University of Southampton

Conceptual Understanding, Scientific Explanations & Arguments

In First Year University Physics:

Development and evaluation of an intervention

Foteini Chaimala

A thesis submitted in fulfillment of the degree of PhD

Faculty of Law, Arts and Social Sciences

October 2009

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF LAW, ARTS AND SOCIAL SCIENCES

SCHOOL OF EDUCATION

PhD Thesis

Conceptual Understanding, Scientific Explanations & Arguments

In First Year University Physics:

Development and evaluation of an intervention

By F. Chaimala

The study was conducted in the context of a 1st year university interactive physics course, aiming to foster students' scientific explanation and argumentation abilities and to promote their conceptual understanding in introductory mechanics. The research was developed and implemented under the mixed-method paradigm. It was formative and participatory evaluation at the 1st stage, resulting in the redevelopment of the course. At the 2nd stage (two years of study), it was summative and non participatory, aiming to investigate students' performance in the following areas: conceptual understanding; provision of correct and concise scientific explanation; ability to identify weakness/fallacies in scientific arguments; quality of written arguments. Analysis showed that – although students were not provided with the correct answers - there was a significant positive change in their conceptual understanding, after participating in the course, for both years of the study. In addition, students' provided significantly more appropriate and concise explanations at the end of the year than at the beginning. Furthermore, better results were obtained after the intervention in relation to students' difficulty in finding incorrect data in an argument, rather than identifying irrelevant or missing components in it. Yet, some students –even after the implementation- had the tendency to focus and respond only to the claim of an argument, rather than addressing and evaluating the whole argument. As for the quality of students' written argument, data analysis resulted in the expansion of the theoretical model initially chosen, so as to consider both oppositional and non-oppositional arguments. The application of the framework to the data showed that for both years of study there was a positive change in the quality of student's arguments after the intervention, with better results being obtained in the 2nd year.

CONTENTS

CHAPTER 1

The background and the context of the research

The importance of students' discourse in facilitating learning in science is widely recognized within a number of theoretical perspectives and has been demonstrated by research studies (Lemke, 1990). This recognition has promoted curricular reforms at all levels of science education. In the introductory university physics level, in particular, during the last two decades a considerable number of teaching programs based on interactive pedagogies has been developed and implemented, aiming to promote students' conceptual understanding. In addition, in recent years higher education has been moving away from an emphasis on knowledge acquisition towards students' development of core skills. No doubt that in educational circles, the aims that university studies should pursue are still debated. However, maybe few would argue that students' abilities to think independently and to possess a critical attitude towards knowledge are not desirable outcomes of higher education studies (Gow & Kember, 1990). Literature suggests that the above-mentioned goals can be promoted through student-to-student interaction and argumentation discourse, as they have been found to be positively linked with critical thinking skills (Kuhn, 1991).

Despite these developments most introductory physics courses in Greek universities are based on lecture-type teaching, which focuses on 'what we know' and on the transmission of scientific knowledge. In the view of the fundamental role that discussion could serve not only in facilitating learning in science, but also in the acquisition of students' core skills, this research study was conducted in the context of an interactive course at the University of Crete in Greece. The course incorporated some major features of university physics teaching programs worldwide based on a constructivist view of learning. It is enhanced by a number of interventions aiming to address both some of the limitations of these pedagogies and the special characteristics of Greek educational system. The main aim of the course is to provide a learning environment, which fosters students' scientific explanation and argumentation abilities and which promotes students' conceptual understanding in introductory mechanics. In such an educational context, this research was conducted as a participatory evaluation study, which was designed to investigate the extent to which the course mentioned above reaches the aims that it seeks to achieve. In particular, this study aims to explore students' achievement in the areas of scientific explanations, scientific argumentation and conceptual understanding.

The object of this chapter is to provide an overview of the research study, by focusing on the background and the context of the research. Given that this study deals with physics higher education in Greece, at first this chapter presents some major features of the Greek educational system (Section 1.1). Following this, attention is turned to physics education as provided by the formal Greek educational system, from the secondary level to the first-year university physics level (Section 1.2). Then, Section 1.3 presents the development, the design and the aims of the introductory physics course, which is the field under research in this study. After the educational context is clarified, this chapter finishes with an overview of the aims of the research study and the research questions that are to be explored (Section 1.4).

1.1. The Greek educational system: a brief description

The present research study took place in the Physics Department at the University of Crete in Greece. To give a general idea about the educational system in Greece, compulsory educatory includes Primary and Lower Secondary Education. Post-compulsory Secondary Education (for sixteen to eighteen years old children) consists of two school types: the Unified Upper Secondary Schools and the Technical Vocational Educational Schools. Tertiary Education, pursuant to Act 2916/2001, is divided into University Education, which is provided by the Universities, and Higher Technological Education, which is provided by the Technological Educational Institutes. Additionally, the Hellenic Open University provides an opportunity for open and distance learning in higher education. In University education, which is the focus of this study, the main aim is to ensure a high level of theoretical and all-round training for the future scientific workforce of the nation. In accordance with the Greek Constitution, Universities are legal entities under public law, which are fully self-governing under the supervision of the State, and they are financed by the State. In an effort to ensure equal opportunities for all, tuition is generally free of charge. However there are some exceptions, mainly in the case of certain post-graduate programmes and studies at the Hellenic Open University. There are 20 universities in Greece located in various towns. The Universities consist of Faculties, which in turn are subdivided into Departments. Students are admitted to these Departments according to their performance at national level examinations, which take place at the third grade of Unified Upper Secondary Schools. University courses last four years, except from certain faculties, where courses last five or six years. The academic year consists of two semesters with thirteen weeks of tuition and three weeks of examinations. The attendance in the courses is not compulsory at the Greek Universities, with an exception to the laboratory courses. In the physics departments in all Greek Universities, the students complete their studies after a minimum period of four years - provided they have passed the examinations both in the compulsory and optional subjects - and they receive a degree.

1.2. An overview of physics education in Greece

After providing a brief description of the structure of the Greek education system, attention is to be turned to physics education in Greece at the first-year university level, which is the area under research in this study. However, to provide just an overview of the first-year university physics scenery in Greece without any reference to physics education at pre-university level would uncover only a part of the picture of the field under investigation in this study. A more thorough investigation of the first-year university physics scenery involves inevitably some reference to the conditions at the educational level before it.

In secondary education in Greece, science is taught as separate subjects, such as physics, chemistry, biology. Physics has traditionally been one of the basic subjects in the curriculum, in both levels of secondary education. More particularly, in the lower secondary level physics is taught in the second and the third grade two hours per week. In the upper secondary level physics is taught in all three grades, from two to four hours per week, depending on the program of study that the students chose to attend. The broad areas of physics that are covered in secondary education are mechanics, electromagnetism, waves, thermodynamics and introductory atomic, nuclear and quantum physics.

Whereas the areas of physics that are taught in secondary schools have remained almost unchanged for more than fifteen years, school science curricula are currently the subject of debate and reform in Greece. Beginning in 2000 and aiming to bring the curriculum up-to-date, the first wave of reforms were concerned with the general aims that school science should pursue. The changes in the rationale for why we want to teach physics resulted in new debates about the suitability of the materials used in the classroom and the recommended activities. As a consequence, in the following years new materials and books were developed, and they have gradually started to be introduced in schools.

In terms of the purposes that school physics should pursue, the revised Greek national curriculum seems in general to follow the worldwide tendency towards curriculum convergence, as identified by Van den Akker (1998). As it has been argued in a documentary analysis of the Greek physics curriculum (Chaimala, 2004), the new physics curriculum is characterized by a tendency to put greater emphasis on making connections among sciences, other disciplines and societal phenomena. Moreover, greater emphasis is put on students' development of social skills and values, as well as on the development of their analytical skills. To acquire scientific knowledge just for the sake of knowing is not considered a goal of science teaching any more. These aims are also emphasized in recent document and reports in the filed of science education, for example in the 'National Science Education Standards' document (NRC, 1995) and the 'Beyond 2000' document (Millar and Osborne, 1998).

Similar conclusions could be reached from a more recent and extensive study on science teaching in schools in Europe (Eurydice, 2006). In a comparative analysis of current official regulations relating to science teaching in thirty European countries, this document provides a detailed breakdown of the desirable outcomes and the recommended activities that are evident in science curricula. Some of the main conclusions of this work for Greece are presented in the following figures. Figure 1.1 summarizes the recommended science activities that are expressed in the physics curriculum for lower-secondary education, while Figure 1.2 presents the desirable learning outcomes in relation to the skills assessed by standardized national examinations at the end of upper-secondary level.

Figure 1.1: Science activities expressed in the physics Greek curriculum (Eurydice, 2006)

Engaging in discussion in relation to :	*Science and society in everyday life*
	Researching information
	Experiments
Experimental work:	*Teacher demonstrations*
	Experiments following a pre-defined protocol
	Making observations
	Verifying a scientific law through experiment
	Formulating and testing hypothesis
	Presenting and communication procedures and results
Using scientific documentation:	*Researching documents for a defined purpose*
	Identifying and summarizing information
	Presenting and communication information
Using electronic technologies:	*Researching the internet for data*
Outside activities:	*Visit to museums, research laboratories*
Projects:	*Science related project work*

Figure 1.2: Outcomes of learning expressed in the physics Greek curriculum in relation to skills assessed by standardized national examinations (Eurydice, 2006)

		Outcomes of learning expressed in the physics Greek curriculum	Skills assessed by standardized national examinations
Knowledge of:	*Scientific concepts/theories*	•	•
	Experimental/investigative techniques	•	•
	And ability to apply Basic mathematical skills	•	•
Practical skills:	*Ability to follow experimental instruction accurately*	•	
Data handling skills:	*Ability to locate and extract information from documents*	•	
	Ability to interpret and evaluate experimental evidence	•	•
	Ability to search and present information from a range of sources	•	•
Scientific thinking:	*Ability to resolve problems formulated in theoretical terms*	•	•
Communication:	*Ability to engage in scientific discussions*	•	
	Ability to communicate procedures and results	•	
	Ability to use ICT	•	

Yet, the information provided in the figures above should be interpreted with caution. It is important to bear in mind that in the above-mentioned study only data available in official documents has been used; therefore, this study does not tell us about what is actually taught in secondary schools. One way to acquire such an insight for the lower secondary level is to turn our attention to the textbooks used in the classroom and to examine whether the aims of teaching as evident in these materials are relevant to the purposes of physics education in the Greek national curriculum. In a comparative analysis between the purposes of physics education as advocated in the national curriculum and as exemplified in the textbooks (Chaimala, 2004), it has been argued that there is a conflict between what it is proposed in the curriculum and the activities found in the textbooks. Indeed, this study indicated that in the physics curriculum most emphasis is put on the development of pupils' social skills and values. In contrast, the analysis of the secondary textbooks has demonstrated that the purpose which was identified more often in them is pupils' intellectual development. The outcome of this study could be seen in relation to the current debate on the suitability of the materials used in the classroom, which has been mentioned in the lines above, and to the fact that the during the last two years new materials and books are being developed are gradually introduced in schools.

On the other hand, as far as upper-secondary level is concerned, it can be argued that the skills assessed in the standardized national examinations at the end of this level could be used as an indicator of what is actually being taught in schools. As it can be viewed from Figure 1.2, among the desirable outcomes of the physics education as expressed in the national curriculum, the knowledge of the scientific concept and theories, and the ability to resolve problems formulated in theoretical terms remain the most important assessment objectives. In contrast, practical and communication skills are not a subject of the examination test. Given that in Greece upper-secondary level is regarded by many teachers and pupils as a stage that prepares students for the national exams, it could be argued that the assessment objectives are the ones that are also emphasized in the school classrooms.

Consequently, as far as secondary education is concerned, it seems that the ongoing reforms have contributed to the fact that there is not an established consensus in the Greek curriculum in terms of the goals that physics education should pursue. That is because of the incompatibility that exists among the ideal curriculum and the other curriculum representations (Goodlad, 1994): the ‘enacted curriculum’ (the actual instructional process in the classroom) and the ‘experiential curriculum’ (the learning experiences of the students). Given this lack of consensus, if we focus on the resulting learning outcomes to students (the attained curriculum), it would be rational to assume that they are closer to what is actually taught in the classroom and to the students’ learning experiences, rather than the vision elaborated in curriculum documents.

Nevertheless, in contrast to secondary education, where a recommended physics curriculum does exist, in tertiary physics education there is not a prescribed national curriculum. This is mainly due to the fact that Universities in Greece are fully self-governed and that each department has the autonomy to develop its own educational program and curriculum. The general educational program, the syllabus of each course, and the specific regulations of each department consist an official document of each University. There are five physics Departments in the Greek Universities, the studying programs of which vary considerably. Yet, in the first year of studies, the physics curricula in all five physics departments share more commonalities than differences. First, the main aim of the introductory physics courses, as evident in the curricula, is to expand students’ understanding of concepts of physics that are taught in the secondary level into a mathematical description. In short, the introductory courses provide an intensive revision of the upper- secondary school physics syllabus, in a more advanced mathematical level. In terms of the learning outcomes of these courses, it should be noted that they are not explicitly stated in the curricula; the physics curricula for the first-year studies contain only the conceptual areas that should be covered, along with the recommended books and the course materials. The absence of such statements, though, does not mean that there are no expected learning outcomes. Rather, the fact that the curricula contain only the areas that are to be taught implies an emphasis on knowledge acquisition and mathematical problem solving, at least in the case of theoretical courses. The above-mentioned emphases are also evident, if we use students’ assessment methods as an indicator of what the expected learning outcomes are. Indeed, in the case of the theoretical courses, students are assessed mainly by means of physics problems, which require except from knowledge of concepts and theories, mathematical problem solving skills. Finally, as far as the teaching and learning methods are concerned, all the introductory physics courses in physics departments in Greece are based on lecturing. Some departments also provide tutorials in which students get more help in problem solving by teaching assistants.

In short, in current years physics education in Greece in the border line between upper-secondary and university level is mainly based on lecture type teaching which emphasizes the transmission of content knowledge. As mentioned above, before entering the university, students experience an educational environment, in which the intellectual development is emphasized in everyday practice, regardless of the vision elaborated in the ideal curriculum. After entering university, in the first year of studies the emphasis remains on the content knowledge, in a more advanced mathematical level; the development of students' abilities, like argumentative and reasoning skills are not a part of the first-year university physics curricula. Yet, it should be noted that the above-mentioned skills are fundamental to what educators call 'critical' thinking (Kuhn, 1991); and that the development of criticality to undergraduate students has long been a central aim of higher education students. Maybe, in the case of first-year studies in the Greek Universities, there is still the assumption criticized many years ago (Rogers, 1948 as quoted in Driver *et al.*, 2000) that '*mere contact with science, which is so critical, will make students think critically'*.

1.3. The educational context of the study

In contrast to the first-year physics courses in Greek Universities, which are based on lecture-type teaching and emphasize the content knowledge, this study was conducted in the context of a first-year physics course that is innovative both in terms of the teaching methodology and the aims that it seeks to achieve. On the one hand, this course is based on a constructivist view of the learning and it is conducted by means of students' discussions, instead of the instructor's monologue. On the other hand, central aims of this course are the development of students' argumentative and scientific explaining abilities, rather than mathematical problem solving. In order to provide the context of this research study, the following lines provide an overview of the development, the design and objectives of this course.

1.3.1. The background

The design of the course was developed in several stages. The direction taken in the development of the unit was initiated by observations made in a first-year University course in the physics department of the University of Crete in Greece which was based on Mazur's (1997) pedagogy. The instructor of the course observed in more than one academic years that although he provided some teaching time for students' discussions based on explaining their answers, many students were at a loss to know whether they were in possession of an explanation, let alone of an 'appropriate' one. Similarly, the results on students' examination papers, in which students were asked to explain their views on conceptual questions on introductory mechanics, were disappointing.

In the first semester of the academic year 2004-2005, the first phase in the development of the course took place, as a pilot work of this study (for a detailed account on the process and the outcomes of the pilot work, view chapter 4). Semi-structured observations of students' dialogues were conducted and students' past exam papers were analyzed. The outcome of this pilot work indicated that, although students were able to find the correct answer to conceptual questions, they confronted difficulties in supporting their views. The observations suggested that students' difficulties were at a more primitive level than the one of conceptual ignorance: some students seemed to be unaware of the facts that an explanation requires more than providing a set of algorithms in mathematical problem solving and that an appropriate explanation cannot be seen independently of the context in which the question is being asked. In addition, focusing on the processes in students' argumentative discourse while the course took place, the observations indicated the poor quality of students' dialogues. For some students being involved in a discourse seemed to be equal to practicing parallel monologues; others exhibited mere inability to perceive logical defects in their peers' arguments or to see that logically defective statements cannot be valid arguments.

In short, the outcomes of the pilot work indicated students' difficulties in the areas of scientific explanation and argumentation. In order to construct a learning environment with the appropriate conditions that support scientific explaining and arguing, the related literature was scanned and provided helpful in the design and the development of the course (view chapter 2).

1.3.2. The design and the aims of the course

In the view of the outcomes of the pilot work and basing on research literature, the following procedures were adopted in the structure of course and were implemented in the first semester of the academic years 2005-06 and 2006-07. At first, like other university programs based on student-to-student interaction (for example Mazur, 1996; Mills *et al.*, 1999) most of the teaching time was provided for students to discuss conceptual questions on the area of mechanics. The questions were in a multiple-choice format, containing, along with the scientifically correct answer, alternative ones derived from literature reports about students' misconceptions. Following a few minutes discussion with their peers the students voted their answers using electronic devices. After the percentages of each answer were shown in a screen, students were asked to explain in the classroom the reasons for supporting their answer and were engaged in debates trying to persuade their peers on their opinions. This final step was the focus of the whole procedure; unlike the above-mentioned teaching programs which focus on students finding the correct answer, this unit emphasized the process of explaining and arguing.

For reasons supported by the literature (view section 2.2.1), in the process of explaining students were not asked to provide reasons for their opinions in general. Instead, the process of '*explaining to someone*' was emphasized as a distinct activity to 'explaining'. Students were introduced to a fictional person named 'Bobos', and were asked by the instructor to address their explanations to him with the aim to promote Bobos' understanding of the situation discussed. Bobos was introduced to the students as a high-school student and he is believed to provide a stimulus for engaging students with the process of scientific explaining. Some of Bobos' characteristics, which are gradually revealed in the educational procedure, are that he may have fundamental questions in mechanics since he does study a lot and therefore needs detailed explanations, he is in the same time smart enough to identify logical fallacies in the explanation procedure and he gets easily bored, especially when the explanations addressed to him contain only mathematical algorithms. In every situation, when a student explained his or her answers to the conceptual questions, the instructor of the unit asked the classroom to comment on whether their peers' explanation is appropriate for Bobos to understand and initiated, in that way, students' discussions concerning the appropriateness of an explanation in a particular context.

Along with the conceptual questions in a multiple-choice format, a set of materials have been developed for the aims of the course which seek to support the appropriateness of argumentation skills and discourse. The framework of these materials has basically the form of a concept map of students' ideas and it presents competing arguments for students to examine, discuss and evaluate. This is an adaptation of the common use of the concept map (Osborne, 1997). Students were given a concept map of statements based on students' perceptions of a science topic derived both from the research literature and students' arguments as expressed in the classroom. The students were then asked to discuss the concepts and the links with their peers in order to decide whether they are scientifically correct or not, while providing reasons for their choices. The instructor asked the students to report where exactly the fallacy in the argument they believed to be incorrect is, initiating conversations as far as the norms of scientific argumentation are concerned.

In the whole teaching process the teacher held the role of the facilitator in a productive dialogic discourse, rather than the instructor of the norms of the argumentative procedure. For example, the structure of arguments was not taught explicitly in the course. Rather, during the procedure of students' discourse about competing arguments, the instructor made use of a set of argumentation prompts such as '*why do you hold this view*', '*can you provide any evidence for your claim*', '*can you think of a piece of evidence that challenges the justification of your classmates argument*'. Moreover, both during and at the end of the process students were not explicitly told by the instructor the correct answers. Instead, students were encouraged to decide on their own, based on the discussions they have participated. This was used in order to avoid rote learning of the 'correct' answers to the questions and as a way to motivate students to take responsibility for their learning. In short, in the development of the unit the conditions that may support explaining and arguing in educational settings as found in the literature were taken into consideration (view section 2.1.3).

The main aim of the above-mentioned procedures was to create a learning environment that fosters independent learning. The specific objectives of the unit, as communicated to the students in the beginning of the course were: a) the promotion of their argumentative abilities, b) the enhancement of their explaining skills on scientific issues and c) deeper understanding of scientific concepts on the area of mechanics.

1.4. The aim of the study and the research questions

This research was designed to investigate the extent to which the course described in the previous section achieves its aims. In other words, this research seeks to determine the extent to which the pedagogical strategies used in the course lead to enhanced explaining and argumentative abilities, and to the improvement of conceptual understanding on the area of classic mechanics. More particularly, the research questions (RQ) that are to be investigated in this study – and in a way form the objectives of it - are the following:

RQ1: What is the change (if any) in students' conceptual understanding on basic mechanics after participating in the course?

RQ2: What is the change (if any) in students' ability to provide correct and concise scientific explanations on basic mechanics after participating in the course?

RQ3: How do students respond to weak or fallacious arguments? What is the change (if any) in the way students respond to such arguments after participating in the course?

RQ4: What is the change (if any) in the quality of students' written arguments after participating in the course?

The research questions presented above signify that this study is related to the following broad areas of investigation and research: the areas of conceptual understanding, scientific explanations and scientific argumentation and the area of the interactive physics courses in

Universities. The following two chapters provide, therefore, a literature review of the above mentioned fields of research, starting with the areas of conceptual understanding, scientific explanations and argumentation, which are presented in the next chapter.

Summary

Students' discourse could serve multiple functions in higher education environments. Despite this, introductory physics courses in Greek universities remain dominated by lecture-type teaching which emphasizes the transmission of content knowledge. This study was conducted in the context of an interactive course in the physics department at the University of Crete, which aims to foster students' argumentation and explaining abilities and their conceptual understanding. This chapter offered at first a brief description of the Greek educational system and an overview of physics education in Greece. It has been argued that physics education in Greece is currently under reform and that – despite the changes in the curriculum – the emphasis remains on content knowledge and on students' intellectual development. Following this, the educational context of this study was clarified, by providing the development, the design and the aims of the course which is the field under research. Finally, the aims and the research questions that are to be explored in this study were presented. In short, this study aims to investigate the extent to which this course reaches the aims that it seeks to achieve; broadly speaking, the areas under investigation in the study are students conceptual understanding, scientific explanations and scientific arguments.

CHAPTER 2

Literature review on the main foci of the study: Conceptual understanding, scientific explanations and arguments

The objectives of this research, which were presented in the previous chapter, signify that this study deals with three broad areas of investigation and research: the areas of conceptual understanding, scientific explanations and scientific argumentation. This chapter examines relevant literature to these areas. Given the extensive body of literature that exists in each one of these areas of interest in science education enquiry, it is beyond the scope of this chapter to provide a thorough investigation in each one of them. Rather, the literature review presented here is limited in respect of its relevance to the objectives of this study. In specific, Section 2.1 is relevant to the first research question and provides a concise literature review on conceptual understanding. Then, in Section 2.2, attention is turned to literature on scientific explanations and arguments, in relation to the research questions RQ2, RQ3 and RQ4. Each section offers the main theoretical trends, the theoretical perspective taken in this study and the research advances that have informed this research. At the end of each of the above-mentioned sections, a summary highlights the main points made before.

2.1. Conceptual understanding

2.1.1. Theoretical issues

Understanding scientific concepts has been traditionally one of the primary goals for science studies, at all levels of formal education. Yet, the notion of conceptual understanding is used in different and sometimes incompatible ways, guided or determined by theoretical beliefs about knowledge and learning. This is mainly due to the fact that ‘understanding’ is closely linked to ‘learning’, and that any account to learning considers inevitably the nature of the knowledge to be taught (Driver *et al*., 1994). Therefore, a theoretical examination of the notion of conceptual understanding should be seen in relation to theories about knowledge and learning.

For most of the twentieth century, behaviorism was the dominant learning theory in science education, deeply embedded in an objectivist epistemology. In short, according to this theory, scientific knowledge is viewed as a specific entity, existing outside the human mind; the aim of science is to discover the true nature of reality, while learning science is about knowing the truth. Learning in a behaviorist perspective is context-independent and it takes place as a result of an external set of stimuli and reinforcements. Under such a learning paradigm, students should exhibit behavioral skills (like knowing to handle equipment) and low-level cognitive skills (such as the ability to repeat definitions and laws, to apply formulae, to resolve standard problems). Therefore, in the behaviorist perspective, understanding of scientific concepts could be seen as synonymous to acquiring information about scientific concepts and being able to repeat it. A common criticism of behaviorism - as applied in educational settings- is that it often results in rote learning and recall of information with limited understanding (McRobbie & Tobin, 1997).

The second half of the twentieth century was marked by changes in theories of learning, from an emphasis on the behavioral towards the cognitive and social nature of thinking. Initially, Piagetian ideas were applied in science education, which acknowledge that learning is an active process of restructuring one's thoughts and of connecting symbols in a meaningful way. Later, constructivist ideas were developed by merging various cognitive approaches, with a focus on viewing knowledge as a changing body, being constructed by the learners, and strongly influenced by the prior knowledge of the learner (Novak, 1977). In short, learning science under a constructivist framework can be considered as restructuring the existing knowledge and constructing new models to fit to new understandings and experiences. Certain limitations of the constructivist ideas of the eighties and early nineties, such as the failure to consider the social dimensions of learning, contributed to the development of social constructivism.

According to social constructivism, knowledge is neither given nor absolute, but is rather an individual construct in the social contexts in which actions occur. In science education, this theory builds mainly on work of Vygotsky, and views learning as a social activity in which learners make meaning through both individual and social activities, like discussions and negotiations with teachers and other learners (Driver *et al.* 1994). In particular, Vygotsky conceptualized speech not as an expression of fully developed thought, but as a means towards the development of thought. Social constructivism as applied in science education settings is not without criticisms, mainly for epistemological issues. A common criticism is that anyone's construction of the world is as viable as another and therefore the word only exists in the mind of each individual (Duit, 1995). In a similar vein, Osborne (1998) has questioned social constructivist epistemology, arguing that it is a misrepresentation of the views and practices of science and scientists and that it confuses the way in which new knowledge is made with the manner in which old knowledge is learned.

Regardless of the criticisms on cognitive and social constructivist learning theories, they had a major contribution to the development of novel meanings to the notion of conceptual understanding in science education. From the early eighties, 'conceptual change' became the term denoting learning scientific concepts under constructivist perspectives (Duit, 1999). Maybe, the most influential theoretical framework in the conceptual change approach was that of Posner *et al.* (1982), which attempted to explain how '*peoples organizing concepts change from one set of concepts to another set, incompatible to the first* (p.211)'. They proposed two types of conceptual change: assimilation (which describes the process where students use existing concepts to deal with new phenomena) and accommodation (which describes when students replace or reorganize their central concepts). In focusing on the second type, they described the conditions that must be fulfilled for this type of change to occur: dissatisfaction with the existing concepts, intelligibility of the new conception, that the new conception must appear plausible, and that the new concept should suggest the possibility of a fruitful programme. This framework was the leading paradigm that guides research and instructional practices, until it became subject to several criticisms. Some of them, under social constructivist perspectives of learning, pointed out that conceptual change happens in a broader educational and socio-cultural context, and that it is affected by motivational and affective variables (Pintrich *et al*., 1993).

Literature reviews on the development of the notion of conceptual change reveal that theorists have viewed changes in students' knowledge status in more than one perspective: an epistemological position (Posner *et al*., 1982), an ontological position (Vosniadou, 1994) and a social/affective position (Pintrich *et al.*, 1993). As a consequence, a range of terms have been used to describe conceptual change except from assimilation and accommodation: weak restructuring and strong restructuring, differentiation and re-conceptualization, enrichment and revision (Tyson *et al.*, 1997). Recently, there have been some efforts to examine changes in students' knowledge in more than one perspective, by using multi-dimensional interpretive frameworks for conceptual change (Venville & Treagust, 1998; Duit & Treagust, 2003). These frameworks

> *are intended to construct a holistic picture of conceptual change by considering not only the changes in knowledge structures that a student requires to construct a scientific view of a concept, but also the ontological, social and affective and epistemological aspects o conceptual change* (Venville &Treagust, 1998; p.1032).

These approaches could be useful in splitting up some components of the process of conceptual change and in providing a more holistic interpretation of conceptual change.

Although there are critical differences between the above-mentioned theoretical perspectives on conceptual change, it can be argued that there is common ground among them: At first, in all these perspectives and in a general sense, conceptual change denotes learning pathways from students' pre-instructional conceptions to the science concepts to be learned (Duit, 1999). Secondly, it can be argued that the above-mentioned perspectives are not contradictory, but complementary to each other, as they deal with different factors which may influence conceptual change (Vosniadou, 2004). Finally, it should be noted that in all these perspectives conceptual change has been mostly considered as a polarization of low and high levels, like assimilation and accommodation, enrichment and revision, addition and restructuring. Therefore, without overlooking the differences in the various theoretical positions, conceptual change could be used at a basic but not oversimplified level, as a term denoting learning under constructivist paradigms.

2.1.2. Theoretical position and definitional issues in this study

In this study, the perspective taken about what learning concepts involves is close to social constructivist views, as applied in science education settings and as described roughly in the previous section. In short, learning concepts in this thesis is generally viewed as an adaptive process - both individual and social - that organizes one's experiential world. Therefore, science learning is considered as conceptual change, in the sense that learning is viewed as a process in which students reorganize their existing knowledge in order to understand concepts and processes of science more completely (Vosniadou *et al*., 2001). The development of the learning environment, which is the context of the research and which was described in the previous chapter, was based on these assumptions about learning, and has taken into consideration the factors that may influence such a process.

However, it should be noted that this study differentiates between the assumptions about the learning process and the views about the nature of knowledge to be learned. Thus, although it is close to social constructivist views about learning, it accepts some of the criticisms about constructivist epistemology and ontology, which were made by Osborne (1998) and were mentioned above. In specific, in this study scientific knowledge is considered as a changing body, socially constructed and validated. Yet, it is acknowledged that this body does exist - even as a social construct - and that for the learners it is discovered as an independent pre-existing word outside their minds. Therefore, in terms of epistemological issues, the position taken in this study is more close to the pragmatism stance that knowledge is both being constructed and based on the reality we experience. Being clear about philosophical issues is important, as the philosophical orientation of the researcher has contributed to the choice of the methodological framework for conducting this research (view chapter 4).

Moreover, this study differentiates between the process of science learning (referred as conceptual change in the lines above) and the actual product of such an activity. The latter is to be referred in this study as conceptual understanding and is one of the subjects under investigation in this research. Tyson *et al*. (1997) made the remark that some empirical studies following conceptual change frameworks have neither used these frameworks for analysing the data, nor focus on the process of students learning; rather, they consider changes in students' conceptual knowledge. In this study, the teaching strategy that was designed and implemented took into consideration the conditions in the process of conceptual change. However, the area under research is mainly the changes in students' conceptions, rather than the process itself. It should be noted here that theoretical frameworks of socially based descriptions of learning often avoid interpreting learning in terms of students' increasing conceptual knowledge. However, this study takes the argument from v. Aufschnaiter (2003, p.342) according to whom:

> *even if knowledge is described as being shared or distributed, there must be at least an individual component, otherwise none of us would be able to re-organize previously experienced situations and practices* .

Under such a rationale, conceptual understanding is mainly viewed in this study as the outcome of the process of conceptual change. In specific, it is considered a complex phenomenon, which comprises of factual knowledge (knowledge about single or more complex concepts), procedural knowledge (rules and algorithmic) and conditional knowledge (the understanding of when to employ procedural knowledge). This definition is similar to Nieswandt's (2007) interpretation of conceptual understanding, with the exception that in that study it is also viewed as students' ability to apply the learned scientific phenomena in everyday life. Moreover, this definition is close to how Alao & Guthrie (1999) use conceptual understanding, which emphasizes 'breadth' of knowledge (the extent of knowledge that is distributed and represents the major sectors of a specific domain) and 'depth' of knowledge (the knowledge of scientific principles that describes the relationship among concepts). Finally, it should be noted that the term conceptual understanding is used instead of conceptual knowledge, given that the second could sometimes be synonymous to acquiring information about concepts, which is not an issue of interest in this research.

2.1.3. Research issues

Concept-learning has received considerable attention of the research literature in science education, during the past three decades. Eylon and Linn's (1988) review of the research literature situated concept-learning among the four perspectives of learning that have emerged from research into science education. According to them, research from concept-learning perspective characterizes students' topic related understanding of scientific concepts and it incorporates

> *the conceptions that students hold, offers ideas about how these conceptions arise and suggests what factors elicit them* (p. 262).

Similarly, in a more recent review, Duit & Treagust (2003) placed research on students' conceptions on various science content domains and their roles in teaching and learning among the most important domains of science education research from the late seventies till nowadays.

Indeed, a historical overview of the research literature reveals that work on conceptual understanding appears in the literature in the late seventies, with the investigation of students' pre-instructional conceptions on various science content domains. This orientation could be attributed to the emphasis that begun to be placed at this time on the role of students' prior knowledge, as a central variable affecting any subsequent learning under constructivist perspectives. The eighties saw the growth of research studies which investigated not only the concepts that students hold before instruction, but also the development of their initial perceptions towards the intended scientific concepts. It was the era of the 'Alternative Frameworks' movement in science education[1]. From the eighties till now, there have been numerous studies describing students' ideas in scientific fields, including physics, chemistry and biology before, during and after instruction. A review on the field can to be found in Limon and Mason (2002). The common consensus of the research findings indicates that many students hold similar perceptions about scientific concepts, regardless of age and gender, which are incompatible with scientific thinking. Although, even from the early nineties considerable attention has also been given in describing students' meta-cognitive conceptions (for example ideas about the nature of science), it is noteworthy that there are still research studies which primarily investigate students' conceptions at the content level (for example Chiu *et al.*, 2007).

---

[1] 'Alternative frameworks', 'alternative perceptions', 'alternative frames of reference', 'children's science' and 'misconceptions' are some of the terms found in the literature to describe students' ideas which are not in harmony with the science views or are even in contrast to them.

The advancement of cognitive science research in the seventies and eighties had a significant role in the emergence and the rapid development of instructional research in the nineties. A large body of research has tried to explain under conceptual change theoretical models how instructional interventions change the learning environment and whether they result in more efficient teaching. According to Limon (2001), three kinds of instructional strategies can summarize many of the instructional efforts made to promote conceptual change: the induction of cognitive conflict as central condition for conceptual change, the use of analogies to guide students' change and cooperative learning to promote collective discussion of ideas. However, it can be argued that these categories are indicative and not mutually exclusive. Indeed, according to Chan *et al.* (1997), the usual cognitive conflict paradigm involves: a) identifying students' current state knowledge, b) confronting students with contradictory information presented through texts and interviewers who make explicit the contradiction or guide the debate among peers and c) evaluating the degree of change between students' prior ideas and a post-test measure after the instructional intervention. Therefore, it can be argued that interactive learning -with the instructor guiding the debate between peers- could be seen as a stage in the cognitive conflict approach. Given the above, the instructional strategy, which is the area under research in this study, could be placed in the category of cognitive conflict instructional processes, using interactive learning as an intermediate process (for a literature review on interactive learning in the area of first year University physics view chapter 3).

From the early nineties a large number of empirical studies have been reported on the application of cognitive conflict strategy, as a means towards conceptual understanding in various scientific domains. The results of such an instructional process have been controversial. On the one hand, many studies have reported positive effects as far as the learning outcomes in students are concerned (for example Dreyfus *et al.*, 1990; Jensen & Finley, 1995; Limon & Carretero, 1997; Mason, 2000). On the other hand, though, it has been pointed out that this instructional strategy seems not to work to the extent that it was expected. The most common 'negative' result of some empirical studies has been that students do not achieve a strong restructuring in their existing knowledge and, therefore, they so no not manage to achieve a deep understanding of the new information they receive. Chan *et al.* (1997, p.2) for example have emphasized that students '*are often unable to achieve meaningful conflict or to become* *dissatisfied with their previous conceptions'*. Moreover, Dreyfus *et al.* (1990) reported that sixteen years old students often failed to reach a stage of meaningful conflict. In a similar vein, Limon & Carretero (1997) reported that such an instructional strategy helped students be aware of contradictions; however, they stressed that this step is a necessary, but not sufficient to achieve conceptual change and that no radical changes were produced as predicted. In reviewing relevant literature, Limon (2001) has also pointed that partial changes are achieved through cognitive conflict strategies, but in some cases they disappear in a short period of time after the instructional intervention.

Without underestimating the positive results that have been reported from research literature on cognitive conflict strategies, it can be argued that for further research and improvement on the instructional strategies it is important to focus the limitations of such a strategy and the factors that may explain why it is not always effective. Chinn & Brewer (1993) pointed out that some of the factors influencing students' achievement of a meaningful conflict are their background knowledge and their ontological and epistemological beliefs. Similarly, Vosniadou (1994) made the point that conceptual change is harder when students' ontological beliefs need to changed. Moreover, Pintrich & Garcia (1991) found out that motivational factors are strongly related to students' use of cognitive strategies, as well as their performance. On the other hand, Kuhn (1991) suggested that students' reasoning abilities are also relevant for them to achieve a meaningful cognitive conflict. In a similar vein, Chinn & Brewer (1993) found out that students who were engaged in a thoughtful, effortful processing of arguments and treated the new concepts as something that needed to be explained were more likely to reach meaningful conceptual conflict.

In short, from the lines above it can be viewed that research on conceptual understanding during the last twenty years has mainly focused on two areas: On the one hand it has been oriented towards investigating students' alternative ideas about scientific concepts. This body of literature could be helpful in constructing a learning environment aiming to foster students understanding of scientific concepts, under constructivist paradigms. On the other hand, research has focused on evaluating instructional interventions aiming to more efficient teaching, like the cognitive conflict approach. As far as this teaching and learning approach is concerned, more than fifteen years of research experience on the field indicates that, although positive outcomes have been reported, a number of factors seem to influence its effectiveness. Some of these factors seem to be students' background knowledge, their ontological and epistemological beliefs, motivational factors and students' explaining and reasoning abilities. In the view of these factors that may influence the effectiveness of the cognitive conflict strategy as a means of achieving conceptual understanding, this study makes a case for developing and evaluating a teaching strategy which aims to foster students' explaining and arguing abilities, along with their conceptual understanding.

