and 12 of the FCI, respectively) is one that gave me great pleasure to see. Very important conclusions are implied with these results. First, it **states the real effectiveness of the tutorial**, since it has absolutely no focus on these other force topics; and second, it proves – what teachers had already given their word for – that they did not keep teaching *force* subjects (confidentially) in the period between both tests. The evidence is amazing: with a difference in means so small that can be disregarded (0,006), a g-factor of zero, and a p-value showing that there is no evidence for the hypothesis that the scores from both tests differ from each other.

After the analysis of the Force-Test results, it is possible to answer part of the research question – *What are the **conceptual effects** of using the Newton's Third Law Open Source Tutorial in Austrian high schools?* – with certainty. It is safe to say that **there was an evident conceptual gain with the use of Newton's third law Open Source Tutorial in Austrian high schools**. Students worked well during the tutorial and developed a better understanding of the physics' concept, achieving better results on the test. Therefore, it is advisable that teachers consider the incorporation of Open Source Tutorials in their lessons' plans.

4.2.2 Epistemological effects

The results from the MPEX survey show, as expected and anticipated by previous research, that physics instruction which does not take epistemology in consideration worsens students' beliefs about physics knowledge and knowing (see chapter Epistemology 2.1.2). Students continued to have traditional instruction between the pre-test and the post-test. Therefore, the expectation that a single 50-minute intervention, with epistemological focus, would improve students epistemology was beyond optimistic. As previously mentioned, modest "reforms" or interventions have little – or no – influence in the development of students epistemology. [21] The changes (in all clusters) were not very large, showing very small normalized movements (all of them in the range of $-0,1$ to $+0,1$). Still, the favorable views deteriorated in all clusters, and the three clusters – as well as Overall – which had little improvement in the unfavorable views are not statistically significant. Unfortunately, this is also true for the Reality link, which had the most positive motion on the unfavorable views. Since most results of the MPEX are likely to have occured randomly (not statistically significant), the criticism of administering its German version in Austria (for example its lack of validation, or confusion with the Likert-scale) should be considered (see chapter 3.2.2).

Since the results are not very revealing, their juxtaposition with others of existing researches could bring some light into the picture. The comparison to other studies is, like before, imprecise. However, neglecting the differences between the two studies (for example what the "pre-test" meant), Wilhelm's [22] results (from high schools in Bavaria), is probably, until now, the best option of comparison. On the table that follows (Figure 36) the results for three students' populations are presented.

Answers in %	1) Austrian 9 high school classes, N=174				2) Bavarian 17 high school classes, N=336				3)Bavarian 3 high school classes, N=66			
	pre-test		post-test		pre-test		post-test		pre-test		post-test	
	fav.	unfav.	fav.	unfav.	fav.	unfav.	fav.	unfav.	fav.	unfav.	fav.	unfav.
Overall	**37**	46	**34**	45	**37**	**38**	**35**	**41**	39	**37**	41	**32**
Independence	**35**	49	**31**	48	33	46	33	44	37	37	40	32
Coherence	**38**	43	**32**	46	43	31	42	33	45	29	48	25
Concepts	32	49	31	46	29	**45**	28	**48**	33	**45**	31	**33**
Reality	47	38	46	34	42	33	42	34	**46**	**30**	**56**	**19**
Math	33	47	30	49	38	**38**	35	**42**	39	32	43	31
Effort	**34**	51	**28**	54	**39**	**37**	**32**	**42**	32	45	29	46

Figure 36: Table - results from both Austrian and Bavarian high schools. The values in bold indicate the statistically significant results ($p < 5\%$). Source: own survey and [22].

The first population (1) is represented by the Austrian students from this research, who had traditional instruction throughout the whole period between pre- and post-tests and one class period with the N3 OST tutorial. The second population (2) is composed of students from Bavaria, who had only traditional instruction. The third population (3) is also composed by Bavarian students, who had traditional instruction but, additionally, they had 5 to 6 class periods with a differentiated teaching method from a modeling approach project. In some of the modeling research epistemic aspects are implicit and in others they are explicit, but they are crucial for productive and meaningful engagement in the practice. Also *"engaging students in scientific modeling* *can help them develop their epistemologies by allowing them to attend to the roles of mechanism and empirical evidence when constructing and revising models."* [63]

The results from populations one (1) and two (2) do not differ very much from each other and both show that students' epistemology was worse on the post-test than on the pre-test. However, population three (3) had their favorable views increased and the unfavorable views decreased in almost all clusters, from the pre-test to the post-test. This suggests that, while not so modest interventions can already show some improvement in students' epistemology, one single intervention is not enough.

A very important result, which should be given special attention – even though its epistemological scope is narrower than the MPEX's scope and it only shows reconciliation on a small subset of FCI questions – is the split task from the Force-Test. It clearly indicates an evident epistemological difference between students' right scientists answers after traditional instruction and students' right scientists answers after the N3 OST intervention. **Not only a bigger percentage of the students knew the correct scientific answer** to the FCI N3 questions, but also **a higher percentage of those students indicated that the correct answers made intuitive sense**. From the students who answered the right answers on the FCI N3 after traditional instruction 41,34% had splits, and after the N3 OST instruction this number was reduced to 28,61%. Really impressive is the dramatic difference that can be seen in the collision question, where these numbers are 62,60% versus 22,07%. Lower split rates indicate that students have reconciled Newton's third law with their intuitive ideas. In this sense, after traditional instruction, students who "learned" Newton's third law did not learn it as "deeply" as after the N3 OST intervention. This result has also been previously observed: *With this new task, comparing traditional to epistemologically focused, reform-oriented instruction, we found that the reform students not only learn the physics' conceptions better in the sense of learning the correct answer, but also reconcile them better with their intuitive ideas.* [43]

Besides the positive results regarding reconciliation (and conceptual improvement), it is also important to state the impression I had when giving the tutorial to all these students. Even though most of them are not used to working in groups and discussing ideas with each other (as previously declared by the teachers, and observed during earlier practical trainings), the tutorial had great acceptance: they participated actively without holding back on their intuitions and reasoning, and, most importantly, they really did seem to enjoy it. Since they did not know that they would have a "post-test" (only that we would have three dates together), many of them asked excited, as they saw me on the last day, if we were going to do another tutorial. As I answered that we wouldn't, they very often expressed disappointment. In the end, after collecting the post-tests, I would ask students to close their eyes and give either a "thumbs-up", "thumbs-down" or a "thumbs-in-the-middle" according to how much they felt that doing the tutorial was helpful. The absolute majority of students showed that the tutorial had helped them do the post-test. One of the students reported that the tutorial did not make the test *"easier"* but it made him *"change his mind"* (indication of reconciliation).

After the analysis of both the MPEX survey and the Force-Test results, it is possible to answer the other part of the research question − *What are the **epistemological effects** of using the Newton's Third Law Open Source Tutorial in Austrian high schools?*. When focusing in one aspect of epistemology, we can say that students had a **first step improving epistemology**, since the split task results show that many of them have reconciled their intuitions with the concepts of Newton's third law, i.e., they indicated that these concepts made intuitive sense. However, when looking at all aspects of epistemology the MPEX results show that there was a **small epistemological deterioration**. This result, though, cannot be attributed to the use of Newton's third law Open Source Tutorial, since the main instruction students had in the mean time was traditional. Therefore, also from the epistemological point of view, it could be advised for teachers to include the tutorials in their lessons' plans, in the attempt to provide students with (at least) the possibility of epistemological development. In any case, in order to see the real epistemological effects of Open Source Tutorials in Austrian high school students and fully answer this question, **further research using OSTs for a longer period is necessary**.

5 Conclusion and further work

In this thesis it was important to find out if the Open Source Tutorials can help students in Austrian high schools achieve a better conceptual understanding as well as a better view about the nature of knowledge and knowing (i.e., epistemology development), compared to their usual lesson methods. Furthermore, it was important to see if the method carried out by the tutorials would have a good acceptance here, since it differs from what students are used to in traditional instruction. The tutorial on Newton's third law was introduced and "tested" in three different schools − 9 classes, 240 students; and through pre-post testing (with a "Force-Test" containing the *Force Concept Inventory*'s Third Law dimension and with the *Maryland Physics Expectations Survey*), students where assessed on both conceptual understanding and epistemology. The pre-tests results presented a good picture of what students had learned, in this topic, with traditional instruction, and the post-test of what they then learned with the tutorial.

The results' analysis shows an **evident gain on conceptual understanding** of Newton's third law. More than that, and most importantly, it shows that students not only achieved a better understanding of the physics' concepts, but **these concepts actually made sense for them**. Many of them have reconciled their intuitive ideas with the correct scientific concepts. This shows that the tutorial's "epistemological plan" to help students understand both science and learning as *"a refinement of everyday thinking"* really does have a positive effect on students.

Students showed excitement and enjoyment during the tutorial classes. They worked actively, had very productive discussions and even expressed disappointment when they realized that it was only this one lesson. Furthermore, most of them gave very positive feedbacks, stating that the tutorial had helped them on the post-test.