Summary

The notion of conceptual understanding is determined by theoretical beliefs about knowledge and learning. In this study, science learning is considered as conceptual change, in the sense that learning is viewed as a process, in which students reorganize their existing knowledge in order to understand concepts and processes of science more completely (Vosniadou *et al*., 2001). Conceptual understanding is generally viewed in this study as the outcome of this process. Research on conceptual understanding during the last twenty years has mainly oriented towards investigating students' alternative ideas about scientific concepts and towards evaluating instructional approaches. Research results on the cognitive conflict approach have been controversial. Positive outcomes have been reported, but it has also been pinpointed that factors like students' background knowledge, their ontological and epistemological beliefs, motivational factors and students' explaining and reasoning abilities may influence its success. In the view of these factors, this study makes a case for evaluating a teaching strategy which aims to foster students' explaining and arguing abilities, along with their conceptual understanding.

2.2. Scientific explanations and arguments

2.2.1. Theoretical issues

Explaining and arguing are mainly referred in science education literature as two distinct processes and areas of study. However, even if the most detailed of definitions are adopted, it seems that the two terms overlap to a considerable extent. Indeed, the term 'argument' is derived from the Latin *arguere,* meaning 'to show, to make clear, to assert, to prove, to accuse'. As for explanation, it is conceived by many theorists as clarification, description, justification or provision of reasons (Norris *et al.,* 2005). Due to the overlapping of the meanings, the concepts of scientific explanations and arguments are not viewed a priori distinct from each other; rather, they are discussed theoretically in parallel, in an effort to illuminate the degree of overlapping and the differences in the meaning of the terms.

As far as explaining is concerned, the literature reports that even within the legitimate scientific discourse, explanation is not a unique mode of activity; therefore the question of what constitutes a scientific explanation cannot have a unique answer. As Solomon (1986) stresses, sometimes an explanation involves the provision of the purpose of the phenomenon, sometimes the provision of the cause of it and at other times explanation is an analogous situation to the phenomenon, which provides a kind of metaphorical illumination. In a similar vein, Martin (1972) has analyzed the complexity of the field, by identifying five meanings for the concept of explanation in science and science education: a clarification of what a phrase means in a scientific context; a justification for some belief or action; a causal account of an event or process; a citation of a theory from which a law may be deduced; an attribution of function to an object. Given the multiple meanings of the term, the literature reports some efforts to develop a typology of scientific explanations, which are based on the different functions that explanations serve in different contexts. Figure 2.1 provides a typology of scientific explanations and some of their characteristics, as constructed by Norris *et al*. (2005).

Just like the notion of scientific explanations is multidimensional, the meaning of an appropriate explanation varies, according to the people who are involved in the process of explaining. Gilbert *et al*. (1998) point out that the actual application of the criteria of

Figure 2.1: A typology of scientific explanations and their characteristics (Norris *et al*., 2005)

Types of explanation	***Characteristics***
Interpretive explanation	clarifies meaning / defines terms, propositions, treatises
Justificatory explanation	explains by justifying why something was done / provides reasons for acting
Descriptive explanation	explains by describing a process or structure
Causal explanation	explains by citing a cause for events or laws
Deductive- nomological explanation	explains particular facts or general laws by deriving the facts or laws from general laws and other facts
Statistical explanation	explains facts by showing them to be highly probable/ basic structure is an inductive argument
Functional explanation	explains a fact by indicating its function
Explanatory unification	explains phenomena by fitting them into a general worldview / aims to derive largest number of facts from smallest number of assumptions
Pragmatic explanation	explains by answering why questions / questions are asked and answers are given in a context
Narrative explanation	explains an event by narrating the events leading up to its occurrence / posits some events as causes of others

explanations' appropriateness depends on the intensions, which the explainee has in respect of a particular situation. In the science community, some of the criteria used in evaluating explanations are the following: plausibility, parsimony, generalizibility, fruitfulness, accuracy (Brewer *et al*., 2000). In the context of science education, different groups (for example scientists, curriculum designers, science teachers and science students) seem to have distinct perceptions about the attributes of an appropriate explanation (Gilbert *et al.,* 1998). In an effort to overcome the difficulty that arises due to the differences in perceptions, Martins' work (1972) provides some helpful insights. In discussing the nature of scientific explanations in the context of teaching, Martin draws a distinction between 'explaining a thing' and 'explaining a thing *to someone'*. According to her, the former is seen as a research activity which aims to seek the truth, to promote new discovery and new understanding in the researcher. The latter is a pedagogical activity which aims to impart knowledge and to promote the understanding of the other person, the explainee. Although both activities take place in educational environments, in the discourse procedures where 'new discoveries' are communicated, the process of explaining a thing to someone plays the major role. Therefore, in the context of science teaching, the notion of an appropriate explanation seems to be twofold: on the one hand it refers to whether or not the explanation is in accordance with the scientific theories or facts (criteria mainly used in the science community); on the other hand, it concerns the degree to which the explanatory act is likely to impact knowledge and to promote understanding in the other person, the explainee (criteria mainly used in educational environments).

Regardless of the variations on the meaning of scientific explanations and on the criteria for their appropriateness, explaining has traditionally had a central place in science and science education. In contrast, only recently did serious discussion begin about the role of argumentation in science learning and about why it is important to enhance students' skills of arguing. A review of the literature on argumentation the last decade reveals that situating argumentation as a central element in the learning of sciences is based on the following theoretical underpinnings: At first, it has been argued that in order for science teaching to address epistemological goals, it is important to establish learning environments, which enable students to engage in modes of discourse that resemble more closely those of scientific community (Lawson, 2003; Duschl & Osborne, 2002). In addition, students' engagement to argumentative discourse is important for the achievement of conceptual goals. It has been argued that during such a process the students have the opportunity to articulate reasons for supporting particular conceptual understandings and they attempt to justify their views to other students who may present alternatives; therefore, a clearer conceptual understanding may emerge (Newton *et al.*, 1999). Furthermore, from a societal perspective, it has been pinpointed that in order citizens to engage in scientific debates it is important to evaluate the validity and reliability of the evidence used in scientific arguments (Simon *et al.*, 2003). Finally, to situate argumentation as a central component of science learning is important, given that it makes students' scientific reasoning visible and therefore enables formative assessment by teachers or instructors (Osborne *et al.*, 2004).

A concise definition of argumentation has been provided by Suppe (1998), who has defined this notion as the coordination of evidence and theory to support or refute an explanatory conclusion model or prediction. Duschl & Osborne (2002), point that -in contrast to the lay perception of argumentation as war to establish a winner - argumentation in science is a social and collaborative process, necessary to solve problems and advance knowledge. Similarly, according to Jimenex-Aleixandre *et al.* (2000), argumentation is a structural element of the language of science, which constitutes strategies for resolving questions, issues and disputes by the use of arguments. Some of these strategies are the processes of reasoning, evaluating and justifying, with the purpose to clarify and refine ideas, so as to come to a decision (Maloney & Simon, 2006). In all the above-mentioned approaches, there is an implicit or explicit differentiation between argumentation and arguments: broadly speaking, argumentation denotes the process of constructing arguments, while arguments refer to the content of such a process. Kuhn (1991) points that that there are two interpretations on the meaning of argument in the educational literature. The first one is described as rhetorical or didactic and could be interpreted as advancing a reason for or against a proposition of action. The second interpretation of argument is involved when different perspectives are being examined aiming to reach an agreement (dialogical argument).

Maybe the most significant impact on how science educators have defined and used argument in educational settings has been made by Toulmin (1958). Based on an analysis of arguments in different contexts, Toulmin presented a model which specifies the components in reasoning from data to a conclusion. The main elements of arguments as identified by Toulmin are presented in Figure 2.2. This pattern has been drawn on increasingly in science education the present years and has provided a way to describe students' arguments (for example Jimenex-Aleixandre *et al.*, 2000; Zohar & Nemet, 2002; Erduran *et al.*, 2004). It should be noted that although this framework could be used to analyze the structure of arguments, it could not lead to judgments about their correctness. In order such judgments to be made, subject knowledge should be incorporated for arguments to be evaluated. Furthermore, as Driver *et al.* (2000) argue, in this framework, argumentation is presented in a de-contextualized way. Therefore, in order to interpret the analysis both cognitive and social dimensions of learning need to be taken into account in the analysis.

Nevertheless, except from being used as a tool to describe and analyze students' arguments, Toulmin's argumentation pattern has provided a theoretical basis for evaluating written and spoken arguments. Recently, Erduran *et al.*, (2004) proposed a well-articulated approach on how this pattern could be adapted for methodological purposes in evaluating the quality of

Figure 2.2: The main components of the argument, as identified by Toulmin (Driver *et al*., 2000)

Components of arguments	***Characteristics***
Data	The facts that those involved in the argument appeal to in support of their claim
Claim	The conclusion whose merits are to be established
Warrants	The reasons (rules, principles) that are proposed to justify the connections between the data and the claim or conclusion
Backing	The basic assumptions that provide the justification for particular warrants
Qualifiers	The conditions under which the claim can be taken for granted, the limitations on the claim
Rebuttals	They specify the conditions when the claim will not be true

students' arguments in oral discourse. Drawing on research literature, they argue that the presence or absence of rebuttals is a significant indicator of the quality of argumentation. In particular, they have perceived low-level argumentation when the opposition between students consisted only of counter arguments that were unrelated. In contrast, they consider higher level argumentation when there was a rebuttal, which was in direct reference to a piece of evidence offered. Another approach on how to use Toulmin's framework for evaluating arguments has been proposed by Cerbin (referred to in Marttunen, 1994). In this approach, the following criteria were used for evaluating arguments: a) the clarity of the claims, b) the relevance and sufficiency of the grounds, c) the relevance of the warrant and d) whether counterarguments have been presented. Kelly & Takao (2002), on the other hand, in developing a model for evaluating students' written arguments modified Toulmin's pattern, by using epistemic levels which allowed to consider the disciplinary-specific knowledge in relation to the students' argument structures. A weakness to this model, as reported by its developers after the empirical application, is that the subject-matter experts differed in their assessment of the students' arguments as compared to the predictions of their model.

In short, the lines above focused on the following theoretical issues, which concern scientific explanations and arguments and which are relevant to the objectives of this research: At first, in respect to what constitutes a scientific explanation and argument, it has been shown that the two notions overlap to a considerable extent, given that they both refer to the process of justification of and reasoning for a position or conclusion. Yet, as far as scientific arguments are concerned, there is an implicit or explicit emphasis on the structure of the reasoning process. This can be viewed from the theoretical frameworks used to analyze and evaluate argumentation, which focused on the structure of the students' arguments. In contrast, scientific explanations mainly refer to the content of a position about how and why a phenomenon occurs. Secondly, it terms of the evaluation of scientific explanations and arguments, it was discussed how theorists have conceived the notion of an appropriate explanation, and how the Toulmin's pattern has provided a useful methodological tool for science educators to analyze and assess students' arguments. Finally, given that the field of argumentation has only recently been a subject of focus and research in science education, the theoretical underpinnings of the importance of argumentation in science learning have been illustrated. Some of the arguments found in the literature for advancing the case for the inclusion of argumentation in science classrooms are the development of conceptual understanding, the understanding of epistemology of science and the public understanding of science.

2.2.2. Definitional issues in this study

Given the multiple meanings and functions of explanations as perceived by different theorists and the overlapping in the meanings between scientific explanations and arguments, it is important to clarify how the terms are defined and used in this study. From the discussion in the previous section it became evident that both processes concern justification, reasoning and support or refutation of a conclusion. In addition, it seems that each process incorporates the other one: indeed, in order to explain (clarify the meaning, justify why or how a phenomenon occurs) one needs to coordinate evidence and theory, therefore needs to argue; on the other hand, during the process of arguing about scientific issues one makes use of different types of scientific explanations (clarifies meanings, states causes or effects). Yet, as mentioned in the discussion in the previous section, it seems that as for argumentation the focus is on the structure of the reasoning process, while scientific explanations mainly refer to the content of how or why a phenomenon occurs. Therefore, in this study explanation refers to the clarification and the justification of a position or conclusion in detail, with the focus turned on the content of the reasoning process. On the other hand, following Suppe's definition (1998), argumentation is seen as the coordination of arguments to support or refute a conclusion, with the emphasis put on the structure of the reasoning process. Under such definitions of explaining and arguing, when talking about the quality of students' scientific explanations in this study attention is turned to the content of the reasoning process. In specific, following the rationale presented in the previous section, an appropriate explanation in this study refers to whether or not the explanation is in accordance with the scientific theories or facts, while the quality of the explanation concerns the degree to which the explanatory act is likely to impact knowledge and to promote understanding in the other person, the explainee. On the other hand, as far as the quality of students' arguments is concerned, the emphasis is put on the structure of their reasoning process. In particular, following Erduran *et al.*'s (2004) approach on Toulmin's model, which was discussed above, the quality of students' arguments, is seen in relation to the presence or absence of rebuttals in their arguments. The way scientific explanations and arguments are defined in this study lays important groundwork for the methodology followed in respect of the research questions RQ2, RQ3 and RQ4 (refer to chapter 4).

2.2.3. Research issues

Over the past few decades a number of projects have promoted the importance of discourse in educational environments (for a discussion on the field in the context of first year university physics view chapter 3). Research on these projects has mainly focused on comparing traditional instruction with cooperative learning, in relation to their outcomes on students' conceptual understanding. Only recently has research focused on the actual process of students' discussion and argumentative reasoning in science education. This section reviews relevant literature, in terms of what it is known from prior research studies about the conditions that could promote explaining and arguing in learning environments, the difficulties that students experience in the process of arguing and the evidence on whether and how argumentation skills can be enhanced by the use of appropriate strategies.

In terms of the conditions that could promote arguing and explaining in learning environments, a review of the literature conducted by Duschl & Osborne (2002) emphasized that a major element for engaging learners in such processes is establishing effective contexts and conditions for such discourse to take place. More specifically, they pointed out that at the core of such contexts is the requirement that students should consider not only singular explanations of phenomena, but plural accounts. In other words students' argumentation can be supported when students spend time considering not only the scientific theory, but also an alternative one, such as a common misconception. On the other hand, Eichilnger *et al.* (1991) focused on the need to develop classroom cultures that support the democratic norms of responsibility and tolerance. In that way the role of the teacher appears to be more as a facilitator in a productive dialogic discourse rather than an instructor of the norms of the argumentative procedure. Last but not least, although some of the research on discourse has stressed the importance of establishing procedural guidelines for the students, a review of research on conditions for productive discourse by Cohen (1994) pointed out that both the lack of guidelines and highly structured guidelines can negatively affect the quality of discourse. The above literature provides some helpful insights in the process of constructing learning environments with the appropriate conditions that may support argumentative discourse and was taken into consideration in the development of the unit which is the context of this research.

To provide the appropriate conditions in order to promote explaining and arguing is of major importance, given these practices do not often come naturally to students. Kuhn (1991) investigated children and adults' ability to construct arguments and found that they often had difficulty in coordinating their claims and evidence. In classroom settings, research findings indicate that –even when scientific explanations is an explicit goal – students have difficulties in using the appropriate evidence (Sandoval & Reiser, 2004) and in providing sufficient evidence for their explanatory conclusions (Sandoval & Millwood, 2005). Research also suggests that students confront difficulties when trying to justify why they choose their evidence to support their claims (Bell & Linn, 2000). Similarly, McNeill & Krajcik (2008) found that students had the most difficulty to use scientific principles and theories to justify why the evidence they use supports their conclusion. Zeidler (1997), on the other hand, focused on students' difficulties in constructing arguments. Drawing on a wide research literature relating to science education, they identified five reasons for students' fallacies in argumentation: *a)* problems with validity (students tend to affirm the consequent); *b)* a naïve conception of the structure of arguments (students tend to have a confirmation bias and select evidence accordingly); *c)* the effects of core beliefs on argumentation (arguments which are consistent with students' beliefs are more convincing than those which are counter to their beliefs); *d)* inadequate sample of evidence (students are not sure about what constitutes enough evidence and come to conclusions before enough data is available); *e)* altering the representation of argument and evidence (students tend to make additional assertions about the context of a problem and they introduce in that way bias in the outcome). Finally, Chinn & Brewer (1993) pointed students' reluctance to modify their views as a consequence of evidence contradictory to their previously held beliefs; rather, they chose to ignore the data contrary to their beliefs, to reject it outright, or to exclude it by declaring that it is irrelevant to the field of study.

Although the findings from research literature indicate that students confront difficulties in the processes of explaining and arguing, evidence does exist that these difficulties could be overcome and that students explaining and arguing skills could be fostered by the use of the appropriate strategies. For instance Kuhn *et al*. (1997), in testing the hypothesis that engagement in thinking about a topic enhances the quality of reasoning about the issue, found that dyadic interaction increased significantly the quality of reasoning in both early adolescents and young adults. In educational settings, Osborne *et al.* (2004) - after developing and using in classrooms a set of materials and strategies to support and facilitate argumentation - found that there was improvement in the quality of students' argumentation. However, they found that the change in students' argumentation was not significant; they argued that this finding suggests that developing the skill to explain and argue effectively is a long term process and difficult to be achieved during the limited period of nine months of their intervention. This point is in accordance with the research findings of Zoller *et al.* (2000), who concluded from their work with first year undergraduates that one semester is too short a period to develop higher order thinking skills. In contrast, Zohar & Nemet (2002) found significant improvements after a short period of implemented their intervention, which consisted of integrating explicit teaching of general reasoning patterns into the teaching of human genetics. In particular, in examining the teaching of argumentation skills in the context of human genetics, they reported an increase both in the frequency of students, who referred to correct knowledge in constructing arguments and in the quality of their argumentation. Significant improvements in the quality of students' argumentation were also reported recently Clark & Sampson (2006). This study focused on supporting scientific argumentation in the classroom through a customized online discourse system. The findings suggest that this intervention results in successful levels of argumentation, particularly in light of the scientific context, which Osborne *et al.* (2004) found it to be more challenging for students than socio-scientific contexts.

In short, in recent years the theoretical discussions about the role of argumentation in educational settings has contributed to the development of research projects aiming to explore and foster students' explaining and argumentative abilities. Research findings indicate that students confront difficulties during such processes, such as the difficulty to coordinate scientific evidence to support their views. Research studies have also provided evidence that under the appropriate conditions and by the use of appropriate strategies these difficulties could be overcome. Yet, some studies report that this could be a short term process, while others pinpoint that that developing the skill to explain and argue effectively is difficult to be achieved during the limited period of implementing an intervention. In the view of these controversial findings, further empirical research on the field could be useful, by providing additional evidence on whether significant changes on students' explaining and arguing abilities could be achieved in a short period of time (research questions RQ2, RQ3 and RQ4).

Summary

The notions of explaining and arguing in the context of science education seem to overlap to a considerable extent, given that they both refer to the process of justification of and reasoning for a position. Yet, a review on the literature on explaining and arguing indicates that in terms of scientific argumentation the emphasis is put on the structure of the reasoning process, while scientific explanations mainly refer to the content of why or how a phenomenon occurs. Moreover, as far the notion of an appropriate explanation is concerned, it seems to be twofold: it refers both to whether or not the explanation is in accordance with the scientific theories or facts and to the degree to which the explanatory act is likely to impact knowledge and to promote understanding in the explainee. On the other hand, students' scientific arguments have mainly been analyzed and evaluated by the use of Toulmin's framework. Research on explaining and arguing indicates that these practices do not often come naturally to students; rather, students confront difficulties in providing sufficient evidence for their explanatory conclusions and in constructing arguments. Research studies have also provided evidence that under the appropriate conditions and by the use of appropriate strategies these difficulties could be overcome. Research findings on the degree to which the quality of students' argumentation abilities could be enhanced are controversial: other studies report significant change in students' explaining and arguing abilities after a short period of teaching using argumentative discourse, while others pinpoint that developing the skill to explain and argue effectively is a long term process and difficult to be achieved during a limited period of time. In the light of these controversial findings, further empirical research on the field could be useful (view RQ2, RQ3 and RQ4).

CHAPTER 3

An overview of the interactive introductory physics units

After providing a review of the literature on the main areas of interest of the study, attention is now turned to first year university physics, which is the educational level under research. Literature suggest that for many years the introductory physics courses in universities worldwide remained dominated by the traditional teaching method based on lecturing. It was not until the early eighties that the first initiatives for teaching reform on the field appeared in the literature. Since then and over the last twenty years, there has been a growing research interest in teaching reforms for introductory university physics worldwide. In the light of this research interest, this chapter offers a literature review on the development and the evaluation of interactive physics units for first-year university students. At first, the main factors that have contributed to the introduction of university physics units, alternative to lectures are identified. This section provides, therefore, the background of the changes in the first-year University physics scenery (Section 3.1). Following this, the research literature on the subject is scanned in the light of the major features that characterize the interactive units. The identification of these characteristics leads to the provision of a definition of the interactive teaching approach as opposed to lecturing (Section 3.2). This chapter then offers a review of the evaluation studies of interactive physics units for first-year University students. The focus here lies on presenting what has already been reported on the field, in terms of both the direction of the research methodology and the main research results of each work (Section 3.3). Some issues, which arise from a critical reflection on the evaluation studies mentioned above and which are based on the literature examined in the previous chapter, are finally pinpointed (Section 3.4). This last section highlights some of the limitations of this body of research and justifies in that way the direction taken in this study and the importance of such an investigation.

3.1. The background

From the early eighties till nowadays there has been a growing research interest in curricular reforms for introductory university physics. Whereas the first efforts for reform are found in universities in the U.S.A., throughout the last twenty years the literature has reported a number of interactive introductory physics units in universities all over the word (for example Booth & James, 2001 in the UK; Mills *et al.*, 1999, in Australia; Chang, 2005 in Taiwan). Given that the changes have been implemented in different regions of the world where the educational conditions vary in a considerable degree, it could be expected that there are different reasons in each case that led to the need for change. However, an overview of the research literature on the subject indicates that three factors have mainly contributed to the development of interactive units in introductory physics.

At first, a number of changes that took place in recent years in higher education seem to have played a role in the introduction of curricular reforms in university physics worldwide. These changes can be summarized as an increase in the number and the diversity of the students entering the universities, along with a large number of poor results in students' performance before the introduction of the teaching reform. Chang (2005), for example, reports that before the introduction of innovation '*the rapid increase in university students led to a deterioration in the quality of entrants (…) not only in their academic background but also in their attitudes towards learning*' (p. 408). Similarly, Sharma *et al*. (1999) refer both to '*large numbers of poor results in first semester physics examinations'* and *'the increasing size of universities'* (p.840), in response to which cooperative learning approaches were introduced along with lecturing. No question that it is hard to advocate a cause and effect relationship between the above-mentioned changes in the first year university scenery and the initiatives for teaching reforms. However, it is likely that a link between the efforts for teaching reforms and the increase in the number, the diversity and the quality of the entrants *does* exist: these changes are likely to have sparked off considerations to educators on how best to deal with them, leading them to modify and adjust their teaching methods to the new conditions.

A second factor that seems to be closely related to the development of innovative courses in University physics is the re-consideration of what the aims of higher education studies are and how best they should be achieved. At least for the western university cultures, maybe one of the less debatable goals of higher education curricula as espoused by both educators and employers has long been the development of criticality. Regardless of the unanimity in this desirable goal, though, it is questionable whether students' criticality can best be achieved in courses in which the conventional teaching of lectures is being followed. Research has shown that traditional units in higher education do not seem to promote to the majority of students skills who are closely linked with critical thinking, such as a deep approach to learning (Gow & Kember, 1990). Furthermore, in recent years there have been changes in the needs and the demands of the employers in relation to the skills and the capacities of their employees. Along with the substantial subject knowledge and the development of criticality, universities are asked to provide society with individuals who have developed a range of transferable skills, such as good communication abilities and teamwork (Light & Cox, 2001). The development of transferable skills, though, is a marginalized feature of the traditional teaching method based on lecturing. Both the demand for students acquiring transferable skills and the re-consideration of how best to achieve the development of core skills, such as criticality, contributed to the recognition that new pedagogies are needed to replace or to complement lecture-type teaching.

Last but not least, it can be argued that the introduction of curricular reforms in first year university physics has been influenced by the increasing emphasis that has been placed in recent years on the role of discussion in facilitating learning in science. Even from the early eighties it has been widely accepted that conceptual understanding cannot be achieved merely through the transmission of knowledge, but within learning environments that support discourse and enquiry (Lemke, 1990). As mentioned in the previous chapter, a number of theoretical perspectives have advocated the importance of discussion in relation to learning. According to the Vygotskyan perspective, for example, which mainly sees knowledge as socially derived, discourse is regarded as the foundation of any subsequent learning. In the Piagetian view, which emphasizes the personal construction of knowledge, discussion is seen as a basic step towards the creation of cognitive conflict in the psychological process of equilibration. In the light of these theoretical developments, the teaching methods in introductory physics started to move away from the lecture-type and teacher-centered teaching towards new pedagogies that involve students' interaction and active participation.

In short, an overview of the literature on the curriculum reforms for introductory physics units over the last two decades - with the focus on the background of the changes - indicates that the increase in the size, the diversity and the quality of the university entrants in recent years seems to have sparked off the need for change in the teaching and learning scenery of the first-year university physics. Moreover, the consideration that students' development of transferable skills is a desirable goal of higher education studies has created the need for new pedagogies to complement or replace lecturing. This need has been reinforced by research results suggesting that traditional teaching methods do not necessarily contribute to students' development of core skills, such as criticality. Finally, the theoretical advances in recent years regarding the importance of students' discussion in facilitating content learning in science seem to have influenced the introduction of innovative pedagogies for first-year university physics which move away from the teacher-centered, lecture-type teaching. It has been a mainly a combination of the above-mentioned factors that have contributed to the development of interactive units in first year university-physics, the main features of which are to be identified and presented in the following lines.

3.2. Major features of the interactive programs

In a first view, it seems that the teaching approaches, which have been proposed as alternatives to lecture-type teaching vary in a considerable degree. This variation is evident both in the terminology used to characterize the innovative units and in the specific implementation of each instructional approach. As far as the terminology used is concerned, except from *'interactive approach/pedagogy'* (Hake, 1998; Booth & James, 2001; Buncick & Horgan, 2001; Meltzer & Manivannan, 2002), some of the innovative programs are reported in the literature as *'active/cooperative learning approach'* (Samiullah, 1995; Gautreau & Novemsky, 1997; Laws, 1997; Heller & Hollabaugh, 1992; Mills *et al*., 1999; Sharma, 1999), while one approach is characterized as *'Peer Instruction pedagogy'* (Mazur, 1997). Regardless of the variation in the terminology and the specific implementation of each instructional approach, though, all these teaching programs have more similar features than differences in common. These major features are identified in the following lines and provide the ground on which the interactive teaching approach is to be defined at the end to this session, as opposed to lecture-type teaching.

At first, all these teaching programs are characterized by a shift in the focus as far as the teaching-learning process is concerned. Whereas lectures are teacher-centered and mainly focus on the process of instruction (teaching), in the core of the above-mentioned programs are the students and the learning process. Actually, the shift in emphasis from teaching to learning has been reported as one of the major features of the process of change in undergraduate education over the last decade, at least in the United States. In a well-informed review of the changes that have occurred in higher education in the United States, Seymour (2001) identified the following inferences that characterize the change of the focus of classroom activities from teaching to learning: carefully defining course objectives, refocusing classroom practice upon enhancing student understanding, clarifying student learning goals, engaging students in their own learning and designing assessment tools to see if the learning goals are being met. According to her, the development of units based on interactive-type teaching could be seen as a response to the demand of appropriate methods which could support an educational environment that focuses on learning rather than on teaching.

Moreover, all the interactive teaching programs are characterized by the provision of teaching time for discussion among students, either in the form of pairs (for example Mazur, 1997; Chang, 2005) or in the form of small groups (for example Mills *et al.*, 1999; Booth & James, 2001). The proportion of time dedicated to students' discussions varies: some approaches include frequent periodic interruptions in teachers' lecturing for students' discourse; in others, the vast majority of the teaching time is provided for students' discussions on conceptual questions or demonstrations. Whichever the case, though, students' discussions is a defining feature of the innovative teaching approaches, given that the lecture-type teaching is dominated by the teacher's monologue.

The provision of teaching time for students' discussions should be seen in relation to two other characteristics that are central to the interactive pedagogies: engagement and inclusivity. As Buncick & Horgan (2001) emphasize, in the core of the interactive programs lies the recognition that all students are – or at least should be - actively engaged with the material in the classroom. In other words, in this approach all students - regardless of their different potentials, needs and resources - are involved actively in their learning through continuing dialogue, negotiation and reflection about their understanding. This is in contrast to lecture-type teaching, which is based on the assumption that students absorb physics concepts simply by being told that they are true and in which the classroom is often dominated by the few students who have the best potentials.

Last but not least, a feature that characterizes the vast majority of the interactive units in introductory physics is that they emphasize conceptual understanding rather than mathematical problem solving. Gautreau & Novemsky (1997), for example, state explicitly that students should first develop an understanding of physics concepts without using mathematics. Similarly, Mills *et al*. (1999) have developed a cooperative learning strategy which mainly seeks to enhance students' understanding of key concepts in physics. This is also the case as far as both Mazur's (1997) pedagogy and Booth & James' (2001) instructional approach are concerned. Sharma *et al.* (1999), on the other hand, advocate a mixture of quantitative problems and concept-based questions, with the focus lying in the latter. Although the specific instructional approaches vary from qualitative concept-tests to in-class demonstrations, in the core of all the above-mentioned teaching programs is the development of students' mastery over concepts in physics rather than solving algorithmic problems.

The identification of the major common characteristics of the interactive introductory physics units that was presented above is essential to this study in an effort to provide a definition for them, as opposed to the traditional or conventional units. It should be noted here that an often quoted definition for interactive courses has been proposed by Hake (1998). According to him:

> *'Interactive Engagement (IE) courses are those reported (…) to make substantial use of IE methods. IE methods are those designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on*

(always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors.' (p.65)

Although this is a rather concise definition, from the lines above it becomes evident that the interactive units exhibit some defining characteristics which are additional to the ones reported by Hake (1998). Therefore, in a more detailed definition, the interactive programs could be defined as the ones containing the main characteristics identified in the lines above and summarized in Table 3.1 below.

Table 3.1: The major characteristics of interactive units vs. lecture-type units

Interactive units	***Lecture-type units***
Focus on learning	Focus on teaching
Provision of teaching time for students' discussions	Dominated by teachers' monologue
Students' learning through active engagement with the material and negotiation	Students' learning by absorbing what they are being told or shown that is true
Inclusion of all students in the classroom activities	Classroom often dominated by the students having the best potentials
Emphasis on conceptual understanding	Emphasis on algorithmic problem-solving

3.3. Overview of the evaluation studies of interactive units

Along with introducing the design and the implementation of the interactive teaching programs, many studies in the literature report research results on the evaluation of these units. As exemplified in the following lines of this section, these research studies were oriented towards evaluating the outcomes of the programs in terms of cognitive and/or affective objectives. The majority of the studies report positive results in terms of both academic and affective objectives of these pedagogies. At the same time, though, the literature pinpoints potential challenges and obstacles that the instructors might need to confront when attempting the teaching reforms. The following lines provide an overview of the evaluation studies of the interactive units, focusing on the evaluation methodology and the research results.

One of the most influential and extensive evaluation studies of interactive units has been reported by Hake (1998). This study reports results of a survey of pre/post-test data for 62 introductory physics units, 14 of which were based on lecturing and 48 of which made use of interactive teaching methods. The main objective of the survey was to compare the lecture-type with the interactive-type method of teaching of basic mechanics, in terms of their effectiveness on students' learning. More particularly, Hake (1998) analyzed statistically the pre and post-test data of three standardized tests. Two of these tests have been designed to capture students' conceptual understanding of Newtonian mechanics and the other has been constructed to measure students' abilities in problem-solving. In this study, the performances in conceptual tests of students following interactive programs were found to greatly surpass those of traditional groups. Moreover, the interactive units showed higher averages in students' performance in problem solving. Based on these research results Hake (1998, p.64) concluded: *'The conceptual and problem-solving test results strongly suggest that the classroom use of interactive engagement methods can increase mechanics-course effectiveness well beyond that obtained in traditional practice.'*

Similar to Hake's (1998) research results were the outcomes of an evaluation study reported by Crouch & Mazur (2001). This study was based on data from ten years of implementing their teaching methodology called Peer Instruction (PI). Crouch & Mazur (2001) compared the cognitive outcomes between the lecture-type method of teaching and their teaching program. In this study the researchers measured students' conceptual understanding and their ability to solve quantitative problems using two standardized tests and traditional examination questions. The data analysis showed that the average post-test scores of the conceptual test increased significantly when changing from traditional instruction to the interactive pedagogy. Moreover, the research results suggested that students taught with Peer Instruction outperformed the students taught traditionally on quantitative problem solving, despite the fact that algorithmic problem-solving is de-emphasized in Peer Instruction according to its developers. This result is consistent with the results of two other research studies (Thacker *et al.*, 1994; Gautreau &

Novemsky, 1997) reported a few years before, which assessed the problem solving performance of students in interactive introductory physics units. In short, the results of these studies indicate that the emphasis on conceptual understanding seems to lead to the improvement of the students' problem solving abilities, as well.

Along with the improvement on students' conceptual understanding and problem solving abilities, Crouch & Mazur (2001, p.970) also reported that their results '*indicate increased student mastery on (…) conceptual reasoning (…) upon implementing PI.'* Yet, they failed to provide a well-defined framework or state the criteria under which they evaluated the students' conceptual reasoning. Indeed, they stated that they implemented a free-response conceptual questions test, requiring students to explain their answers along with stating the correct answer. They then stressed that '*the number of students who explained their answers correctly (…) is comparable to the number of students who answered the standardized concept test correctly after discussions and significantly greater than the number who answered the concept test correctly before discussion'* (p. 973). It is noteworthy, though, that the researchers did not provide either the appropriate data or the methodology they analyzed the students' scientific explanations in order to reach this conclusion.

While the Crouch & Mazur's (2001) study reported results based on the implementation of PI by the developers of the pedagogy, Fagen *et al.* (2002) investigated the implementation of PI by instructors other than its developers. In particular, they conducted a web-based survey to collect data on the effectiveness of this interactive teaching method in terms of conceptual understanding. Out of thirty units using PI and implementing the Force Concept Inventory (FCI) standardized test both before and after the courses, twenty-seven (27) units showed an average normalized conceptual gain over 30%. Fagen *et al.* (2002) referred to Hake's survey (1998) to pinpoint the effectiveness of PI as compared to the traditional lecture type teaching, since none of the traditionally taught courses there showed conceptual gains in this range.

Another study of the effectiveness of the interactive teaching methods for large-enrollment classes has been reported by Meltzer & Manivannan (2002). Similar to the above-mentioned reports, the main objective of this study was to assess students' conceptual understanding. In contrast to the other studies, though, which assessed students' understanding of basic mechanics, the subject area of their unit was electromagnetism. The researchers implemented for five academic semesters two standardized tests and some tests containing concept questions and quantitative problems. The analysis of the results revealed high normalized learning gains in the standardized tests, while the students' ability to solve quantitative problems in electromagnetism was found to be maintained or even slightly improved. These results suggest that the effectiveness of the interactive teaching methods on conceptual understanding does not seem to be related with the specific subject area, given that positive outcomes have been reported from courses on both mechanics and electromagnetism.

Along with evaluating students' learning outcomes, Meltzer & Manivannan (2002) reported some results on the students' attitudes towards the unit. Based on end-of-course surveys, the researchers noted that approximately 30% of the students gave the highest ratings on the course evaluations, while less than 10% seemed to despise this method of teaching. A more thorough research, though, concerning the students' attitudes towards interactive teaching methods has been conducted by Mills *et al.* (1999). In this report, both quantitative (close questions) and qualitative (interviews, observations, open questionnaire) methods were used to capture students views on a learning strategy based on students' interactions. This study showed that the interactive method seemed to benefit students' affective learning outcomes and their attitudes towards learning physics. More particularly, some of the positive aspects of the unit that were frequently mentioned by students were the following: high level of enthusiasm, confidence, opportunity to discuss and modify prior beliefs, awareness of the value of understanding concepts and of how they were learning and opportunity to explore implications of concepts.

While most of the above-mentioned research studies took place in the western world, Chang (2005) reported on three years of experience of action research on an interactive-based unit in Taiwan. In line with the majority of previous research studies on the field, the main aim of this work was to compare conventional with innovative teaching methods in terms of students academic performances and attitudes towards the teaching method. The researcher used three standardized tests to examine the students' understanding on mechanics and electromagnetism and their problem-solving abilities. The researcher also used open-ended questions to capture students' attitudes on the strengths and weaknesses of the program. In short, the data analysis in the area of conceptual understanding revealed that the gain percentage of the innovative courses was higher than the traditional ones. However, this gain appeared to be lower than the majority of the conceptual gains reported by studies in the western countries. This result may be due to the cultural differences and the differences in the specific implementation of the programs. More studies in different countries and cultures are needed to elucidate this important point.

Nevertheless, the above study gave a new insight on the field: it mainly focused on *the process* of change from conventional to interactive teaching and on the difficulties that the instructor faced during the implementation of the teaching program. It should be noted here that the consideration of the potential challenges when performing interactive teaching, can be traced back in some of the reports presented before in this section. In Fagen *et al.*'s (2002) study, for example, the instructors who used PI pedagogy stated that some of the difficulties that they faced were the following: spending too much time in preparing appropriate materials and in covering the material in the classroom, confronting their colleagues' skepticism, dealing with some students' resistance to the method and their reluctance to participate in class discussions and managing the classroom when it became too noisy. However, Chang's (2005) study moved one step forward: it reported empirical data which revealed that due to such difficulties the process of change from traditional to interactive teaching was neither linear nor straightforward. Indeed, in the first year that the innovation was introduced, the program did not lead to the learning outcomes anticipated by the researcher, neither in developing students' conceptual understanding nor in enhancing their interest in learning physics. In contrast, the results during the following two years of implementing the program appeared to improve considerably, indicating that '*one year of implementation may be too short to indicate the outcomes of the innovative program, especially if it is a radical departure from the existing teaching style and culture'* (Chang, 2005; p.420).