The biggest challenge for a teacher is, I believe, to listen to students, respond to their difficulties and find ways to help them use the productive cognitive resources they possess. The Open Source Tutorial help teachers do that very well. Being able to "actively listen" requires practice, but it comes with a great satisfaction of seeing students work and succeed.

Still, there is much work to be done.

Since the results of the MPEX did not reveal an improvement of students' epistemology, I suggest that further work be done in Austrian high schools with the use of Open Source Tutorials for a longer period, for example, throughout the school year and with many physics' topics. This will give a much better perspective of students' epistemological development.

Further work is also urgent in the validation of the German version of the MPEX. To have a validated test instrument for students' epistemology could be a great benefit for the physics education research community.

An even further research could be studying the effectiveness of Open Source Tutorials considering the different student population (schools or classes with focus in math and science, in sports, in languages, etc.) and seeing which has larger gains.

And last, a focus on the development of Newton's third law tutorials that deal with pushing and pulling situations would be of great value helping students understand this topic even better.

References

[1] Albert Einstein, *Physics and Reality.* J. Franklin Inst. 221, 1936.

[2] OECD (2016), "Students' attitudes towards science and expectations of science–related careers", in *PISA 2015 Results (Volume I): Excellence and Equity in Education*, OECD Publishing, Paris,
https://doi.org/10.1787/9789264266490-7-en.

[3] OECD iLibrary, PISA Results, ISSN: 19963777, Online: https://doi.org/10.1787/19963777; accessed 31.05.2020.

[4] Hartmut Wiesner, Horst Schecker and Martin Hopf, *Physikdidaktik kompakt.* Aulis Verlag, 2015 Edition, 2011.

[5] Walter Jung, *Begründung und Zielsetzung.* (In Fachdidaktik Physik, W. Bleichroth, H. Dahncke, W. Jung, W. Kuhn, G. Merzyn and Weltner, K.(Hrsg.)) Aulis Verlag, Köln, 17-63, 1999.

[6] John Amos Comenius, *The great didactic.* (1896) "The great didactic of John Amos Comenius" Translated by M. W. Keatinge, A. and C. Black, London, 1896. Online:https://archive.org/details/cu31924031053709/page/n253/mode/2up; accessed 3/2/2020.

[7] Suzanne M. Bourgoin and Paula K. Byers, *Encyclopedia of world biography.* John Amos Comenius, Volume 4 (Chippendale-Dickinson), 2^{nd} Edition, Gale group, 1998.

[8] Bernd Hackl, *Lernen Wie wir werden, was wir sind.* Julius Klinkhardt Verlag, 2017.

[9] Aristotle, *Nicomachean Ethics.* Translated by W. D. Ross, 1956, Book II, 1, §2, Online: http://classics.mit.edu/Aristotle/nicomachaen.html; accessed 8/2/2020.

[10] David Hestenes, Malcolm Wells and Gregg Swackhamer, *Force Concept Inventory.* The Physics Teacher 30, 141-151, 1992.

[11] Eric Mazur, *Peer Instruction: A User's Manual.* Prentice-Hall, Upper Saddle River, 1997.

[12] Edward F. Redish, *Teaching Physics With the Physics Suite.* 2003.
Online: http://www2.physics.umd.edu/~redish/Book/; accessed 15/2/2020.

[13] David P. Ausubel, *Educational Psychology: A Cognitive View.* Holt, Rinehart and Winston Inc., New York, 1968.

[14] Horst Schecker, Thomas Wilhelm, Martin Hopf, Reinders Duit and Helmut Fischler. *Schülervorstellungen und Physikunterricht.* Springer Berlin Heidelberg, 2018.

[15] Salomon F. Itza-Ortiz, N. Sanjay Rebello, Dean A. Zollman and Manuel Rodriguez-Achach, The vocabulary of introductory physics and its implications for learning physics. The Physics Teacher 41, no.6, 330-336, 2003.

[16] Horst Schecker, Thomas Wilhelm, Martin Hopf and Reinders Duit, *Schülervorstellungen-Forschungsstand, Konsequenzen Desiderata.* GDCP-Tagungsband zur Jahrestagung in Kiel, 201-204, 2019

[17] Michael M. Hull, Beth A. Lindsey, Matthew Archambault, Kathleen Davey and Amy Y. Liu, *Unexpected attitudinal growth in a course combining reformed curricula.* Physical Review Physics Education Research 12, no.1, 2016.

[18] David Hammer, *Student resources for learning introductory physics.* American Journal of Physics 68, no.S1, S52-S59, 2000.

[19] David Hammer and Andrew Elby, *Tapping epistemological resources for learning physics.* The Journal of the Learning Sciences 12, no. 1, 53-90, 2003.

[20] Detlef Urhahne and Martin Hopf, *Epistemologische Überzeugungen in den Naturwissenschaften und ihre Zusammenhänge mit Motivation, Selbstkonzept und Lernstrategien.* Zeitschrift für Didaktik der Naturwissenschaften 10, no.1, 71-87, 2004.

[21] Andrew Elby, *Helping physics students learn how to learn.* American Journal of Physics 69, no.S1, 54-64, 2001.

[22] Thomas Wilhelm, *Ansichten von Elftklässlern über Physik und Lernen von Physik–Ergebnisse beim „Maryland Physics Expectations Survey".* Didaktik der Physik-Kassel, 2006.

[23] Sir Isaac Newton, *Philosophiae Naturalis Principia Mathematica*, "Sir Isaac Newton's Principia." Reprinted for Sir William Thomson LL.D. and Hugh Blackburn M.A. Glasgow: James Maclehose, publisher to the Univeristy, printed by Robert Maclehose, 1871. Online: `https://oll.libertyfund.org/titles/1254`; accessed 15/4/2020.

[24] Sir Isaac Newton, *The Mathematical Principles of Natural Philosophy.* Translated Into English by Andrew Motte, 1729.

[25] Paul G. Hewitt, *Conceptual Physics.* Pearson Education Inc., 12th edition – Global edition, 2015.

[26] Thomas Wilhelm, Verena Tobias, Christine Waltner, Martin Hopf and Hartmut Wiesner. *Zweidimensionaldynamische Mechanik–Ergebnisse einer Studie.* Chemie-und Physikdidaktik für die Lehramtsausbildung, Jahrestagung der GDCP in Potsdam, 438-440, 2010.

[27] Reinders Duit, Walter Jung, Helga Pfundt and Kiel Institut für die Pädagogik der Naturwissenschaften, *Alltagsvorstellungen und naturwissenschaftlicher Unterricht.* Aulis-Verlag Deubner, 1981.

[28] Richard P. Feynman, *The Feynman Lectures on Physics, Volume I: mainly mechanics, radiation, and heat.* Part 1, Chapter 10: Conservation of Momentum, BILINGUA-Ausgabe, R. Oldenbourg Verlag und Addison-Wesley Publishing Company, 1974.

[29] Trevor I. Smith and Michael C. Wittmann, *Comparing three methods for teaching Newton's third law.* Physical Review Special Topics - Physics Education Research 3, no.2, 2007.

[30] Online:`http://umdperg.pbworks.com/w/page/10511238/Tutorials%20from%20the%20UMd%20PERG`; accessed 15/5/2020.

[31] Michael C. Wittmann, Mindi Kvaal Anderson and Trevor I. Smith. *Comparing three methods for teaching newton's second law.* In AIP Conference Proceedings, vol. 1179, no.1, 301-304, American Institute of Physics, 2009.

[32] Online:`http://umdperg.pbworks.com/w/page/10511167/121-122%20Facilitating%20in%20Tutorial`; accessed 15/5/2020.

[33] Online:`http://www2.physics.umd.edu/~elby/CCLI/index.html`; accessed 15/5/2020.

[34] Online:`https://www.physport.org/curricula/MD_OST`; accessed 15/5/2020.

[35] Michael Malvern Hull, *Do students have cultural scripts? Results from the first implementation of open source tutorials in Japan.* Doctoral dissertation, 2013.

[36] Michael M. Hull and Andrew Elby, *A conceptual physics class where students found meaning in calculations.* In AIP Conference Proceedings, vol. 1513, no.1, pp. 190-193, American Institute of Physics, 2013.

[37] Online:https://ufind.univie.ac.at/de/person.html?id=64710&teaching=true; accessed 15/5/2020.

[38] Online:http://www.physics.umd.edu/perg/OSTutorials/04_Newton_Three/Tutorial_04_Newton3.doc; accessed 15/5/2020.

[39] Online:http://www.physics.umd.edu/perg/OSTutorials/04_Newton_Three/Newton_Three_Guide.pdf; accessed 15/5/2020.

[40] Online:http://www.iolab.science/; accessed 15/5/2020.

[41] Jing Han, Lei Bao, Li Chen, Tianfang Cai, Yuan Pi, Shaona Zhou, Yan Tu and Kathleen Koenig, *Dividing the Force Concept Inventory into two equivalent half-length tests.* Physical Review Special Topics-Physics Education Research 11, no.1, 2015.

[42] Edward F. Redish, Jeffery M. Saul and Richard N. Steinberg. *Student expectations in introductory physics.* American Journal of Physics 66, no.3, 212-224, 1998.

[43] Timothy L. McCaskey and Andrew Elby, *Probing students' epistemologies using split tasks.* In AIP Conference Proceedings, vol.790, no.1, 57-60, American Institute of Physics, 2005.

[44] Douglas Huffman and Patricia Heller, *What does the force concept inventory actually measure?* The Physics Teacher 33, no.3, 138-143, 1995.

[45] David Hestenes, Ibrahim Halloun, *Interpreting the Force Concept Inventory: A response to Huffman and Heller.* The Physics Teacher, 33, 502-506, 1995.

[46] Jing Wang and Lei Bao, *Analyzing force concept inventory with item response theory.* American Journal of Physics 78, no.10, 1064-1070, 2010.

[47] Jun-ichiro Yasuda, Naohiro Mae, Michael M. Hull and Masa-aki Taniguchi. *Analyzing false positives of four questions in the Force Concept Inventory.* Physical Review Physics Education Research 14, no.1, 2018.

[48] Christopher J. Omasits and Doris J. Wagner, *Investigating the Validity of the MPEX Survey.* In AIP Conference Proceedings, vol.818, no.1, 145-148, American Institute of Physics, 2006.

[49] U. Wutchana and N. Emarat, *Student effort expectations and their learning in first-year introductory physics: A case study in Thailand.* Physical Review Special Topics-Physics Education Research 7, no.1, 2011.

[50] Adrian Madsen, Sarah B. McKagan and Eleanor C. Sayre. *How physics instruction impacts students' beliefs about learning physics: A meta-analysis of 24 studies.* Physical Review Special Topics-Physics Education Research 11, no.1, 2015.

[51] Timothy Lee McCaskey, *Comparing and contrasting different methods for probing student epistemology and epistemological development in introductory physics.* Doctoral dissertation, 2009.

[52] Jeffrey D. Marx and Karen Cummings, *Normalized change.* American Journal of Physics, 75, 87-91, 2007.

[53] Christian Heumann and Michael Schomaker Shalabh, *Introduction to Statistics and Data Analysis: With Exercises, Solutions and Applications in R.* Springer, Switzerland, 2016.

[54] Bertie Vidgen and Taha Yasseri, *P-Values: Misunderstood and Misused.* Frontiers in Physics, Volume 4, Article 6, 2016.

[55] Larry Wasserman, *All of Statistics: A Concise course in Statistical Inference.* Springer, New York, 2004.

[56] Online:`http://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bildung/schulen/schulbesuch/122126.html`; accessed 22/5/2020.

[57] Austrian Federal Ministry of Education, Science and Research (Bundesministerium Bildung, Wissenschaft und Forschung), Online:`https://www.bmbwf.gv.at/`; accessed 22/5/2020.

[58] Austrian official curriculum for general high schools: BGBl. II Nr. 216 (2018). Verordnung des Bundesministers für Bildung, Wissenschaft und Forschung, mit der die Verordnung über die Lehrpläne der allgemeinbildenden höheren Schulen geändert wird; Änderung der Bekanntmachung der Lehrpläne für den Religionsunterricht sowie Bekanntmachung der Lehrpläne für den Religionsunterricht. Online:`https://www.ris.bka.gv.at/eli/bgbl/II/2018/216/20180828`; accessed 22/5/2020.

[59] Austrian official curricula for vocational high schools: Lehrpläne: Höhere Land- und Forstwirtschaftliche Lehranstalten, Forstfachschule,
Online:`https://www.abc.berufsbildendeschulen.at/downloads/?kategorie=19`; accessed 22/5/2020.

[60] Michael Wittmann, *Excel template for the MPEX survey.* Public release version 2001, With help from Jeff Saul (University of Central Florida), Apriel Hodari (University of Maryland), Edward F. Redish (University of Maryland), and Chris van Breen (University of Maryland), also input provided by Brant Hinrichs (Drury).

[61] Edward F. Redish and David Hammer, *Reinventing college physics for biologists: Explicating an epistemological curriculum.* American Journal of Physics 77, no.7, 629-642, 2009.

[62] Antti Savinainen, Philip Scott and Jouni Viiri. *Using a bridging representation and social interactions to foster conceptual change: Designing and evaluating an instructional sequence for Newton's third law.* Science Education 89, no.2, 175-195, 2005.

[63] Hamin Baek and Christina V. Schwarz. *The influence of curriculum, instruction, technology, and social interactions on two fifth-grade students' epistemologies in modeling throughout a model-based curriculum unit.* Journal of Science Education and Technology 24, no.2-3, 216-233, 2015.

Additional Literature

David Hammer, *Two approaches to learning physics.* The Physics Teacher 27, no.9, 664-670, 1989.

Dominik Ertl, *The nature of science.* Plus Lucis 1, 5-7, 2010.

Edward F. Redish, *Introducing students to the culture of physics: Explicating elements of the hidden curriculum.* In AIP conference proceedings, vol. 1289, no. 1, pp. 49-52, American Institute of Physics, 2010.

Ibrahim Abou Halloun and David Hestenes. *The initial knowledge state of college physics students.* American journal of Physics 53, no.11, 1043-1055, 1985.

Jeffrey Marx and Karen Cummings, *What factors really influence shifts in students' attitudes and expectations in an introductory physics course?* In AIP Conference Proceedings (Vol. 883, No. 1, pp. 101-104), American Institute of Physics, 2007.

Lei Bao, *Theoretical comparisons of average normalized gain calculations.* American Journal of Physics 74, 917-922, 2006.

Richard R. Hake, *Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses.* American journal of Physics 66, no.1, 64-74, 1998.

Thomas Wilhelm and Dieter Heuer, *Ansichten deutscher Elftklässler über Physik und Lernen von Physik-Ergebnisse beim „Maryland Physics Expectations Survey".* Oberländer, A (Hrsg.): Tagungsbeiträge der Frühjahrstagung des Fachverbandes Didaktik der Physik in der Deutschen Physikalischen Gesellschaft. Tagungs-CD Berlin, 2005.

Vincent P. Coletta and Jeffrey A. Phillips, *Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability.* American Journal of Physics 73, no.12, 1172-1182, 2005.

Attachments

Abstract

The aim of this master thesis is to research Newton's third law *Open Source Tutorial* in Austrian high schools in order to see if students achieve a better conceptual understanding of the physics concepts as well as a more developed epistemology, compared to traditional instruction. The research was carried out in nine classes from three different schools (a total of 240 students). Pre-post testing was done with a "Force-Test" containing the *Force Concept Inventory*'s Third Law dimension and with the *Maryland Physics Expectations (MPEX) Survey.* All classes had already had their lessons in mechanics by the time of the pre-test. Therefore, the pre-tests results presented a good picture of what students had learned with traditional instruction. Between pre- and post-tests students had their normal classes with traditional instruction, which did not include mechanics, and only one 50-minute intervention with the Open Source Tutorial on Newton's third law. Subsequently they had the post-test, which showed what they learned with the tutorial. The results' analysis shows an evident gain on conceptual understanding of Newton's third law's concepts (g-factor=0,45). More than that, and most importantly, it shows that students not only achieved a better understanding of the physics' concepts, but that these concepts actually made sense for them. Many of them have *reconciled their intuitive ideas with the correct scientific concepts.* The MPEX results showed that students' beliefs or views about the nature of knowledge and knowing had gotten worse, proving that only one short intervention with focus on students' epistemology is not enough to prevent its typical deterioration caused by traditional instruction.