Regardless of the challenges to be confronted in order for the innovation to be successful, it should be pinpointed here that all of the research studies cited above report positive outcomes as far as the cognitive and affective objectives of the interactive pedagogies are concerned. However, the literature also reports a few cases where the results of the evaluation of interactive units were not satisfactory. Samiullah (1995), for example, investigated the effectiveness of an interactive unit in first-year physics in terms of students' conceptual understanding and their attitudes towards learning physics. In line with the results of other research, his analysis showed that the student–student interaction led to an improvement of students' attitudes towards the course and increased their motivation to learn. Yet, in this study the statistical analysis of the standardized conceptual test failed to reveal a significant difference in the mean scores between the control group (lecture-type teaching) and the experimental group of students (cooperative learning). This is a notable result given that it contrasts the outcomes of the other statistical studies of the effectiveness of the interactive pedagogies as compared to lectures presented above (for example Hake, 1998). However, it should be added, that Samiullah's (1995) study was a small scale one with only a total of 33 participants, while Hake's (1998) research contained a sample of 6542 students and that in statistical analysis larger samples enhance the reliability of the outcomes.

Poor learning outcomes after the implementation of an interactive unit were also reported by Booth & James (2001). In this study, the evaluation of an interactive program was conducted by means of a standardized conceptual test, a questionnaire aimed at establishing a deep learning 'index' for each student and a standardized test on students' preferred learning style. The analysis of the data showed that that there was no significant improvement in either deep learning or conceptual understanding of mechanics as a result of the interactive teaching. In addition to this, the analysis showed that more than half of the class had a preferred learning style that coped well with traditional rather than interactive classes. There is no question that the design of this study does not allow the establishment of a cause and effect relationship between the students preferred learning style and the poor outcomes after the implementation of the interactive pedagogy. Nevertheless, this study draws attention to an aspect which is often neglected when studying the effectiveness of the interactive units: that cooperative learning may not have the expected learning outcomes for all students, given that some of them may have a transmission view of learning and therefore regard cooperative teaching as ineffective in terms of knowledge accumulation.

3.4. Remarks on the evaluation studies of the interactive programs

With the focus being turned on the main research results of the studies presented above, it could be argued that they provide compelling evidence on the efficiency of the interactive pedagogies as opposed to lecturing in introductory physics units. Indeed, the vast majority of the evaluation studies report better outcomes compared to more traditional approaches in students' conceptual understanding and the affective domain. Moreover, the overview of the literature shows that these studies constitute a body of evidence of more than twenty years of research in different regions of the world and that most of them have followed a large scale research design. These are elements that enhance the reliability of the conclusion that interactive methods in introductory physics have been proven more efficient than lecturing. No question that some studies report potential challenges and obstacles that need to be confronted when implementing the interactive teaching approaches. Even in this case, though, the literature provides evidence to support that the barriers can be confronted and that this is a worthwhile endeavor given the positive outcomes after the implementation of the interactive teaching programs.

In the light of the above and from an educational practice perspective, it could be assumed that the evaluation studies of introductory physics units constitute a well-documented area of research. Given the large number of studies reported and the rather high degree of consensus among the various studies in terms of the positive outcomes of the interactive pedagogies, the following question could be reasonably raised: What else do we need to investigate in this area of research, when there is a compelling body of evidence on the efficiency of interactive methods in introductory physics courses as opposed to lecturing? Yet, from an educational research perspective, it can be argued that the evaluation studies reported so far are limited in a number of respects. These limitations are not evident, unless the focus is moved from the research results to the objectives of the evaluation studies and the methodologies followed so far.

First of all, the research studies reported so far have were orientated towards evaluating the outcomes of the interactive programs in terms of conceptual understanding and algorithmic problem solving objectives. In other words, they have mainly addressed aims concerning the students' cognitive development on the science *content* level. This orientation can be attributed to the fact that content understanding objectives have traditionally been the focus of any type of education. However, a crucial question that arises at this point is whether objectives on content level are or should be the main focus of higher education studies. Isn't students' development of core skills (like criticality) and transferable skills (like communication abilities) equally important targets of university education? Actually, as discussed earlier in this chapter (view section 3.1), it has been mainly the demand for students acquiring transferable skills and the re-consideration of how best to achieve the development of core skills, such as criticality, that have partially contributed to the introduction of interactive teaching in universities. Under such a rationale, it seems that evaluation studies of introductory interactive physics units have reported outcomes in a limited area of the spectrum of the potential objectives of the interactive pedagogies.

Nevertheless, even if we consider aims which exclusively concern students' cognitive development on the science content level, the evaluation approaches reported so far seem to be limited in the following respects. On the one hand, from the previous section it became evident that students' content understanding has been mainly assessed by the use of conceptual questions. The focus, therefore, is put on whether student's views about science concepts and principles change after the implementation of the teaching programs. One question that arises here, though, is what content understanding involves. Doesn't it include knowledge *of* science concepts and principles and *about* science content knowledge as well? As discussed in the previous chapter, recently there is an increasing body of literature that advocates the importance of students understanding not only the concepts of science, but also the epistemology of it (for example Driver *et al*., 2000). Moreover, it has been argued that in order for science teaching to address epistemological goals, it is important to establish learning environments, which enable students to engage argumentative discourse that resemble more closely the discourse of scientific communities (view section 2.2.1). If we assume, therefore, that knowledge of science content involves knowledge about science, then it seems that the studies reported so far are limited in terms of focusing on isolated scientific concepts, rather than addressing also the processes that led to the scientific evolutions.

On the other hand, even if we equate cognitive development on the science content level with conceptual understanding, attention should be drawn on how conceptual understanding has been evaluated. As mentioned earlier, contemporary research on the field has assessed students' conceptual understanding by using multiple choice standardized tests. In these tests, the focus lies on students finding the correct answer in conceptual questions, therefore they are suitable for investigating *what* students believe about various physics concepts. The format of these tests, though, does not allow investigating *why* students hold such beliefs on science concepts, in a way that they can justify to themselves or to others. Maybe, implicit to the methodology followed is the assumption that students' beliefs on science topics are reasoned views. Yet, should cognitive skills on scientific explaining be taken for granted? At least this body of the literature does not provide any evidence that this could be the case[1]. In contrast, as exemplified in the previous chapter, research findings suggest that students confront difficulty to use scientific principles and theories to justify why the evidence they use supports their conclusion (view section 2.2.3). Therefore, a limitation that becomes evident at this point is that the studies on the field have drawn attention only to students' views about physics concepts, overlooking that conceptual understanding may also involve students' capacity in conceptual reasoning.

Finally, it should be noted that some of the research studies presented in the previous section have also focused on evaluating the outcomes of the interactive pedagogies in terms of affective objectives. Implicit to this orientation is the acknowledgement of the important role that motivational factors play in conceptual learning. Are affective measures, though, the only factors that may influence learning in the content level? It should be reminded, here, that conceptual understanding in interactive approaches is mainly seen as an act socially driven and as the desirable outcome of students' discursive interactions. In this process, doesn't the

[1] Indeed, as can be viewed in Section 2.3 in this chapter, Crouch & Mazur's (2001) study turned its attention to students' capacity in conceptual reasoning. However, as argued earlier, this study did not provide either the data or the methodology with which they analyzed students' scientific explanations.

students’ ability to participate in a meaningful scientific dialogue play a role to the range of the potential outcomes that discourse may have? Indeed, as mentioned earlier in this study (view section 2.1.3) research findings indicate that students’ explaining and reasoning abilities are factors than may influence students’ conceptual change. Therefore, another limitation of the research studies appears: it is the fact they have not considered factors embedded in the process of dialogic discourse, such as students’ argumentation and explanation skills, which may influence the final outcome of the interactive teaching methods.

Consequently, research has revealed that interactive methods in introductory physics are more efficient as compared to traditional approaches of teaching and learning in terms of conceptual and affective objectives. Yet, a critical reflection on the literature cited in the previous section of this chapter indicates that the evaluation studies on the field seem to be limited in terms of the focus of the research and the methodologies followed. More particularly, in the lines above it has been noted that research studies on the field have mainly addressed aims concerning cognitive development on the science content level. However, they have overlooked aims concerning students’ development of core skills, like criticality, or of transferable skills, like communication abilities. Furthermore, it has noted that students’ cognitive development on the science content level has been seen exclusively as conceptual understanding, rather than including as well understanding of the scientific processes that led to the invention of these concepts. Moreover, it has been pinpointed that conceptual understanding has been evaluated in terms of what students believe in regard of various physics concepts, rather than including as well the students’ ability to provide scientific explanations and arguments. Finally, it has been argued that research studies on the field have considered affective measures as potential influences to cognitive development; however, they have neglected factors embedded in the process of dialogic discourse, which may affect the outcome of the interactive teaching methods, like students’ difficulties to participate to a scientific dialogue.

In the view of the above-mentioned limitations of this body of research literature, this study makes a case of considering areas of research additional to the conceptual and the affective domain, in the context of interactive first year university units. Based on the theoretical and research advances presented in the previous chapter, it has been justified that - while developing and evaluating an interactive introductory physics unit - it is it important to focus not only in students' conceptual understanding, but on students' abilities to explain and to participate in argumentative discourse, as well. Under such a rationale, this study investigates an introductory interactive physics unit in a Greek University in terms of students' conceptual understanding, explaining and arguing abilities. The following chapter provides the way by which the above mentioned objectives were methodologically tackled.

Summary

This chapter offered a literature review on the development and the evaluation studies of interactive introductory physics units. Focusing on the background of the changes in the first-year physics scenery, it was first argued that three factors have contributed to the development of such units, namely a) the change in the size, the diversity and the quality of the entrants, b) the reconsideration of the desirable aims of higher education studies and c) the theoretical and research advances regarding the importance of discussion in facilitating learning in science. Following this, the major characteristics of the interactive units were identified (focus on learning, students' discussions, active engagement, inclusivity, emphasis on conceptual understanding). This chapter then offered a review of the evaluation studies on interactive introductory physics units, focusing on the methodology and the main research results of each work. A critical reflection on the evaluation studies revealed that interactive approaches have been found more efficient than lecturing in regard of conceptual and affective domains. However, it was noted that this body of the literature is limited in the following respects: It has not addressed aims central to university studies, like students' development of criticality; it has identified cognitive development on the science content level with students' views *of* science concepts, rather than also including knowledge *about* science; it has not addressed conceptual reasoning; it has not considered factors embedded in dialogic discourse which may influence content understanding, like students' ability to participate in scientific dialogues. Based on these limitations and the literature review presented in the previous chapter, the direction taken in this research is justified.

CHAPTER 4

Methodology, methods of data collection and analysis

When dealing with methodological issues in this study, two areas of focus are implied: the first concerns the research paradigm that has influenced how the research was formulated and tackled; the second has to do with the specific techniques used in the study, as a means of achieving the research objectives. In relation to the first area, this chapter makes a case for the use of a mixed research methodology, as a framework for designing and conducting research (Section 4.1). The fundamental principles and the key features of the mixed-method research paradigm are described in brief, while the rationale for choosing a mixed-method research paradigm in this study is presented. Following this – and in relation to the second area of focus – Section 4.2 offers an outline of the research design. In this section, the stages of the research are described, the specific procedures followed to investigate the research questions are presented, and the type of the research study is identified. After the brief outline of the design of the study, attention is to be turned to a more detailed examination of the procedures followed in the research. Specifically, Section 4.3 presents the pilot work done for this study. Then, this chapter presents the main part of the research (Section 4.4). At first, this section provides an insight to the setting and participants (Section 4.4.1); then, it discusses the methods of data collection (Section 4.4.2); finally, it describes the procedures of data analysis in relation to each one of the research questions (Section 4.4.2). A discussion about the rigor of the study follows in Section 4.5, while some comments about ethical concerns and procedures followed in this research finish the chapter (Section 4.6).

4.1. Mixed-method research paradigm

Traditionally, research studies in education have been orientated towards either the quantitative or the qualitative research paradigm. Behind this polarization lies, explicitly or implicitly, the incompatibility thesis, according to which these two research paradigms cannot or should not be mixed (Howe, 1988). Recently, though, the mixed- methods research has emerged as an alternative research paradigm, aiming to bridge the gap between qualitative and quantitative research. Mixed-methods research could be defined as '*the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study'* (Johnson & Onwegbuzie, 2004).

Philosophically, the mixed-methods research paradigm is embedded on pragmatism, which has attempted to find a middle ground between philosophical dogmatisms. Indeed, pragmatism itself rejects traditional dualisms (for example objectivism versus subjectivism) and endorses pluralism as a way to gain an understanding of the world from different perspectives (Johnson & Onwuegbuzie, 2004). For example, in pragmatist epistemology, knowledge is viewed as being both constructed and based on the reality we experience. From an ontological perspective, on the other hand, each notion should be interpreted by tracing its respective practical consequences (Murphy, 1990). For research methodology, this last remark could be interpreted as putting the appropriateness of the methods as a point of reference (Flick, 2002). In short, the pragmatist philosophy is not overly concerned with longstanding philosophical arguments; it rather makes a case about doing what works in a given situation.

The mixed-methods research paradigm adheres to pragmatism as a philosophy, given that it is an attempt to legitimate the use of multiple perspectives, theories and research methods in answering research questions. According to Johnson & Turner (2003), the fundamental principle of this paradigm is that multiple approaches, strategies and methods are *combined*, in a way that the resulting combination is likely to result in complementary strengths and non-overlapping weaknesses. In other words, the researcher can use one method to overcome the weaknesses of another method, without being constrained by qualitative or quantitative research orientations. In mixed-methods research, the objectives and questions are fundamental, while the research methods should follow in a way that offers the best opportunities for answering the questions.

This study was conducted within the mixed-method research paradigm. This orientation can be attributed to the fact that the philosophy that underpins this paradigm is close to the researcher's ontological and epistemological beliefs. Indeed, the researcher's background in both physics and educational studies has contributed to her preference for a more moderate version of philosophical dualism, based on how well they work in solving problems. This philosophical orientation was evident previously in this study, when discussing the attributes of constructivist theory (view Section 2.1.2). It has been argued, there, that although this study is based on constructivist assumptions about learning, it rejects some of its epistemological assumptions. Rather, it is closer to the pragmatic stance that knowledge is both constructed and based on experience. The philosophical pragmatism orientation of the researcher, therefore, has played a role in choosing a mixed-methods framework for conducting this research.

Apart from philosophical reasons, the nature of the research problem for this study has determined the research methodology. Traditionally, as exemplified in Chapter 3, introductory physics units have been investigated mainly using a quantitative research approach. However, these studies have focused on investigating students' conceptual understanding by comparing traditional with interactive teaching methods. In contrast, this study investigates the areas of conceptual understanding, scientific explanations and argumentation, and the links among them, while it does not aim to compare traditional with interactive instruction or to establish cause and effect relationship. Under such a rationale, it is believed that a quantitative research paradigm is not an appropriate reference for conducting this research.

Given the objectives of this research, the mixed-methods research paradigm provided an appropriate framework for designing this research. Some of the strengths of this paradigm that have found an application in this study are the following (Johnson & Christensen, 2004): Under such a paradigm words have been used to add meaning to numbers, while numbers added precision to words. Moreover, the strengths of one method have been used to overcome the weaknesses of another. Furthermore, mixed-methods research is believed to provide stronger evidence for the conclusion, through convergence of the findings (triangulation). Apart from the strengths, some of the weaknesses of this paradigm were taken into consideration when designing and conducting the research: that it is time consuming; it requires the researcher to understand how to mix the methods appropriately; that there are difficulties on how to interpret conflicting results.

In summary, both the philosophical orientation of the research and the nature of the research problem for this study have determined a mixed-method research methodology, as the appropriate framework for conducting the research. Under the influence of this theoretical framework, the appropriate techniques to investigate the research objectives were chosen and are presented below.

4.2. Overview of the research design

As mentioned earlier in this study (view Section 1.4), the main purpose of this to investigate the interactive first-year university physics unit, which was described in Section 1.3, in terms of students' conceptual understanding, scientific explanations and scientific arguments. To achieve this aim, the study was conducted under the mixed-method research paradigm - as described in the previous section of this chapter- by means of constructing a mixed research design. The research design is illustrated in Figure 4.1, which summarizes the methods of the data collection and the methodology under which these data were analyzed for each one of the research questions.

The research was developed and implemented in several stages (see Figure 4.2). A more detailed description of the procedures followed within each stage is presented in the following two sections of this chapter. As it can be viewed in Figure 4.2, the first phase of the research took place in the first semester of the academic year 2004-05, as a pilot work of this study. The main aim of this part of the research was to gain an insight to the class environment and the teaching-leaning procedures and in that way to conceptualize and frame the aims of this research. Based on the results of this phase of the research, the unit was redesigned taking the form that was described in Section 1.3.3 of the thesis, while the objectives of this research were formed and the methods of data collection were chosen. The second phase of the research was conducted in the first semester of the academic year 2005-06, in the context of the interactive

Figure 4.1: The research design of this study

RESEARCH QUESTIONS (OBJECTIVES)		METHODS OF DATA COLLECTION	METHODOLOGY OF DATA ANALYSIS
RQ1:	What is the change (if any) in students' conceptual understanding on basic mechanics after participating in the course?	Multiple choice test	Quantitative analysis
		Observations	Qualitative analysis
RQ2:	What is the change (if any) in students' ability to provide correct and concise scientific explanations on basic mechanics after participating in the course?	Open-ended questionnaire	Quantization of qualitative data
		Observations	Qualitative analysis
RQ3:	How do students respond to weak or fallacious arguments? What is the change (if any) in the way students respond to such arguments after participating in the course?	Open-ended questionnaire	Quantization of qualitative data
		Observations	Qualitative analysis
RQ4:	What is the change (if any) in the quality of students' written arguments after participating in the course?	Open-ended questionnaire	Quantization of qualitative data
		Observations	Qualitative analysis

Figure 4.2: The outline of the stages of the research

MAIN STAGES OF THE RESEARCH		RESEARCH INTERESTS
When	**What**	
2004-05	Pilot work	Gain an insight to the teaching-learning environment- Recommend improvements- Conceptualize and frame the research
2005-06	1st year of the research	Achieve the research objectives (view Figure 4.1)
2006-07	2nd year of the research	Achieve the research objectives (view Figure 4.1) Compare to the 1st year of study

unit (as it had been redeveloped) and it aimed to investigate the research objectives that were presented in Figure 4.1. Finally, in the first semester of the academic year 2006-07 the third phase of the research took place. This phase was almost identical to the second one, in terms of the research objectives and procedures followed, but with different participants (the new students who entered the university in this academic year). The decision to repeat the main research was made for two reasons: First, it has been acknowledged that some of the teaching and learning strategies used were novel for the instructor; therefore, it might need more than one semester of implementation to achieve its potentials. On the other hand, a two-year design of the main study was adopted, so as to gain a more insightful description and interpretation of the implemented innovation, within two different sets of participants. It is believed that the comparison between the results of the two years, may contribute to the rigor of this study.

From above, it becomes evident that this investigation could be characterized as an evaluative research study. Indeed, the first stage of the research, which was done in the first year of study, could fall into the category of 'formative evaluation research'. In this part of the research the focus lay on identifying the strengths and weaknesses of the unit under investigation and it served the purpose of recommending some improvements to the intervention. In this part of the research the researcher had an active role in the development and the redesigned of the unit; therefore, in terms of the researcher's role, this stage was a participatory evaluative research. As far as the second main stage of the study is concerned (second and third year of study), the research took mainly the form of a 'summative evaluation', given that it aimed to determine the effectiveness of the intervention, as it had been redeveloped. This part of the evaluation was goals-based, in the sense that it aimed to investigate the extent to which the unit had attained its objectives. In this part of the research, the researcher tried to obtain an 'outsider's view' of the project, as far as the data collection and the data analysis is concerned, and to minimize the degree of her engagement to the settings. In this respect, the research was a non-participatory evaluation. In short, the researcher had the role of a 'researcher participant', as she participated in a social situation, but she was personally only partially involved so as to function as a researcher.

Figure 4.3: The type of the research and research features

	Type of the research / Characteristics	
	In terms of purpose	***In terms of the researchers' role***
1st stage (pilot work)	Formative evaluation	Participatory evaluation (design and observations)
2nd stage (main research)	Summative evaluation	Non-participatory (in the data collection and analysis)

In short (view Figure 4.3), this research study is an evaluation study; it is formative/participatory in the first stage and summative/non participatory in the second stage; it was conducted by the use of a mixed-methods design; it was developed and implemented in two main stages. The following two sections provide a detailed description of the stages of the research and the procedures followed, starting with the pilot work done for this study.

4.3. The pilot work of the study

4.3.1. The background and the aims

The pilot work of this study took place in the academic year 2004-05, in the context of a first year physics interactive unit, based on Mazur's (1997) pedagogy. The area of physics covered in the unit was introductory classical mechanics. The main aim of the unit was to enhance students' conceptual understanding in introductory mechanics and to contribute to the improvement of students' communication abilities. The lessons were conducted on the first semester of the year, twice a week (four hours per week). The teaching method was based on student-to-student- interaction. In particular, the students were given a number of multiple choice questions containing, along with the scientifically correct answer, alternative ones derived from literature reports about students' misconceptions. The students were given a few minutes to discuss the questions with their peers and then they voted the answers they believed to be the correct ones, using electronic devices. The percentages of each answer were shown on a screen and then the instructor was asking students to support their views and explain their answers to their peers. In order for the instructor to help students in the process of explaining, he had introduced to them 'Bobos', a fictional high school student. Students were asked to form their explanations in such a way, so as to promote Bobos' understanding. The assessment included two examinations, at the middle and at the end of the semester, containing open-ended questions, in which students were asked to answer a number of conceptual questions and to explain their answers.

In such an educational context, a formative participatory evaluation study was conducted, in order to investigate the strengths and weaknesses of this teaching program and to recommend improvements to the intervention. The main role of this part of the investigation in the overall study was to help the researcher gain an insight to the class environment and conceptualize and frame the research investigation.

4.3.2. The participants

In Greek Universities attendance is not compulsory; therefore, not all of the students, who were enrolled, were present in each lesson. Thirty nine students attended the majority of the sessions, as they participated in more than half of the lessons and they constitute the research population in the pilot study.

4.3.3. The methods of data collection

The following data sources were used in this part of the study, in order to conduct a formative evaluation of the unit: the Force Concept Inventory (FCI)[1]- a standardized conceptual test - both at the beginning and the end of the course (Hestenes *et al*., 1992); students' exam papers; unstructured and semi-structured participant observations of the lessons and field notes; informal interviews with the instructor. In particular, the FCI was implemented in the beginning of the second week of the unit and at the last week of it, aiming to investigate students' conceptual understanding of basic mechanics after participating in the unit. Out of the thirty-nine students, a sample of thirty one students completed both the pre- and the post- test. For these students, two open-ended questions were selected from the exam papers, and they were examined qualitatively, in terms of how the students explained their answers in conceptual scientific questions. Moreover, important insight to the class environment was gained through observations of all the lessons: at the beginning of the unit, the observations were unstructured,

[1] View Appendix A

so as to help determine the subsequent patterns of observations (students' abilities to participate in a meaningful dialogue, students' explaining abilities, and their attitudes towards the intervention). The field notes that were taken during the lessons were typed, right after the completion of each lesson. During this semester, a number of informal interviews were conducted with the instructor, in order to exchange views about the process of the lessons. In an effort to make the instructor feel more comfortable and given that these interviews were informal, they were not recorded; in contrast the researcher took notes during them, which were typed right after the end of these meetings.

4.3.4. Outcomes of the pilot work

As far as Hestenes *et al*'s (1992) standardized test is concerned, the data was analyzed by means of the gain percentage, as proposed by Hake (1998). The gain percentage is defined as <g> = (% <post-test> – % <pre-test>) / (100- % <pre-test>) and it indicates the ratio to what has been gained to what remains unlearned. The average score was 63.3% in the pre-test and 79.2% in the post-test and the <g> gain percentage calculated as 0,43. According to Hestenes *et al.* (1992), a score of 60% is regarded as being the 'entry threshold' to Newtonian physics and below that limit students' grasp of Newtonian concepts is considered insufficient, while a score of 85% is considered to be the Newtonian 'mastery threshold'. Moreover, on the basis of the data Hake (1998) gathered from a survey of 14 lecture type courses and 48 interactive engagement ones, he classified introductory mechanics courses in three regions: a) high-g courses (<g> > 0.7), b) medium-g courses (0.7> <g> > 0.3) and c) low-g courses (<g> < 0.3). No unit from this study reached the high-g region, as the best was 0.69. Most of the innovative teaching programs based on students' discussions, which were mentioned earlier in this thesis, used Hake's classification in assessing the outcomes of students' academic achievements and reported that they fall in the medium-g courses category. The results on FCI indicated an increase in students' conceptual understanding after participating in the unit. However, the observations made during the lessons indicated that, although students were in many cases able to find the correct answers in the conceptual questions, they confronted difficulties in the process of explaining their answers to their peers. The observations suggested that students' difficulties could not entirely be contributed to their conceptual ignorance. In contrast, they suggested that some students seemed to be unaware of the facts that an explanation requires more than providing a set of algorithms in mathematical problem solving and that an appropriate explanation cannot be seen independently of the context in which the question is being asked. The following exacts come from the field notes and could be used to exemplify the points made above:

> *Many students -when asked by the instructor to explain their answers to their peer in conceptual questions - try to formulate the question in mathematical terms and to provide the correct set of algorithmic so as to come to a result. It seems that for them a scientific explanation is synonymous to mathematical problem solving.*
>
> *Some students -when encouraged by the instructor to explain their answers, in such a way so as Bobos could understand - they provided the facts of the question and then the result, without referring to the intermediate process/thinking which led them to this conclusion.*
>
> *Students rely mostly on the instructor to provide them the correct explanation, so as to note it down. When a student provided a correct explanation in a conceptual question, the instructor asked the others to comment on whether they agree or disagree with it and why. The class seemed reluctant to evaluate their classmates' explanation and asked the instructor to comment on whether it is correct or not.*
>
> *Some students seem to evaluate their peer's explanation, mostly in terms of whether he/she has reached a conclusion with which they agree, rather by the facts and the thinking processes.*

Students' difficulties in providing scientific explanations were also identified in their exam papers, which were of the form of open-ended conceptual questions, requiring explanation. At first, students' grades in the exams could be used as an indicator for this difficulty: For the sample of thirty one students who participated in more than half the lectures and completed both the pre and post test, the average grade was 59% in the first exam at the middle of the semester and 47% in the second exam at end of it. The instructor, in an informal interview that was conducted, attributed his low grading to the fact that many students, despite giving the correct answer to the conceptual questions, provided incorrect explanations. The qualitative examination of two questions of the final exam for this sample of students indicated that - apart from the above-mentioned observation made by of the instructor- some of the students' explanations were incomplete, as they lacked the appropriate justification to back up the conclusion they had reached.

Moreover, the observations made during the lessons indicated that some students confronted difficulties in the process of the discourse. For some students being involved in a discourse seemed to be equal to practicing parallel monologues; others exhibited mere inability to perceive logical defects in their peers' arguments or to see that logically defective statements cannot be valid arguments. The following extracts from the researcher's field notes exemplify some of the points made above:

> *During many debates, it seems that the students just participate in parallel monologues. They respond only to the conclusion of their classmate, either just by stating their conclusion, or by giving their own argument. In that way the dialogue comes to an end, without any actual effect in either of the students.*
>
> *A student stated that she disagrees with the argument of her classmate. The instructor asked her why, and where he had made the mistake. She stated that he must be wrong because he had reached a different conclusion. No reference was made about where the fallacy was.*
>
> *Students' entering debates keep on judging their classmates arguments only by means of the conclusion. The dialogues do not seem to be productive.*

The outcomes of the pilot work that were presented above had effects in two levels that concern this study: On the one hand, they led to the modification the redevelopment of the unit. After negotiations with the instructor, the aims of the unit were reformed, new materials were developed to help the students in the arguing and explaining process and the unit took the form described in Section 1.3. On the other hand, the outcomes of the pilot work, contributed to the design of the main research. During the second semester of this academic year, the research was conceptualized, taking the form outlined in section 4.2. At the same time, the methods of data collection were chosen and developed. The following section provides a detailed description of the procedures followed within this stage of the research.

4.4. The main stage of the research

4.4.1. The participants and their background

Both in the first and the second year of study, the total number of students enrolled in the unit exceeded eighty, the vast majority of them being first-year students. In the first year of the research (2005-06), an average of forty students was present during the course, while thirty-five students participated in more than half of the lessons. During the second year of the research (2006-07), an average of forty five students participated in each lesson and forty students participated in more than half of the lessons. Similar to the pilot stage of the research, for the purpose of this research the students who participated in more than half of the lessons constitute the population of the study – i.e. 35 in 2005-06; 40 in year 2006-07.

As far as the background of the first-year students is concerned, it should be noted that they came from a higher secondary environment, which was under reform at the time. As it has been discussed in Section 1.2, in higher secondary education, the main focus was on the acquisition of scientific knowledge and on resolving problems using mathematics, regardless of the learning objectives presented in the ideal curriculum. Students were taught mechanics in the first grade of upper secondary school, while in the third grade they were taught the areas of periodic motions, waves, oscillations and rigid body kinematics. The participants of this study entered the university after achieving a minimum grade of 15 out of 20 in the national exams at the end of upper high school. Moreover, apart from being enrolled in the interactive unit, all of the students took the compulsory course 'Physics I.', which offered an intensive revision of the upper- secondary school physics syllabus, in a more advanced mathematical level. The course 'Physics I' was based on lecturing and students are assessed by means of physics problems, which require mathematical problem solving skills.

4.4.2. The methods of data collection

The following methods of data collection were used in this part of the study, in order to achieve the research objectives: the Force Concept Inventory (FCI) (Hestenes *et al*., 1992); an open ended questionnaire about explaining and arguing; students' exam papers; unstructured and semi-structured participant observations of the lessons and field notes; students' evaluation sheets; semi-structured interviews with the participants

The Force Concept Inventory (FCI)[2]: The FCI is a widely used multiple choice test designed to measure students' understanding of force and related kinematics concepts. It is given as a

[2] view Appendix A

pre-test at the beginning of a course and then as a post-test at the end of it and it has been used by many instructors and researchers as a diagnostic tool and as a means for evaluating instruction. The FCI is based on the Mechanics Diagnostic Test (MDT), which was designed around common misconceptions on the concept of force. The reliability and the validity of the MDT were established by interviews and statistical analysis[3]. About half of the FCI questions are essentially the same as those in the MDT, while a revised version of the FCI was developed later, involving mainly the clarification of some ambiguities (Mazur, 1997).

It should be noted that the developers of the FCI did not repeat the lengthy procedures to establish test reliability and validity for the FCI. This has caused some criticism in terms of the reliability and the validity of the test (for example Dancy, 2000 - quoted in Savinainen & Scott, 2002). Savinainen & Scott (2002), though, support that the reliability of the versions of the FCI has been well established through extensive use of the test - for example Hake's (1998) large survey-, while face and content validity have been established through the support of the numerous physics instructors who have used the test. Another criticism of the FCI has been made by Huffman & Heller (1995): After conducting a factor analysis (based on groups of questions proposed by the developers of the test[4]), they concluded that this test can be used as a means for evaluating instruction, but it does not measure the coherence of students' understanding of concepts. Recently, though, another study provided evidence that the FCI can be used to evaluate students' conceptual coherence, especially the contextual coherence, which is the ability to apply a concept across a variety of contexts (Savinainen & Viiri, 2008). Another concern raised by Huffman & Heller (1995) -after the factor analysis they conducted- is that the FCI should not be decomposed into the six dimensions originally proposed by its developers; they advised caution in analyzing any of the six conceptual dimensions separately. Hestenes & Halloun (1995) have responded to this argument, by emphasizing that the FCI should be administered and interpreted as a whole.

---

[3] The Kuder-Richardson reliability coefficients were determined to be 0.86 for pre-test use and 0.89 for post-test use, which are indicative of a highly reliable test.

[4] The groups of the questions, which are proposed by the authors and are called 'conceptual dimensions' are the following six: kinematics, Newton's first law, Newton's second law, Newton's third law, superposition principle and kinds of forces.

For the purpose of this research the FCI was used as a means for evaluating the intervention, in terms of students' conceptual understanding. It was implemented in the beginning of the second week of the unit and at the last week of it, both in first and the second year of the research. The last version of the test was chosen (Mazur, 1997) with thirty multiple choice items, given that it has fewer ambiguities and it is shown to be relatively free from the tendency towards false positives (correct answers for incorrect reasons) (Savinainen & Scott, 2002). The test was translated by another researcher from English to Greek; then the researcher translated it back into English. A comparison between this translation and the original test was conducted so as to correct any ambiguities. The test was used and was interpreted as a whole, following *Huffman & Heller's (1995)* concerns, which were discussed above.

Open-ended questionnaire on explaining and arguing[5]: In order to investigate students' explaining and arguing abilities, a test containing a number of open-ended items was developed by the researcher. The questionnaire was implemented at the beginning and at the end of the unit both in the first and the second year of the research. The test is divided in two parts: In Part A of the test the students were given a statement made from one of their classmates in the area of mechanics; they were asked to explain to their classmate why they agree or disagree with this statement. This part of the test contains six items. The items 1-5 derived from the FCI, while the item 6 was constructed for the purposes of this research. The following criteria were used in the selection of the items of this part of the test:

a) *Fitness to the purpose of the research:* Some items of the FCI were used, given that they have been selected from the vast literature on students' common alternative perceptions. Given that the main aim of this part of the test was to investigate students' scientific explanations, from the thirty items of the FCI, the ones which require design of forces were excluded.
b) *Difficulty of the items:* The analysis of the FCI given to the student when the pilot work took place (2004-05) revealed that items 2 and 4 were the ones in which students confronted most difficulty, both in the pre- and the post-test. As for the items 1, 3 and 5, although the majority of the students gave an incorrect answer in the pre-test, most of them gave the correct answer in the post-test; therefore, they

[5] View appendixes B and C, for the 1st and the 2nd part of the questionnaire accordingly

must be considered easier for the students after instruction. The item 6, derived from the exam papers of this year, and it is a question in which the majority of the students confronted difficulty in answering and explaining correctly.

c) *The effect of the content on students' explanations:* As it has been pinpointed in other studies (for example Osborne *et al.*, 2004), the lack of knowledge of any relevant theory or concept often constrains students' ability to reason effectively. In an effort to control the effect of the content on the individual performance to explain, the items which were chosen were from different conceptual dimensions in mechanics.

The second part of the questionnaire contains four items (items 7-10). In this part, the students were given an argument made by one of their classmates; they were asked to comment on whether they are persuaded by it or not, and to explain the reasons for their choice; in other words they were asked to provide their counter-arguments. The items were constructed under the following rationale:

Item 7: The item derived from a past- exam paper and it was an argument given by a former student. It aims to test whether students accept a conclusion, when it is not a necessary consequence of the premises, regardless of the truth or falsity of the content of these premises. This is one of the fallacies common to argumentation (Zeidler, 1997).

Item 8: As it has been observed during the pilot work of this study, many students seem to be unable to split an argument to its premises and decide on the scientific correctness of each one of them. As a consequence, they appear agnostic about the validity of their classmates' conclusions. This argument derived from a past-exam paper. The student reaches the wrong conclusion, because she claims that N = W, despite the fact that the man on the scales accelerates. In order to control the fact that students may not be able to find the fallacy to the argument because of ignorance of the relevant theory, this item should be interpreted in relation to the item 6, in which students are explicitly asked on whether N = W in this situation. By comparing the data of the two questions, it would be possible to investigate whether students are able to identify an incorrect premise in an argument, which leads to a wrong conclusion.

Item 9: This item contains a fallacy common to argumentation (Zeidler, 1997), which is to affirm the consequence in deductive arguments. The content of the item is not related to physics, in an effort to exclude the possibility of students deciding to be agnostic, because of lack of any relevant theory in physics.

Item 10: In many studies, authority has been pinpointed as a crucial factor affecting argumentative discourse. This item is used to investigate the degree in which students are affected in their judgment by the authority of the person forming an argument.

Semi-structured observations and field notes: Observations of all the lessons took place both in the first and the second year of the research. Following the outcomes of the pilot work, the patterns of the observations were the following: students' ability to find the correct answers to conceptual questions, students' scientific explanations (correctness and their quality), students' abilities to participate in a meaningful dialogue (difficulties and fallacies to argumentation), students' attitudes towards the interventions (students' comments during the lessons). The field notes that were taken during the lessons were typed right after the end of each session. During the observations the researcher neither participated in the students' dialogues, nor did she provide them the correct answer to the conceptual questions or correct explanations. Especially after the second week of the unit, students seemed to be become comfortable with the presence of the researcher as an observer and not be distracted by it.

4.4.3. The procedures of the data analysis

In relation to each one of the research questions and for the two years of study, the data collected by the sources described above was analyzed in the following ways:

RQI: *What is the change (if any) in students' conceptual understanding on basic mechanics after participating in the course?*

Data deriving from the FCI were used and analyzed by means of the gain percentage, as proposed by Hake (1998). The gain percentage is defined as <g> = (% <post-test> – % <pre-test>) / (100- % <pre-test>) and it indicates the ratio to what has been gained to what remains unlearned. A statistical was also implemented to investigate whether there is a significance difference between the pre- and post- test achievement of the students. The field notes from the structured observations were used to interpret the results.

RQ2: *What is the change (if any) in students' ability to provide correct and concise scientific explanations on basic mechanics after participating in the course?*

The first part of the open-ended questionnaire on explaining and arguing was used (questions 1-6). Students' pre and post-test answers in each one of the six questions of Part A of the test were categorized both in terms of the content and the quality.

As far as the classification in terms of the content is concerned, typical responses were identified and placed in a hierarchy in terms of scientific reasoning content. Three broad categories were formed for each one of the six questions: *a)* appropriate explanations (which include scientifically accepted ideas about phenomena), *b)* inappropriate explanations (including students' alternative ideas) and *c)* no explanation (including cases where no answer was given or where the response was not referring to the question asked). Then, for each student, his or her pre- and post-test response to each question was identified in terms of the category in which they fell and the overall numbers of the students' responses in each category were calculated. Comparing the pre and post students' performance on the field, the change in students' ability to provide correct and concise scientific explanations was identified, while the field notes were used to interpret the results.

The students' responses were also classified in terms of the quality of their explanations. The focus here was turned on the scientific appropriate explanations and more specifically on how complete their explanations are. In other words, quality in students' explanations was seen as the degree to which the information provided by the student is enough to back up his or her conclusion. Coding categories were formed for each one of the questions. The properties on each category varied in different questions in relation to their different content. However, for all questions Category 3 stands for the most complete explanation, Category 2 refers to an explanation where the student needed to provide explicitly more information to match with the conclusion reached, while Category 1 stands for scientifically correct response, where only the correct conclusion is provided by the student. For each one of the questions, the overall number of students' explanations in each category was identified both in pre- and post-test and the differences in the numbers of the students in each category were calculated (posttest – pretest). The field notes from the observations were used to interpret the results.

RQ3**:** *How do students respond to weak or fallacious arguments? What is the change (if any) in the way students respond to such arguments after participating in the course?*

The second part of the open-ended questionnaire on explaining and arguing was used (items 7-10) in order to identify students' ability to recognize fallacies common to argumentation. For each one of the items, students' responses were analyzed qualitatively and typical patterns on responses were identified. The primary aim of the analysis has been to investigate whether students manages to identify the weakness or the fallacy in each argument. The overall numbers of the students' responses in each category were calculated both for the pre- and the post-test and the numbers were compared.

RQ4**:** *What is the change (if any) in the quality of students' written arguments after participating in the course?*

In order to investigate the quality of students' arguments, attention was turned to items 7, 8 and 9 which were analyzed under the following rationale: Research evidence suggests that the skill of argument is linked to an understanding of how to rebut another's point of view (for example Kuhn, 1991). Therefore, the quality of an argument could be judged on the presence or absence of rebuttals. Under such a rationale, students' arguments were classified in three broad categories: low-level arguments, middle-level arguments and high-level arguments. When an opposition consisted of only counter arguments that were a simple claim versus a counter claim, it was perceived to be a low-level argument. When an argument consisted of claims with data or backings but with no rebuttals, it was regarded as a middle-level argument. On the other hand, when rebuttals were evident and were in direct reference to a piece of evidence (data, warrant or backing) offered - thereby engaging with the present argument - it was considered a higher level of argumentation. This theoretical framework for analysing scientific arguments in terms of the quality is similar to Erduran *et al.*'s (2004) methodological approach. In Erduran *et al.*'s (2004) study, though, this framework was used to analyze interactive discourse in small groups and five levels on the quality of argumentation were formed, given that data from interactive discourse could be richer (more than one rebuttals could be possible).

4.5. The rigor of the study

Following the description of the research design, attention is turned to the rigor of this study, which mainly concerns the degree to which the procedures of data collection, analysis and interpretation inform the research questions. In any type of research, while discussing the rigor of the study, issues about validity and reliability are prominent. Cohen *et al.* (2000), suggest that these terms can be applied in both the qualitative and quantitative research, but the way by which they are addressed in the two approaches varies. This study was conducted under the mixed research paradigm; therefore, a combination of the ways by which validity and reliability are addressed in qualitative and quantitative paradigms is used, in order to establish the trustworthiness of the research.

In terms of the research design, threats to validity are minimized in this study by selecting the appropriate methodology to answer the research questions, without being constrained by quantitative or qualitative research orientations. In the mixed methods design that has been used, the research questions are fundamental, while the methods follow in a way that offers the best opportunities for answering the questions. This is believed to introduce more rigor into the study, in comparison to mono-method investigations (Johnson & Christensen, 2004). Moreover, the credibility of the findings and the interpretations is established through data and methods triangulation, the prolonged engagement of the researcher to the field and persistent observation (Cohen *et al.*, 2000). Finally, reliability as a measure of consistency over similar samples is tried to be established by conducting the research in two years, for two different sets of participants having similar educational backgrounds. It should be noted, though, that the findings of this study do not claim to be universal or be generalized to the wider population; rather they should be interpreted as dependable on the educational and cultural context, which has been discussed earlier in this study (view Section 1.3).

In the stage of the data gathering, a standardized test was used (FCI), for which reliability has been established through extensive use by other researchers. For this test face and content validity have been established through the support of the numerous physics instructors, who have used the test (Savinainen & Scott, 2002). To minimize the threats to validity from ambiguities in the translation from English to Greek, peer debriefing took place, as mentioned before. As for the open questionnaire that has been used, the clarity of the items, instructions and layout was checked by peer debriefing during the pilot stage of this research. The criteria and the rationale under which the items were selected have been discussed in detail before, in an effort to demonstrate that it covers comprehensively the domain that it purports to cover (establishment of content validity).

As far as the observations and the fieldwork are concerned in the stage of the data gathering, an issue that might have affected the rigor of this study is the extent to which the researcher has affected what has been observed by her presence. It should be noted, though, that especially after the first week of the unit, the students started to be confident with the presence of the researcher and not being distracted by it. In addition, although the participants were informed about the general aims of the research, they were not aware of the patterns of the observation; this increases the confidence that the participants did not respond to the evaluation conditions and they did not regulate their behavior from feedback gained from the observer. Another issue that concerns validity during fieldwork is the personal bias of the researcher, especially in the view of her double role as a participant in one stage of the research and an observer in another stage. Any possible effort has been made by the researcher to function separately in these two roles; however, it should be noted that this is a marginal position and difficult to sustain. The prolonged engagement of the researcher to the field, though, gives confidence that the researcher –like any other data collection instrument- has been refined, through training, to be attentive and responsive to data gathering through this method.

Finally, in the stage of the data analysis, triangulation methods have been used, in a way that the weaknesses of one method are minimized by the use of another method (for example standardized questionnaire and field notes for RQ1). Methods triangulation is believed to establish credibility in the research (Lincoln & Guba, 1985). The credibility criterion has also been addressed in the stage of data analysis through negative case analysis, which involves the refining of the findings until all known cases are accounted for. In the analysis of the open questionnaire, data was analyzed inductively, rather than using a priori categories, while during the analysis regular attempts were made to find data that refuted preliminary assertions. In Chapters 7 and 8, which provide the results of the data analysis in the area of scientific argument, the attempts to address the credibility criterion will be exemplified.

4.6. Ethical concerns

The American Education Research Association standards (AERA) have guided this research study, as far as ethical issues are concerned. Students' participation in the study was voluntary in all stages of the research. In the first day of the unit each year, the researcher described to the students the features and the aims of the study. The prospective participants of the study were adults; therefore consent was obtained by them and not by their parents or their legal guardians. Given the large number of students who were enrolled in the unit, for the use of the data from the FCI and the open-ended questionnaire passive consent was asked. Before the implementation of the tests (both pre and post), students were reminded that this procedure was not a part of the exams, that it had nothing to do with their grades and that they had the freedom not to participate in the procedure. The students were also reminded that they were free to withdraw from the study any time without prejudice. Moreover, students were assured about how confidentiality would be achieved in the research. The names of the participants remained confidential during study, while the instructor of the unit had no access either to the names of students who had given consents to participate or to the data. All data collected are secured and only the researcher has access to them.

Summary

This study was conducted under the mixed-method research paradigm, according to which the research objectives and questions are fundamental, while the research methods follow in a way that offer the best opportunities for answering the questions. A mixed-method research design was constructed to achieve the research objectives, which was developed and implemented in three stages: the pilot work, the first year of research and the second year of research. In short, this research is an evaluation study, formative/participatory in the first stage and summative/non participatory in the other two stages. In the pilot work done for the study, the main aim was to gain an insight to the class environment and to conceptualize the research. Thirty nine students constituted the research population in this stage and the following data sources were used: the FCI, students' exam papers, unstructured and semi-structured participant observations of the lessons and field notes; informal interviews with the instructor. The outcomes of the pilot work led to the modification the redevelopment of the unit and contributed to the design of the main research. In the main stage of the research, thirty five students constituted the population of the study in year 2005-06 and forty students in year 2006-07. The following methods of data collection were used in this part of the study, in order to achieve the research objectives: the FCI; an open ended questionnaire about explaining and arguing; unstructured and semi-structured participant observations of the lessons and field notes; This chapter provided a detailed account on the procedures followed to analyze the data collected by the above sources, in relation to each one of the research questions. The trustworthiness of the research was established by a combination of the ways by which validity and reliability are addressed in quantitative and qualitative research paradigms. Finally, the AERA standards guided this research, as far as ethical issues are concerned.

CHAPTER 5

Results on the area of conceptual understanding

This chapter addresses students' conceptual understanding on introductory mechanics and presents the research results for this area of study. In particular, the main research question, which is relevant to this chapter, is given below:

> *RQ1:* What is the change (if any) in students' conceptual understanding on basic mechanics after participating in the course?

In order to investigate this research question, data deriving from the FCI was used and analyzed for the two years of study, as explained in Section 4.4.3. The first part of this chapter (Section 5.1) focuses on the results of the first year of study, which relate to the above-mentioned question. Then, in Section 5.2 attention is turned to the second year of implementing the innovation and the according results are given. Finally, in Section 5.3 the results of the two years of study are compared, so as to obtain an overview of the outcome of the implementation as far as students' conceptual understanding is concerned.

5.1. Results of the first year of study

In the academic year 2005-06, thirty five students participated in more than half of the lessons and constitute the population of the study. A sample of thirty four of these students completed the FCI both in the beginning of the second week of the unit (pre-test) and in the last week of it (post-test). For each student, the correct answers in the pre- and post-test were calculated and are presented in Table 5.1. It is reminded that the FCI consists of 30 items, therefore students' scores range from 0 (no correct answer) to 30 (all the items answered correctly). First, data was analyzed using SPSS12, by calculating the mean scores in the pre- and the post-test. The outcome of this analysis is shown in Figure 5.1. As viewed in this figure, the mean score for the pre-test is 18.47 (61%), while the mean score for the post-test is 20.56 (68.5%). This result suggests a positive change in students' conceptual understanding, after participating in the course. The standard deviation both in the pre- and the post-test is small relative to the mean, which suggests that the mean represents the data well. In addition, the standard error both for the pre- and the post-test is not relative to the sample mean, which indicates that the sample is representative of the population.

Table 5.1: Students' scores in the FCI (pre- and post-test, 1st year of study)

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Pre-score	25	25	18	22	17	14	18	18	15	18	8	29	18	24	20	20	24	17
Post-score	28	29	26	17	22	18	18	21	17	13	19	28	20	21	21	16	28	15

Student	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Pre-score	17	7	14	20	30	8	23	10	21	26	21	14	28	22	7	10
Post-score	17	17	13	23	30	17	26	17	23	28	23	16	28	13	11	20

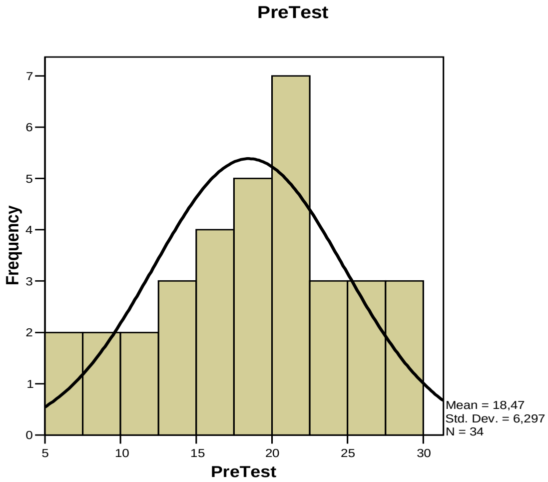
Figure 5.1: The output of the SPSS analysis, concerning the mean scores in the pre- and post-test (1st year of study).

Descriptive Statistics

	N	Mean		Std.
	Statistic	Statistic	Std. Error	Statistic
PreTest	34	18,47	1,080	6,297
PostTest	34	20,56	,917	5,350
Valid N (listwise)	34			

In order to investigate whether the positive change in students' conceptual understanding is significant, the differences between the sample means were compared. First, in order to decide whether a parametric or a non parametric test is appropriate for the comparison, the data was tested for normality: the histograms of the pre- and the post distributions were constructed and compared qualitatively with a normal curve (view Figure 5.2 for the output of the SPSS program), while the Kolmogorov-Smirnov and the Shapiro-Wilk tests were implemented to see if the distributions are significant different from normal distributions. The output of the implementation of the tests of normality is shown in Figure 5.3. As viewed from the histograms, the distributions of the pre- and post-data do not seem to be close to normal distributions. However, as seen in Figure 5.3, the normality tests conducted are non significant (p >0.05); this outcome suggests that the distributions of the two samples are not significantly different from normal distributions. Therefore, a parametric test can be used to compare the two sample means (between the pre and the post-test).

Figure 5.2: The histograms of the distributions of the data (pre and post-test, 1st year of study)



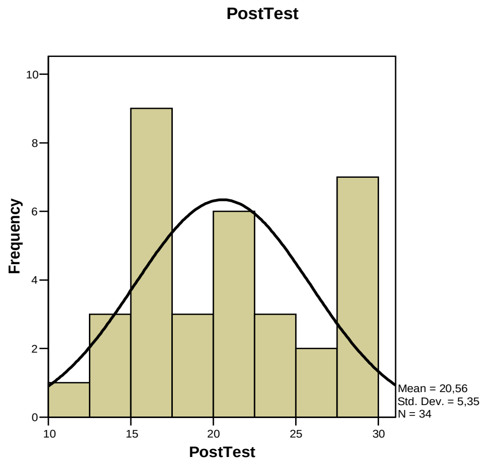


Figure 5.3: The output of the SPSS program, in terms of testing the data for normality (1st year of study)

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
PreTest	,114	34	,200*	,965	34	,335
PostTest	,129	34	,161	,943	34	,075

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

A dependent t-test was implemented to investigate whether the difference in the mean scores between the pre- and the post-test is statistically significant. The output of the SPSS analysis is shown in Figure 5.4. In this figure, the first table shows a summary of statistics for the pre- and post-data (similar to Figure 5.1), which has been discussed previously in this section. The second table presents the Pearson correlation between the two conditions. The correlation coefficient is fairly large (r = 0,708) and the significance value is $p < 0.05$, a result which shows that the pre- and post-scores of the students are significantly correlated, meaning that there is consistency in students' pre- and post-responses. The third table shows whether the difference between the means is large enough not to be a chance results, therefore it shows whether the difference is statistically significant. As viewed in this table, the two-tailed probability is very low (p = 0.011), meaning that there is only 1.1% probability that a value of t this big could happen by chance. In addition, the effect size - defined as $r = \{t^2 / (t^2+df)\}^{1/2}$ – was calculated r = 0.42, which a fairly large effect, as it is close to 0.5, the threshold for a large effect according to Cohen (1992). Therefore, as well as having a statistically significant positive change in students' performance in the FCI, the effect is large and so represents a substantive finding.

Figure 5.4: The output of the SPSS analysis (dependent t-test, 1st year of study)

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PreTest	18,47	34	6,297	1,080
	PostTest	20,56	34	5,350	,917

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	PreTest & PostTest	34	,708	,000

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 PreTest - PostTe	-2,088	4,535	,778	-3,671	-,506	-2,685	33	,011

Data from FCI was also analyzed as proposed by Hake (1998). Given that the vast majority of other interactive units have evaluated students' conceptual understanding by means of the gain percentage (view Chapter 3), such an analysis permits the comparison of the outcomes between this unit and other reported in the literature. The gain percentage <g>, defined as <g> = (% <post> - % <pre>) / (100 - % <pre>), was calculated 0.18; therefore, according to Hestenes' classification, the course falls in the region of low-g courses (view Section 4.3.4). On average, for the 1st year of study, the output of the t-test suggests that students scored significantly better in the post-test (M = 20.56, SE = 0.917), than in pre-test (M = 18.47, SE = 1.080, t (33) = -2.685, p <0.05, r = 0.42). Therefore, there was a significant positive change in students' conceptual understanding. In addition, the analysis of the data, as proposed by Hake, signifies that the course falls in the category of low-g courses (<g> = 0.18). These results will be discussed in relation to the ones obtained of the second year of study, right after the presentation of the results for 2nd year in the following section.

5.2. Results of the second year of study

During the second year of study (academic year 2006-07), although forty students participated in more than half of the lessons, only twenty three of them completed the FCI both at the beginning and at the end of the program. This was due to the fact that at the last week of this semester (when the post-test was implemented) students were exercising abstinence from courses demonstrating against new laws concerning higher education; as a consequence most of them did not come to the lessons. The small sample of students for this year is a limitation of this study and will be taken into consideration when the results will be discussed in Chapter 9. Nevertheless, for the sample of the 24 students, who completed both the pre- and post-test, procedures of data analysis similar to the previous year were implemented. First, the correct answers in the pre and post-test were calculated and are presented in Table 5.2. The mean scores in the pre- and the post-test were calculated using SPSS12 and the results are presented in Figure 5.5. Given that the mean score for the pre-test is 20.50 (68.3%) and the mean score for the post-test is 22.5 (75%), there a positive change in students' conceptual understanding in the second year of study. Given that the standard deviation is small relative to the mean both in the pre and the post-test, the means represent the data well.

Table 5.2: Students' scores in the FCI (pre- and post-test, 2nd year of study)

Student	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Pre-score	24	20	13	15	14	28	21	16	22	15	21	16	11	23	29
Post-score	21	19	18	21	16	29	20	17	29	21	19	13	17	26	30

Student	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Pre-score	28	20	29	30	18	14	26	26	13
Post-score	26	28	30	29	25	15	28	26	17

Figure 5.5: The output of the SPSS analysis, concerning the men scores in the pre- and post-test (2nd year of study).

Descriptive Statistics

	N	Mean		Std.
	Statistic	Statistic	Std. Error	Statistic
PreTest	24	20,50	1,222	5,985
PostTest	24	22,50	1,110	5,437
Valid N (listwise)	24			

Similar to the procedures followed for the 1st year of study, the data for the 2nd year of the implementation was tested for normality. Figure 5.6 shows the histograms of the distributions of the pre- and post-data, while Figure 5.7 presents the output of the SPSS tests of normality. As viewed there, the tests are non significant ($p > 0.05$), therefore a t-test was implemented to investigate whether the change between the pre- and post scores is significant. The output of the SPSS analysis is shown in Figure 5.8. As viewed in the second table in this figure, the correlation coefficient is large ($r = 0{,}827$) and the significance value is $p < 0.05$; therefore, the pre- and post-scores of the students are significantly correlated, meaning that there is consistency in students' pre- and post-responses. In addition, as viewed in the third table of this figure, the two-tailed probability is very low ($p = 0.008$), therefore there is only 0.8% probability that a value of t this big could happen by chance. In addition, the effect size - defined as $r = \{t^2 / (t^2+df)\}^{1/2}$ – was calculated $r = 0.47$, which a fairly large effect, representing a substantive finding. On average, the output of the t-test for the second year of study suggests that students scored significantly better in the post-test (M = 22.50, $SE = 1.110$), than in pre-test ($M = 20.50$, $SE = 1.222$, $t(23) = -2.881$, $p < 0.05$, $r = 0.47$). Data was also analyzed by means of the gain percentage (Hake, 1998), similar to the 1st year of study. The g percentage was calculated $\langle g \rangle = 0.21$, which denotes that the course falls in the low-g courses according to Hestenes' classification, similar to the first year of study.

Figure 5.6: The histograms of the distributions of the data (pre- and post-test), 2nd year of study

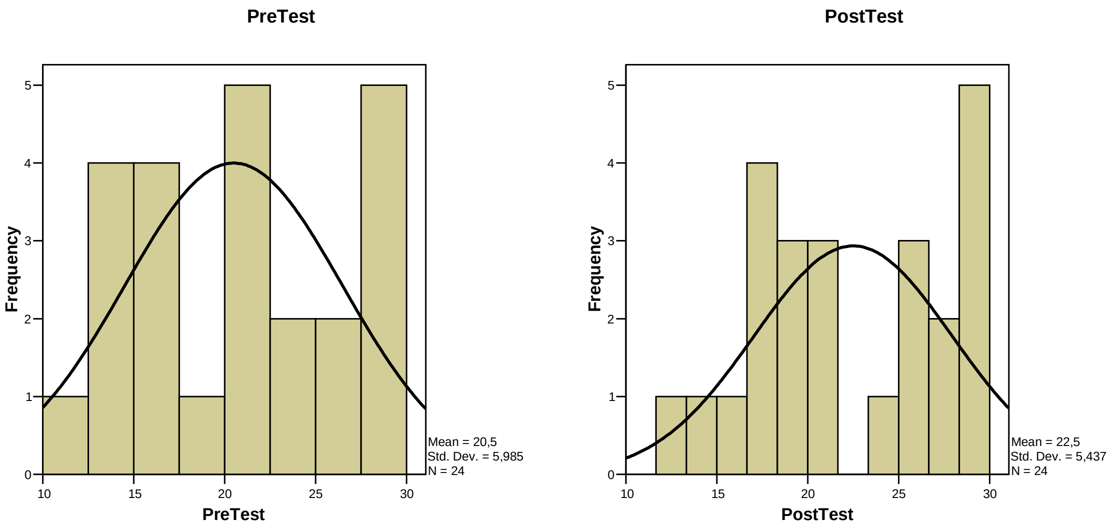


Figure 5.7: The output of the SPSS program, in terms of testing the data for normality (2nd year of study)

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
PreTest	,149	24	,181	,934	24	,120
PostTest	,157	24	,131	,915	24	,046

a. Lilliefors Significance Correction

Figure 5.8: The output of the SPSS analysis (dependent t – test, 2nd year of study)

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PreTest	20,50	24	5,985	1,222
	PostTest	22,50	24	5,437	1,110

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	PreTest & PostTest	24	,827	,000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	PreTest - PostTest	-2,000	3,401	,694	-3,436	-,564	-2,881	23	,008

5.3. Comparing the results between the two years of study

In order to compare the results on the change of students' conceptual understanding between the 1st and the 2nd year of study, it is important to investigate at first whether there is a significant difference in the two groups of participants before participating in the unit, in terms of their conceptual knowledge. If one of the two groups performed significantly better in the pre-test than the other group, then this could be a factor affecting a possible difference in the results between the two years of study. Given that the focus now is on differences between the overall means of two different samples, an independent t-test was implemented. The output of the SPSS analysis is presented in Figure 5.9.

Figure 5.9: The output of the SPSS analysis concerning the comparison of means in the pre-test between the two years of study

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
PreTest	0	34	18,47	6,297	1,080
	1	24	20,50	5,985	1,222

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
PreTest	Equal variances assumed	,015	,902	-1,234	56	,223	-2,029	1,645	-5,325	1,266
	Equal variances not assumed			-1,245	51,196	,219	-2,029	1,631	-5,303	1,244

The first table in this figure shows that the mean score of the pre-test in the 1st year of study is 18. 47 with a standard deviation 6.297, while the according numbers for the 2nd year is *M* = 20.50 with a standard deviation of 5.985. Although there is difference in students' pre-score between the two years of study, the second table of Figure 5.9 suggests that it is not significantly important. Indeed, the significance of the test is $p = 0.223$[1], which is bigger than 0.05, therefore the means of the two groups are not significantly different. On average, the students of the 2nd year of study achieved better scores in the pre-test (*M*= 20.50, *SE*=1.222) in

[1] Given that the significance for the Levene's Test of equality is $p > 0.05$, equal variances can be assumed; therefore the results reported are the ones in the first row of the table, labeled Equal variances assumed.

comparison to the pre-scores of students in the 1st year (*M*= 18.47, *SE*=1.080); however this difference is not statistically significant (t (56) = -1.234, p>0.05). Therefore, it can be assumed that students' conceptual knowledge before participating in the unit in the two years of study was similar.

As evident from the presentation of the results in the previous sections of this chapter, both for the two years of study and on average, students' performed significantly better in the post-test than in the pre-test; therefore, there was a statistically significant positive change in students' conceptual understanding. Indeed, comparing the mean paired differences between the two years (view Figures 5.4 and 5.8), it is evident that they are almost identical, i.e. *M*=-2.088 and *SE*=0.778 for the 1st year and *M*=-2.000 and *SE*=0.694 for the 2nd year. In addition, the <g> scores were similar, calculated as 0.18 and 0.21 for the 1st and the 2nd year accordingly, signifying that for both years the course falls in the low g category according to Hestenes' classification. From the lines above it becomes evident that there is consistency in the results of the study between the two years, in regard of the first research question. This increases the confidence on the reliability of the research, if defined as consistency and replicability over different groups of students, with no significant difference in their conceptual knowledge before participating in the unit.

Yet, although the increase is students' conceptual understanding is statistically significant, it should be noted that it is rather low. Indeed, the average increase in students' scores for both years is almost 7%, while the outcome of the pilot work (view Section 4.3.4) indicated an increase of almost 16% in students' conceptual understanding[2]. In addition, focusing in the <g> scores, the course falls in the region of low-g courses, while most of the interactive teaching programs, which were discussed earlier in the thesis (view Chapter 3), reported that they fall in the medium-g courses category- similar to the course in the pilot work. However, it should be reminded that both in these interventions and during the pilot work of the study, students were provided by the teacher with the correct answers in the conceptual questions. In contrast, in the two years of implementing the intervention, the instructor neither took sides,

[2] The average in students' pre- scores in the pilot work was 19 out of 30, comparable to the average pre-scores in the two years of the main research; therefore, the difference in the increase in conceptual knowledge cannot be attributed to differences in the conceptual knowledge between the samples, before entering the course.

when competing views were discussed, nor provided the students with the correct answers in the multiple choice items discussed in the classroom. In such a learning environment some students may be led to confusion concerning what the scientifically accepted theory is. This may be an explanation of why the conceptual gain in this unit is lower than the one in units, in which students are provided with the correct answers.

However, it should be noted that when the correct answers are provided to the students, there is a danger that some students memorize the information provided by the instructor at the end of the discourse, without achieving a meaningful conceptual conflict. As a result, these students may not manage to achieve deep understanding of the concepts and theories discussed. In contrast, when students decide on their own about accepting or rejecting the competing theories discussed in this classroom, the danger of rote learning of the ‘correct theory’ is minimized. The point made here is that the progress in students’ conceptual understanding may be rather low in comparison to other interactive units, but it is statistically significant and it is the outcome of the students’ interactions, rather than the rote learning of the scientific knowledge provided by the instructor, without deep understanding. Given that the issue of achieving deep understanding through interactive and cognitive conflict pedagogies has been raised in a number of other research studies (view Section 2.1.3), the point made above will be discussed more extensively in Chapter 9.

Summary

This chapter relates to the area of students’ conceptual understanding and provides the research results on RQ1. Results for the two years are similar; this increases the confidence on the reliability of the research, defined as consistency over different groups of students, with no significant difference in conceptual knowledge before participating in the intervention. For both years of study, the data analysis showed that there is a significant positive change in students’ conceptual understanding, after participating in the course. The increase, though, is rather low in comparison to other interactive courses, while for both years of study the course falls in the category of low-g courses according to Hestenes’ classification. This could be explained by the fact that students in the course are not provided by the correct answers by the instructor, as in other interactive courses reported in the literature. Yet, the progress in students’ conceptual understanding may be rather low in comparison to other interactive units, but it is statistically significant and it is the outcome of the students' interactions, rather than the rote learning of the scientific knowledge provided by the instructor, without deep understanding.

CHAPTER 6

Results on the area of scientific explanations

This chapter focuses on the area of study, which concerns students' scientific explanations on introductory mechanics. Following the procedures of data analysis, as presented in Chapter 4, it provides the research results, which relate to the second objective of this study:

> *RQ2:* What is the change (if any) in students' ability to provide correct and concise scientific explanations on basic mechanics after participating in the course?

In order to investigate this research question, the first part of the open-ended questionnaire was used (items 1 to 6). Students' responses to these items were categorized in terms of both the content and the quality, as described in Section 4.4.3. At first, this chapter presents the coding categories that emerged from the students' answers for each one of the questions (Section 6.1). Then, Section 6.2 focuses on the content of students' explanations and presents the research results on students' ability to provide correct scientific explanations on introductory mechanics, both before and after participating in the unit, for the two years of study. Finally, in Section 6.3 attention is turned to the quality of students' explanations; this section gives the results on students' ability to provide concise explanations both in the pre- and post-test, for the two years of implementing the innovation. Comparing pre and post students' performance on the field, any change in students' ability to provide correct and concise scientific explanations is identified.

6.1. Coding categories in students' explanations

The first step on the analysis of students' scientific explanations involved their classification in terms of scientific reasoning content. As mentioned in Chapter 4 (view Section 4.4.3), for each one of the six items of the questionnaire, students' responses were classified in four broad categories:

a) *Appropriate explanations*: they include scientifically accepted ideas about the phenomena;

b) *Inappropriate explanations*: they include student's alternative ideas about phenomena;

c) *No explanations with justification*: the students comment that they do not know and explain the reasons;

d) *No explanations*: no explanation is given or the response is irrelevant to the question asked.

Then, attention was turned to the appropriate explanations, which were classified in terms of their quality (defined as the degree to which the information provided by the student is enough to back up the explanatory conclusion reached). As exemplified in the following lines of this section, the properties of this category varied, in relation to the different content of each item. Yet, for all the items the categories that were formed involved the following:

a) *Category 3:* It stands for the most complete appropriate explanations, where all the needed information is provided by the student to back up the explanatory conclusion;

b) *Category 2:* It involves the cases where more information is needed to back up the explanatory conclusion;

c) *Category 1:* It refers to appropriate explanations, where only the correct claim is provided by the students.

The following lines present the categories that were formed from students' responses to the six items of the questionnaire. A brief discussion on how students' responded to each item is provided, while extracts from students' answers exemplify the attributes of each category.

Item 1: The first item of the questionnaire (view Table 6.1) addresses a common misconception among students, in the area of the free fall of bodies. The time that a body needs to fall is not related to its mass, under the condition that is no resistance from the air or when the air resistance is negligible. Yet, a common alternative conception among students is that heavier objects fall faster than the lighter ones. It should be noted that even when the resistance of the air is not negligible, the time of falling does not only depend on the mass, but on the shape of

Table 6.1: The coding categories that were formed from students' responses to Question 1

Question 1		
Categories in terms of the content	**Categories in terms of the quality**	**Description**
Appropriate explanations	*Category 3*	Containing law/principle AND the appropriate data AND the correct conclusion (AND the conditions) Example: *'If the resistance of the air is negligible, the balls reach the ground at the same time, as $x=1/2gt^2$, x is the same and g the same for the two balls.'*
	Category 2	Containing law/principle AND inefficient data AND correct conclusion OR law/principle AND correct conclusion OR appropriate data AND correct conclusion Example: *'The balls reach the ground at the same time, as g is the same for the two ball and fall from the some height.'*
	Category 1	Containing only the correct conclusion *Example: 'The balls reach the ground at the same time.'*
Inappropriate explanations		Containing misconceptions Examples: *' The force that accelerates the two balls is the same, therefore the acceleration is the same, therefore they reach the ground at the same time'* *'The heavier ball needs half the time to reach the ground, because the force from the air is the same on the two balls'* *'The heavier ball needs half time to reach the ground, because the gravitational force is double. Therefore the acceleration is double'*
No explanation with justification		The student stated that he/she does not know and explained the reasons
No explanation		No answer was given or the response was not referring to the question

the object as well. The categories that were formed from students' responses in this item are presented in Table 6.1. As illustrated there, in the appropriate explanations in this item the students state that the balls reach the ground at the same time *and* provide scientifically correct facts and reasons to back up their claim (when the explanation contained more than the conclusion). Inappropriate explanations in this item include responses containing misconceptions, which are evident either in the claim or in the reasoning, or in both. For example, some students respond correctly that the time is the same, yet they argue that this is because the force exerted in the two balls is the same, which is incorrect given the difference in the mass of the two balls. Another misconception evident in some students' explanations is that the heavier ball needs less time to reach the ground. In terms of the quality of appropriate explanations, the most complete explanations (*Category 3*) include the cases where the students provide the correct claim (the time is the same), a correct principle or law (in free fall x depends on the time of falling and on g or $x=1/2\ g\ t^2$) , the appropriate data to back up the claim (x is the same, given that they fall from the same high and g is the same) and the conditions under which the claim is correct (when the resistance of the air is negligible). An example of a students' response, which falls in this category, is provided in Table 6.1. In *Category 2* explanations, more information is needed to back up the conclusion reached. For example, in the explanation '*The balls reach the ground at the same time, as g is the same for the two balls and the fall from the same high'*, (which is provided as an example in the table for this category), a law or a principle is needed for the explanation to be complete.

Item 2: This question (view Table 6.2) is more demanding than the first one, as it refers to motion in a plane. It requires from the students to 'split' the motion of the object in two unrelated movements (on a vertical and horizontal axis) and then combine the characteristics of two movements to one, so as to come up with the correct conclusion. Although the horizontal distance that the two balls fall is the same (it is unrelated to their mass), a common misconception among the students is that the heavier ball reaches the ground at a closer point than the lighter one (yet, not necessarily in half the distance). As illustrated in Table 6.2, in student's responses both the scientifically correct ideas and alternative conceptions can be identified. In terms of the quality, the most complete appropriate explanations are the ones in which students offer the appropriate principle (independence of motion in the two axes),

Table 6.2: The coding categories that were formed from students' responses to Question 2

Question 2		
Categories in terms of the content	**Categories in terms of the quality**	**Description**
Appropriate explanations	*Category 3*	Containing laws/principles AND the appropriate data AND the correct conclusion (AND the conditions) Example: *'The trajectory results from the independence of motion along the horizontal and vertical directions (the vertical motion is unaffected by the horizontal motion of the balls). In x-axis there is no force, therefore the velocity of the ball is stable and x=ut (1). In the y-axis, if there is no air-resistance, then the object falls freely, therefore $y=1/2gt^2$. The balls fall from the same high, therefore the time of falling is the same. Therefore, from (1), it is concluded that x is the same, therefore the two balls reach the ground at the same horizontal distance).*
	Category 2	Containing laws/principles AND inefficient data AND correct conclusion OR laws/principles AND correct conclusion OR appropriate data AND correct conclusion Example: *'The balls reach the ground at the same horizontal distance, as x=ut and the time of falling is the same, as I mentioned in the previous question.'*
	Category 1	Containing only the correct conclusion *Example: 'The balls reach the ground at the same horizontal distance'*
Inappropriate explanations		Containing misconceptions Example: *'The gravitation force is bigger in the heavier ball. Therefore it will reach the ground in a closer horizontal distance than the heavier one, but not the half'.*
No explanation with justification		The student stated that he/she does not know and explained the reasons Example: *'I don't know, it seems complicated'*
No explanation		No answer was given or the response was not referring to the question

provide the appropriate formulas for each motion, reach the correct explanatory conclusion and acknowledge the conditions under which the conclusion is valid. An example of a students' answer in this category is provided in Table 6.2. In *Category 2* explanations, a premise or more from the above is missing. For example, the student's response illustrated in the table was regarded as a *Category 2* explanation, given that there is no reference to the independence of motion in the two axes, and the condition under which the conclusion is valid.

Item 3: Question 3 (view Table 6.3) can be regarded as a simple item, both in terms of scientific knowledge demand and in terms of complexity. In contrast to the previous item, which required a combination of principles and data so as to result to a correct explanatory claim, this one requires the knowledge and the ability to apply properly the 3^{rd} Newton's law. As viewed in Table 6.3, students' appropriate explanations in this item are the ones mentioning that the two forces have the same magnitude, regardless of the difference in masses between the two vehicles. In terms of quality, the less complete appropriate explanations consist only of the conclusion (for example: '*I agree, the two forces have the same magnitude and opposite directions*'); in *Category 2* explanations, the students state that they agree and made simple reference to the principle used, while in the most complete explanations (*Category 3*) the students apart from the above made reference to the content of the principle used (for an example view Table 6.3). Student's inappropriate explanations contain the following misconceptions: some students believe that the magnitude of the two forces depends on the relevant motion of the two bodies: the bigger force is exerted by the vehicle which pushes the other one during the collision. (In the first example provided in Table 6.3 for this category, the misconception is that force depends on the velocity). Other students make an inappropriate application of the 2^{nd} Newton's law ($F = m \cdot a$), arguing that the mass of the lorry is bigger; therefore, the force that it exerts to the car is bigger than the one acting on it. Another misconception evident in this item is that the 3^{rd} Newton's law can not be applied in this situation, as it stands only when the forces are acting on the same body. What is actually the case is the opposite, given that the 3^{rd} Newton's law refers to the interaction of two bodies.

Table 6.3: The coding categories that were formed from students' responses to Question 3

Question 3		
Categories in terms of the content	**Categories in terms of the quality**	**Description**
Appropriate explanations	*Category 3*	Containing reference to law/principle AND the content of it AND the correct conclusion Example: *'According to the third Newton's law, when a body exerts a force a second body, then the second body exerts a force on the first one with the same magnitude (F_1= F_2), Therefore, for this example during the collision the lorry exerts force on the car which has the same magnitude as the one that the car exerts on the lorry.'*
	Category 2	Containing law/principle AND correct conclusion Example: *'I agree, because of the third Newton's law.*
	Category 1	Containing only the correct conclusion *Example: 'F_1= F_2'*
Inappropriate explanations		Containing misconceptions Examples: *'It depends on whether the lorry is moving or not. If it does not move, then the car exerts bigger force than the force that exerts the lorry to the car'.* *' I disagree because the third Newton law is applied only for two forces that are exerted on the same body'* *'According to the second Newton's law the force depends on the mass and the acceleration. The lorry has bigger mass than the car, therefore it exerts bigger force'*
No explanation with justification		The student stated that he/she does not know and explained the reasons
No explanation		No answer was given or the response was not referring to the question

Item 4: In relation to the fourth item of the questionnaire (view Table 6.4), as mentioned in Chapter 4 (view Section 4.4.2) it is an item in which students confront difficulty in answering (a result deriving from the pilot work of this study). This difficulty can to be attributed to the

fact that in order to explore this question, students need to apply the 3^{rd} Newton's law for the system, and the 2^{nd} Newton's law for the lorry to explain why the lorry is moving. Indeed, similar to the previous item, the two forces have the same magnitude, as a result of the 3^{rd} Newton's law. The motion of the lorry can be attributed to the fact that the force from the car is bigger than the force of friction, therefore according to the 2^{nd} Newton's law it accelerates. The categories that were formed from students' responses in this item are illustrated in Table 6.4. As it can be viewed there, the most complete appropriate explanations contain the correct conclusion, refer to the 3^{rd} Newton's law and the content of it (most of the students in this category made a cross-reference to the previous question for the content of this law) and explain correctly why the truck is moving. In *Category 2* explanations, one or more of the above mentioned premises is missing. Most of the answers in this category include cases, where the students do not explain why the lorry is moving. It could be argued here that only by referring to the 3^{rd} Newton's law is enough to back up the claim that the two forces are the same, similar to the previous item (question 3); therefore, these answers could be regarded as *Category 3* explanations. The decision not to include these cases in the most complete appropriate explanations was based on the following rationale: in the claim given to the students there is the underlying assumption that the car pushes the track, therefore the force exerted on the truck should be bigger so that the track can move. Third Newton's law is enough to back up that the two forces are of equal magnitude, but the question of why the lorry is moving remains. This argument is evident in some students' responses (view the example in the category '*No explanation with justification*' in Table 6.4). Given that in this study the notion of an appropriate explanation is both an explanation that is correct and that it is likely to promote understanding in the other person, the explainee (view Section 2.2.1), it was decided that only the responses, which also explain why the track is moving fall in *Category 3* explanations. As for the inappropriate explanations in this item, some of them contain the misconception that the force acting on the lorry from the car is bigger, otherwise the lorry could not move. Another misconception evident here, is that the lorry exerts bigger force, as it has bigger mass than the car. Finally, misconceptions are found in the procedure of reasoning why the lorry is moving. In the second example provided for the category in Table 6.4, the student correctly comments that the two forces have the same magnitude; however, she attributes the movement of the lorry to the fact that no friction force acts on it, which is incorrect.