Zusammenfassung

Ziel dieser Masterarbeit ist es, das dritte Newtonsche Gesetz *Open Source Tutorial* an der Sekundarstufe II in österreichischen Schulen zu untersuchen, um festzustellen, ob die Schülerinnen und Schüler durch das Tutorial ein besseres konzeptionelles Verständnis der Physikkonzepte sowie eine weiter entwickelte epistemologische Überzeugungen erreichen, im Vergleich zum traditionellen Unterricht. Die Forschung wurde in neun Klassen von drei verschiedenen Schulen (insgesamt 240 Schülerinnen und Schüler) durchgeführt. Vor- und Nach-Tests wurden mit einem *"Kräftetest"* durchgeführt, welcher die Dimension des dritten Gesetzes des *Force Concept Inventory* enthielt sowie die *Maryland Physics Expectations (MPEX) Survey*. Zum Zeitpunkt des Vor-Tests hatten alle Klassen bereits Unterricht in Mechanik. Daher zeigten die Ergebnisse der Vor-Tests ein gutes Bild davon, was die Schülerinnen und Schüler mit traditionellem Unterricht gelernt hatten. Zwischen den Vor- und Nach-Tests fand bis auf eine 50-minütige Intervention mit dem *Open Source Tutorial* zu Newtons drittem Gesetz kein Mechanik Unterricht statt. Anschließend hatten sie den Post-Test, welcher zeigte, was sie durch das Tutorial gelernt hatten. Die Analyse der Ergebnisse zeigt einen offensichtlichen Gewinn für das konzeptionelle Verständnis der Konzepte des dritten Newtonschen Gesetzes (g-Faktor = 0,45). Darüber hinaus und vor allem zeigt es, dass die Schülerinnen und Schüler nicht nur die Konzepte der Physik besser verstanden haben, sondern, dass diese Konzepte für sie tatsächlich Sinn machten. Viele von ihnen haben ihre intuitiven Ideen mit den richtigen wissenschaftlichen Konzepten in Einklang gebracht. Die MPEX Ergebnisse zeigten, dass sich die Überzeugungen oder Ansichten der Schülerinnen und Schüler über die Natur von Wissen und Wissenserwerb verschlechtert hatten, was beweist, dass nur eine kurze Intervention mit Schwerpunkt auf epistemologische Überzeugungen nicht ausreicht, um ihre typische Verschlechterung durch den traditionellen Unterricht zu verhindern.

Figure 37: Table - results of the Force-Test by group of questions (1)

All N3 questions (total = 7)									FCI N3 questions (total = 4)								
Student	Pre	Post	Student	Pre	Post	Student	Pre	Post	Student	Pre	Post	Student	Pre	Post	Student	Pre	Post
1A002	4	4	2A074	4	5	2D146	0	5	1A002	2	2	2A074	1	2	2D146	0	3
1A003	4	7	2B077	4	4	2D148	2	5	1A003	2	4	2B077	1	2	2D148	0	2
1A005	7	7	2B078	5	6	2D149	4	4	1A005	4	4	2B078	2	3	2D149	2	2
1A006	3	4	2B080	4	5	2D151	6	7	1A006	0	2	2B080	1	2	2D151	3	4
1A007	5	7	2B081	4	6	2D152	4	6	1A007	2	4	2B081	1	3	2D152	2	3
1A008	3	5	2B082	4	5	2D153	3	7	1A008	1	2	2B082	1	2	2D153	0	4
1A009	3	2	2B083	2	3	2D156	5	6	1A009	1	1	2B083	0	2	2D156	3	4
1A010	4	7	2B085	6	7	2D158	4	6	1A010	2	4	2B085	3	4	2D158	2	3
1A011	1	3	2B087	2	7	2D159	3	7	1A011	1	2	2B087	1	4	2D159	2	4
1A012	4	7	2B088	2	3	2D160	4	5	1A012	2	4	2B088	0	3	2D160	1	4
1A013	2	4	2B089	3	4	2D162	4	6	1A013	1	1	2B089	1	3	2D162	1	3
1A014	2	7	2B090	3	3	3A168	5	7	1A014	1	4	2B090	0	2	3A168	2	4
1A015	3	5	2B091	1	3	3A169	4	7	1A015	1	2	2B091	0	1	3A169	1	4
1B017	0	6	2B092	2	4	3A170	3	7	1B017	0	3	2B092	0	1	3A170	1	4
1B018	4	3	2B093	3	5	3A171	5	5	1B018	1	1	2B093	1	3	3A171	2	2
1B019	6	7	2B094	3	3	3A173	6	6	1B019	3	4	2B094	0	0	3A173	3	3
1B021	2	7	2B095	4	4	3A174	7	6	1B021	1	4	2B095	1	2	3A174	4	3
1B023	3	5	2B096	3	3	3A176	7	7	1B023	0	2	2B096	0	1	3A176	4	4
1B024	4	7	2B097	2	5	3A177	1	3	1B024	1	4	2B097	1	2	3A177	0	3
1B025	4	5	2B100	3	0	3A178	3	5	1B025	1	3	2B100	2	0	3A178	1	2
1B026	2	6	2C102	3	7	3A180	2	3	1B026	0	3	2C102	0	4	3A180	0	1
1B027	3	4	2C103	5	6	3A181	7	6	1B027	1	2	2C103	2	4	3A181	4	3
1B029	3	7	2C104	4	3	3A183	3	5	1B029	1	4	2C104	3	2	3A183	2	2
1B031	6	7	2C106	4	2	3A185	2	7	1B031	4	4	2C106	2	0	3A185	0	4
1B032	6	3	2C107	3	6	3A186	7	7	1B032	3	2	2C107	0	3	3A186	4	4
1B033	3	5	2C108	3	4	3A187	3	6	1B033	2	3	2C108	1	1	3A187	0	3
1B035	3	4	2C109	3	7	3B192	3	3	1B035	2	2	2C109	1	4	3B192	0	1
1B036	1	5	2C110	3	3	3B193	3	3	1B036	1	2	2C110	0	2	3B193	2	1
1B037	5	7	2C111	4	5	3B194	1	7	1B037	3	4	2C111	1	2	3B194	1	4
1B20	4	4	2C112	3	4	3B196	4	5	1B20	2	2	2C112	2	2	3B196	1	2
2A041	2	4	2C113	6	7	3B197	1	1	2A041	0	2	2C113	4	4	3B197	0	0
2A042	3	4	2C114	3	3	3B199	1	6	2A042	1	1	2C114	2	1	3B199	1	4
2A043	3	5	2C115	4	5	3B200	5	5	2A043	1	2	2C115	1	2	3B200	2	2
2A044	3	2	2C116	3	5	3B201	5	5	2A044	1	1	2C116	1	2	3B201	3	2
2A045	4	7	2C117	2	4	3B202	5	4	2A045	1	4	2C117	0	1	3B202	2	2
2A046	2	3	2C118	1	6	3B203	2	4	2A046	1	1	2C118	1	3	3B203	1	1
2A047	3	3	2C119	4	4	3B204	5	6	2A047	0	0	2C119	2	1	3B204	2	3
2A048	3	5	2C120	2	0	3B205	4	5	2A048	1	3	2C120	0	0	3B205	2	3
2A049	2	5	2C121	1	1	3B206	2	3	2A049	0	2	2C121	0	0	3B206	1	0
2A051	2	4	2C122	2	7	3B207	3	3	2A051	1	3	2C122	0	4	3B207	1	1
2A052	3	1	2C123	5	4	3B208	2	6	2A052	1	1	2C123	3	2	3B208	0	3
2A053	2	7	2C124	6	7	3B209	2	5	2A053	0	4	2C124	3	4	3B209	1	3
2A055	5	6	2C125	2	3	3B210	2	3	2A055	2	3	2C125	0	1	3B210	0	0
2A056	4	4	2C126	5	7	3B211	2	5	2A056	1	1	2C126	2	4	3B211	0	2
2A057	1	4	2C127	5	5	3B212	3	6	2A057	0	1	2C127	2	2	3B212	1	3
2A058	3	4	2C128	3	3	3B213	4	4	2A058	1	1	2C128	0	0	3B213	2	3
2A059	3	7	2C129	1	4	3B214	2	1	2A059	0	4	2C129	1	2	3B214	0	1
2A060	3	4	2C130	6	7	3C217	2	6	2A060	1	3	2C130	4	4	3C217	1	3
2A061	4	6	2C131	3	7	3C218	3	5	2A061	1	3	2C131	1	4	3C218	1	2
2A062	4	4	2C132	4	6	3C219	7	7	2A062	1	1	2C132	1	4	3C219	4	4
2A063	3	2	2C133	3	7	3C223	4	7	2A063	2	1	2C133	0	4	3C223	1	4
2A064	4	6	2C134	4	5	3C224	3	3	2A064	1	3	2C134	1	3	3C224	0	0
2A065	3	3	2C135	4	5	3C225	1	7	2A065	1	2	2C135	2	2	3C225	0	4
2A066	2	4	2D139	3	4	3C226	3	6	2A066	2	2	2D139	1	2	3C226	0	3
2A067	3	3	2D140	3	6	3C230	3	7	2A067	1	0	2D140	1	3	3C230	0	4
2A068	1	3	2D141	3	6	3C231	7	7	2A068	1	1	2D141	0	3	3C231	4	4
2A069	3	6	2D142	4	4	3C233	5	6	2A069	0	3	2D142	2	2	3C233	2	3
2A071	3	7	2D143	3	6	3C234	3	7	2A071	0	4	2D143	0	4	3C234	2	4
2A072	3	4	2D144	6	7	3C235	3	3	2A072	1	1	2D144	3	4	3C235	0	0
2A073	5	7	2D145	4	4	3C237	3	7	2A073	3	4	2D145	2	1	3C237	1	4
						3C238	1	6							3C238	0	3

Figure 38: Table - results of the Force-Test by group of questions (2)

FCI N3 collision question (total = 1)									Non N3 questions (total = 3)								
Student	Pre	Post	Student	Pre	Post	Student	Pre	Post	Student	Pre	Post	Student	Pre	Post	Student	Pre	Post
1A002	1	1	2A074	0	1	2D146	0	1	1A002	1	1	2A074	1	1	2D146	1	0
1A003	0	1	2B077	0	1	2D148	0	1	1A003	1	3	2B077	1	0	2D148	1	0
1A005	1	1	2B078	1	1	2D149	0	0	1A005	2	2	2B078	2	2	2D149	0	2
1A006	0	1	2B080	0	0	2D151	0	1	1A006	1	1	2B080	1	1	2D151	0	1
1A007	0	1	2B081	1	1	2D152	0	1	1A007	0	0	2B081	0	1	2D152	0	0
1A008	0	1	2B082	0	1	2D153	0	1	1A008	1	1	2B082	1	1	2D153	1	1
1A009	1	0	2B083	0	1	2D156	1	1	1A009	0	2	2B083	0	1	2D156	2	2
1A010	1	1	2B085	1	1	2D158	0	1	1A010	1	0	2B085	2	2	2D158	1	1
1A011	0	1	2B087	1	1	2D159	1	1	1A011	2	1	2B087	1	2	2D159	0	0
1A012	0	1	2B088	0	1	2D160	0	1	1A012	2	2	2B088	0	1	2D160	0	0
1A013	0	1	2B089	0	1	2D162	1	1	1A013	1	0	2B089	1	0	2D162	2	1
1A014	0	1	2B090	0	1	3A168	1	1	1A014	1	1	2B090	2	0	3A168	0	1
1A015	0	1	2B091	0	1	3A169	0	1	1A015	2	0	2B091	1	1	3A169	2	2
1B017	0	1	2B092	0	1	3A170	1	1	1B017	1	1	2B092	1	0	3A170	1	0
1B018	0	1	2B093	0	1	3A171	0	1	1B018	1	1	2B093	1	1	3A171	1	2
1B019	0	1	2B094	0	0	3A173	1	1	1B019	2	1	2B094	0	0	3A173	0	0
1B021	1	1	2B095	0	1	3A174	1	1	1B021	2	1	2B095	1	1	3A174	2	2
1B023	0	1	2B096	0	1	3A176	1	1	1B023	1	0	2B096	1	1	3A176	2	2
1B024	1	1	2B097	1	1	3A177	0	1	1B024	0	0	2B097	0	0	3A177	1	0
1B025	0	1	2B100	0	0	3A178	0	0	1B025	1	0	2B100	1	0	3A178	1	1
1B026	0	1	2C102	0	1	3A180	0	0	1B026	1	0	2C102	2	1	3A180	1	1
1B027	0	1	2C103	0	1	3A181	1	1	1B027	0	1	2C103	2	3	3A181	0	2
1B029	1	1	2C104	1	0	3A183	1	1	1B029	0	1	2C104	2	1	3A183	1	1
1B031	1	1	2C106	1	0	3A185	0	1	1B031	0	1	2C106	0	1	3A185	0	1
1B032	1	0	2C107	0	1	3A186	1	1	1B032	2	0	2C107	1	1	3A186	1	1
1B033	1	1	2C108	0	0	3A187	0	0	1B033	0	1	2C108	1	0	3A187	1	1
1B035	1	1	2C109	0	1	3B192	0	1	1B035	0	0	2C109	2	1	3B192	0	0
1B036	1	1	2C110	0	1	3B193	0	0	1B036	0	2	2C110	1	0	3B193	1	2
1B037	1	1	2C111	0	1	3B194	1	1	1B037	0	0	2C111	0	1	3B194	1	1
1B20	0	1	2C112	1	1	3B196	0	1	1B20	0	0	2C112	2	2	3B196	1	1
2A041	0	1	2C113	1	1	3B197	0	0	2A041	1	1	2C113	2	3	3B197	1	1
2A042	0	0	2C114	1	1	3B199	1	1	2A042	1	1	2C114	0	0	3B199	0	0
2A043	0	1	2C115	0	0	3B200	1	1	2A043	1	1	2C115	0	0	3B200	2	0
2A044	0	0	2C116	0	1	3B201	1	1	2A044	1	1	2C116	1	2	3B201	0	0
2A045	0	1	2C117	0	0	3B202	1	0	2A045	0	0	2C117	0	1	3B202	1	0
2A046	0	0	2C118	0	1	3B203	1	1	2A046	1	1	2C118	0	1	3B203	1	1
2A047	0	0	2C119	1	1	3B204	1	1	2A047	1	2	2C119	1	1	3B204	0	1
2A048	0	1	2C120	0	0	3B205	0	1	2A048	1	3	2C120	1	2	3B205	1	1
2A049	0	1	2C121	0	0	3B206	1	0	2A049	0	1	2C121	1	0	3B206	1	1
2A051	0	1	2C122	0	1	3B207	1	1	2A051	1	2	2C122	2	1	3B207	1	2
2A052	0	0	2C123	1	1	3B208	0	1	2A052	0	0	2C123	2	1	3B208	0	1
2A053	0	1	2C124	1	1	3B209	1	0	2A053	0	1	2C124	1	1	3B209	1	1
2A055	0	1	2C125	0	0	3B210	0	0	2A055	1	0	2C125	1	0	3B210	1	1
2A056	0	1	2C126	0	1	3B211	0	1	2A056	2	1	2C126	1	0	3B211	0	0
2A057	0	0	2C127	1	1	3B212	1	1	2A057	1	3	2C127	1	1	3B212	0	1
2A058	0	1	2C128	0	0	3B213	1	1	2A058	0	0	2C128	0	1	3B213	2	1
2A059	0	1	2C129	0	1	3B214	0	1	2A059	2	2	2C129	0	1	3B214	0	0
2A060	0	1	2C130	1	1	3C217	1	1	2A060	0	0	2C130	1	2	3C217	1	1
2A061	0	1	2C131	0	1	3C218	1	1	2A061	2	0	2C131	0	1	3C218	0	0
2A062	0	1	2C132	1	1	3C219	1	1	2A062	2	1	2C132	1	0	3C219	0	0
2A063	1	0	2C133	0	1	3C223	0	1	2A063	2	1	2C133	1	0	3C223	1	0
2A064	0	1	2C134	0	1	3C224	0	0	2A064	2	1	2C134	1	1	3C224	2	1
2A065	0	1	2C135	1	1	3C225	0	1	2A065	1	1	2C135	2	2	3C225	1	1
2A066	0	1	2D139	0	1	3C226	0	1	2A066	1	1	2D139	0	1	3C226	0	0
2A067	1	0	2D140	0	1	3C230	0	1	2A067	1	2	2D140	2	1	3C230	0	0
2A068	0	0	2D141	0	1	3C231	1	1	2A068	2	2	2D141	1	2	3C231	1	1
2A069	0	1	2D142	1	1	3C233	1	1	2A069	1	1	2D142	1	1	3C233	1	0
2A071	0	1	2D143	0	1	3C234	1	1	2A071	1	1	2D143	1	1	3C234	1	2
2A072	0	1	2D144	0	1	3C235	0	0	2A072	1	0	2D144	2	1	3C235	1	1
2A073	1	1	2D145	0	0	3C237	1	1	2A073	1	1	2D145	1	1	3C237	1	1
						3C238	0	1							3C238	2	2

Figure 39: Table - T-test for the Force-Test

All N3 questions - Null H.	*Post-test*	*Pre-test*
Mean	4,977900552	3,370165746
Variance	2,832842234	2,156660528
Observations	181	181
Pearson Correlation	0,358454006	
Hypothesized Mean Difference	0	
df	180	
t Stat	12,05854017	
P(T<=t) one-tail	3,16681E-25	
t Critical one-tail	2,347242646	
P(T<=t) two-tail	**6,33E-25**	
t Critical two-tail	2,603418229	

All N3 questions - Nill H.	*Post-test*	*Pre-test*
Mean	4,977900552	3,370165746
Variance	2,832842234	2,156660528
Observations	181	181
Pearson Correlation	0,358454006	
Hypothesized Mean Difference	0,96	
df	180	
t Stat	4,858224225	
P(T<=t) one-tail	1,28246E-06	
t Critical one-tail	2,347242646	
P(T<=t) two-tail	**2,56493E-06**	
t Critical two-tail	2,603418229	

FCI N3 questions - Null H.	*Post-test*	*Pre-test*
Mean	2,491712707	1,243093923
Variance	1,584653161	1,218354819
Observations	181	181
Pearson Correlation	0,309325701	
Hypothesized Mean Difference	0	
df	180	
t Stat	12,05002524	
P(T<=t) one-tail	3,35316E-25	
t Critical one-tail	2,347242646	
P(T<=t) two-tail	**6,71E-25**	
t Critical two-tail	2,603418229	