Table 6.4: The coding categories that were formed from students' responses to Question 4

Question 4		
Categories in terms of the content	**Categories in terms of the quality**	**Description**
Appropriate explanations	*Category 3*	Containing reference to law/principle AND the content of it AND the correct conclusion AND explain why the lorry is moving <u>*Example:*</u> *'According to the third Newton's law, when a body exerts a force a second body, then the second body exerts a force on the first one with the same magnitude ($F_1 = F_2$), Therefore the car exerts force on the lorry which has the same magnitude as the one that the lorry exerts on the car. The lorry is moving because the force F is bigger than the friction.'*
	Category 2	Containing reference to law/principle AND the content of it AND correct conclusion <u>OR</u> Containing to law/principle AND the correct conclusion <u>OR</u> correct conclusion AND explain why the lorry is moving <u>*Example:*</u> *'I disagree, because of the third Newton's law.*
	Category 1	Containing only the correct conclusion <u>*Example:*</u> *'$F_1 = F_2$'*
Inappropriate explanations		Containing misconceptions <u>*Example:*</u> *'The lorry is moving, therefore the force that the car exerts to the lorry is bigger than the force that the lorry exerts to the car'*
No explanation with justification		The student stated that he/she does not know and explained the reasons <u>*Example:*</u> *'I am sure that the 3rd Newton's law applies in this situation; therefore the two forces have the same magnitude…But, if this is the case, how can the car push the truck, so as the track can move? I do not know. I am confused.'*
No explanation		No answer was given or the response was not referring to the question

Item 5: As viewed in Table 6.5, this item addresses a common misconception among students, according to which when two bodies are at the same position, they have the same velocity. As it can be viewed in the figure given to the students, the body that moves above the axis accelerates, given that that its displacement increases at equal and successive time intervals. In contrast, the body that moves below the axis has stable velocity, as the displacement of the body is the same at equal time intervals. The categories that were formed from students' responses in this item are presented in Table 6.5. Students' appropriate explanations in this item involve responses, in which the students state that the bodies do not have the same velocity and provide scientifically correct facts and reasons to back up their claim (when the explanation contained more than the conclusion). As it can be viewed in the examples provided in this table, in the most complete explanations (*Category 3*) students provide enough and appropriate data and reasons in order to explain why the two bodies cannot have the same velocities at positions 2 and 5. For example, some students commented that the velocity is $v = \Delta x/\Delta t$, when Δt is small; therefore, the velocity of the bodies is the same when the displacement of the two bodies is the same at equal and small time intervals; they add that this not the case at the above-mentioned positions. Others argue that one of the two vehicles has stable velocity, while the other accelerates; therefore, it is not possible that the two of them have the same velocity in two positions. Students' inappropriate explanations involved mainly the misconception that the instant velocity is defined as $v = x/t$, or that the instant velocity is $v= \Delta x/\Delta t$, without the limitation that Δt is small. Some examples of students explanations is each category are given in Table 6.5.

Item 6: The last item of the questionnaire (Question 6, view Table 6.6) is a rather difficult one, as the results of the pilot work indicate. It addresses some students' misconception that the weight of the body (W) always equals the nominal force from the ground (N). Actually, W=N is not a general law as some students believe, but the outcome of the 1st Newton's law in some cases (for example in horizontal ground and when no other forces are exerted on the body). In this question the body accelerates, therefore the 1st Newton's law does not apply here. Students' most complete explanations in this items involved accepting the statement given to them and referring either to the second Newton's law or to the first one to back up their conclusion (view

Table 6.5: The coding categories that were formed from students' responses to Question 5

Question 5		
Categories in terms of the content	**Categories in terms of the quality**	**Description**
Appropriate explanations	*Category 3*	Students provide enough and appropriate data and reasons in order to explain, why the two bodies cannot have the same velocities at positions 2 and 5. Example: *'The fact that the two bodies have the same position does not mean that they have the same velocity. The first body has stable velocity given that has equal displacement in equal time intervals. The second body accelerates, given that its displacement increases in equal time intervals. The instant velocity is defines as $v = \Delta x/\Delta t$ when $\Delta t \rightarrow 0$. At position 2, $\Delta x_1 < \Delta x_2$, therefore $v_1 < v_2$. At position 5, $\Delta x_1 > \Delta x_2$, therefore $v_1 > v_2$'. (note: the student has drawn correctly the displacement Δx_1 and Δx_2 in the figure)*
	Category 2	More information is needed to back up why the two bodies cannot have the same velocities at positions 2 and 5. Example: *'At positions 2 and 5 the bodies have the same positions, but not the same velocities. The first one has stable velocity, while the other one has stable acceleration'.*
	Category 1	Containing only the correct conclusion Example: *'The bodies have the same positions there, not the same velocities'*
Inappropriate explanations		Containing misconceptions Example: *'Given that $v=x/t$, it becomes evident that the two bodies do not have the same velocity at positions 2 and 5.' (note: the student has drawn the positions x_1 and x_2 in the figure)*
No explanation with justification		The student stated that he/she does not know and explained the reasons
No explanation		No answer was given or the response was not referring to the question

Table 6.6, *Category 3*). In *Category 2* explanations, although most students stated that N does not equal W because the person accelerates, they failed to provide the law which relates the forces with the acceleration. As far as inappropriate explanations in this question are concerned, a number of misconceptions are evident: some students' state that N= W as a result of the 3rd Newton's law, which is incorrect given that in this law the forces are exerted on different objects; other students come to the correct conclusion that the two forces are unequal, but argue incorrectly that this is because of another force that is exerted on the body from the elevator. Some examples of students' responses in each category are presented in Table 6.6.

It should be noted that the above - mentioned categories were refined after the data of both years was gathered, in an effort to address the credibility criterion. During the analysis, regular attempts were made to identify data that refuted preliminary assertions, such as a priori attributes that were given to the categories formed. In addition, regular discussions with academics from the physics department from the University of Crete in Greece provided feedback while analysing students' explanations in terms of scientific reasoning content and quality. In short, the above procedures were made to minimize threats to validity at the stage of data analysis, such as subjective interpretation of the data, and the halo effect, where the researcher's knowledge of the participants or other data about them influence subsequent judgments.

Table 6.6: The coding categories that were formed from students' responses to Question 6

<table>
<tr><td colspan="3">Question 6</td></tr>
<tr><td>Categories in terms of the content</td><td>Categories in terms of the quality</td><td>Description</td></tr>
<tr><td rowspan="3">Appropriate explanations</td><td>Category 3</td><td>Containing law/principle AND the appropriate data AND the correct conclusion
<u>Examples:</u>
'N≠W because according to the second Newton's law ΣF = ma. The elevator accelerates, therefore a≠0, therefore ΣF≠0.'
'If N=W, then ΣF= 0, which according to the 1st Newton's law means that the elevator is not moving or that it moves with constant velocity. However, here the elevator accelerates, therefore ΣF≠0, therefore N≠W'.</td></tr>
<tr><td>Category 2</td><td>Containing law/principle AND inefficient data AND correct conclusion
<u>OR</u>
law/principle AND correct conclusion
<u>OR</u>
appropriate data AND correct conclusion
<u>Example:</u> ''N≠W because the elevator accelerates'</td></tr>
<tr><td>Category 1</td><td>Containing only the correct conclusion
Example: ' I agree, in this case 'N≠W'</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td>Inappropriate explanations</td><td></td><td>Containing misconceptions
<u>Examples:</u>
' N=W, otherwise the man would not touch the floor of the elevator'
'N≠W, because there is another force that moves the elevator'
N≠W, and if the elevator is moving up, N>W'</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td>No explanation with justification</td><td></td><td>The student stated that he/she does not know and explained the reasons</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td>No explanation</td><td></td><td>No answer was given or the response was not referring to the question</td></tr>
</table>

6.2. Results on the content of explanations

Based on the categories presented in the previous section of this chapter, for each one of the items 1-6 of the open-ended questionnaire, each student explanation was identified in terms the category which falls, as far as the content is concerned. The overall number of students' explanations fallen in each category for each item was counted, both in pre- and post-test, for the two years of study. The results of the data analysis are presented below in tables. In particular, Table 6.7 shows the number of students' explanations in each category in terms of the content for each item both in the pre- and post-test for the 1st year, while Table 6.8 shows the according numbers for the 2nd year of study. Given the difference in the sample size between the two years of study, the percentages in each category are also provided below the absolute numbers, for facilitating the comparison between the two years.

Table 6.7: The number of students' explanations in each category in terms of the content, for each item, for pre – and post-test (1st year of study – 33 students). The percentages are provided below the absolute numbers.

	Question 1		Question 2		Question 3		Question 4		Question 5		Question 6	
	Pre	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**
Appropriate	23 (70%)	28 (85%)	14 (42%)	19 (58%)	24 (73%)	32 (97%)	24 (73%)	27 (82%)	21 (64%)	22 (67%)	5 (15%)	14 (42%)
Inappropriate	8 (24%)	3 (9%)	7 (21%)	7 (21%)	4 (12%)	1 (3%)	7 (21%)	5 (15%)	8 (24%)	10 (31%)	21 (64%)	16 (49%)
No explanation	2 (6%)	0 (0%)	10 (31%)	7 (21%)	3 (9%)	0 (0%)	0 (0%)	0 (0%)	4 (12%)	1 (3%)	7 (21%)	3 (9%)
Explain why they do not know	0 (0%)	2 (6%)	2 (6%)	0 (0%)	2 (6%)	0 (0%)	2 (6%)	1 (3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Table 6.8: The number of students' explanations in each category in terms of the content, for each item, for pre – and post-test (2nd year of study – 23 students). The percentages are provided below the absolute numbers.

	Question 1		Question 2		Question 3		Question 4		Question 5		Question 6	
	Pre	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**
Appropriate	18 (78%)	23 (100%)	10 (43%)	14 (61%)	19 (83%)	22 (96%)	11 (48%)	15 (65%)	16 (70%)	20 (87%)	9 (39%)	15 (65%)
Inappropriate	4 (18%)	0 (0%)	5 (22%)	5 (22%)	4 (18%)	1 (4%)	11 (48%)	6 (26%)	6 (26%)	3 (13%)	13 (57%)	8 (35%)
No explanation	1 (4%)	0 (0%)	8 (35%)	4 (18%)	0 (0%)	0 (0%)	1 (4%)	1 (9%)	1 (4%)	0 (0%)	1 (4%)	0 (0%)
Explain why they do not know	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Focusing on the appropriate explanations provided by the students in each item before and after instruction, data from the tables above are also presented in the format of charts, so as to facilitate the interpretation of the results. Figure 6.1 below shows the percentages of students' appropriate explanations, in the pre-and post-test in the six items of the questionnaire, for the 1st year of study, while Figure 6.2 presents the according percentages for the 2nd year of research. A first notable feature of these data is that there seems to be consistency in terms of students' appropriate responses in the pre-test across items between the two years. Indeed, focusing on the pre-scores in Figures 6.1 and 6.2 it becomes evident that for both years of the study the easiest items were the third and the first one, while the most difficult items were the second and the sixth question. These results are also consistent to the pilot work done for the study. Only in item 4, the pre-scores between the two years differ considerably (73% and 48% of appropriate explanations for the 1st and the 2nd year of study accordingly), while the results of the pilot work are in accordance with the 2nd year results. The essential issue raised here is that for five out of six questions of the test, the item difficulty does not seem to be dependent on the sample; this result increases the researcher's confidence on the piloting of the test.

Figure 6.1: The percentages of students' appropriate explanations, in the pre- and post test for the 1st year of study for each item.

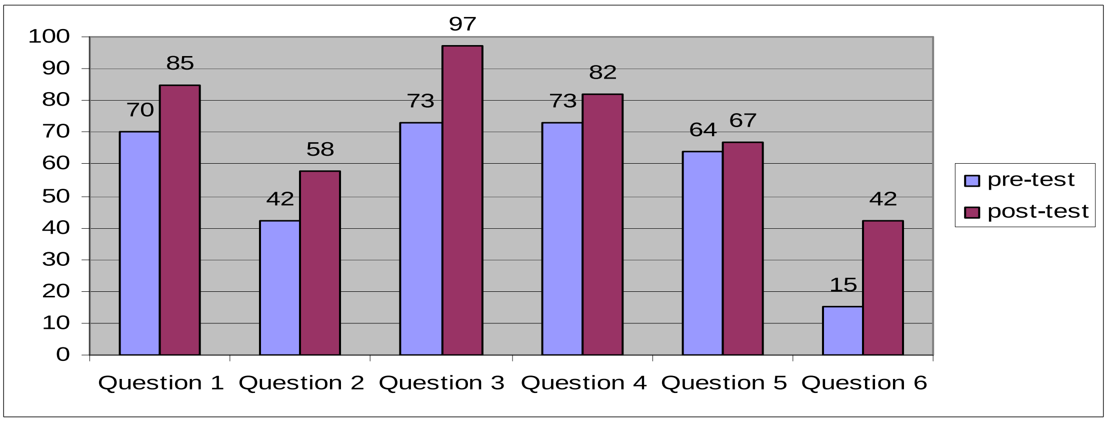
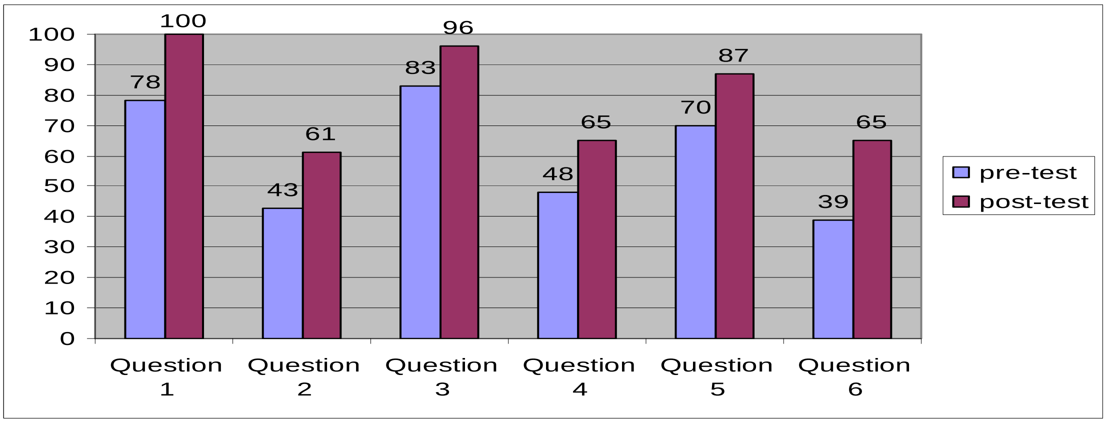


Figure 6.2: The percentages of students' appropriate explanations, in the pre- and post test for the 2nd year of study for each item.



Focusing on 1st year of research, as viewed in Figure 6.1 , the pre-test percentages of scientifically correct responses range from 15% (in item 6) to 73% (in items 3 and 4), a variance that can be attributed to the difference in the difficulty of the items; the according percentages for the post-test range from 42% (item 6) to 97% (in item 3). Turning attention on the change of students' performance on the field in each item, from Figure 6.1 it becomes evident that in all the six items of the questionnaire the overall number of students, who provided a scientifically correct explanation in the post-test exceeds the number of students who provided an appropriate explanation in the pre-test. Yet, differences in the progress in scientific correct reasoning across the items are considerable, varying from 3% (in item 5) to 27% (in item 6). As far as students' performance in item 5 is concerned (in which there is only a 3% increase in the number of appropriate explanation), it is worth noticing that this is the only item of the test, in which more students gave an inappropriate explanation in the post- than in the pre-test, as viewed in Table 6.7. A triangulation of these results with the researcher's field notes suggests that the variance in the progress in terms of scientific reasoning content across the items may reflect the outcome of students' discourse, when similar items were debated in the classroom. For example, focusing on item 5 (in which students appeared to progress only 3% in terms of scientifically correct content and it is the only item in which more students in the post-test gave an inappropriate explanation than in the post-test) the field notes suggest that, when a similar item to this was discussed in the classroom, at the end of the discourse students seemed confused in regard of the scientifically correct explanation. Although the voting system showed that there was general consensus in regard of the correct conclusion, students disagreed about the correct explanation. The researcher's impression was that the class seemed to be influenced by the opinion of one of the best students of the class, who held an incorrect explanation on this item. Given that the instructor held the role of the facilitator in the dialogue and did not take sides or provided the correct answers while competing explanations were discussed, he tried to initiate further conversation on the item both at this lecture and at the following one; yet, the other students were not willing to rebut the opinion of a good student. What is important about this finding is that it provides an indication of the role of authority that a person has during discourse, an issue that will be discussed further later in Chapter 9 of the study.

Turning attention to the 2nd year of study, Figure 6.2 illustrates that in all the items the post-test percentages of scientifically correct explanations outnumber the according percentages in the pre-test, a result consistent with the one of the 1st year of study. In addition, similar to the 1st year of study the easiest items were the third and the first one, while the most difficult items were the second and the sixth question, as expected from the pilot work of the study. Yet, as it can be viewed in Figure 6.2, the progress in scientific correct reasoning for the 2nd year varies from 13% (in item 3) to 26% (in item 6), which is bigger than the progress of the previous year. In addition, in contrast to the 1st year results, for the 2nd year of research, in none of the six items of the questionnaire the inappropriate explanations of the post-test outnumber the ones of the pre-test. Indeed, as shown in the second row of Table 5.8, in five out of six items fewer students gave less inappropriate explanations in the post- than in the pre-test, while in one item (question 2) the number of inappropriate explanations remains unchanged. From the lines above and in a first reading of the data it seems that students performed better in terms of providing correct explanations in the 2nd year than in the 1st year of research. A possible reason for this could be the difference between the two groups of participants before participating in the unit, in terms of their ability to provide correct scientific explanations. This is an issue which was tested and will be discussed later in this section, when the results of the analysis of the questionnaire as a whole are provided.

After viewing each item of the questionnaire separately, and aiming to obtain an overview of the change in students' ability to provide correct scientific explanations, data from the questionnaire was analyzed in the following way: At first, counts of the number of explanations of each category were made for all the items of the pre-and the post-test for both years of the study. This kind of analysis permits the comparison of the total numbers of explanation of each category provided by the students, regardless of the scientific area and the difficulty of each item. The results of this analysis are presented in Figures 6.3 and 6.4 for the 1st and the 2nd year accordingly[1]. As evident from these figures, there is a shift at the end of the intervention towards more appropriate explanations and less inappropriate ones for both years of study. In order to see whether the shift towards more appropriate explanations at the end of the course is statistical significant the following procedure was followed. For each student, the numbers of appropriate explanations for all the items in the pre - and post-test was calculated and are presented in Tables 6.9 and 6.10. There are 6 items in the test, therefore the numbers range from 0 (no appropriate explanation in the test) to 6 (appropriate explanations in all items).

[1] It should be noted that given the difference in the sample sizes, the number of total explanations obtained the 1st year are 198 (33students X 6items), while for the 2nd year are 138 (23students X 6items).

Figure 6.3: Chart showing numbers of explanations in each category in terms of content, as provided in all the items of the questionnaire for the 1st year of study (n = 198)

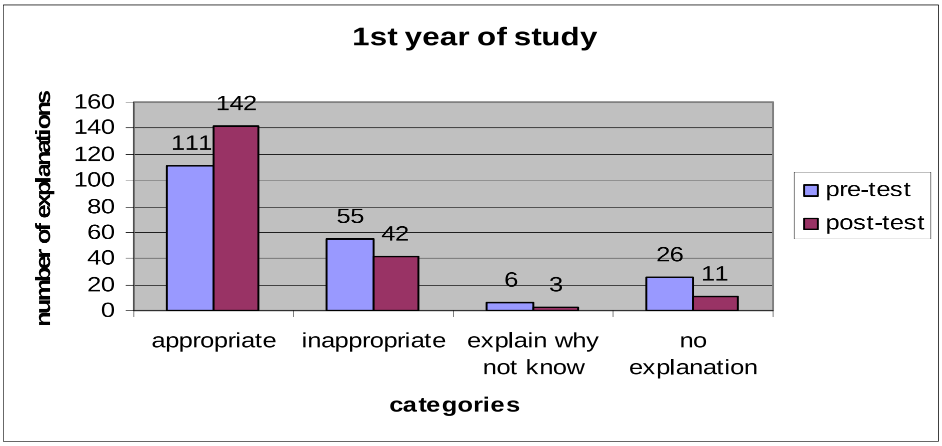
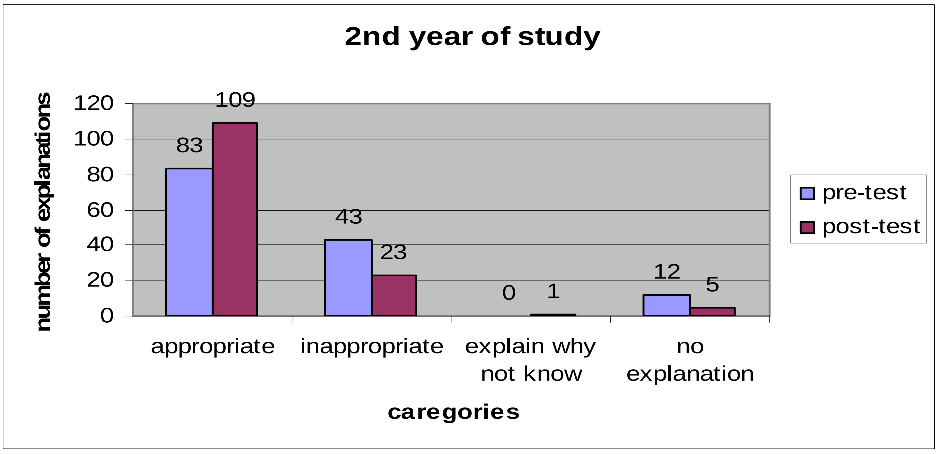


Figure 6.4: Chart showing numbers of explanations in each category in terms of content, as provided in all the items of the questionnaire for the 1st year of study (n = 138)



A non parametric test was then implemented[2] - the Wilcoxon Signed ranks test-, so as to see if there is a significant difference in the number of appropriate explanations provided by the students between pre – and post-test. Figures 6.5 and 6.6 illustrate the output of the SPSS program for the 1st and the 2nd year accordingly.

[2] The data was tested for normality; the distributions violated the assumption of normality therefore a parametric test is not appropriate for the analysis.

Table 6.9: The number of appropriate explanations provided by each student (Pre- and post-test, 1st year of study)

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Pre-score	**6**	**1**	**3**	**1**	**5**	**4**	**4**	**4**	**3**	**5**	**4**	**4**	**2**	**5**	**5**	**3**	**3**	**2**
Post-score	**4**	**5**	**4**	**2**	**4**	**5**	**3**	**4**	**4**	**5**	**3**	**5**	**2**	**5**	**5**	**5**	**4**	**4**

Student	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Pre-score	**4**	**6**	**2**	**5**	**2**	**4**	**4**	**6**	**5**	**3**	**2**	**1**	**2**	**1**	**1**
Post-score	**4**	**6**	**4**	**4**	**4**	**4**	**6**	**4**	**5**	**4**	**4**	**4**	**5**	**4**	**6**

Figure 6.5: The output of the SPSS program – Wilcoxon Signed Ranks Test – 1st year of study

Ranks

		N	Mean Rank	Sum of Ranks
post - pre	Negative Ranks	6[a]	9,17	55,00
	Positive Ranks	18[b]	13,61	245,00
	Ties	9[c]		
	Total	33		

a. post < pre

b. post > pre

c. post = pre

Test Statistics[b]

	post - pre
Z	-2,758[a]
Asymp. Sig. (2-tailed)	,006

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

For the first year of the study (view Figure 6.5), the first table provides information about the ranked scores: it tells us the number of negative ranks (these of students who provided less appropriate explanations in the post-test than in the pre-test) and the number of positive ranks (these of students who gave more appropriate explanations in the post- than in the pre-test). As viewed in the table, only 6 out of 33 students provided less appropriate explanations in the post-test than in the pre test, while there are 9 tied ranks (i.e. students who gave the same number of appropriate explanations in the pre- and post-test). The test statistic T is the lowest value of the two types of ranks, i.e. T = 9.17, while the z score is -2.758 and this value is significant at p = 0.006. The size effect r was calculated r = -0.342, which represents a medium change in the number of appropriate explanations provided by the students, as it is between Cohen's criteria of 0.3 and 0.5 for medium change. Therefore, it can be concluded that for the 1st year of study the number of appropriate explanations provided by the students is significantly higher in the post-test than in the pre-test and that the change is of medium size (z = -2.758, p< 0.05 and r = -0.342). For the 2nd year of the study (refer to Figure 6.6), it becomes evident that only 2 out of 23 students provided less appropriate explanations in the post-test than in the pre test, while there are 4 tied ranks. The test statistic T is the lowest value of the two types of ranks, i.e. T = 5, while the z score is -3.493 and this value is highly significant (p = 0.000). The size effect r was calculated r = -0.515, which represents a large change in the number of appropriate explanations provided by the students, as it is above Cohen's criterion of 0.5 for large change.

Table 6.10: The number of appropriate explanations provided by each student (Pre- and post-test, 2nd year of study)

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Pre-score	**4**	**4**	**3**	**5**	**2**	**6**	**3**	**3**	**3**	**3**	**3**	**4**	**2**	**3**	**6**	**4**	**4**	**5**	**1**	**4**	**5**	**4**	**1**
Post-score	**5**	**5**	**5**	**5**	**4**	**6**	**3**	**4**	**5**	**3**	**5**	**5**	**5**	**5**	**5**	**6**	**5**	**6**	**5**	**3**	**6**	**6**	**3**

Figure 6.6: The output of the SPSS program – Wilcoxon Signed Ranks Test – 2nd year of study

Ranks

		N	Mean Rank	Sum of Ranks
post - pre	Negative Ranks	2[a]	5,00	10,00
	Positive Ranks	17[b]	10,59	180,00
	Ties	4[c]		
	Total	23		

a. post < pre
b. post > pre
c. post = pre

Test Statistics[b]

	post - pre
Z	-3,493[a]
Asymp. Sig. (2-tailed)	,000

a. Based on negative ranks.
b. Wilcoxon Signed Ranks Test

Therefore, for the 2nd year of study the number of appropriate explanations provided by the students is significantly higher in the post-test than in the pre test and the change is of large size (z = -3.493, p< 0.05 and r = -0.515). In the whole, what the numbers above indicate is that for both years of the study the students gave significantly more appropriate explanations in the post- than in the pre-test; therefore –in relation to the research question investigated in this chapter- there is a significant positive change in their ability to provide correct scientific explanations after participating in the unit for both years of the study.

Apart from viewing each year of the study separately, the kind of analysis made above permits the comparison of the change in students' ability to provide correct scientific explanations between the two years. As it was mentioned earlier in this section, while analyzing each item of the questionnaire separately, in a first look it seemed that students in the 2nd year performed better than in the 1st year. The analysis made above is in accordance with this preliminary reading of the data; indeed, the statistical analysis indicate that for the 1st year of the study the change is of medium size (r = -0.342), in contrast to the 2nd year for which there is a large change (r = -0.515). A possible reason for this could be the difference in the two groups of participants before participating in the unit, in terms of their ability to provide correct scientific explanations. To test this hypothesis, attention was turned to the numbers of appropriate explanations given by the students in the pre-test for both years. For this data – which are not normally distributed as mentioned before - a Mann-Whitney test was conducted to investigate if the two groups (1st and 2nd year) differ considerably in terms of the number of appropriate explanations the participants provided at the beginning of each year. The output of the SPSS analysis is shown in Figure 6.7. As shown in this figure, the z value is smaller than 1.96 (ignoring the minus sign), while the p value is far bigger than 0.05 (p = 0.792); this result signify that there is no significant difference between the 1st and the 2nd year, in terms of the number of appropriate explanations the students were able to provide before the beginning of the course. Therefore, the data here do not support the hypothesis that the difference in performance between the two years is due to the differences in such ability between the samples before the intervention.

Figure 6.7: The output of the SPSS program – Mann-Whitney test

Ranks

	group	N	Mean Rank	Sum of Ranks
pre	0	33	28,03	925,00
	1	23	29,17	671,00
	Total	56		

Test Statistics (a)

	pre
Mann-Whitney U	364,000
Wilcoxon W	925,000
Z	-,263
Asymp. Sig. (2-tailed)	,792

a Grouping Variable: group

Another possible reason as to why the group of students of the second year improved more in providing correct scientific explanations than the group of the first year could be the instructor's accumulation of experience in the innovative teaching approach. As mentioned earlier in this study (view Section 1.3.2), one of the main modifications of the unit – which resulted from the pilot work - was the new role of the instructor as a facilitator of students' discourse and not the provider of the correct answers. At the early stages of the implementation of the program, the researcher's field notes suggest that this role of the instructor was neither easy to be followed by him, nor always led to the desired results. In many cases students put pressure on him to give them the correct answers after the debate. A characteristic students' comment from the 1st year of study, which was noted in the researcher's notes during the observations and exemplifies the pressure put on the instructor is the following: '*I guess it is your job to teach us, isn't it? All I am asking for you is to do your job…*' In such cases at the beginning of the intervention the instructor's negotiations with the students while trying to highlight the importance of their autonomy in learning were not always successful. In other cases, students were expressing insecurity in regard of what they have learned, or in terms of if what they have learned is actually correct. Typical comments by the students at the beginning of each semester were the following: '*How can I be sure that I am correct?*' or '*I pretty sure that I am correct, but still I would like to have reassurance at the end of the discourse by someone who really knows… the professor*'. It took time and modifications in the instructors strategies during the semester, so as the students to accept his new role. The above indicate that the better results acquired during the 2^{nd} than the 1^{st} year of research may be a reflection of the development of the teacher's ability to support his new role in the classroom and of the accumulation of the new teaching approach. This hypothesis is supported by other studies on interactive units (view Section 3.3), which highlight similar difficulties confronted on the 1^{st} year of implementing an innovation - which were overcome in the following years- and the sophisticated nature of interactive teaching approaches.

In summary, this section has presented the results of the data analysis, which relate to the change in students' ability to provide correct scientific explanations after participating in the course. At first, analysis of each item of the test before and after instruction for the two years of study has shown that the item difficulty does not depend on the sample; this increases the researcher's confidence on the piloting of the test. In addition, this kind of analysis suggested variance in the progress in terms of scientific reasoning content across the items, which was related to the outcome of students' discourse, when similar items were debated in the classroom. A triangulation of these results with the researcher's field notes has provided indication on the importance of the role of authority that a person holds during discourse. Secondly, analysis of the test as a whole showed that for both years of study, students gave significantly more appropriate explanations in the post- than in the pre-test; therefore, there is a significant positive change in their ability to provide correct scientific explanations after participating in the unit: for the 1^{st} year, the change is of medium size ($r = -0.342$), while for the 2^{nd} year the change is of large size ($r = -0.515$). Given that no significant difference was found in the two groups of participants before participating in the unit, in terms of their ability to provide correct scientific explanations, the better results of the 2^{nd} year were attributed to the development of the instructor's accumulation of experience in the innovative teaching approach. In the following section, the data analysis that relate to the change in students' ability to provide concise scientific explanations is presented.

6.3. Results on the quality of explanations

In order to investigate the possible change in the quality of students' explanations after participating in the unit, attention was turned to students' appropriate explanations. Based on the categories presented in Section 6.1, for each one of the items 1-6 of the open-ended questionnaire, the appropriate explanations provided by the students were identified in terms of the category into which they fall. Counts of the numbers of explanations in each category for each item, in pre- and post-test for the two years of study were made and the results are shown in the tables below. In particular, Table 9.11 shows the numbers of explanations in each category in the pre-and post-test along with the percentages for the 1st year of study, while Table 9.12 illustrates the according numbers for the 2nd year of research.

Table 9.10: The number of students' explanations in each category in terms of the quality, for each item, for pre – and post-test (1st year of study – 33 students). The percentages are also provided.

	Question 1		Question 2		Question 3		Question 4		Question 5		Question 6	
	Pre	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**
Category 3	3 (10%)	15 (45%)	4 (12%)	4 (12%)	19 (58%)	16 (49%)	5 (15%)	3 (9%)	7 (21%)	8 (25%)	2 (6%)	3 (9%)
Category 2	14 (42%)	13 (40%)	7 (21%)	14 (42%)	5 (15%)	16 (48%)	16 (48%)	23 (70%)	10 (31%)	8 (24%)	3 (9%)	10 (31%)
Category 1	6 (18%)	0 (0%)	3 (9%)	1 (4%)	0 (0%)	0 (0%)	3 (10%)	1 (3%)	4 (12%)	6 (18%)	0 (0%)	1 (2%)
TOTAL (Appropriate Explanations)	(70%)	(85%)	(42%)	(58%)	(73%)	(97%)	(73%)	(82%)	(64%)	(67%)	(15%)	(42%)

Table 9.11: The number of students' explanations in each category in terms of the quality, for each item, for pre – and post-test (2nd year of study – 23 students). The percentages are also provided.

	Question 1		**Question 2**		**Question 3**		**Question 4**		**Question 5**		**Question 6**	
	Pre	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**
Category 3	6 (26%)	11 (48%)	3 (13%)	7 (30%)	9 (39%)	13 (57%)	0 (0%)	5 (22%)	7 (30%)	9 (39%)	1 (4%)	8 (35%)
Category 2	10 (43%)	12 (52%)	7 (30%)	7 (30%)	8 (35%)	8 (35%)	10 (43%)	10 (43%)	6 (26%)	9 (39%)	7 (30%)	6 (26%)
Category 1	2 (9%)	0 (0%)	0 (0%)	0 (0%)	2 (9%)	1 (4%)	1 (4%)	0 (0%)	3 (13%)	2 (9%)	1 (4%)	1 (4%)
TOTAL (Appropriate Explanations)	18 (78%)	23 (100%)	10 (43%)	14 (61%)	19 (83%)	22 (96%)	11 (48%)	15 (65%)	16 (70%)	20 (87%)	9 (39%)	15 (65%)

Focusing on *Category 3* explanations –defined as the most complete ones, which contain all the needed claims to back up the explanatory conclusion - data from the tables above are presented below in the format of chart bars. Figure 6.8 refers to the pre- and post percentages of *Category 3* explanations for the 1st year of study. A first feature evident from this figure that is worth commenting is that the percentages of students' *Category 3* explanations at the beginning of the year are low: with an exception to the third item, in all the other items the best quality explanations range from 6% to 21%. More encouraging seem to be the according percentages for the 2nd year of study (view Figure 6.9), which are low as well, but they come up to 39%. As for item 3 (in which for both years the pre-percentages are the highest among all the items),

Figure 6.8: The percentages of students' *Category 3* explanations, in the pre- and post test for the 1st year of study.

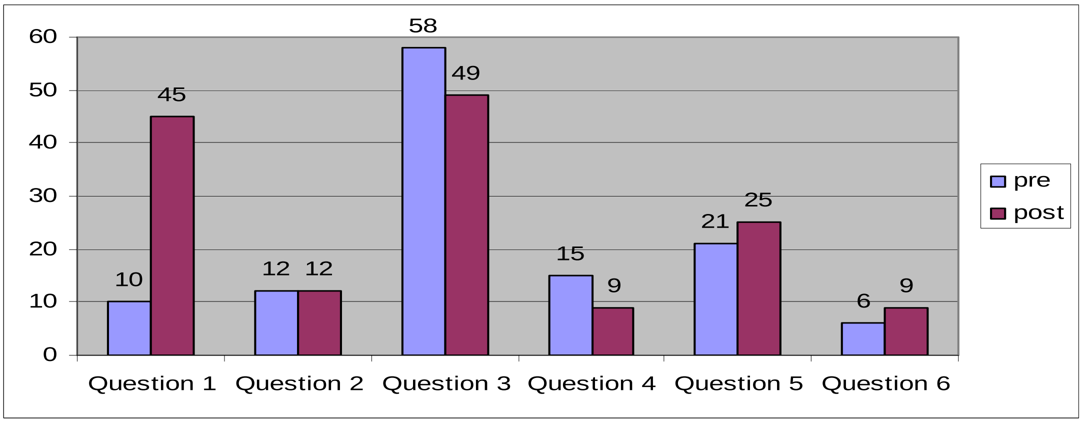
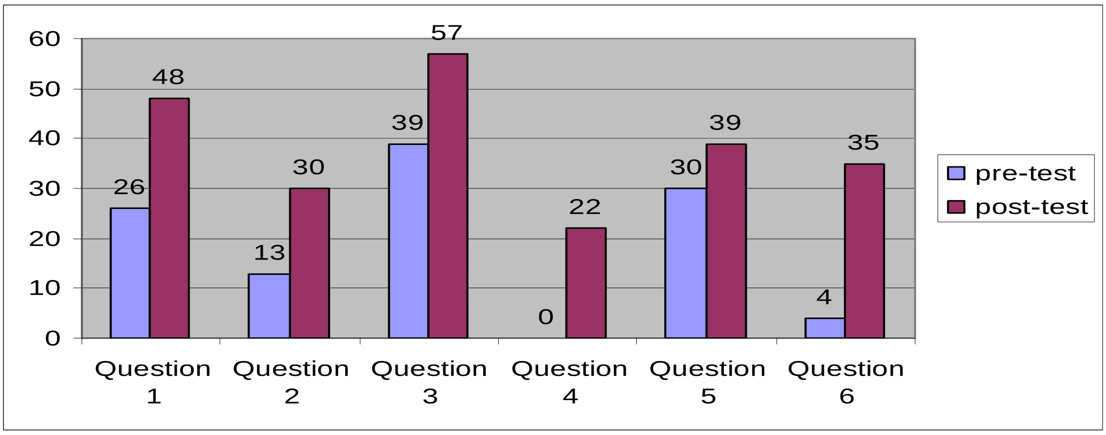


Figure 6.9: The percentages of students' *Category 3* explanations, in the pre- and post test for the 2nd year of study.



it is the least complex item of the test- as mentioned in the first section of this chapter; therefore, it should be expected that students wouldn't find much difficulty to provide a concise explanation (refer to Section 6.1 for examples in students' *Category 3* explanations for each item). In all the other items, though, students' pre-scores are rather disappointing, as it is the minority of students, who provided both a correct and a concise scientific explanation. This outcome is in line with the pilot work done for the study, where it had been pinpointed that one of the entrants' major difficulty is on communicating their explanations to their classmates, rather the ignorance of scientific facts and theories (view Section 4.3.4). It should be reminded that the entrants of these two years are the first graduates of the upper-secondary school, after the curriculum reforms (view Section 1.2) and that one of the subjects of these reforms was to put more emphasis on students' communication abilities. Although no generalizibility can be established by these results, students' performances on the area before entering the course can provide an indication of the attained curriculum (what the actual learning outcomes of the reforms were). Data here shows that -at least this sample of students- the vision elaborated in the ideal curriculum and the actual learning outcomes of the reforms are incompatible.