FCI N3 questions - Nill H.	*Post-test*	*Pre-test*
Mean	2,491712707	1,243093923
Variance	1,584653161	1,218354819
Observations	181	181
Pearson Correlation	0,309325701	
Hypothesized Mean Difference	0,94	
df	180	
t Stat	2,978382344	
P(T<=t) one-tail	0,001648663	
t Critical one-tail	2,347242646	
P(T<=t) two-tail	**0,00330**	
t Critical two-tail	2,603418229	

FCI N3 collision - Null H.	*Post-test*	*Pre-test*
Mean	0,801104972	0,35359116
Variance	0,160220994	0,229834254
Observations	181	181
Pearson Correlation	0,10796565	
Hypothesized Mean Difference	0	
df	180	
t Stat	10,19695116	
P(T<=t) one-tail	7,34062E-20	
t Critical one-tail	1,653363013	
P(T<=t) two-tail	**1,47E-19**	
t Critical two-tail	1,973230823	

Non N3 questions - Null H.	*Post-test*	*Pre-test*
Mean	0,91160221	0,906077348
Variance	0,492142419	0,574462861
Observations	181	181
Pearson Correlation	0,339545174	
Hypothesized Mean Difference	0	
df	180	
t Stat	0,08849211	
P(T<=t) one-tail	0,464791966	
t Critical one-tail	2,347242646	
P(T<=t) two-tail	**9,30E-01**	
t Critical two-tail	2,603418229	

Figure 40: Table - T-test for the MPEX: Overall, Independence, Coherence, Concepts

Overall - Favorable	*Pre-Test*	*Post-Test*
Mean	12,72988506	11,62068966
Variance	23,79366819	20,00558102
Observations	174	174
Pearson Correlation	0,644909758	
Hypothesized Mean Difference	0	
df	173	
t Stat	3,697491828	
P(T<=t) one-tail	0,00014605	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,00029**	
t Critical two-tail	1,973771337	

Overall - Unfavorable	*Pre-Test*	*Post-Test*
Mean	15,51149425	15,3045977
Variance	24,39004053	20,65234868
Observations	174	174
Pearson Correlation	0,630971349	
Hypothesized Mean Difference	0	
df	173	
t Stat	0,667437127	
P(T<=t) one-tail	0,25269114	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,50538**	
t Critical two-tail	1,973771337	

Independence - Favorable	*Pre-Test*	*Post-Test*
Mean	2,074712644	1,856321839
Variance	1,468374194	1,476347087
Observations	174	174
Pearson Correlation	0,337110055	
Hypothesized Mean Difference	0	
df	173	
t Stat	2,061891245	
P(T<=t) one-tail	0,020356735	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,04071**	
t Critical two-tail	1,973771337	

Independence - Unfavorable	*Pre-Test*	*Post-Test*
Mean	2,942528736	2,885057471
Variance	1,592053684	1,99827254
Observations	174	174
Pearson Correlation	0,349518483	
Hypothesized Mean Difference	0	
df	173	
t Stat	0,49521411	
P(T<=t) one-tail	0,310538697	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,62108**	
t Critical two-tail	1,973771337	

Coherence - Favorable	*Pre-Test*	*Post-Test*
Mean	1,908045977	1,597701149
Variance	1,425021593	1,282306823
Observations	174	174
Pearson Correlation	0,331667578	
Hypothesized Mean Difference	0	
df	173	
t Stat	3,042304105	
P(T<=t) one-tail	0,001356855	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,00271**	
t Critical two-tail	1,973771337	

Coherence - Unfavorable	*Pre-Test*	*Post-Test*
Mean	2,172413793	2,298850575
Variance	1,334263504	1,401501561
Observations	174	174
Pearson Correlation	0,414395625	
Hypothesized Mean Difference	0	
df	173	
t Stat	-1,317528715	
P(T<=t) one-tail	0,094701826	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,18940**	
t Critical two-tail	1,973771337	

Concepts - Favorable	*Pre-Test*	*Post-Test*
Mean	1,58045977	1,563218391
Variance	1,250714238	1,218523686
Observations	174	174
Pearson Correlation	0,347027298	
Hypothesized Mean Difference	0	
df	173	
t Stat	0,179105054	
P(T<=t) one-tail	0,429032401	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,85806**	
t Critical two-tail	1,973771337	

Concepts - Unfavorable	*Pre-Test*	*Post-Test*
Mean	2,431034483	2,304597701
Variance	1,159956149	1,276626138
Observations	174	174
Pearson Correlation	0,271491842	
Hypothesized Mean Difference	0	
df	173	
t Stat	1,251547979	
P(T<=t) one-tail	0,106212132	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,21242**	
t Critical two-tail	1,973771337	

Figure 41: Table - T-test for the MPEX: Reality, Math, Effort

Reality - Favorable	*Pre-Test*	*Post-Test*
Mean	1,885057471	1,83908046
Variance	1,212145372	1,002856953
Observations	174	174
Pearson Correlation	0,423515733	
Hypothesized Mean Difference	0	
df	173	
t Stat	0,535823802	
P(T<=t) one-tail	0,296384272	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,59277**	
t Critical two-tail	1,973771337	

Reality - Unfavorable	*Pre-Test*	*Post-Test*
Mean	1,517241379	1,373563218
Variance	1,12975882	1,044614976
Observations	174	174
Pearson Correlation	0,401083746	
Hypothesized Mean Difference	0	
df	173	
t Stat	1,660365608	
P(T<=t) one-tail	0,04932617	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,09865**	
t Critical two-tail	1,973771337	

Math - Favorable	*Pre-Test*	*Post-Test*
Mean	1,672413793	1,488505747
Variance	1,759118995	1,650156136
Observations	174	174
Pearson Correlation	0,488020447	
Hypothesized Mean Difference	0	
df	173	
t Stat	1,835744784	
P(T<=t) one-tail	0,034056072	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,06811**	
t Critical two-tail	1,973771337	

Math - Unfavorable	*Pre-Test*	*Post-Test*
Mean	2,367816092	2,471264368
Variance	1,794565145	2,030961398
Observations	174	174
Pearson Correlation	0,46578996	
Hypothesized Mean Difference	0	
df	173	
t Stat	-0,953750349	
P(T<=t) one-tail	0,170770373	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,34154**	
t Critical two-tail	1,973771337	

Effort - Favorable	*Pre-Test*	*Post-Test*
Mean	1,701149425	1,402298851
Variance	1,736761677	1,374792373
Observations	174	174
Pearson Correlation	0,41492852	
Hypothesized Mean Difference	0	
df	173	
t Stat	2,91469098	
P(T<=t) one-tail	0,00201574	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,00403**	
t Critical two-tail	1,973771337	

Effort - Unfavorable	*Pre-Test*	*Post-Test*
Mean	2,545977011	2,701149425
Variance	2,064347884	1,817686532
Observations	174	174
Pearson Correlation	0,553211435	
Hypothesized Mean Difference	0	
df	173	
t Stat	-1,552264795	
P(T<=t) one-tail	0,061213167	
t Critical one-tail	1,653709184	
P(T<=t) two-tail	**0,12243**	
t Critical two-tail	1,973771337	

Figure 42: N3 OST in German

Datum: ____________________ Schule: ______________________________ Name: ______________________________

Auftrag für die Masterarbeit in Physikdidaktik – Iva Nunes Sampaio-Kronister, Bed

universität wien

Open Source Tutorial

Universität Maryland

Code: ________________________

Ideen entgegen der Intuition: das 3. Newtonsche Gesetz

Das Hauptziel dieses Tutorials ist es dir zu helfen mehr Strategien für Physikkonzepte, die Alltagsvorstellugnen zu widersprechen scheinen, kennen zu lernen.

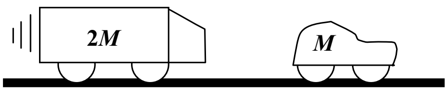
I. Das Newtonsche Gesetz und Intuitionen

Laut dem 3. Newtonschen Gesetz gilt:

Wenn zwei Objekte miteinander wechselwirken, ist die Kraft ausgehend von Objekt A auf Objekt B gleich groß (aber in entgegengesetzter Richtung) wie die Kraft ausgehend von Objekt B auf Objekt A.

Oft macht dieses Gesetz Sinn. Aber in einigen Fällen scheint es nicht so zu sein.

Stell dir vor, ein schwerer LKW rammt ein geparktes nicht besetztes Auto.



A. *(Gruppenarbeit)* Welche Kraft ist bei der Kollision nach eurer Intuition größer: Die Kraft ausgehend vom LKW auf das Auto oder die Kraft ausgehend vom Auto auf den LKW? Begründet eure intuitive Antwort.

B. *(Gruppenarbeit)* Diese Frage wurde vielen Studenten/innen gestellt und hier ist eine typische Antwort: *Gefühlsmäßig reagiert das Auto stärker während der Kollision (Man würde lieber im LKW sein!) : Also erfährt das Auto die größere Kraft.*

Ist die Erklärung eurer Gruppe aus Teil A ähnlich oder anders als diese Aussage? Begründet.

C. *(Gruppenarbeit)* Welche dieser Kräfte ist laut dem 3. Newtonschen Gesetz größer?

D. *Experiment.* Trifft in diesem Fall das 3. Newtonsche Gesetz nicht zu? Im Klassenzimmer hat der Betreuer/in ein Experiment aufgebaut, das die Kollision eines LKWs mit einem Auto simuliert. Begebt euch zum Versuchsaufbau und notiert die Ergebnisse hier. Ihr könnt zusätzlich überprüfen, ob das 3. Newtonsche Gesetz auch für andere Zusammenstöße gilt.

Übersetzung: Felix Scharf, Rupert Bärenthaler-Pachner

II. Wie man mit dem Gegensatz zwischen Intuition und dem 3. Newtonschen Gesetz umgeht?

Bevor wir mit dem nächsten Teil unserer Einheit beginnen, lasst uns den Gegensatz zwischen Physik und Alltagvorstellungen, den wir soeben gefunden haben, berücksichtigen.

A. *(Einzelarbeit)* Das 3. Newtonsche Gesetz widerspricht, nach der Meinung der meisten Leute, der Intuition, dass das Auto während der Kollision stärker reagiert. Welcher der folgenden Ausdrü cke entspricht am besten deiner Meinung bezüglich dieses Widerspruchs?

   i. Wir sollten diese Art von Widerspruch nicht beachten und stattdessen genau lernen, wann das 3. Newtonsche Gesetz anwendbar ist.

   ii. Wahrscheinlich gibt es eine Möglichkeit die Intuition zu verdeutlichen, um es mit dem 3. Newtonschen Gesetz in Einklang zu bringen, aber ich weiß nicht wie.

   iii. Obwohl die Physik normalerweise mit Intuition vereinbart werden kann, ist hier der Widerspruch zwischen Physik und Intuition so offensichtlich, dass man es akzeptieren muss.

   Erkläre kurz warum du diese Antwort gewählt hast.

B. Besprich deine Antwort in deiner Gruppe. Seid ihr euch alle einig oder nicht?

III. Eine neue Strategie: Verfeinern einer Intuition

Bevor wir einfach hinnehmen, dass das 3. Newtonsche Gesetz und unsere Intuition im Falle eines LKWs, der mit einem Auto kollidiert, einander widersprechen, wollen wir eine Strategie zur *Verfeinerung deiner Intuition* ausprobieren.

A. *(Gruppenarbeit)* Wir stellen uns eine neue Frage: Die Masse des LKWs beträgt 2000 kg und die Masse des Autos 1000 kg. Und wir stellen uns vor, dass der LKW während der Kollision um 5m/s langsamer wird. Wie viel Tempo nimmt das Auto während der Kollision laut eurer Intuition zu (verwende die Intuition, dass das Auto während der Kollision mehr reagiert und berücksichtige, dass der LKW doppelt so schwer ist). Erklärt eure intuitive Antwort.

✯ *Wendet euch an eine/n Betreuer/in bevor ihr fortfahrt.*

B. Stimmt eure Antwort zum Teil A mit dem 3. Newtonschen Gesetz überein? Um es herauszufinden führen wir euch durch einige schnelle Rechenschritte.

1. Angenommen das Auto und der LKW berühren sich für 0.5 Sekunden, bevor sie sich gegenseitig abstoßen. Berechnet:
   i. die Beschleunigung des LKWs während der Kollision.

   ii. die Beschleunigung des Autos während der Kollision.
   (vorausgesetzt eure Annahme der Geschwindigkeitsänderung ist korrekt)

2. In guter Näherung sind die Kräfte, die das Auto und der LKW auf den jeweils anderen ausüben, die horizontalen Gesamtkräfte, die während dem Zusammenstoß wirken. Ausgehen von den bereits errechneten Beschleunigungen, verwendet das 2. Newtonsche Gesetz (welches die Gesamtkraft mit der Beschleunigung in Verbindung setzt) um folgendes herauszufinden:
   i. die Kraft, die während der Kollision auf den LKW wirkt.

   ii. die Kraft, die während der Kollision auf das Auto wirkt.

3. Die Beschleunigungen und Kräfte basierten alle auf euren Annahmen bezüglich der Tempozunahme des Autos. Diese Annahme basiert auf der Intuition, dass das Auto während der Kollision stärker reagiert. Stimmt diese intuitive Annahme mit dem 3. Newtonschen Gesetz überein? Wie könnt ihr dies begründen?

✯ *Wendet euch an eine/n Betreuer/in bevor ihr fortfahrt.*

IV. Was ist soeben passiert?

A. *(Gruppenarbeit)* Wir müssen herausfinden, wie wir mit der „das-Auto-reagiert-mehr-Intuition“ umgehen, sozusagen die Idee, dass das Auto mehr reagiert als der LKW. Als ihr am Beginn dieses Tutorials eine Frage über die Kräfte zwischen Auto und LKW beantwortet habt, führte euch eure Intuition zu einer falschen Antwort, die nicht mit dem 3. Newtonschen Gesetz übereinstimmte. Aber in Teil 3 führte euch die gleiche Intuition über die Änderung der Geschwindigkeit zu einer richtigen Antwort, die mit dem 3. Newtonschen Gesetz übereinstimmt. Also, wie verhält es sich mit eurer „das-Auto-reagiert-mehr-Intuition“? Ist sie falsch? Ist sie richtig? Ist sie etwas anderes? Begründet.

B. *(Gruppenarbeit)* Wir sehen nun, dass das Auto während der Kollision stärker reagiert im Sinne, dass sich die Geschwindigkeit mehr verändert. Das heißt es erfährt eine stärkere Beschleunigung. Gebt eine Erklärung mittels Intuition, warum das Auto während der Kollision stärker reagiert (beschleunigt), obwohl die Kraft auf Auto und LKW gleich groß bleibt. Fragt eure/n Betreuer/in um einen Hinweis, wenn nötig. Dies ist die zweitwichtigste Frage in dieser Einheit.

C. *(Gruppenarbeit)* Dieses Tutorial ist nicht nur für das 3. Newtonsche Gesetz gedacht, sondern auch für Strategien zum Erlernen physikalischer Konzepte, die scheinbar Intuitionen widersprechen. Welche allgemeinen Strategien werden hier vorgeschlagen - Strategien, die ihr vielleicht später bei nicht intuitiven Konzepten im Unterricht gebrauchen könnt? Dies ist die wichtigste Frage in diesem Tutorial.

Figure 43: Split task in German

Für jede Testfrage:
– Bitte **kreise** jene Antwort ein, welche dir gefühlsmäßig am sinnvollsten erscheint;
– Bitte mache ein **Quadrat** um jene Antwort, welche, deiner Meinung nach, ein/e Wissenschaftler/in geben würde.

Beispiel:
A) Antwort, welche dir gefühlsmäßig am sinnvollsten erscheint
B) eine beliebige Antwort
C) Antwort, die ein/e Wissenschaftler/in geben würde
D) eine beliebige Antwort
E) eine beliebige Antwort

Falls die Antwort, welche dir gefühlsmäßig am sinnvollsten erscheint, mit jener Antwort, welche ein/e Wissenschaftler/in geben würde, **übereinstimmt**, **kreise** die Antwort ein und mache ein **Quadrat** um dieselbe Antwort.

Beispiel:
A) eine beliebige Antwort
B) Antwort, welche dir gefühlsmäßig am sinnvollsten erscheint und die ein/e Wissenschaftler/in geben würde
C) eine beliebige Antwort

Figure 44: Force-Test

Datum: ________________ Schule:_________________________ Name: ___________________________

Test für die Masterarbeit in Physikdidaktik – Iva Nunes Sampaio-Kronister, BEd

Kräftetest - Physik

Code: __________________

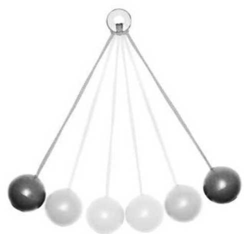
Für jede Testfrage:
– Bitte **kreise** jene Antwort ein, welche dir gefühlsmäßig am sinnvollsten erscheint;
– Bitte mache ein **Quadrat** um jene Antwort, welche, deiner Meinung nach, ein/e Wissenschaftler/in geben würde.

Beispiel:
A) Antwort, welche dir gefühlsmäßig am sinnvollsten erscheint
B) eine beliebige Antwort
C) Antwort, die ein/e Wissenschaftler/in geben würde
D) eine beliebige Antwort
E) eine beliebige Antwort

Falls die Antwort, welche dir gefühlsmäßig am sinnvollsten erscheint, mit jener Antwort, welche ein/e Wissenschaftler/in geben würde, **übereinstimmt**, **kreise** die Antwort ein und mache ein **Quadrat** um dieselbe Antwort.