Turning attention to the change in students' ability to provide concise explanations after participating in the unit, which is one of the main foci of this chapter, results of the 1st year of study are mixed. As it can be viewed in Figure 6.8, in three out of six items of the test there is a positive change in the number of *Category 3* explanations, in one item the numbers remain unchanged and in two items there is regress. In contrast, for the 2nd year of the study in all the items students provided more *Category 3* explanations in the post- than in the pre-test. It could be argued that mixed results for the 1st year might have been expected: it is maybe unrealistic to expect radical changes from the 1st year of implementing an innovation, especially when we deal with such a complex skill, as the provision of correct and concise explanations. Maybe one semester is insufficient time for the instructor to articulate the innovative approach in such a degree so as to foster the quality of students' explanation. On the other hand, though, it is worth noticing that items 3 and 4, in which fewer students gave a *Category 3* explanation in the post-that in the pre-test, are the two items which was the least difficult for the students of this year. Indeed, if we view back the pre- scores at Figure 6.1, these are the items in which 73% of the students gave an appropriate explanation in the pre-test, which is the maximum score among all the items. This observation may suggests that the easier or the more obvious an explanation is for some students, the less degree of information is provided by them to back up their conclusion.

There is no question that the data here must be interpreted with caution. First, the above hypothesis is not supported by the data of the 2nd year of study - where students' provided more *Category 3* explanations after the intervention than before, in all the items. Second, the sample size is small; therefore, it would be unrealistic to try to establish any generalization of the results. If, however, the outcomes of this study are seen as the starting point for further investigation and research -rather than regarded as firm conclusions- it is worth to emphasize the observation made in the lines above. Data of the 1st year of study provide an indication that the difficulty and the degree of conciseness of an explanation maybe *negatively* correlated; in other words, maybe for some students the more easy and obvious an explanation is regarded, the less information they provide to back up the conclusion reached. If this is the case, then an explanation that is regarded as concise by the explainers (as some data is *obvious* for them, therefore they are not worth to be mentioned explicitly) may be seen by the explainee as one that lacks important information. As a result, the goal of the explanatory act – if defined as the promotion of understanding in the explainee – is unlikely to be fulfilled. Given that the factors that constrain students' effective reasoning is an area of study of both theoretical and practical interest (view Section 2.2.3), the above observation will be discussed thoroughly later in this study (view Chapter 9).

Nevertheless, after analyzing each item of the questionnaire separately, data from the questionnaire was analyzed as a whole, with an aim to acquire an overview of students' ability to provide concise explanations. Specifically, in order to compare the total number of explanations in each category between the pre- and post-test, counts of the number of students' explanations of each category were made for all the items, for the two years of study. The results of this kind of analysis are presented in Figures 6.10 and 6.11.

Figure 6.10: Chart showing numbers of each category of appropriate explanations, for the 1st year of study, before and after the intervention

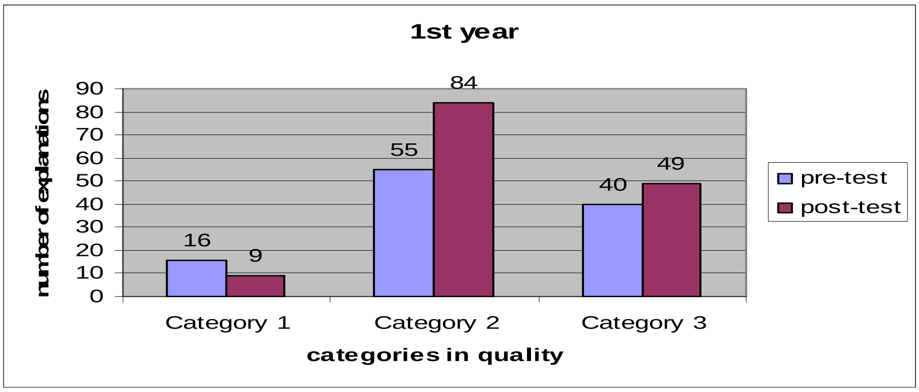
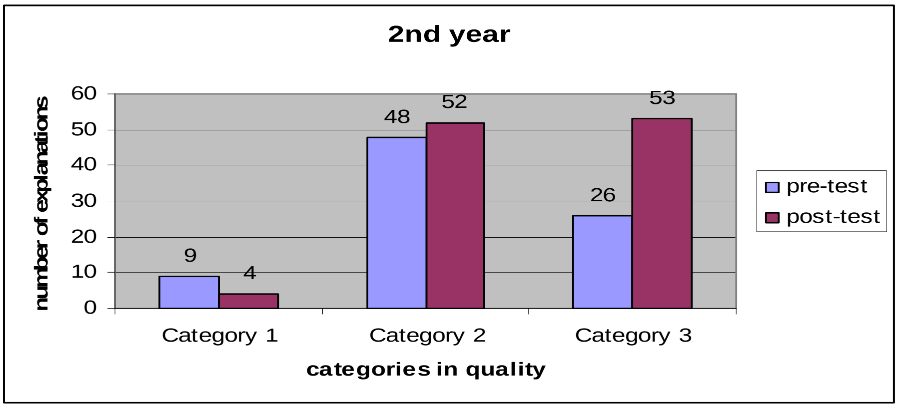


Figure 6.11: Chart showing numbers of each category of appropriate explanations, for the 2nd year of study, before and after the intervention



For the 1st year of study, Figure 6.10 shows that the largest number of explanations emerging from the data at both the beginning and the end of the semester is at *Category 2* (55 and 84 respectively). In addition, from this figure it becomes evident that at the end of the intervention there is a shift towards more *Category 3* explanations, while the number of *Category 1* decreased from 15 to 9.The decrease in the number of *Category 1* explanations is important, as it signifies that only a small minority of the explanations provided by the students did not offer a rationale to back up the explanatory conclusion. However, the statistical analysis conducted failed to provide that the shift towards more *Category 3* explanations and less *Category 1* ones is statistical significant. Indeed, as shown in Figure 6.12 – presenting the outcome of the Wilcoxon Signed Ranks Test for comparing the number of explanations in each category before and after intervention – the *p* value is bigger than 0.05. Only for the increase of *Category 2* explanations statistical significance is established, given that $p < 0.05$, while the size effect was calculated $r = -0.355$ signifying a marginally medium change. More encouraging seems to be the outcome for the 2nd year of study on the field. As evident from Figure 6.11, at the beginning of the year the larger number of explanations is at *Category 2*; on the contrary, at the end of the intervention, the *Category 3* explanations outnumber the explanations in the other two categories. In addition, comparing the pre- and post-numbers in the three categories, from Figure 6.11 it becomes evident that –similarly to the 1st year of study- students at the post- test gave more *Category 3* and less *Category 1* explanations than in the pre-test. In contrast, though, to the previous year the change in the number of *Category 3* explanations is statistically

Figure 6.12: The outcome of SPSS analysis (Wilcoxon Signed Ranks Test) in comparing the number of explanations in each category between the pre- and the post-test (1st year of study)

Ranks

		N	Mean Rank	Sum of Ranks
cat1post - cat1pre	Negative Ranks	9[a]	8,67	78,00
	Positive Ranks	6[b]	7,00	42,00
	Ties	18[c]		
	Total	33		
cat2post - cat2pre	Negative Ranks	4[d]	11,25	45,00
	Positive Ranks	19[e]	12,16	231,00
	Ties	10[f]		
	Total	33		
cat3post - cat3pre	Negative Ranks	10[g]	10,20	102,00
	Positive Ranks	13[h]	13,38	174,00
	Ties	10[i]		
	Total	33		

a. cat1post < cat1pre

b. cat1post > cat1pre

c. cat1post = cat1pre

d. cat2post < cat2pre

e. cat2post > cat2pre

f. cat2post = cat2pre

g. cat3post < cat3pre

h. cat3post > cat3pre

i. cat3post = cat3pre

Test Statistics[c]

	cat1post - cat1pre	cat2post - cat2pre	cat3post - cat3pre
Z	-1,107[a]	-2,888[b]	-1,134[b]
Asymp. Sig. (2-tailed)	,268	,004	,257

a. Based on positive ranks.

b. Based on negative ranks.

c. Wilcoxon Signed Ranks Test

significant: the Wilcoxon Signed Ranks Test that was conducted showed that for *Category 3* explanations $p<0.05$ and the size effect was calculated $r = -0.445$ (view Figure 6.13). Therefore, for the 2nd year of study the students provided a significantly larger number of the most concise explanations at the end of the intervention, and the change is of medium size. As for the other two categories, no statistical significance was established for the change between pre- and post-performances. In short, the numbers above indicate that at the end of each year of the intervention, there is a shift towards less *Category 1* explanations and more *Category 2* and *Category 3* ones. This shows a positive development in students' ability to provide concise explanations for both years. Yet, better results are obtained during the 2nd year of study: on the one hand at the end of the semester the best quality explanations outnumber the explanations in the other categories (while for the 1st year the larger number of explanations in the post-test is

Figure 6.13: The outcome of SPSS analysis (Wilcoxon Signed Ranks Test) in comparing the number of explanations in each category between the pre- and the post-test (2nd year of study)

Ranks

		N	Mean Rank	Sum of Ranks
cat1post - cat1pre	Negative Ranks	6[a]	4,00	24,00
	Positive Ranks	1[b]	4,00	4,00
	Ties	16[c]		
	Total	23		
cat2post - cat2pre	Negative Ranks	7[d]	8,79	61,50
	Positive Ranks	11[e]	9,95	109,50
	Ties	5[f]		
	Total	23		
cat3post - cat3pre	Negative Ranks	3[g]	6,00	18,00
	Positive Ranks	15[h]	10,20	153,00
	Ties	5[i]		
	Total	23		

a. cat1post < cat1pre
b. cat1post > cat1pre
c. cat1post = cat1pre
d. cat2post < cat2pre
e. cat2post > cat2pre
f. cat2post = cat2pre
g. cat3post < cat3pre
h. cat3post > cat3pre
i. cat3post = cat3pre

Test Statistics[c]

	cat1post - cat1pre	cat2post - cat2pre	cat3post - cat3pre
Z	-1,890[a]	-1,108[b]	-3,024[b]
Asymp. Sig. (2-tailed)	,059	,268	,002

a. Based on positive ranks.
b. Based on negative ranks.
c. Wilcoxon Signed Ranks Test

at Category 2). On the other hand, for the 2nd year students provided significantly more *Category 3* explanations in the post-test than in the pre-test (p<0.05, r = -0.445). In contrast, for the 1st year of the implementations, statistical significance was established only for the increase in *Category 2* explanations (p<0.05, r = -0.355).

In order to explain why the group of the 2nd year performed better in providing concise explanations after the intervention than the group of the 1st year of study, students' pre performances were tested for statistical difference. Attention was turned to the number of the explanations in each one of the three categories each student was able to give at the pre-test. For this data, a Mann-Whitney test was conducted[3] for each category to explore if the two groups (1st and 2nd year) differ considerably in relation to the number of explanations the participants gave at the beginning of each year. As shown in Figure 6.14 (which present the output of the SPSS analysis), no significant difference was found between the two groups in the number of explanations in each category before participating in the unit (z < 1.96 ignoring the minus sign, while p > 0.05 for all the categories). Therefore, the better results on the field for the 2nd year can not be attributed to differences between the groups before the intervention.

Figure 6.14: The output of SPSS analysis (Mann-Whitney Test) in comparing the number of explanations in each category in the pre-test, between the two groups (1st and 2nd year of study)

Ranks

	group	N	Mean Rank	Sum of Ranks
pre_1	0	33	29,20	963,50
	1	23	27,50	632,50
	Total	56		
pre_2	0	33	26,48	874,00
	1	23	31,39	722,00
	Total	56		
pre_3	0	33	28,47	939,50
	1	23	28,54	656,50
	Total	56		

Test Statistics[a]

	pre_1	pre_2	pre_3
Mann-Whitney U	356,500	313,000	378,500
Wilcoxon W	632,500	874,000	939,500
Z	-,474	-1,140	-,018
Asymp. Sig. (2-tailed)	,636	,254	,986

a. Grouping Variable: group

[3] Data violated the assumptions for parametric test; therefore a non-parametric test was implemented.

Rather, the difference between the two years in terms of students' is likely to be a reflection of the development of the instructor's ability in implementing the pedagogic practices in the classroom. In should be reminded here that in order to foster students' abilities in providing concise explanations, the instructor used a fictional person – Bobos- to whom students should address their views with an aim him to understand (view Section 1.3.2 for more details). As resulting from the field notes of the observations conducted, the researcher's impression is that for the 2nd year of the implementation, Bobos' role was more successful than in the 1st year. Specifically, during the 1st year of the study it took the instructor more time to make the students discover gradually Bobos' characteristics and as a result to understand his role in the teaching-learning procedure. Some characteristic comments of the students as evident in the researcher field notes, towards the end of the semester which exemplify the above, are the following:

> *I really cannot understand what Bobos really needs to get what I am saying…I mean… my explanation is obvious. Is he stupid or what? (Laughs),*

or addressing to the instructor:

> *I believe that the important thing is that* you *understand my argument. At the end of the day you* ar*e a professor, you can get what I am saying.*

In contrast, during the 2nd year the accumulation of experience helped the instructor to make students discover Bobos' role even from the first weeks of the unit. A crucial incident that took place in the 3rd week of the semester and it is noted is the researcher's diary is the following: At the end of a discourse between two pupils, one of them commented:

> *You know, I am a first-year university student and still I cannot get your explanation…how can you expect Bobos to understand? Try to explain in a more simple way…it is too complicated…*

Based on this comment, the instructor seized the opportunity even from the early stages of the semester to initiate conversations concerning the appropriateness of an explanation in a particular context. It is important to note that this student's comment soon became a kind of a 'slogan' among the pupils during their discourse in the classroom, initiating discussions about what more the explainee needs from the explainer so as to promote his or her understanding.

In summary, this section has dealt with the results of the data analysis on the change in students' ability to provide concise scientific explanations after participating in the intervention. A first notable feature of the analysis is that the number of student's most concise explanations in the pre-test is low for the vast majority of the items of the questionnaire for both years of study. Given that the participants belong to the first group of graduates after the curriculum reforms in upper-secondary school, their performances in the pre-test provide an indication of the learning outcomes of the reforms. In addition, the analysis of each item of the questionnaire separately has provided indication that maybe the more obvious an explanation is for some students, the less degree of information is provided by them to back up the conclusion. Although this hypothesis is supported only by the 1st year of data, whether the difficulty and the degree of conciseness of an explanation as perceived by students are correlated, is an open question. Focusing on the change in students' ability to provide concise explanation after the unit, data analysis showed that for both years of study at the end of the intervention students provided less *Category 1* explanations and more *Category 2* and *Category 3* ones than at the beginning. This shows a positive development in students' ability to provide concise explanations for both years. Yet, the analysis showed that for the 1st year of study statistical significance is obtained for the increase in *Category 2* explanations, in contrast to the 2nd year of study for which students gave significantly more *Category 3* explanations in the post-test than in the pre-test. The better results obtained in the 2nd year can not be attributed to difference between the two groups of participants in terms of the ability to provide concise explanations before participating in the unit, as no significant difference was established. Rather, the triangulation of these results with the researcher field notes has suggested that the difference between the two years is likely to be a reflection of the development of the instructor's ability in implementing the pedagogic practices in the classroom.

Summary

This chapter provides the results on RQ2 and relates to students' change in providing correct and concise scientific explanations after participating in the unit. For the 1st year of study, data analysis showed that students provided significantly a larger number of appropriate explanation in the post-test than in the pre-test, with the change being of medium size. In addition, at the end of the intervention there was a shift towards less *Category 1* explanations and more *Category 2* and *Category 3* ones; yet, statistical significance was established only for the increase in *Category 2* explanations, which was of medium size. Likewise, for the 2nd year of study, the analysis indicated that there was a significant positive change in students' ability to provide correct explanations at the end of the unit; in contrast, though, to the previous year, the change for the 2nd year is of large size. Moreover, for this year the pattern again suggests that there were more high-quality explanations at the end of the intervention than at the beginning. Yet, for this year of study students provided significantly more *Category 3* explanations in the post-test than in the pre-test, with the change being of medium size. The above indicate a positive development in students' ability to provide correct and concise explanation for both years of the study, while better results were obtained during the 2nd year of the implementation. Given than no significant difference was found between the two groups (1st year-2nd year) in the pre-test on the field, the better results for the 2nd year were attributed to the development of the instructors ability in implementing the pedagogic practices in the classroom. Finally, from the data analysis a couple of issues emerged, worthy to be further discussed and investigated: on the one hand the analysis has provided indication on the importance of the role of authority that a person has during the discourse; on the other hand, the data has suggested that the degree of difficulty of an explanation and the degree of information the students give to back up their claims might be negatively correlated. The points made above will be discussed thoroughly in Chapter 9.

CHAPTER 7

Results on students' responses to weak or fallacious arguments

This chapter provides the results of data analysis in RQ3, which falls in third area of interest in this study, the one of scientific arguments. In particular, the main research question, which is relevant to this chapter, is given below:

> *RQ3:* How do students respond to weak or fallacious arguments? What is the change (if any) in the way students respond to such arguments after participating in the course?

The chapter comprises of two main sections, in connection with the two parts of the research question under examination. At first, in relation to the first part of RQ3, Section 7.1 focuses on the way that students respond to arguments, which contain weaknesses or fallacies, both before and after participating in the course. Students' responses were analyzed qualitatively: typical patterns of students' responses were identified in relation to how students responded to each argument provided to them, while extracts from students' answers provide examples of the outcomes. Then, in relation to the second part of RQ3, Section 7.2 provides the quantitative results, regarding pre- and post-test students' performances on the field. Comparing the pre- and post-test students' responses, the change in the way they respond to weak or fallacious arguments, for the two years of implementing the intervention, is identified.

7.1. The way that students respond to weak or fallacious arguments

This section presents the results of data analysis in terms of how students respond to weak or fallacious arguments. As mentioned earlier in this study (view section 4.4.3), to achieve this aim data deriving from the second part of the questionnaire was used (*Items 7-10*). It is reminded that in these items students were given arguments containing weaknesses or fallacies and they were asked to state and to justify whether they are persuaded or not by them. As mentioned in Section 4.4.3, the primary aim of the analysis of students' responses has been to explore whether students managed to identify the weakness or the fallacy in each item.

However, the data analysis revealed cases, which fall beyond the polarization 'identification' - 'no identification' of the weakness or the fallacy of the argument. This was the outcome of an effort to address the credibility criterion, as the analysis was conducted in the base of identifying data, which refuted a priory categories formed. The following lines present the categories that were formed in each item, as grounded on the data, in terms of how students responded to each argument. The outcome of the analysis provides evidence for some difficulties that students confront, when responding to weak or fallacious arguments.

A) How did students respond to an argument, in which the conclusion is not an outcome of the premises? [1]

Item 7 aims to test whether students accept a conclusion, when it is not a consequence of the premises, regardless of the truth or falsity of the elements or the conclusion reached. The first part of the argument given to the students provides elements, which are irrelevant to the conclusion reached[2]. In addition, a missing component of the argument is that when the ice melts, the distribution of the mass will change, therefore the moment of inertia will change. Given that the angular momentum should be reserved, the rotational speed will change. The analysis of students' responses in this item resulted in the following categories:

A1. Explicit identification of the weakness of the argument: Some students state that – even if they agree with the conclusion - they are not persuaded by the argument, because it comprises of irrelevant or/and missing components. An example of such an answer is provided bellow (2nd year, post-test):

> *I am not persuaded by the argument. Even if I agree with the conclusion, I believe that Tom provides irrelevant evidence at the beginning of the argument. In addition, he should add that the moment of inertia changes, therefore the angular speed changes in order for L to remain stable.*

Cases like the above are regarded as the most successful attempts to identify the weakness of the argument: on the one hand, in these responses there is a distinction between being

---

[1] View Item 7 of the open-ended questionnaire, Appendix C

[2] Indeed, even if the ice which floats in the sea melts, then there would be no change in the level of the sea. In contrast, the level of the sea would rise, if parts of the ice on the ground melt.

convinced by an argument as a whole and accepting the conclusion reached. On the other hand, in such responses it becomes evident that the student has the ability to identify that the conclusion is not an outcome of the evidence provided. Likewise, successful attempts to identify the weakness of the arguments are the ones, in which students state that they are not persuaded by the argument, as there are elements missing or/and others that are not relevant to the conclusion, without, though, commenting on the conclusion. For example a student responded (1^{st} year, pre-test):

> *(I am not convinced). What has Pascal's principle to do with the fact that the rotational speed of the earth changes, if the ice in the North Pole melts and the level of the sea rises? It is irrelevant.*

In such cases, although no information is provided by the students in regard of the conclusion, it is clear that they evaluated the argument as a whole and they managed to identify the weaknesses of the argument.

A2. Implicit identification of the weakness of the argument: Some students display a more implicit response to the weakness of the argument: in a first view, they seem to focus on the conclusion - which they accept - without an explicit reference to the weaknesses of the argument, like in the cases above. Yet, it is worth noticing that these responses *do* contain the elements that are missing from the initial argument, so as to back up the conclusion reached. For example, a student from the 1^{st} year of study (pre-test) states:

> *(I am convinced) I agree. As we know, L = Iw. When the ice melts, the distribution of the mass in the earth will change. As a result, the moment of inertia will change. Given, though, that L remains constant, w will change as well.*

It could be argued that in such responses the students accept a conclusion, which is not an outcome of the elements provided; therefore, the attempt to identify the weakness of the argument should be regarded as unsuccessful. If this was the case, though, then the students wouldn't feel the need to complement the argument with the missing components, so as to back up the conclusion that they accept. Rather, it seems that the students acknowledge that there are missing components in the argument and provide them, even without an explicit reference to the weaknesses of the argument. Under such a rationale, in this study responses like the above are regarded as *partially* successful in identifying the weakness of the argument: they are not fully successful, given that the students mainly focus on the conclusion reached, rather than addressing explicitly the whole argument. However, they are considered successful, given that the students have examined the line of reasoning of the initial argument and address its weakness, even in an implicit way.

A3. No identification of the weakness of the argument: In responses in this category it is evident that the students did not manage to identify that in this argument the conclusion is not an outcome of the data provided. For example, some students state that they are convinced, without an explicit or implicit reference to the weakness of the argument. In the following response, it is evident that the student (1st year of study, pre-test) evaluated only the conclusion (which is indeed correct), without reflecting on the correctness or the conciseness of the argument.

> *(I am convinced) It is a correct argument. I agree that the rotational speed of the earth will change if the ice melts.*

Actually, some students' tendency to focus on the conclusion of an argument and accept or reject it without examining the line of reasoning or the data provided has been observed in the other items of the test, as well. This tendency will be discussed further, while providing the analysis of *Item 10* of the test. Other students reject the argument under the rationale that there is a scientifically incorrect premise in it: they state that ice floats because of Archimedes' principle and not because of Pascal's principle. Although these students have a point, it should be noted that this incorrect premise of the argument does not relate in any way to the conclusion reached; therefore, it does not affect the truth of falsity of the claim. The main weakness found in these students' responses, is the inability to identify which of the evidence provided relates to the conclusion and which is not. As a consequence, in this study such responses are regarded as examples of the least successful attempts to identify the weakness in an argument, in which the conclusion is not an outcome of the elements provided.

A4. Plead agnostic on the argument: Some students plead agnostic on the argument, because they do not know the relevant theory. Others do not justify their response at all. There are also a few cases, in which students state that they do not know and question whether the change in the rotational speed of the earth is significant. In such responses students' tendency to focus on and reflect about only the conclusion of an argument is again evident. Similar to the previous category, responses in this category are regarded as the least successful attempts to find the weakness of the argument, as there is no evidence that the students have followed the line of reasoning of the argument and have identified irrelevant or missing components in it.

B) How did students respond to an argument, in which scientifically incorrect elements lead to a wrong conclusion?[3]

Item 8 aims to test whether students are able to identify an incorrect premise in an argument, which leads to an incorrect conclusion. As evident from the argument given to the students, Nik reaches the wrong conclusion, because he assumes that N=W (which is incorrect, given that the person accelerates). In order to control the possibility that students are unable to find the incorrect data in the argument because of ignorance of the relevant theory, this item was analyzed in relation to the *Item 6,* in which students are explicitly asked whether $N \neq W$ in such a situation. By comparing the data of the two questions, it is possible to investigate, whether the students, who know the relevant theory, are able to identify an incorrect premise in an argument, which leads to a wrong conclusion. The parallel analysis of *items 6* and *8* resulted in the following categories:

B1. Knowledge of the theory **and** ***identification of the incorrect premise***: Responses in this category include cases where the students know that $N \neq W$, given that the system accelerates (as evident from their answers in *Item 6*). In addition, in *Item 8* they respond that they are not convinced by the argument, because the nominal force does not equal the weight; in other words, they comment explicitly on the incorrect premise of the argument, which leads to the wrong conclusion. A students' response, which exemplifies the above is the following (2nd year, post-test):

> *(I am not convinced) The nominal force changes in relation to the acceleration of the elevator. Therefore, the force that is exerted on the scales changes as well and it is not equal to the weight of the person.*

---

[3] View Item 8 of the questionnaire Appendix C

As evident from the example, the student manages to identify the falsity of the argument; therefore, responses like the above are regarded as successful attempts to find what is wrong with the argument. In addition, in such responses, there seems to be a positive relationship between knowing scientific facts and being able to recognize the according scientifically incorrect claim in an argument, which leads to a wrong conclusion.

B2. Knowledge of the theory* but *no identification of the incorrect premise: Students whose responses fall in this category state in *Item 6* that the nominal force does not equal the weight; therefore, they are regarded as having the scientific knowledge to address the argument in *Item 8*. Yet, in *Item 8* they do not manage to identify the wrong element in the argument, which leads to an incorrect conclusion. Therefore, responses in this category are considered unsuccessful attempts to identify the fallacy in the argument. The vast majority of these students claim that they are not persuaded by the conclusion reached; instead, they provide their own, *i.e*. that the reading of the scale changes. However, instead of opposing to the existing claims of the initial argument, they provide data to back up their conclusion. A characteristic response is the following (2^{nd} year, pre-test):

> *(I am not convinced) When the elevator moves up, the reading of the scales should be bigger than the actual weigh of the person; when it moves down, it should be less than the weight.*

Regardless of the scientific incorrect elements in the students' argument[4], the essential weakness found in this response in terms of argumentation is that the student does not address the data of the initial argument. Rather, the student seems to evaluate only the conclusion, without examining the correctness of the data provided. As it had been observed during the pilot work done for this study, counter arguments like the above are problematic, as they do not contribute to productive dialogues, rather they lead to parallel monologues (refer to Section 4.3.4). Nevertheless, other students whose responses fall in this category plead agnostic on the argument, while in a very few cases students state that they are persuaded by it, even if they have stated in *Item 6* that the nominal force does not equal the weight. In these two cases, no

[4] The student confuses movement i.e. velocity with acceleration. The reading of the scales does change indeed, but in relation to the acceleration of the elevator (positive or negative) and not to the velocity (up or down).

further justification was given by any of the students. In short, from students' responses in this category, it becomes evident that knowledge of the scientific theory does not *necessarily* relate to the ability to recognize a relevant scientifically incorrect claim in an argument.

B3. Ignorance of the theory* and *no identification of the incorrect premise: In the responses that fall in this category, students have stated that N=W in *Item 6*, therefore they possess incorrect scientific knowledge; in addition, as it could have been expected, they do not manage to find the incorrect premise in the argument in *Item 8*. For example, a student from the 2nd year in the pre-test, who commented in *Item 6* that N=W because of the 1st Newton's law, replies in this item that she is not persuaded, but fails to identify the wrong data that led to the wrong conclusion.:

> *'(I am not convinced.) The force that is exerted on the scales changes in relation to the movement of the elevator, i.e. it is different when we go up and when we go down.'*

This response is similar to the one quoted as an example in the previous category, given that the student evaluates only the conclusion, without addressing the data of the argument. Yet, in this case the inability to identify the incorrect premise could be attributed to lack of knowledge of the relevant theory, in contrast to the previous category, in which the students seem to have the knowledge required to deal with the argument.

B4. Ignorance of the theory* but *identification of the incorrect premise: Theoretically, it could have been expected that some students, who hold an incorrect view as evident in *Item 6*, might manage to identify the fallacy in *Item 8*. Yet, from the data obtained from this study, no such case was found in both years of this study. This provides a strong indication that the ignorance of scientific facts is a crucial factor affecting the ability to identify incorrect premises in scientific arguments.

C) How did students respond to a deductive argument which affirms the consequence? [5]

According to Zeidler (1997), one of the most common fallacies of the deductive arguments is affirming the consequence. Errors in syllogisms such as the above, might seem naïve, however

---

[5] View item 9 of the questionnaire in Appendix C

they do occur even within the scientific community[6]. Nevertheless, the analysis of students' responses in Item 9 resulted in the following categories:

C1. Not convinced **and** ***identifying that this is an invalid syllogism:*** The students whose responses fall in this category comment that this is an invalid form of argument; therefore, they are not persuaded by it. These students characterize the argument or the syllogism as '*invalid'*, '*irrational*', *'not logical'* or *'wrong'*. Others move one step forward, trying to explain or to exemplify why the argument is invalid. A quote from a student's response is the following (1st year of study, pre-test):

> *(I am not convinced.) It is not a valid syllogism. The fact that dinosaurs are extinct cannot necessarily be attributed to the melting of the ice; other reasons might have led to their extinction.*

In a similar vein, a student from the 2nd year of study, stated in the pre-test:

> *(I am not convinced.) It is possible, but not necessary that dinosaurs are extinct because of this reason. For example, just say that if I have a certain disease, then I will definitely die. Well, one day I die; does this prove that I have the disease?*

Responses like the above are regarded as successful attempts to identify the falsity of the argument.

C2. Not convinced **but** ***not identifying that this is an invalid syllogism:*** Data suggests that there are students, who - despite stating that there are not convinced by the argument - fail to recognize validity problems in it. A small minority of them offered no justification on why they are not convinced by the argument. The majority of them seem to focus on the conclusion and reject it, on grounds of 'truth' or personal beliefs. For example, the student who gave the following response (2nd year of study, post-test) rejects the conclusion, being based on personal beliefs:

> *(I am not convinced.)I know that dinosaurs are extinct because of another reason, and not because the earth flooded.*

[6] Barnes (1985), as quoted in Zeidler (1997), reports of different studies in which 25-33% of scientists accepted the following argument as valid: 'If the scientific hypothesis His correct, then the empirical event E will be observed. The event E is observed. Therefore, hypothesis H is correct.' Actually, the syllogism behind this argument and the one given to the students is the same.

Another example of a student's response, in which validity issues in the argument is not a matter of concern, is the following (1st year, pre-test):

> *(I am not convinced.)This is not possible, given that nowadays there are vast expanses of land. In addition, we have found fossils of dinosaurs, which could not have been preserved in the water. Therefore, logically, dinosaurs must have been extinct in the ice ages. Fossils can be preserved in the ice.*

This student examines the truthfulness of the conclusion and rejects it *('this is impossible'),* without addressing the line of reasoning in the initial argument; rather, she provides grounds to support her alternative theory. Another student chose to present an alternative theory about the extinction of the dinosaurs, as an explanation for why she is not convinced by the argument (1st year, pre-test):

> *(I am not convinced.) Their extinction does not have to do with the melting of the ice. In order for them* (the dinosaurs) *to survive, food is a major factor. The argument is not correctly documented.*

It could be argued that this student provided another possible factor, which could lead to the extinction of the dinosaurs; therefore, she might have commented on the validity of the argument, in an implicit way. A careful reading of the student's response, though, suggests that this hypothesis should be rejected: one the one hand, the student does not provide *a* possible reason, but *the* reason for the extinction, according to her beliefs; on the other hand, she comments on the 'correctness' of the argument, confusing truth with validity. Given that problems concerning the distinction between truth and validity in deductive arguments have been reported in other studies (for a review refer to Zeidler, 1997), this point will be discussed further in Chapter 9.

C3. Convinced by the argument: From the data gathered during the two years of study, both in the pre- and the post-test, there have been only two cases, where the students commented that they are persuaded by the argument: in one of them, the student offered no justification for his choice; in the other response, the student commented that *'this is a valid syllogism'*. This is an encouraging outcome, as it indicates that only a small minority of the sample accepts or characterizes as valid, deductive arguments, which affirm the consequence.

C4. Plead agnostic on the argument: The vast majority of students, who answered '*I do not kn*ow' in this item, provided no further justification for their choice. Except from not being able to identify validity issues in this item, maybe some of them did not feel confident enough to provide an answer, given that this is the only item, the content of which does not relate to physics. Nevertheless, what is worth commenting is that some of these students *did* question the validity of it; yet, they plead agnostic on the argument. For example, a student from the 2nd year (post-test) states:

> *(I do not know.) They might have been extinct because of other reasons, as well, like earthquakes; or because of lack of food.*

In responses such the above, it becomes evident that for some students even if validity concerns do exist, they are not enough to reject an argument. As Nickerson *et al*. (1985, p.112) comment, although people make the distinction between truth and validity, they '*fail to appreciate that in evaluating the logical soundness of a deductive argument validity alone is relevant.'*

D) How did students respond to an argument expressed by an expert, which provides a claim without any justification? [7]

As observed during the pilot work done for the study and as it turn up in the results of data analysis (refer to Section 6.2), the way students evaluate arguments is sometimes influenced by the expertise of the person, who expresses the argument. Item 10 aims to examine how students deal with arguments held by a person, who is regarded by them as 'experts' on the field. The argument is weak, as it comprises of a claim without any justification. An extreme case was purposefully selected (argument containing only a claim), so as exclude the possibility that students accept or reject the argument on grounds of evaluating the data provided. The following lines provide the outcome of the analysis of students' responses.

D1. Not convinced by the argument: Students whose responses fall in this category state that they are not convinced, because there is no justification for the claim provided. Most of these students comment briefly that the professor *'does not give evidence'*, that she *'does not explain what she supports'*, or that she *'does not provide scientific justification'*. These students do not

[7] View Item 10 of the questionnaire Appendix C

comment explicitly on the expertise of the person, who expresses the argument; yet, the professor's expertise does not seem to influence the way they deal with the argument. Other students add as well that they cannot accept such an argument just because it is expressed by an expert on the field. For example, a student from the second year of study (pre-test) comments:

> *(Not convinced.) She does not give any justification for what she supports. The fact that she is a professor of the Harvard University is not enough for me to accept her claim.*

In a similar vein, another student of the first year of study (post-test) comments:

> *(Not convinced.) Should I be convinced by her argument in which she explains nothing at all, just because she is a professor? I do not think so.*

In short, in such responses the students identify that this is a weak argument, as no justification is provided by the person who holds it. In addition, the expertise of the person on the field does not seem to be an adequate reason for them to accept such an argument.

D2. Examining and responding to the truthfulness of the claim: In responses fallen in this category the students focus on the claim and provide reasons for accepting or rejecting it. What differentiates these responses from the ones in the previous category is that students examine the correctness of the claim, rather than treating the argument given to them as a whole. For example, a student from the first year of study (post-test), who rejects the claim, states:

> *(Not convinced.) The force that the people will exert on the earth is equal with the force that the earth will exert on the people, because of the third Newton's law. Given that $\Sigma F=0$, I do not think that the orbit of the earth around the sun will change.*

A first notable feature in these responses is that they provide evidence about a confusion that some students make: the confusion between examining the correctness of a claim and examining a soundness of an argument. In other words, for some students the question '*are you convinced by the argument?*' is interpreted as '*do you agree with the claim of the argument?*'. Actually, such confusion has been observed in the other items of the test, as well. For example, in *Item 7*, it have exemplified that some students tend to focus only on the conclusion of the argument and to provide reasons for accepting or rejecting it, rather than examining the data of the argument provided. Likewise, in *Item 9*, it has been pinpointed that some students examine only the truthfulness of the conclusion, without addressing the line of reasoning in the argument provided. In chapter 9, the students' tendency to examine and respond only to the claim of the argument will be discussed more extensively. Nevertheless, the encouraging finding here is that students, whose responses fall in this category, have reflected on the claim of the professor. Regardless of whether they have accepted or rejected the claim, their choice has been the outcome of examination and based on justification; therefore, it can be assumed that the authority of the person who holds the argument is not an adequate reason for them to accept a claim without any evidence provided.

D3. Convinced by the argument: From the data gathered during the two years of study, both in the pre- and the post-test, there are have been cases, in which students comment that they are persuaded by an argument, which comprises only of a claim. Some of these students comment explicitly that it is the expertise of the person, who expresses the argument, which has driven their choice. A characteristic response from the second year of study (post-test), which exemplifies the above, is the following:

> *(Convinced.) I cannot question the view of a professor. I am I first year university student.*

In other responses fallen in this category, students offer no further justification for their choice; as a consequence, it is not possible to make assertions on whether their response is influenced by the authority of the professor. Whichever the case, though, for these students the fact that the argument consists of no evidence to back up the claim is not enough to reject the argument.

D4. Plead agnostic on the argument: Responses in this category comprise of the cases where students reply that they do not know. Most of them gave no justification, while a few of them commented that they do not know the relevant theory.

In summary, Table 7.1 provides the categories that they were formed for all the items of the test, as grounded on the data. Apart from summarizing the points made above, this table will be helpful in the following section, where qualitative result will be provided, regarding pre- and post-test students' performances on the field.

Table 7.1: The categories in students’ responses to weak or fallacious arguments

Fallacy/ weakness examined	***Categories of students’ responses***	
	Code	***Description***
A. *The conclusion is not a necessary outcome of the elements provided*	***A1***	Explicit identification of the weakness of the argument
	A2	Implicit identification of the weakness of the argument
	A3	No identification of the weakness of the argument
	A4	Plead agnostic on the argument
B. *Scientifically incorrect elements lead to a wrong conclusion*	***B1***	Knowledge of the theory *and* identification of the incorrect element
	B2	Knowledge of the theory *but no* identification of incorrect element
	B3	Ignorance of the theory *and no* identification of the incorrect element
	B4	Ignorance of the theory *but* identification of the incorrect element
C. *Deductive argument which affirms the consequence*	***C1***	Not convinced *and* identifying that this is an invalid syllogism
	C2	Not convinced *but no*t identifying that this is an invalid syllogism
	C3	Convinced by the argument
	C4	Plead agnostic on the argument
D. *An argument hold by an expert, which provides a claim without any justification*	***D1***	Not convinced by the argument
	D2	Examining and responding to the truthfulness of the claim
	D3	Convinced by the argument
	D4	Plead agnostic on the argument

7.2. The change in how students respond to weak or fallacious arguments

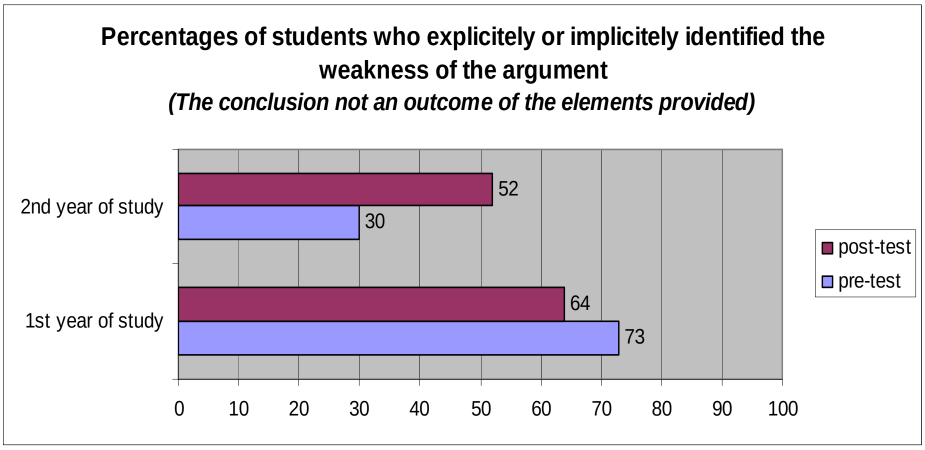
Based on the categories presented in the previous subsection, for each one of the *Items 7 - 10* of the open-ended questionnaire, each student response was identified in terms the category in which it fell. The overall number of students' responses in each category for each item was counted, both in pre- and post-test, for the two years of study. The quantitative results of the data analysis for each item are presented below in the format of tables and chart bars, along with a brief discussion about students' performances.

At first, Table 7.2 presents the numbers and the percentages of students' responses in each category for *Item 7* of the questionnaire, which sought to examine how students responded to an argument, in which the conclusion is not an outcome of the elements provided. Focusing on the pre-test students' performances, as evident from the figures in the table, it is the minority of students that identified explicitly the weakness of the argument (18% and 13% for the 1st and the 2nd year accordingly). In addition, it is worth commenting that for the 2nd year almost half of the students pleaded agnostic on the argument (48%). The above figures indicate the degree of the difficulty that students confront before entering the course in identifying the elements that are not necessary or that are absent for supporting the conclusion in the argument. Turning attention to the post-test students' performances, the responses in the category *'Implicit identification of the weakness'* outnumber the answers in the other categories for both years of study (40% in the 1st year and 35% in the 2nd year).

Table 7.2: The number of students' responses in each category, for Item 7, for pre – and post-test, for the two years of study. The percentages are provided below the absolute numbers.

Fallacy/ weakness examined	***Categories of students' responses***	**1st year (33 students)**		**2nd year (23 students)**	
		pre	**post**	**pre**	**post**
A. The conclusion is not a necessary outcome of the elements provided	A1. Explicit identification of the weakness of the argument	6 (18%)	8 (24%)	3 (13%)	4 (17%)
	A2. Implicit identification of the weakness of the argument	18 (55%)	13 (40%)	4 (17%)	8 (35%)
	A3. No identification of the weakness of the argument	4 (12%)	8 (24%)	5 (22%)	7 (31%)
	A4. Plead agnostic on the argument	5 (15%)	4 (12%)	11 (48%)	4 (17%)

Figure 7.1: The percentages of students who explicitly or implicitly identified the weakness of the argument, in the pre- and post-test for the two years of study.



On the one hand, this could be considered a fairly encouraging outcome, as it signifies that most of the students followed the line of reasoning of the initial argument and addressed its weakness, even in an implicit way (view the previous subsection for the rationale under which responses in this category are regarded as partially successful). On the other hand, though, post – test figures in the table show that one third of the students of the 1st year and almost half of the students in the 2nd year either did not identify the weakness of the argument or pleaded agnostic on it. This indicates that for a fairly large proportion of students in both years of study the difficulty in identifying unnecessary or missing components in an argument is still present after the intervention.

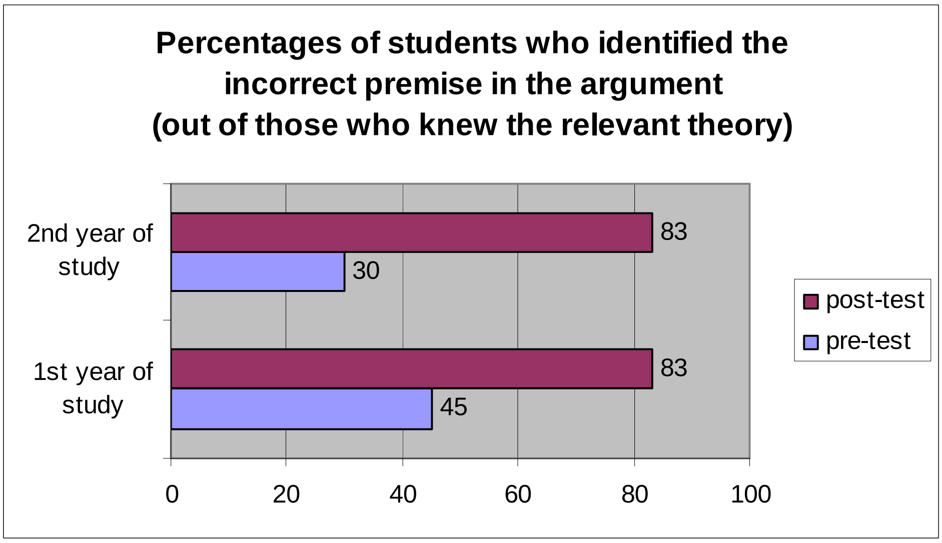
Nevertheless, in order to facilitate the comparison between pre- and post- students' performance on the item, data from Table 7.2 are presented in the format of chart bar in Figure 7.1. The chart bars illustrate the percentages of students, who explicitly or implicitly identified the weakness of the argument, in the pre- and post-test for the two years of study. As evident from this figure, in the 1st year of study fewer students in the post-test than in the pre-test managed to identify the weakness of the argument, with the decrease being of 9%. In contrast, the results for this item of the 2nd year are more encouraging, as 22% more students identified the weakness in the post- than in the pre-test. As it will be evident in the following lines -where the quantitative results for the other items are provided- this is the only item of the test in which there is such a controversy between the 1st and the 2nd year of study, and in which less students in the post- than in the pre-test found the weakness in the argument. One possible reason for this is that during the 2nd year of study some of the materials that had been developed to help students evaluate arguments (view Section 1.3.2) were enhanced: the experience of the previous year of implementing them in the classroom resulted in modifying some items so as to clarify some ambiguities, while a few additional items complemented the existing materials. Maybe the better results of the 2nd year could be attributed in some degree to these modifications; this does not explain, though, why in this particular item and not in the others as well, such a controversy between the two years is observed (regress in the 1st year and progress in the 2nd year). In short, for the controversy between the two years of study in this item, the researcher has no explanation supported by classroom observations and field notes. Rather, the researcher's impression is that for both years of study, students confronted difficulty in identifying unnecessary or missing components in an argument even at the last stages of the intervention.

In contrast to *Item 7*, quantitative results for *Item 8* of the test are consistent between the two years of study and indicate that for both years more students in the post-test than in the pre-test identified the fallacy. Results are provided in Table 7.3, which shows numbers and percentages of students' responses, in terms of how students responded to an argument, in which an incorrect premise leads to a wrong conclusion. A first worth noticing feature in this table is that for both years of study no response was found, in which students, who hold an incorrect view in *Item 6*, managed to identify the fallacy in *Item 8*. As it might have been expected, this finding

Table 7.3: The number of students' responses in each category, for Item 8, for pre – and post-test, for the two years of study. The percentages are provided below the absolute numbers.

Fallacy/ weakness examined	***Categories of students' responses***	**1st year (33 students)**		**2nd year (23 students)**	
		pre	**post**	**pre**	**post**
B. Scientifically incorrect elements lead to a wrong conclusion	B1. Knowledge of the theory *and* identification of the incorrect element	5 (15%)	20 (60%)	4 (17%)	15 (65%)
	B2. Knowledge of the theory *but no* identification of incorrect element	6 (18%)	4 (12%)	9 (40%)	3 (13%)
	B3. Ignorance of the theory *and no* identification of the incorrect element	22 (67%)	9 (28%)	10 (43%)	5 (22%)
	B4. Ignorance of the theory *but* identification of the incorrect element	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Figure 7.2: The percentages of students who managed to identify the incorrect premise in the argument, out of those who knew the relevant theory, in the pre- and post-test, for both years of study



indicates that the ignorance of scientific facts seems to be a crucial factor affecting the ability to identify incorrect premises in scientific arguments. However, a careful reading of the data in the table shows that the opposite is not always the case. In other words, the knowledge of the scientific theory does not seem to be necessarily related to the ability to address the correctness of the data in the argument. Indeed, as evident from Table 7.3, before the intervention, in the 1st year of study, 6 out of the 11 students, who replied correctly to *Item 6*, did not managed to identify the incorrect premise in the argument; for the 2nd year of study, the according number is 9 students out of 13. The above figures signify that for both years of study and before the intervention, less than half of the students, who had the scientific knowledge to address the argument, identified the wrong element in it (view Figure 7.2 for the actual percentages). As discussed in the previous subsection, most of these students responded only to the conclusion of the argument, without evaluating the data that supported it. In contrast, at the end of intervention, the vast majority of the students, who knew the relevant theory, managed to find the fallacy in the argument. Indeed, Figure 7.2 shows that at the end of intervention and for both years of study 83% of the students, who knew the theory found the fallacy in the argument. This signifies an increase of 38% for the 1st year and of 53% for the 2nd year in the number of students, who recognized the incorrect premise - out of those who knew the relevant theory.

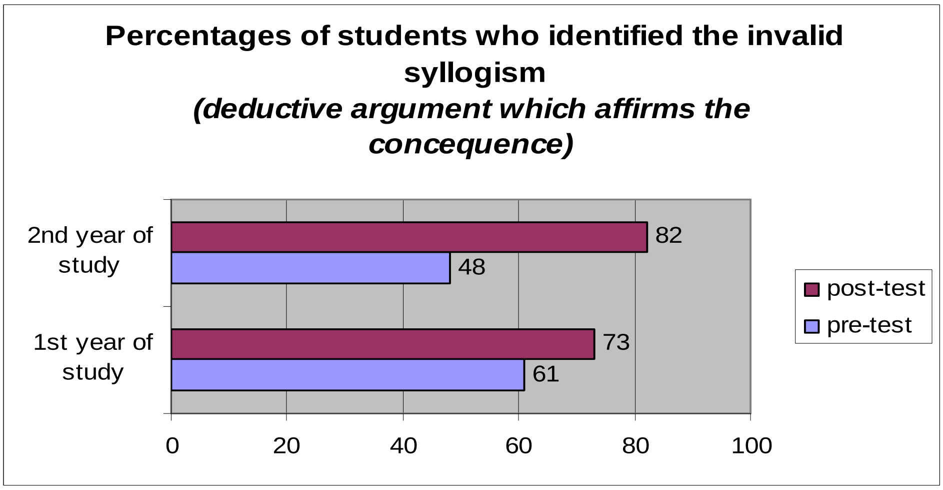
Comparing the outcome of the analysis between *Item 7* and *Item 8*, it becomes evident that better results are obtained in relation to students' difficulty in finding significantly incorrect data in an argument, rather than in identifying irrelevant and missing components in it. This outcome is supported by the observations in the classroom during both years of study. As it has been explained previously (view Section 1.3.2), in order to help students evaluate their classmates arguments, a set of materials in the form of concept-map were implemented in the classroom. In each item the instructor asked the students to evaluate the argument in terms of *a)* the scientific correctness of each component of the argument and *b)* the way the components are linked. The researcher's field notes suggest that it was easier for students to identify wrong components in an argument, rather than follow the line of reasoning in it and comment on possible unnecessary or missing components, even during the last weeks of the intervention. Indeed, when the argument under examination contained wrong evidence, in most of the cases even at the beginning of the intervention this fallacy became the centre of students' discussion. In contrast, when the argument was weak in terms of the line of reasoning, students tended to focus on the conclusion and debate on whether they agree with the claim. Many instances have been noted in the researcher's diary, in which the instructor's attempts to change the focus towards the line of reasoning of the argument were not successful, even in cases where the gap between the data provided and the conclusion reached was obvious. Given that students' difficulties in argumentation is an issue of prominent research interest (view Section 2.2.3), the above finding are discussed more extensively in Chapter 9.

Turning attention to *Item 9*, which sought to examine how students responded to a deductive argument that affirms the consequence, quantitative results are presented in Table 7.4. Although the syllogism under evaluation might be regarded as naïve, pre-test results show that only 61% and 48% for the 1st and the 2nd year of study accordingly, stated that this is an invalid syllogism. As discussed in the previous section of the chapter, apart from the students, who pleaded agnostic on the argument, the others confused truth and personal beliefs with validity. This outcome is in line with other studies on the field (view previous subsection), which highlight that such errors in syllogism occur even in the scientific community. As for the results at the end of the intervention, the percentages of the students who found validity problems in the argument is increased to 73% and 83% for the 1st and the 2nd year of the study accordingly (view Figure 7.3). This increase is rather encouraging especially for the 2nd year of study. Yet, a number of caveats should be placed on the interpretation of these data. At first, it should be noted that during the lessons the students did not have the chance to evaluate many arguments like the above, given that a very few arguments with validity problems were expressed by students during discourse. Secondly, as it can be seen in Table 7.4, it is mainly the decrease in the number of students, who pleaded agnostic in the pre-test (which is rather high in both years of study), which contributed to the increase in the numbers of the students, who found the fallacy in the argument after the intervention. As it has been pinpointed in the

Table 7.4: The number of students' responses in each category, for Item 9, for pre – and post-test, for the two years of study. The percentages are provided below the absolute numbers.

Fallacy/ weakness examined	*Categories of students' responses*	**1st year (33 students)**		**2nd year (23 students)**	
		pre	**post**	**pre**	**post**
C. Deductive argument which affirms the consequence	C1. Not convinced *and* identifying that this is an invalid syllogism	20 (61%)	24 (73%)	11 (48%)	19 (82%)
	C2. Not convinced *but not* identifying that this is an invalid syllogism	4 (12%)	3 (9%)	4 (17%)	2 (9%)
	C3. Convinced by the argument	1 (3%)	1 (3%)	1 (4%)	0 (0%)
	C4. Plead agnostic on the argument	8 (24%)	5 (15%)	7 (31%)	2 (9%)

Figure 7.3: The percentages of students who identified the invalid syllogism in a deductive argument, which affirms the consequence in the pre- and post-test, for both years of study



previous section, we cannot exclude the possibility that some students in the pre-test pleaded agnostic on this item because they did not feel confident enough to provide an answer (this is the only item of the test that does not relate to physics). The above indicate that the better results obtained in the post-test for this item might be a reflection of students' confidence in evaluating arguments after the intervention, rather than regarded as a change in their ability to evaluate deductive arguments.

Finally, moving to the last item of the test (*Item 10*), Table 7.5 presents the numbers and the percentages of students' responses in an argument, which contains a claim without any justification, as expressed by someone who has expertise on the field. It is reminded that the primary aim of this item has been to examine how the expertise of the person, who holds an argument influences the way that students accept or reject it. In order to support this aim, it was judged necessary to minimize the possibility that students evaluate the argument in terms of the evidence the expert provided. That is the reason why such an extreme case was selected, i.e. an argument containing only a claim without any justification. Regardless of the primary aim of the item, though, the qualitative analysis of students' responses revealed another issue worth commenting: the confusion that some students make between evaluating the soundness of an argument and examining the truthfulness of its central claim. As evident in Table 7.5 almost one third of the responses of the 1st year and approximately half of the ones of the 2nd year (both in the pre- and post-test) fall in the category '*examining and responding on the truthfulness of the claim*'. This signifies that a fairly large proportion of students in both years

Table 7.5: The number of students' responses in each category, for Item 10, for pre – and post-test, for the two years of study. The percentages are provided below the absolute numbers.

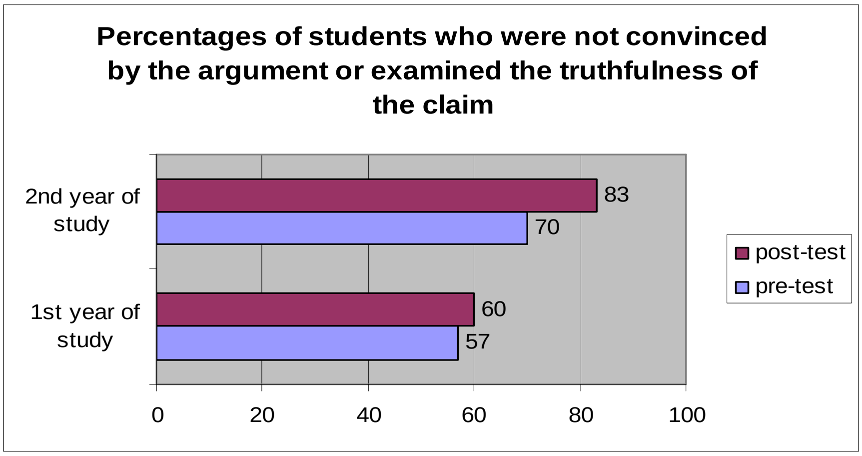
Fallacy/ weakness examined	***Categories of students' responses***	**1st year (33 students)**		**2nd year (23 students)**	
		pre	**post**	**pre**	**post**
D. An argument hold by an expert, which provides a claim without any justification	D1. Not convinced by the argument	10 (30%)	10 (30%)	4 (18%)	8 (35%)
	D2. Examining and responding on the truthfulness of the claim	9 (27%)	10 (30%)	12 (52%)	11 (48%)
	D3. Convinced by the argument	1 (3%)	4 (13%)	0 (0%)	1 (4%)
	D4. Plead agnostic on the argument	13 (40%)	9 (27%)	7 (30%)	3 (13%)

of the study focused on the claim and justified whether they agree or disagree with it, rather than rejecting an argument in which the claim is supported by no evidence.

Actually, as discussed previously, students' tendency to evaluate only the conclusion of an argument is evident in the other items of the test and has been observed during students' discourse in the classroom. What differentiates this item from the others, though, is that in the previous arguments of the test the line of reasoning and some evidence to support the conclusion were evident. In these cases, the students' tendency to focus only to the conclusion rather than addressing the whole argument could be attributed to the fact that such strategy requires less investment of cognitive energy; therefore, it might be favoured by the students. Another possible reason that has been pinpointed by Zeidler (1997) is that maybe students find ways to ignore evidence in an argument, if it conflicts with their initial beliefs–similar to scientific misconceptions. The outcomes of the analysis in this item, though, signify that even when no evidence is provided to back up a claim – therefore issues of ignoring evidence or investing less energy do not apply- , students follow the same strategy. For the moment, the question about the possible reasons for why some students tend to evaluate only the conclusion rather than addressing the whole argument remains open; it will arise again in the last chapter of the thesis.

Nevertheless, returning to the main aim of *Item 10*, i.e. the way that the expertise of the person, who expresses an argument influences the way students evaluate it, data from Table 7.5 are presented below in the format of chart bar. Figure 7.4 presents the percentages of responses, which fall in the first two categories of the table. These responses are considered successful attempts to reply to the argument – when examining the issue of the expertise - under the following rationale: As explained in the previous section, in the first category the students stated explicitly that they are not convinced by such an argument, given that it provides no justification. Either they have commented explicitly on the authority of the professor who expressed the argument or not, the professor's expertise did not constrain these students from rejecting an argument that contains only a claim. In the second category, students seem to have

Figure 7.4: The percentages of students who were not convinced by the argument or examined the truthfulness of the claim in the pre- and post-test, for both years of study

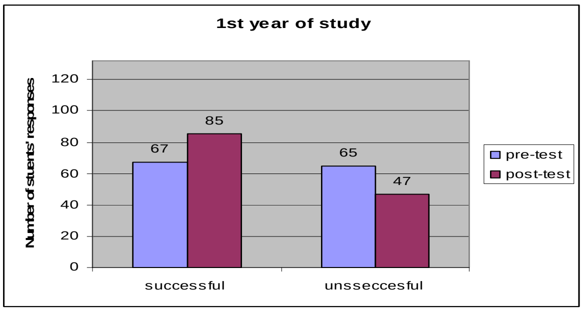


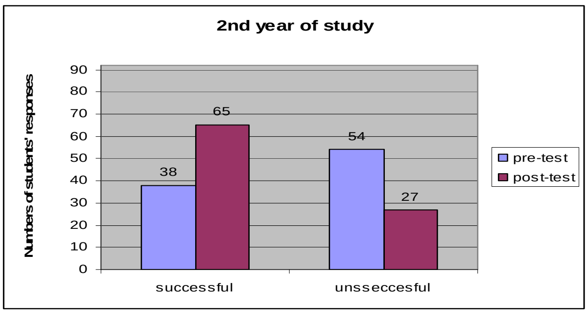
reflected on the claim of the professor. Given that their statements on whether they agree with it or not are justified, it can be assumed that they have examined the soundness of the claim. In other words it seems that the expertise of the professor is not an adequate reason for them to accept a claim without any evidence. There is no question that the data in the other two categories do not allow to be certain that all of the students, who were either convinced by the argument or pleaded agnostic to it, were influenced by the expertise of the professor (given that not all of them justified their answers or commented on the authority of the professor somehow). Considering, though, the extreme case that has been selected for this item (no evidence is given at all in this argument) it is highly possible that the expertise of the person, who holds the argument influenced in some degree the students, who were convinced or did not provided a justified reply to the argument. In any case, the results presented below, should be read having in mind the above rationale in the interpretation of the data. .

Nevertheless, as evident from Figure 7.4, in the 1st year of study the percentages of the students' responses in the first two categories remained almost unchanged (around 60%). This signifies that 4 out of 10 students were either convinced or pleaded agnostic to an argument that contains only a claim. This is a very big percentage – especially in terms of the post-test results- if we consider that the instructor's prompts towards students to justify their view and to evaluate evidence were prominent in any lecture. Better results are obtained in the 2nd year of study, as there is an increase of 13% in students' responses who fall in the first two categories, with the post-test figures increase to 83%. Given that in the 2^{nd} year of study pre-test percentages in this item are higher than in the 1^{st} year (70% and 57% accordingly), it could be assumed that the difference between the two years of study in this item is may be attributed to the difference between the samples before entering the course.

After providing the results for each item of the questionnaire separately, and aiming to obtain an overview of the change in the way students respond to weak or fallacious arguments, data from the questionnaire was analyzed in the following way: Attention was turned to each student performance in the *Items 7-10* of the questionnaire and counts of the numbers of successful responds to the arguments were made in the pre and post-test. Based on the rationale that have been presented in the previous section and bearing in mind the assumptions provided in the lines above, for each item successful responses are considered the following: for *Item 7*, categories A1 and A2; for *Item 8*, category B1; for *Item 9*, category C1 and for *Item 10*, categories D1 and D2 (refer to Table 7.1 for the description of each code, and in the previous section for the rational under which the responses in this category are considered successful). Figure 7.5 shows the numbers of successful and unsuccessful responds to all the items, as provided by the sample of students, before and after the intervention for both years of study.

Figure 7.5: The numbers of successful and unsuccessful responses to the arguments, as provided in all the items of the questionnaire for the 1^{st} year (n = 132) and the 2^{nd} year (n = 92)





It is reminded that given the difference in the sample between the two years, the total numbers of responses between the two years are different: 132 (33students X 4items) and 92 (23students X 4items) for the 1st and the 2nd year accordingly. A first worth commenting outcome of this analysis is that for both years of the study pre-test performances are rather disappointing: indeed for the 1st year the total numbers of successful and unsuccessful response are practically the same, while for the 2nd year the unsuccessful attempts outnumber the ones in the other category. This is an indication of the degree of the difficulty that students confront before entering the unit to respond to weak or fallacious arguments, which has also been pinpointed while providing the results of each item separately. Moving the focus on the change between the pre- and the post-test performances, as evident from the figure above for both years of the study there is a shift at the end of the intervention towards more successful responses and less unsuccessful ones.

In order to see whether this change is statistical significant the following procedure was followed. For each student, the numbers of successful responses for all the items in the pre - and post-test was calculated and are presented in Tables 7.6 and 7.7. There are 4 items in the test, therefore the numbers range from 0 (no successful response) to 4 (successful responds in all items). The Wilcoxon Signed ranks test was implemented for this data, so as to see if there is a significant difference in the number of successful responses provided by the students between pre – and post-test. Figures 7.6 and 7.7 illustrate the output of the SPSS program for the 1st and the 2nd year accordingly. The analysis showed that for the 1st year of study the number of successful responds provided by the students is significantly higher in the post-test than in the pre-test, but the change is of small size (z = -2.087, p< 0.05 and r = 0.256). Statistically significant results are also obtained for the 2nd year of study; yet, the change in the number of successful responses for this years is of large size (z = -3.453, p< 0.05 and r = 0. 509). In short, the above signify that for both years students provided significantly more successful responses to weak and fallacious arguments after the intervention than before. Better results are obtained in the 2nd year, in which the change in the number of responses is of large size, in contrast to the 1st year, where the change is of small size.

Table 7.6: The number of successful responses provided by each student

(Pre- and post-test, 1st year of study)

Student	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Pre-score	0	1	3	3	3	2	1	2	2	4	2	1	1	3	3	2	3	2
Post-score	2	4	3	2	1	3	3	4	2	4	1	3	2	3	1	1	3	3

Student	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**
Pre-score	1	4	1	3	0	1	2	3	2	3	1	3	2	1	2
Post-score	3	3	2	3	3	2	2	1	3	4	1	4	4	2	3

Figure 7.6: The output of the SPSS program – Wilcoxon Signed Ranks Test – 1st year of study

Ranks

		N	Mean Rank	Sum of Ranks
post - pre	Negative Ranks	7[a]	12,43	87,00
	Positive Ranks	18[b]	13,22	238,00
	Ties	8[c]		
	Total	33		

a. post < pre

b. post > pre

c. post = pre

Test Statistics[b]

	post - pre
Z	-2,087[a]
Asymp. Sig. (2-tailed)	,037

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Table 7.7: The number of successful responses provided by each student

(Pre- and post-test, 2nd year of study)

Student	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**
Pre-score	2	2	0	0	1	3	0	2	0	3	1	2	3	3	2	2	3	3	0	2	0	2	2
Post-score	3	2	3	2	3	4	3	3	3	3	2	4	2	4	3	3	4	3	2	2	3	2	2

Figure 7.7: The output of the SPSS program – Wilcoxon Signed Ranks Test – 2nd year of study

Ranks

		N	Mean Rank	Sum of Ranks
post - pre	Negative Ranks	1[a]	5,00	5,00
	Positive Ranks	16[b]	9,25	148,00
	Ties	6[c]		
	Total	23		

a. post < pre

b. post > pre

c. post = pre

Test Statistics[b]

	post - pre
Z	-3,453[a]
Asymp. Sig. (2-tailed)	,001

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Similar to the area of scientific explanations (for which the data analysis revealed better results for the 2nd year), the difference between the two years in this case cannot be attributed to differences between the groups before the intervention. This was the outcome of a Mann-Whitney test, which was conducted to investigate if the two groups (1st and 2nd year) differ considerably in terms of the number of successful responds the participants provided at the beginning of each year. The output of the SPSS analysis is provided in Figure 7.8. Although the mean rank for the 1st year is bigger for the 1st year than in the 2nd, as shown in this figure, the z value is smaller than 1.96 (ignoring the minus sign), while $p > 0.05$. The above numbers signify that there is no significant difference between the 1st and the 2nd year, in terms of the number of successful responds the students were able to provide before the beginning of the course. Rather, the better results for the 2nd year of study might be attributed in a degree to the enhancement of the materials given to the students to help the in the process of argumentation. As mentioned previously, the implementation of the materials in the format of concept-maps with competing arguments in the first year showed that some items had ambiguities, so required modifications. In addition, during the 2nd year, additional to the existing items were developed and implemented in the classroom, as the experience of the 1st year showed that students might need more occasions to reflect on and evaluate competing arguments in such a format. On the other hand, field notes suggest that in the 2nd year of study the instructor made more use of argumentation prompts, while competing arguments were discussed, such as ‘ *what is the evidence for rejecting this argument’*, ‘*can you think of any reason for supporting your view’*, ‘*what do you think is the main problem with this argument’*. In short, the researcher’s

Figure 7.8: The output of the SPSS program – Mann-Whitney test - for comparing the numbers of successful responds to weak or fallacious arguments, before the intervention between the two years of study.

Ranks

	group	N	Mean Rank	Sum of Ranks
pre	1	33	30,33	1001,00
	2	23	25,87	595,00
	Total	56		

Test Statistics[a]

	pre
Mann-Whitney U	319,000
Wilcoxon W	595,000
Z	-1,046
Asymp. Sig. (2-tailed)	,295

a. Grouping Variable: group

impression - based on the observations and informal discussions with the instructor- is that for the 2nd year of study the instructor himself had conceptualized better the norms of argumentation, so he made better use of the materials in the classroom. This could have been expected, given that the previous year was the first time that the instructor was asked to implement such materials. The point made here is that the ability to argue effectively is a long-term process not only for the students – as current literature suggest (for example Osborne *et al*, 2004). Teachers and instructors themselves need time to develop a theoretical understanding of the nature of scientific arguments and they need experience and practice to implement successfully strategies aiming at eliciting satisfactory arguments from the students. It is important to have this point in mind, when significant improvements are expected after a short time of implementing interventions towards developing argumentative abilities.

Summary

This chapter provides the research results on the third objective of this study, which relates to the way that students respond to weak or fallacious arguments and to the change in their responses after participating in the course. Qualitative analysis of students' responses resulted in the formation of categories showing how students dealt with arguments: *a)* in which the conclusion is not an outcome of the data provided; *b)* in which an incorrect element leads to a scientifically wrong conclusion; *b)* deductive one which affirms the consequence and *d)* containing a claim by an expert with no justification. After counts of students' responses in each category for each item were made, the following issued emerged, which will be furthered discussed in the last chapter of the thesis. First, better results were obtained after the intervention in relation to students' difficulty in finding incorrect data in an argument, rather than identifying irrelevant or missing components in it. Secondly, problems concerning the distinction between truth and validity in deductive arguments were evident in some students' responses. Third, it was observed that some students –even after the implementation- have the tendency to focus and respond only to the claim of an argument, rather that addressing and evaluate the whole argument. In relation to the change in the way students' respond to weak or fallacious arguments, statistical analysis revealed that for both years of study, students provided more successful responds to such arguments after the intervention than before. Yet, better results were obtained for the 2nd year as the change was of large size, in contrast to the previous year, in which the change was of small size. Given that no statistical significance was found in the area between the groups before the intervention, the better results of the 2nd year were attributed to a) the modification of the materials given to the students to help the in the process of argumentation in this period; b) the instructor's better conceptualization of the norms of argumentation, and the better use of the materials in the classroom.

CHAPTER 8

Results on the quality of students' arguments

In this final chapter – in relation to the results of data analysis - the interest remains on the area of scientific arguments, but the focus is shifted towards the quality of students' written arguments. The main objective of the research, which is relevant to this chapter, is provided below:

> *RQ4:* What is the change (if any) in the quality of students' written arguments after participating in the course?

At the early stages of the research, students' arguments were intended to be analyzed by the use of Toulmin's Argumentation Pattern (TAP), as implemented by Erduran *et al.*'s (2004) and as adapted for the purposes of this study (view Section 4.4.3 for further details). However, the data analysis revealed that the theoretical framework chosen for evaluating the quality of arguments needs to include features, additional to the ones initially planned. Given the limitations that were found during the analysis, this chapter begins with a brief discussion about the need for modification and of the theoretical framework initially chosen (Section 8.1). Following this, the expanded theoretical framework for evaluating the quality of written arguments is presented, while examples of students' arguments are provided to illustrate how it has been applied to the data (Section 8.2). Finally, in Section 8.3 the results of the data analysis after the application of the framework are provided, in a view of identifying the change in the quality of students' arguments after participating in the course.

8.1. Modification of the theoretical framework: the rationale

As mentioned in Chapter 4 of this study, the theoretical framework chosen initially for analysing scientific arguments in terms of the quality is an adaptation of Erduran *et al.*'s (2004) methodological approach. It is reminded that the quality of an argument was planned to be judged on the presence or absence of rebuttals and that students' arguments were intended to be classified in three broad categories:

- *Low-level arguments* - when an opposition consisted of counter arguments that were a simple claim versus a counter claim.
- *Middle-level arguments* - when an argument consisted of claims with data or backings but with no rebuttals.
- *High-level arguments* - when rebuttals were evident, in direct reference to a piece of evidence (data, warrant or backing) offered - thereby engaging with the present argument.

As evident, the above mentioned framework can be applied, only when there are oppositional episodes in argumentation, in other words when students are not convinced by their classmates' arguments and offer an oppositional claim. The decision to focus only on oppositional arguments had been informed by recent research findings, which suggested that non-oppositional episodes tend to be unsophisticated in terms of scientific discourse structure. For example, Clark & Sampson (2006) when analysing online argumentation found a number of non-oppositional episodes in their research (140 out of 416), and classified them (in terms of whether they consisted or not of claims, data, warrants). Yet, they argued that in these arguments students tended to accept what it is written and move onward. Under such a rationale, they decided not to use these episodes, when evaluating the quality in students' argumentation.

During the analysis for this research, though, data revealed that the above is not always the case. Similar to Clark & Samson's (2006) results, a considerable number of non-oppositional arguments were identified in this study: these included the cases, when the response to an argument offered further support to the claim or the student chose to be agnostic about the claim by questioning some piece of evidence in the argument. Yet, in this study, not all non-oppositional arguments, which were provided by the students, appeared to be unsophisticated. In contrast, data analysis showed that they differed in terms of quality. More specifically, some students, who stated that they were persuaded by the argument, offered further support to the claim or further clarification; other students just accepted or rephrased the claim of the initial argument. Given the differences observed between the ways students provide non-oppositional arguments, the researcher decided that such arguments should merit further attention. This resulted to the consideration that the theoretical framework for evaluating arguments should be expanded, so as to include non-oppositional episodes along with oppositional ones. The following lines present briefly the rationale, under which the framework was modified.

A review of current literature does suggest that the assessment of the overall quality of an extended written argument or dialogic argument seems to be problematic. Apart from the difficulty to make objective distinctions between the various Toulmin components, some researchers report that the structural analysis of dialogic argumentation needs to include additional to Toulmin's argumentative operations, such as clarification, query, or support for another's claim (for a review, view Erduran, 2008). This last remark concerns arguments where students do not oppose to an argument. Some studies on evaluating the quality of students' arguments have acknowledged this deficiency, but chose to focus only in oppositional arguments, as described above (for example, Clark & Sampson, 2006).

Yet, it can be supported that when evaluating the quality of students' argumentation, non-oppositional arguments could and *should* be used, along with oppositional ones. Maybe, implicit to focusing only in oppositional arguments, while evaluating scientific discourse is the everyday sense of argumentation, which suggest that it is a competitive interaction with 'winners' and 'losers' and which contains mainly oppositions. Yet, both science educators and philosophers of science have extended this everyday sense of argumentation: they seem to agree that scientific discourse is a collaborative procedure, in which participants work together to resolve an issue (Osborne *et al.*, 2004). Under this perspective, the procedure of resolving an issue does not only contain controversies, but involves as well sharing in the construction of explanations or questioning. Therefore, apart from the ability to offer rebuttals when opposing to an argument, the ability to clarify the meaning or offer further support to another's claim are also important aspects in the ability to argue effectively in educational environments.

In addition, observations of students' discourse in the classroom for the two years of this study reinforce the theoretical position supported above, about the significance of non-oppositional episodes in the process of argumentation. Indeed, in cases when two students, who participating in a debate offered two opposing arguments and the instructor asked the other students to comment and take sides, the following remarks had been noted in relation to which arguments contributed to an effective dialogue: Some students chose to oppose to one of the arguments, while others choose to support one of them. In the second case, when the student supplemented the missing components of the argument, or clarified the meaning when there was an ambiguity, he or she contributed positively to the process of argumentation. In contrast, when the student rephrased the claim or repeated the data when asked to say why he is convinced by the argument, in most of the times the discourse came to an end.