Beispiel:
A) eine beliebige Antwort
B) Antwort, welche dir gefühlsmäßig am sinnvollsten erscheint und die ein/e Wissenschaftler/in geben würde
C) eine beliebige Antwort
…

1. Bei diesem Spiel stoßen einander zwei exakt gleiche Kugeln immer wieder in der Mitte ab. Sei F_{LR} die Größe der Kraft, die die linke Kugel auf die rechte ausübt und F_{RL} die Größe der Kraft, die die rechte Kugel auf die linke ausübt.



Es ist richtig zu sagen, dass:

A) $F_{LR} = \frac{1}{2} F_{RL}$

B) $F_{LR} = F_{RL}$

C) $F_{LR} = 2 \cdot F_{RL}$

D) $F_{LR} = 3 \cdot F_{RL}$

E) Keine davon.

2. Eine Person übt auf eine große Kiste eine konstante horizontale Kraft aus. Infolgedessen bewegt sich die Kiste mit konstanter Geschwindigkeit v_0 über den Boden.

Die von der Person ausgeübte konstante horizontale Kraft ist

A) genau so groß wie das Gewicht der Kiste.

B) größer als das Gewicht der Kiste.

C) genau so groß wie die Summe aller Kräfte, die der Bewegung entgegenwirken.

D) größer als die Summe aller Kräfte, die der Bewegung entgegenwirken.

E) größer als das Gewicht der Kiste und größer als die Summe aller Kräfte, die der Bewegung entgegenwirken.

Die nächsten beiden Fragen (3 und 4) beziehen sich auf den folgenden Text und die zugehörige Skizze.

Ein Lastwagen bleibt unterwegs mit Motorschaden liegen und wird, wie in der Skizze gezeigt, von einem Mittelklassewagen (Auto) zur nächsten Werkstatt geschoben.

3. Während das Auto **beschleunigt**, um beim Schieben auf eine bestimmte Geschwindigkeit zu kommen, gilt:

A) Der Betrag der Kraft, mit der das Auto gegen den Lastwagen drückt, ist gleich groß wie der Betrag der Kraft, mit der der Lastwagen gegen das Auto zurückdrückt.

B) Der Betrag der Kraft, mit der das Auto gegen den Lastwagen drückt, ist kleiner als der Betrag der Kraft, mit der der Lastwagen gegen das Auto zurückdrückt.

C) Der Betrag der Kraft, mit der das Auto gegen den Lastwagen drückt, ist größer als der Betrag der Kraft, mit der der Lastwagen gegen das Auto zurückdrückt.

D) Da der Motor des Autos läuft, übt das Auto eine Kraft auf den Lastwagen aus. Da der Motor des Lastwagen nicht läuft, kann der Lastwagen nicht gegen das Auto zurückdrücken. Der Lastwagen wird einfach deshalb vorwärts geschoben, weil er dem Auto im Weg ist.

E) Weder das Auto noch der Lastwagen üben aufeinander irgendeine Kraft aus. Der Lastwagen wird einfach deshalb vorwärts geschoben, weil er dem Auto im Weg ist.

4. Nachdem das Auto die gewünschte Geschwindigkeit erreicht hat, schiebt es den Lastwagen mit konstanter Geschwindigkeit vor sich her. In diesem Fall gilt:

A) Der Betrag der Kraft, mit der das Auto gegen den Lastwagen drückt, ist gleich groß wie der Betrag der Kraft, mit der der Lastwagen gegen das Auto zurückdrückt.

B) Der Betrag der Kraft, mit der das Auto gegen den Lastwagen drückt, ist kleiner als der Betrag der Kraft, mit der der Lastwagen gegen das Auto zurückdrückt.

C) Der Betrag der Kraft, mit der das Auto gegen den Lastwagen drückt, ist größer als der Betrag der Kraft, mit der der Lastwagen gegen das Auto zurückdrückt.

D) Da der Motor des Autos läuft, übt das Auto eine Kraft auf den Lastwagen aus. Da der Motor des Lastwagen nicht läuft, kann der Lastwagen nicht gegen das Auto zurückdrücken. Der Lastwagen wird einfach deshalb vorwärts geschoben, weil er dem Auto im Weg ist.

E) Weder das Auto noch der Lastwagen üben aufeinander irgendeine Kraft aus. Der Lastwagen wird einfach deshalb vorwärts geschoben, weil er dem Auto im Weg ist.

5. Eine Tennisspielerin trifft einen Ball mit dem Schläger, so dass er trotz eines starken Windes über das Netz ins gegnerische Feld geschlagen wird.

Welche der folgenden Kräfte wirkt (wirken) auf den Tennisball, nachdem er sich vom Schläger entfernt hat und bevor er den Boden berührt?

1. Eine nach unten gerichtete Schwerkraft
2. Die Kraft des Schlags
3. Eine Kraft, die durch die Luft ausgeübt wird

A) Nur Kraft 1.

B) Kräfte 1 und 2.

C) Kräfte 1 und 3.

D) Kräfte 2 und 3.

E) Kräfte 1, 2 und 3.

6. Ein schwerer LKW stößt frontal mit einem Kleinwagen zusammen. Für den Zeitraum des Zusammenstoßes gilt:

A) Der LKW übt eine größere Kraft auf den Kleinwagen aus als der Kleinwagen auf den LKW.

B) Der Kleinwagen übt eine größere Kraft auf den LKW aus als der LKW auf den Kleinwagen.

C) Die beiden Fahrzeuge üben keine Kräfte aufeinander aus. Der Kleinwagen wird einfach deshalb zerdrückt, weil er dem LKW im Wege ist.

D) Der LKW übt eine Kraft auf den Kleinwagen aus, aber der Kleinwagen übt keine Kraft auf den LKW aus.

E) Der LKW übt die gleiche Kraft auf den Kleinwagen aus wie der Kleinwagen auf den LKW.

7. Ein Mädchen und ein Bursch halten sich und drehen einander im Kreis. Sei F_{MB} die Größe der Kraft, die vom Mädchen auf den Bursch ausgeübt wird und F_{BM} die Größe der Kraft, die vom Bursch auf das Mädchen ausgeübt wird. Richtig ist:

A) $F_{MB} = 3 \cdot F_{BM}$

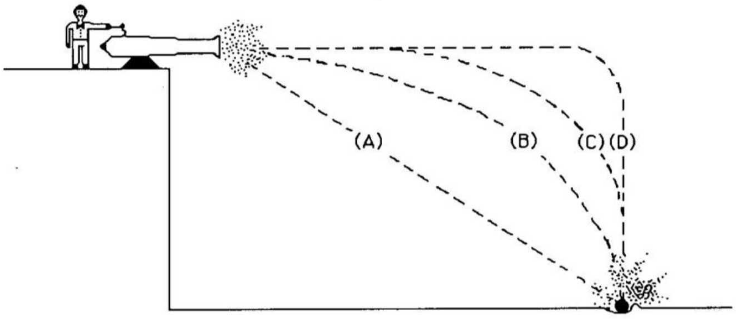
B) $F_{MB} = 2 \cdot F_{BM}$

C) $F_{MB} = \frac{1}{2} F_{BM}$

D) $F_{MB} = F_{BM}$

E) Keine davon.

8. Eine Kugel wird aus einer Kanone oberhalb eines Abhanges abgefeuert (vgl. Skizze). Welche der eingezeichneten Bahnkurven beschreibt die Flugbahn der Kugel am besten? (Markieren Sie Ihre Antwort in der Skizze.)



9. Zwei Schüler, *a* mit einer Masse von 95 kg und *b* mit einer Masse von 77 kg, sitzen sich auf zwei gleichen Bürostühlen gegenüber.

Wie die Skizze zeigt, stellt Schüler *a* seine Füße auf die Knie von Schüler *b*. Schüler *a* streckt plötzlich seine Beine, wodurch beide Stühle in Bewegung versetzt werden.

Für den Zeitraum, während Schüler *a* seine Beine ausstreckt und sich die beiden Schüler noch berühren, gilt:

A) Keiner der beiden Schüler übt eine Kraft auf den anderen aus.

B) Schüler *a* übt eine Kraft auf Schüler *b* aus, aber nicht *b* auf *a*.

C) Jeder der beiden Schüler übt eine Kraft auf den anderen aus, aber *b* übt die größere Kraft aus.

D) Jeder der beiden Schüler übt eine Kraft auf den anderen aus, aber *a* übt die größere Kraft aus.

E) Die beiden Schüler üben gleich große Kräfte aufeinander aus.

10. Bei einem Fußballspiel jagen Spieler A und Spieler B nach dem Ball, der gerade in der Luft ist, und prallen aufeinander. Genau im Moment des Zusammenstoßes ist die Größe der Kraft, die der Spieler A auf B ausübt (F_{AB}), im Vergleich zu der Größe der Kraft, die der Spieler B auf A ausübt (F_{BA}):

A) $F_{AB} = 3 \cdot F_{BA}$

B) $F_{AB} = 2 \cdot F_{BA}$

C) $F_{AB} = \frac{1}{2} F_{BA}$

D) $F_{AB} = F_{BA}$

E) Keine davon.

List of Figures