In short, the analysis of students' written arguments and the observations of students' discourse suggest that – although the theoretical framework initially chosen to evaluate the quality of arguments is useful – it does not address all the aspects evident in the data in this study. That is because it considers only oppositional arguments. In addition, from a theoretical perspective, it has been supported that non-oppositional episodes in argumentation should merit some significance. That is because in educational settings argumentation does not only include contradictions, but also collaboration towards better understanding. Under such a rationale and as an outcome of the data analysis, in the following section an expanded theoretical framework is presented and proposed for evaluating written arguments.

8.2. Presenting and exemplifying the expanded theoretical framework

Based on the rationale presented previously, this section offers the schema of analysis of students' written arguments in terms of their quality, as applied to the data of this study. For evaluating the quality of students' arguments, data deriving from the second part of the questionnaire was used (*Items 7-9*). In these items students' were given arguments, which offered justification, in contrast to the other items, which comprised only by a claim. At first, students' arguments were identified as being *A) oppositional* arguments (contradicting to the claim of the offered argument) or *B) non- oppositional* ones (supporting or questioning the claim of the offered argument). For each of these two categories three levels of quality were formed. Tables 8.1 and 8.2 present the levels of the quality of arguments in each category, along with the description of each level of students' arguments.

For non-oppositional episodes, as evident in Table 8.1, the quality in students' arguments was judged by the ability to offer rebuttals. Arguments with rebuttals are considered of better quality than those without, given that oppositional arguments without rebuttals '*have the potential to continue for ever without no change of mind or evaluation of the quality of the substance of an argument'* (Osborne *et al.*, 2004). In this study, which involves students' written arguments as a response to a provided argument, rebuttals are considered as statements that provide criticism to a piece of evidence or the line of reasoning of the initial argument. High-level arguments consist of responses in which it is evident that both the provided and the alternative 'theory' have been considered by the student, arguing that the alterative theory is more correct or the provided one is wrong. As Kuhn (1991) has argued, this is the most complex skill in terms of argumentation. In other words, these responses consist of statements that are direct reference the provided argument. In contrast, in middle-level arguments, there is

Table 8.1: Levels of arguments in oppositional episodes in terms of quality

Levels of argument	**Code**	**Description**
Low-level	Op1	When the opposition consisted of only counter arguments, that were a simple claim versus a counter claim
Middle-level	Op2	When an argument consisted of claims with data or backings, but no rebuttals
High- level	Op3	When rebuttals were evident (statements in direct reference to a piece of evidence, thereby engaging with the present argument)

Table 8.2: Levels of arguments in non- oppositional episodes in terms of quality

Levels of argument	**Code**	**Description**
Low-level	Nop1	When the response consisted of arguments that the claim was re-stated or rephrased
Middle-level	Nop2	When an argument consisted of highlighting a piece of evidence of the initial argument
High- level	Nop3	When a clarification or additional features were provided by the students to support or query the claim

no evidence that the students have integrated the two positions (the one provided and the one supported), given that there is no statement in their responses, which engage to the provided argument, except from the claim. In other words, middle-level oppositional arguments consist of a counter-claim supported with evidence (data or backings), but no rebuttals. Finally, in low-level oppositional arguments there is only a claim versus a counter-claim, without any justification. The examples that follow are provided to illustrate how the analysis has been applied to the data, for oppositional episodes.

Examples of low-level oppositional argument: The following extract is taken from the pre-test in the 2nd year or study; it is a students' response to *Item 7* of the test:

> *I don't agree. I believe that - even this happens - the change in the rotational speed would be small and non-significant.*

In this response the students rejects the claim of the initial argument (that the rotational speed will change), without, though, offering any support for this position. In a similar vein, a student who rejects the claim in *Item 8* of the test (1st year pre-test) states:

> *(I am not convinced.) I believe that the reading of the scales changes, when the elevator moves.*

The above arguments consist only of a claim versus a counterclaim, with either data to support it, nor any rebuttal; therefore, they are considered low- level oppositional arguments.

Examples of middle-level oppositional argument: In the following example (2nd year, pre-test), the student provides some evidence for the counter-claim:

> *(I am not convinced) The reading of the scales is not the same, when the elevator moves. When the elevator is going up, the reading of the scales is bigger than the actual weigh of the person; when it is going down, it is less than the weight.*

Another example of a student's response, which provides a piece of data to back up the position supported, is the following (1st year, pre-test, *Item 9*):

> *(I am not convinced.)This is not possible, given that nowadays there are vast expanses of land. In addition, we have found fossils of dinosaurs, which could not have been preserved in the water. Therefore, logically, dinosaurs must have been extinct in the ice ages. Fossils can be preserved in the ice.*

In both these arguments, the students examine the truthfulness of the claim and reject it without offering a statement engaging to the provided argument, except from the claim. Given that no rebuttals are evident, these are considered as middle-level arguments.

Examples of high-level oppositional argument: In the following examples, students' responses are engaged with the provided argument, given that there are statements in direct reference to the initial argument. The following extract is taken from the 2nd year, post-test (*Item 8*):

> *(I am not convinced) The nominal force changes in relation to the acceleration of the elevator. Therefore, the force that is exerted on the scales changes as well and it is not equal to the weight of the person.*

In this argument, the student opposes to the piece of evidence of the initial argument, that the nominal force is equal to the weight; in addition, he provides evidence to support it (there is acceleration). Similarly, the following response is engaged with a piece of evidence of the provided argument (1st year, pre-test, *Item 7*):

> *(I am not convinced). 'What has Pascal's principle to do with the fact that the rotational speed of the earth changes, if the ice in the North Pole melts and the level of the sea rises? It is irrelevant.*

In short, in the examples above the students rebut to the argument given to them, therefore they are regarded as high-level oppositional arguments.

For non-oppositional episodes (view Table 8.2), the quality in students' argument in this study is judged by the ability to clarify or to offer additional features that support /query the claim. Based on the rationale presented in the previous section, low- level non-oppositional arguments are considered the ones, in which the students repeat or rephrase the claim that they accept from the initial argument. As supported before, it is unlikely that such responses might contribute to a successful dialogue. Middle-level arguments are regarded the ones, which offer a kind of justification by highlighting some piece of evidence provided in the initial argument; by doing so, there is an increased possibility to contribute to the construction of knowledge, given that this evidence might have been overlooked by the other students, when evaluating the initial argument . Finally, in higher-level non-oppositional arguments, the students complement the initial argument with additional evidence, so as to reinforce or to question the claim of the initial argument. The extracts of students' responses that follow exemplify the above-mentioned categories.

Examples of low-level non-oppositional argument: In the following example (1st year of study, pre-test, *Item 7*) the student just repeats the claim of the initial argument:

> *(I am convinced) It is a correct argument. I agree that the rotational speed of the earth will change if the ice melts.*

Similarly, a students' response to *Item 8* (pre-test, 1st year of study), in which the student rephrases the claim of the provided argument is the following:

> *(I am convinced.) The reading of the scales remains the same, while the elevator is moving.*

In responses like the above it might be possible that the students have focused on the conclusion and accept it without following the rationale of the provided argument, given that there is no evidence in their answers that engages with the provided argument-except from the claim. Students provide no justification for their response; therefore such responses are regarded as low-level arguments.

Examples of middle-level non-oppositional argument: In contrast to the examples above, the following responses provide some evidence that the students have engaged with the initial argument; apart from repeating the claim, they highlight a piece of evidence of the initial argument, either to support the claim or to question it. The response to *Item 7* below is taken from the 2nd year of study, post test:

> *(I do not know). I agree that if the ice melts the level of the sea will rise and as a consequence the rotational speed of the earth will change. That is because the angular momentum of the earth must be conserved. But is this change significant or observable? I don know.*

In this example, the student queries the claim of the initial argument, repeating some of the data of it, without, though, offering additional to the existing justification for questioning the claim. In a similar vein, in the following argument (as a response to *Item 8*, 1st year pre-test) the student repeats a piece of evidence of the initial argument:

I agree. We know that the nominal force always equals the weight, which does not change.

As mentioned before, by highlighting a piece of data of the initial argument, rather than just repeat or question its claim, there is a higher possibility to contribute to a successful dialogue: it gives the opportunity to the other students to focus on the data of the argument rather the claim and – in that way – facilitate the process of offering justification that engages to the initial argument.

Examples of high-level non-oppositional argument: In the following responses the students reinforce the claim that they accept or question, by providing further evidence. For example, a student from the 1st year of study (pre-test) states in *Item 7*:

(I am convinced) I agree. As we know, L = Iw. When the ice melts, the distribution of the mass in the earth will change. As a result, the moment of inertia will change. Given, though, that L remains constant, w will change as well.

In the above response the student complement the initial argument with further elements, to back up the conclusion reached. Similarly, a student from the 2nd year of study, stated in the pre-test in *Item 8*:

(I am convinced.)The nominal force always equals the weight. The elevator is accelerating because of the force, which is exerted to it by the ropes. Given that N=W, and the weight does not change the reading of the scales remains the same.

In this argument, the additional evidence provided by the student is that the acceleration of the elevator is attributed to a force exerted by the ropes; this piece of evidence –although it is scientifically incorrect – reinforce the data of the initial argument that N=W, and as a consequence provide further backing to the claim. In short, responses like the above are considered high-level non-oppositional arguments, under the rationale that they clarify or reinforce the claim that they accept and –in that way- contribute to a successful dialogue.

The framework for analyzing the quality of written arguments as presented and exemplified above, was applied to the data of the study. The following section presents the results of the analysis, which was made to investigate the change in the quality of students' written arguments after participating in the course.

8.2. Applying the expanded theoretical framework: the results

The application of the expanded theoretical framework for evaluating the quality of students' written arguments involved the following stages: First, each student response to *Items 7-9* of the questionnaire was identified as being a) an oppositional argument and b) a non-oppositional argument. Then, for each of the above categories, each student response was identified in terms of the level it achieved (view Tables 8.1 and 8.2). The overall number of students' arguments fallen in each level for all the items was counted, both in pre- and post-test, for the two years of study. The results of the data analysis are presented below in tables.

Table 8.3 provides the numbers of oppositional and non- oppositional arguments in the pre- and post-test for the two years of study, in *Items 7, 8* and *9* of the questionnaire. The total number of arguments differs between the two years of study, given the difference in the sample size (33 students X 3 items = 99 episodes for the 1st year; 23 students X 3items = 69 episodes for the 2nd year). As evident in this table for both years of study the majority of the episodes were oppositional ones (42% and 60% in the pre- and post-test accordingly for the 1st year; 43% and 75% in the pre- and post test for the 2nd year). A considerable number of non-oppositional arguments were also provided by the students (varying from 16%-24%). A notable feature about these data, though, is that the percentages of total responses in which students did not provide any argument, as a justification for the choice 'convinced' / 'not convinced', is considerable high: for the first year of study 33% of the total responses in the pre-test fall in the category 'no argument', with the according percentage in the post-test falling to 21%.

Table 8.3: The numbers of oppositional, non-oppositional and no arguments that were provided by the students in Items 7, 8 and 9 of the questionnaire, in the pre and post-test for the two years of study- the percentages are provided below the absolute numbers

	1st year of study		**2nd year of study**	
	pre	**post**	**pre**	**post**
Oppositional	42 (42%)	60 (60%)	30 (43%)	52 (75%)
Non-oppositional	24 (24%)	18 (18%)	12 (17%)	11 (16%)
No argument	33 (33%)	21 (21%)	27 (40%)	6 (9%)
Total arguments	99 (100%)		69 (100%)	

More encouraging seem to be the results for the second year: although in the pre-test 40% of the responses did not contain any argument, in the post-test the according figure falls to 9%. The high percentages in this category in the pre-test for both years can be attributed to students' lack of confidence to deal with questions in such a format, rather than to mere inability to respond to arguments. This is possible the case, given in the last class of the upper-secondary school (which is the previous educational level of the participants) students were assessed by test in the format of the national exams papers, which included algorithmic problems and multiple choice questions. As for the post-test percentages, results on the 1st year are rather disappointing, as almost 20% of the total responses did not contain an argument. If issues of confidence are the reason for students' not providing any argument in these items, then it seems that for the 1st year, the intervention did not have the expected outcomes on the field; in contrast, for the 2nd year the according number falls below 10%, signifying that only a small minority of responses did not entail an effort to provide an argument.

Focusing to students' oppositional arguments, Table 8.4 presents the results of data analysis in terms of the levels in the quality of arguments. Data suggest that for both years of study the total number of the low-level arguments was decreased after the intervention and represents the small minority of the total number of arguments. This decrease in important as it signifies that only a few episodes entailed a claim versus a counter-claim; in other words, in a very few instances the students did not offer any data to back up the conclusion supported.

Table 8.4: The numbers of oppositional arguments in terms of quality in the pre and post-test for the two years of study- the percentages are provided below the absolute numbers are out of the total oppositional arguments

	1st year of study		2nd year of study	
	pre	**post**	**pre**	**post**
High-level	20 (48%)	31 (52%)	14 (47%)	32 (62%)
Middle-level	15 (35%)	26 (43%)	13 (43%)	18 (34%)
Low-level	7 (17%)	2 (3%)	3 (10%)	2 (4%)
Total oppositional arguments	42	60	30	52

Better results obtained in the 2nd year of study: the percentages of middle-level arguments decreased, while the ones of high-levels increased, signifying that more counter-arguments were engaged with the provided argument. In contrast, for the 1st year of year, a rather small increase was found for both the percentages of high-level and middle-level arguments. The increase in middle-level arguments is a fairly positive outcome, given that more episodes in the post- than the pre-test were found in which students provided data to support their view. Yet, as argued in the previous section, counter-arguments with no rebuttals are likely to lead to parallel monologues.

Turning attention to non-oppositional arguments, Table 8.5 presents the numbers of students' arguments that fall in each level in terms of quality. Similar to the case of oppositional episodes, the pattern again suggests a decrease of low-level arguments, for both years of the study. For starters, this is a positive outcome as it signifies that the number of post-test students' responses that contained some piece of evidence is increased. In addition, a marginal increase is observed for high-level arguments, for both years of study – defined as the ones in which students offered a clarification or additional features to support or query the claim of the provided argument. Mixed results are obtained for middle-level category: for the 1st year the percentages of total arguments remained practically the same, in contrast to the 2nd year for which they are increased after the intervention. Arguments in this category involved highlighting a piece of data of the initial argument, rather than just repeat or question its claim.

Table 8.5: The numbers of non-oppositional arguments in terms of quality in the pre and post-test for the two years of study- the percentages are provided below the absolute numbers are out of the total oppositional arguments

	1st year of study		**2nd year of study**	
	pre	**post**	**pre**	**post**
High-level	14 (58%)	11 (61%)	6 (50%)	6 (55%)
Middle-level	8 (34%)	6 (34%)	4 (33%)	5 (45%)
Low-level	2 (8%)	1 (5%)	2 (17%)	0 (0%)
Total non-oppositional arguments	24	18	12	11

As argued in the previous section, this gives the opportunity to the other students to focus on the data of the argument rather the claim and – in that way – facilitate the process of offering additional justification that engages to the initial argument.

In short, for both categories (oppositional and non-oppositional episodes) and after the intervention data analysis showed that it was the small minority of arguments developed by the students that did not attempt to offer a rational or some grounds for their claims. In an educational context, the predominance of arguments containing only claims would be problematic, in that they do not offer fruitful ground for the establishment of better understanding. In addition, for both years of study an increase was observed for high-level arguments: for oppositional ones this signified that more arguments developed by the participants rebutted the provided argument and they contained statements in direct reference to a piece of evidence; for non-oppositional ones, this increased signified that more arguments clarified or reinforced the claim with additional to the existing evidence. The above provide an indication that for both years of study there was a positive change in the quality of student's arguments after the intervention. Better results seem to have been obtained in the 2nd year: one the one hand for the 1st year in the post-test 20% of the responses did not contain any argument, which is a negative outcome-in contrast to the following year for which the according figure falls to 9%. On the other hand, for the 2nd year it was a greater increase in the percentages of arguments with rebuttals, than in the first year. This is important, given that the ability to rebut is considered the most complex skill in argumentation (Kuhn, 1991).

However, a number of caveats should be placed on the interpretation of these data: First, the number of arguments given to the students to evaluate and respond is small –only three items -, and it may not represent all the potential types of arguments that students could confront in real –time discourse. For example, the quality of students counter arguments in episodes when the initial argument contains only a claim, or when the claim is implicitly supported was not investigated. Second, the total number of students' arguments in each category was rather small. As a result no investigation was possible to be made, in terms of whether the changes in the quality are statistically significant. Provided that a greater number of arguments were evident, a similar statistical analysis to the one made for the quality of scientific explanations (view Chapter 6), would permit the investigation for statistical significance on the outcome. Nevertheless, regardless these limitations, an important outcome of the above analysis is that it exemplified the application of an expanded theoretical framework for evaluating the quality of students' written arguments – the one that considers oppositional and non-oppositional episodes as well. In the following chapter, the theoretical implications of this analysis will be further discussed, as well as recommendations for future research on the field will be made.

Summary

This chapter addressed the fourth objective of the study and provided the results of the data analysis in relation to the change of the quality of students' arguments after participating in the intervention. Data analysis reveled that the theoretical framework initially chosen for evaluating students' written arguments is useful, but limited as it considers only non-oppositional episodes. This resulted to the modification of the framework, so as to address the data evident in this study. The expanded theoretical framework – considering both oppositional and non-oppositional episodes- was presented and its suitability was supported, both from a theoretical perspective and as grounded to the data. The application of the framework to the data showed that for both years of study there was a positive change in the quality of student's arguments after the intervention, with better results being obtained in the 2nd year. Yet, the low number of arguments in each category did not allow the investigation of statistical significance in the outcome. Some strengths and limitations regarding the outcomes of the analysis are finally pinpointed.

CHAPTER 9

Discussion on the findings, summary and implications of the study

The primary aim of the study has been to explore the extent to which the course described in the first chapter of the thesis achieved its aims. In short, the research has sought to determine the extent to which the pedagogical strategies used in the course contributed to enhanced students' explaining and argumentation abilities, and to their improvement of conceptual understanding on the area of classical mechanics. In the first part of this final chapter, the findings regarding the main objectives of the research are summarized and discussed (Section 9.1). Apart from the main findings of the study, though, during the data analysis several issues emerged, worthy to be further discussed and investigated. Section 9.2 considers and summarizes these findings, which mainly concern the areas of scientific explanations and arguments, in the view that they might provide the ground for further research on the field. Following this, in Section 9.3 some limitations of this study are provided. The chapter finishes by considering the theoretical and the research implications of the investigation and by recommending directions for further research (Section 9.4).

9.1. Summary and discussion on the main findings of the study

This study was conducted in the context of a first year university physics unit, innovative both in the teaching methodology and the aims it sought to achieve. The research was guided by four research questions, which have formed its objectives:

RQ1: What is the change (if any) in students' conceptual understanding on basic mechanics after participating in the course?

RQ2: What is the change (if any) in students' ability to provide correct and concise scientific explanations on basic mechanics after participating in the course?

RQ3: How do students respond to weak or fallacious arguments? What is the change (if any) in the way students respond to such arguments after participating in the course?

RQ4: What is the change (if any) in the quality of students' written arguments after participating in the course?

These objectives have been realized under the mixed-method research paradigm, by means of constructing a mixed research design. Broadly speaking, findings have suggested that the intervention achieved the positive outcomes that had been expected, albeit not to the extent that it was hoped in all of the areas of investigation. The following lines discuss and summarize the main findings, in relation to the three main areas of interest.

9.1.1. Students' conceptual understanding on classical mechanics

In relation to the first objective of the research (RQ1), results of the data analysis showed that there was a significant positive change in students' conceptual understanding, after participating in the course, for both years of the study. On average, for the 1^{st} year of study, the outcome of the t-test suggested that students scored in the FCI significantly better in the post-test (M = 20.56, SE = 0.917), than in pre-test (M = 18.47, SE = 1.080, t (33) = -2.685, p <0.05, r = 0.42). In addition, the analysis of the data, as proposed by Hake, signified that the course falls in the category of low-g courses (<g> = 0.18). Similarly, for the 2^{nd} year of study and on average, the output of the t-test showed that students provided more correct answers in the post-test (M = 22.50, SE = 1.110), than in pre-test (M = 20.50, SE = 1.222, t (23) = -2.881, p <0.05, r = 0.47), while the g percentage was calculated <g> = 0.21, denoting that the course falls in the low-g courses.

The encouraging finding here is that students' conceptual understanding was significantly increased for both year of study after participating in the course. Yet, the course fell in the region of low-g courses, according to Hake's classification, in contrast to most of the interactive teaching programs, presented in Chapter 3, which reported outcomes on the medium-g courses category. A possible reason for this difference is that in the other interactive programs students were provided by the teacher with the correct answers in the conceptual questions. In contrast, in the two years of implementing the intervention, students decided on their own about the correctness of the views expressed, without the instructor taking sides, when competing theories where discussed. As Osborne *et al*. (2004) have argued, the exposure to plural explanatory theories might confuse some students, or lead to the development of a scientifically incorrect idea. Especially when at the end of discourse students do not get reassurance by an 'expert' about the correct view, the danger that some of them are led to confusion is increased. This may be an explanation of why the conceptual gain in this unit is lower than the one in units, in which students are provided with the correct answers.

A question that arises here, though, is whether students, who are provided with the correct answers in interactive programs, achieve deep conceptual understanding, or just memorize the ‘correct’ views that are expressed by the instructor after the discourse. In Chapter 2 it had been supported that the issue of achieving deep understanding through interactive pedagogies has been raised in a number of other research studies. It is reminded that according to Chinn & Brewer’s findings (1993) students are more likely to reach meaningful conceptual conflict, if engaged in a thoughtful and effortful processing of arguments. This process, though, requires actual engagement by the students, when competing arguments are expressed. In contrast, the rote learning of the correct answers –if they are provided - requires less cognitive energy and it might be favoured by some students, who are reluctant to take responsibility of their learning. In such a process, though, it is unlikely that deep understanding might occurs.

In short, there is no question that the strategy of not providing the correct answers to the students entails the danger of confusion or development of incorrect views to the students. On the other hand, though, when the instructor provides explicitly the correct scientific view, there is the danger that the students are not engaged actively in the argumentative process, rather memorize what the teacher advocates. The point made here is that in this course the progress in students’ conceptual understanding was rather low in comparison to other interactive units, maybe because students were not provided with the correct answers by the instructor. Yet, data analysis revealed a statistically significant positive change in their conceptual understanding. Most importantly, is has been argued that this change is the outcome of the students’ interactions, rather than the rote learning of the scientific knowledge provided by the instructor, without deep understanding. Under such a perspective, one of the aims of the course – the one relating to students’ acquisition of deep conceptual understanding – is believed to have been achieved.

9.1.2. Students’ scientific explanations

In relation to the second objective of the study (RQ2), findings suggest that for both years of the implementation, the students gave significantly more appropriate explanations in the post- than in the pre-test. Specifically, for the 1st year of study the number of scientifically correct explanations provided by the students was significantly higher in the post-test than in the pre-test, with the change being of medium size (z = -2.758, p< 0.05 and r = -0.342). Better results are obtained in the 2nd year: again, appropriate explanations in the post-test outnumbered significantly the according ones in the pre-test, but the change in this year was of large size (z = -3.493, p< 0.05 and r = -0.515). What is important about the increase in the number of correct explanations is that students did not only hold correct views about scientific phenomena in the area of classical mechanics, but that their views were *reasoned*, in a scientifically correct manner. As exemplified in Chapter 3, contemporary research on interactive 1st year university courses has mainly focused on investigating what students believe about various physics concepts. These studies, though, did not investigate *why* students hold such beliefs on science concepts, in a way that they can justify to themselves or to others. Maybe, implicit to the direction followed in this body of research is the assumption that students' beliefs on science topics are reasoned views. Kuhn (1991), though, has questioned this assumption, by making the case that cognitive skills on reasoning should not be taken for granted. This research adds to the existing body of literature on interactive units by investigating students' scientific reasoning. Most importantly this study showed that -through the interactive strategies used in the course- it is possible to contribute positively to students' ability to provide correct scientific explanations, as well as to their understanding of scientific concepts.

On the other hand, data analysis showed a positive development in students' ability to provide concise explanations. For both years of the study, at the end of the intervention students provided less *Category 1* explanations – defined as the ones in which only the explanatory conclusion is given. The decrease in the number of *Category 1* explanations is important, as it signifies that only a small minority of the explanations provided by the students did not offer a rationale to back up the claim. Better results were again obtained during the 2nd year of study, as students provided significantly more *Category 3* explanations in the post-test than in the pre-test (p<0.05, r = -0.445) - defined as the most complete ones. In contrast, for the 1st year of the implementation, although students provided more Category 3 explanations in the post- than in the pre-test, statistical significance was established only for the increase in *Category 2* explanations (p<0.05, r = -0.355) – defined as the ones in which more evidence is needed to back up the explanatory conclusion.

In should be reminded that in order to foster students' abilities in providing concise explanations in the course, the attributes of appropriate explanations were not taught explicitly. Instead, the process of 'explaining *to someone'* was emphasized as a distinct activity to 'explaining' – defined as a pedagogical activity, which aims to impart knowledge and to promote the understanding of the other person, the explainee. To do so, the instructor used a fictional person – Bobos - to whom students were asked to address their views, with an aim him to understand. The improvement in the quality of students' explanations provided indication that this was a successful strategy: the decrease in *Category 1* explanations signified that it has probably contributed to students' understanding that explaining is more than providing a set of algorithms to solve a problem or offering a conclusion without making explicit the rational; in addition, the increase in *Category 3* explanations –although not statistically significant in the 1st year of study – gives confidence that the students have gained an insight that explaining in educational contexts entails the consideration to promote understanding in the other person, the explainee.

There is no question that this study does not allow the establishment of a cause and effect relationship between the strategies used in the course and the outcomes in the area of scientific explanations to the students, given that it was not conducted under an experimental design. Therefore, it can not be concluded that the use of the fictional person Bobos in the classroom necessarily *results to* the enhancement of students' explaining skills. However, the prolonged engagement of the researcher to the field by observing the lessons for the two years of study has provided indication that such a strategy has positive outcomes in engaging students in the process of providing concise explanations. It should be reminded that the better results obtained in the 2nd year, cannot be attributed to difference between the two groups of participants in terms of the ability to provide concise explanations before participating in the unit, as no significant difference was established (view chapter 6). Rather, the triangulation of these results with the researcher field notes suggested that the difference between the two years is likely to be a reflection of the development of the instructor's ability in implementing the pedagogic practices in the classroom. The above provide the researcher with confidence about the positive contribution of the strategies used in the classroom to the observed positive outcomes of the study in the area of scientific explanations.

In summary, in relation to the second objective of the research, data analysis showed that the course has achieved the objective of enhancing students' explaining abilities, in terms of students providing correct and concise scientific explanations. The significance of this finding is that students do not only hold correct views about scientific concepts, but *reasoned* views as well, in a scientifically correct manner. In addition, the findings of this study indicated that students gained an insight about what explaining involves in an educational context; in other words, that it inevitably involves the consideration to promote understanding to the other persons, the explainees. Better results were obtained in the 2nd year of study on the field. A triangulation with the field notes from observations of the lessons suggested that the better results for this year can be attributed to the development of the instructors' ability in implementing more successfully the pedagogic strategies in the classroom.

9.1.3. Students' scientific arguments

In the area of scientific arguments, the main objectives of this research are: to explore how students respond to weak or fallacious arguments and to investigate the change in the way students deal with such argument (RQ3); to identify the change in the quality of their written arguments after participating in the course (RQ4). Although these objectives were realized, the findings discussed below must be read having in mind a couple of limitations that emerged at the stage of data analysis: First, in the process of forming categories in the way students' respond to weak or fallacious arguments, issues of bias may be evident in terms of subjective interpretation of the data. Although during the analysis, regular attempts were made to identify a priori attributes that were given to the categories formed (as exemplified in Chapter 7), it is acknowledged that the use of more researchers to code students' responses would provide greater reliability of the outcome. Second, in terms of identifying the change in the quality of students' arguments, the low number of arguments in each category did not allow the investigation of statistical significance in the outcome.

Nevertheless, as far as the first objective is concerned, qualitative analysis of students' responses resulted in the formation of categories showing how students dealt with arguments: *a)* in which the conclusion is not an outcome of the data provided; *b)* in which an incorrect element leads to a scientifically wrong conclusion; *b)* deductive one which affirms the consequence and *d)* containing a claim by an expert with no justification. Apart from exemplifying the way that students' deal with such arguments, data analysis showed a couple of issues worth mentioning on the field: First, it was shown that even at the end of the intervention and for the two years of research, students confronted more difficulty in identifying irrelevant or missing components in an argument, rather than finding incorrect data in it. This signifies that maybe the process of following the line of reasoning in an argument and of evaluating the adequacy and the appropriateness of evidence is a more complex skill in comparison to the ability to evaluate the correctness of a piece of evidence in an argument– provided that the students have the scientific knowledge to address it. Given that no generalizibility can be established from these data, the above indication provides ground for further research. Another outcome of the analysis was that issues concerning the distinction between truth and validity in deductive arguments were evident in some students' responses. Such issues have been pointed in other studies as well (for a review, see Zeidler, 1997); it is important, though, to note that in this study were also evident, in the view that errors in syllogism such as affirming the consequence in deductive arguments might be regarded as naïve and not worth being addressed.

In relation to the change in the way students' respond to weak or fallacious arguments, statistical analysis revealed that for both years of study, students made more successful attempts to respond to such arguments after the intervention than before. Yet, better results were obtained for the 2nd year as the change was of large size (z = -3.453, p< 0.05 and r = 0. 509), in contrast to the previous year, in which the change was of small size (z = -2.087, p< 0.05 and r = 0.256). Given that no statistical significance was found in the area between the groups before the intervention, the better results of the 2nd year were attributed to a) the modification of the materials given to the students to help the in the process of argumentation in this period; b) the instructor's better conceptualization of the norms of argumentation, and the better use of the materials in the classroom.

As for the quality of students' written arguments, the application of the theoretical framework that was developed showed that both years of study there was a positive change in the quality of student's arguments after the intervention: both for oppositional and non-oppositional arguments it was the small minority of arguments after the intervention that did not attempt to offer a rational for their claims. In addition, for both years of study an increase was observed for high-level arguments: for oppositional ones this signified that more arguments developed by the participants contained statements in direct reference to a piece of evidence; for non-oppositional ones, this increased signified that more arguments clarified or reinforced the claim with additional to the existing evidence. Better results seem to have been obtained in the 2nd year as there was a greater increase in the percentages of arguments with rebuttals, than in the first year – a skill that is considered as the most complex one in argumentation. Although the data in the study did not allow the investigation for statistical significance on the outcomes – something that represents a limitation of the research– the above analysis provided an indication that the intervention has achieved the positive outcome that were desired on the field.

9.2. Further issues emerged from the data analysis

Along with the findings regarding the main objectives of the study, during the data analysis some issues emerged, worth to be considered and summarized. Although they are out of the main focus of the research, the remarks below have been an outcome of the data analysis and provide directions for further research. The first remark relates to the area of scientific explanations and the factors that may affect effective reasoning. Koslowski (1996) has argued that the lack of knowledge of any relevant theory often constrains young's people ability to explain effectively. Data of the 1st year of study (view Chapter 6) provided an indication that another factor may play a role in effective reasoning: maybe for some students the more easy and obvious an explanation is regarded, the less information they provide to back up the conclusion reached. In this case, though, an explanation that is regarded as concise by the explainers (given that some evidence is obvious) may be seen by the explainee as one that lacks important information; therefore, the goal of the explanatory act is unlikely to be fulfilled. Given that the analysis made does not allow any generalization of the outcomes, whether the difficulty and the degree of conciseness of an explanation are *negatively* correlated, is an open question for further research.

Another issue that emerged during the data analysis is that students tend to focus and evaluate only the conclusion of an argument, rather than addressing the rationale and the evidence of the argument. This tendency was evident both in all the items of the test, as well as during the observation of students' discourse. As discussed in Chapter 7, focusing only to the claim could be attributed to the fact that such strategy requires less investment of cognitive energy; therefore, it might be favoured by the students. Another possible reason that has been pinpointed by Zeidler (1997) is that maybe students find ways to ignore evidence in an argument, if it conflicts with their initial beliefs. However, data analysis showed that even in arguments that contain a claim with no justification, students follow the same strategy, rather than rejecting the argument as weak. It is possible that the students confuse the notions of 'examining the correctness of a claim' and 'examining a soundness of an argument'. In other words, maybe for some of them the question '*are you convinced by the argument?*' is interpreted as '*do you agree with the claim of the argument?*'. Nevertheless, further investigation is needed, in relation to gaining an understanding about the reasons for which students focuses and evaluate only the claim of argument.

Finally, data in this study provided an indication about the role of authority that a person has during discourse (view Chapter 6). A triangulation of the analysis of the test and the field notes suggested that the way students evaluate their classmates' views is maybe influenced by the expertise of the person, who expresses the explanation or argument. Actually, being influenced by the expertise of the arguer is not irrational and might have been expected. Although, ideally cognitive autonomy requires pointing away from reliance on experts (Walton, 1997), inevitably individual judgment is influenced by experts authority. After all, as Kolsto & Ratcliffe (2008) have argued, in principle students' scientific knowledge has been mainly built through trust in the authority of the teacher, or the science textbooks. This point highlights the importance of social aspects that are embedded in the argumentation process. Further research would be useful in identifying social factors that influence the way students evaluate their peers' arguments, as well as the educational conditions under which the ideal cognitive autonomy could be achieved.

9.3. Limitations of the study

Although during the stage of the research design, any possible effort was made to ensure validity and reliability for this study (refer to Chapter 4), some issues that emerged during the implementation of the research might limit the strength of evidence for the outcome. A first limitation of this research regards sampling: as mentioned earlier (view Chapter 5), during the 2^{nd} year of study, although forty students participated in more than half of the lessons, only twenty three of them completed the tests, as at the last week of this semester (when the post-test was implemented) students were exercising abstinence from courses. Another limitation may be found in the process of forming categories in the way students' respond to weak or fallacious arguments, as it might entail subjective interpretation of the data. It is acknowledged that the use of more researchers to code students' responses would provide greater reliability of the outcome; however this was not possible given the conditions under which this study was conducted. Finally, for RQ4 no statistical significance was investigated for the change in the quality of students' arguments, given that the number of arguments in each category did not allow such an investigation.

Apart from the above limitations that emerged during the data gathering and analysis, a number of interrelated issues were not explored in this study due to time and space constrains. First, in the early stages of the study, it has been planned to explore students' attitudes towards the pedagogy, along with their performance in the three areas of investigation. Although data from interviews and questionnaires focusing on students' views on the course was gathered, time constrains limited any reports of the outcomes. This issue, though, is planned to be a focus of future investigation. Secondly, it is acknowledged that it would be interesting to investigate the research objectives, not only for the sample of students who participated in the majority of the lessons, but for those who did not come regularly to the sessions, as well as to those who dropped out. Unfortunately, it was not possible to gather some data for an adequate number of these students (pre- and post test); in addition, these students were reluctant to provide an interview, that would allow the researcher to investigate the reasons for not coming regularly to the course or for deciding to drop out. Finally, data of this study allows the investigation of the three areas of interest (conceptual understanding-explanations-arguments) in relation to each other, and the search for potential links among them. Although this issue was not addressed in this study due to time and space limitations, it is planned to be the focus for future research.

9.4. Implications of the study and directions for further research

The study has sought to examine the extent to which the course – in the context of which the research took place – achieves its aims. In Section 9.1, the main findings of the study were summarized: broadly speaking, it was concluded that the intervention achieved positive outcomes, albeit not to the extent that it was hoped in all of the areas of investigation. A question that might arise here, though, relates to the significance of such an outcome: it could be argued that when implementing an intervention in educational contexts, positive outcomes to the students are expected up to a degree – or at least negative ones are not frequently reported. Under such a rationale, which could be the implications of this study in the research domain?

It should be reminded, though, that in this course neither the correct scientific views were provided by the teacher, nor the norms of argumentation were explicitly taught. In such a context, the improvement of students' conceptual understanding, the increase in the content and quality of their explanations and the enhancement in the quality of written arguments should not be taken for granted. The point made here is that this study made a case that even in an educational environment, in which the students decide on their own about the correctness and appropriateness of scientific concepts and explanations, positive outcomes in these areas can be observed. On the other hand, it provided evidence that even when argumentation is not explicitly taught, improvement in the quality of students written arguments can occur with the aid of more 'modest' strategies. In addition, in relation to the implications of this study for further research, data analysis has revealed a number of issues, worthy further investigation (for a summary view Section 9.2). These issues provide directions for further research.

Apart from the research implications, this study has contributed theoretically in the broad area of argumentation research: it proposed and exemplified a theoretical framework for evaluating the quality of students' arguments, which includes both oppositional and non-oppositional episodes. This was an outcome of the analysis of students' written arguments and of the observations of students' discourse, which suggested that the theoretical framework initially chosen – an adaptation of an adaptation of Erduran *et al.*'s (2004) methodological approach - is useful but limited, as it considers only non- oppositional arguments. The framework was successfully applied to the data, in terms of investigate the change in the quality of students' written arguments; yet, the limited number of arguments in each category in this study did not allow investigating statistical significance for the change. A more extended research is needed to provide ground for the implementation of the proposed framework, so as to investigate more thoroughly its strengths and weakness. Given that the way of assessing the quality of arguments is a major issue in the study of argumentation in either written or verbal data and yields difficulties and challenges (Erduran, 2008), the proposed theoretical framework that resulted from this study is believed to constitute the main theoretical implication of it.

APPENDIX A

Force concept Inventory: The Greek translation

1.Δυο μεταλλικές μπάλες έχουν το ίδιο μέγεθος αλλά η μια ζυγίζει διπλάσια από την άλλη. Αφήνουμε και τις δυο μπάλες ταυτόχρονα να πέσουν στο έδαφος από την ταράτσα ενός μονώροφου κτιρίου. Ο χρόνος που χρειάζονται οι μπάλες για να φτάσουν στο έδαφος είναι

(1) περίπου ο μίσος για την πιο βαριά μπάλα από ότι για την πιο ελαφριά
(2) περίπου ο μίσος για την πιο ελαφριά μπάλα από ότι για την πιο βαριά
(3) περίπου ο ίδιος και για τις δυο μπάλες
(4) αρκετά μικρότερος για την πιο βαριά μπάλα, αλλά όχι απαραίτητα ο μίσος
(5) αρκετά μικρότερος για την πιο ελαφριά μπάλα, αλλά όχι απαραίτητα ο μίσος

2.Οι δυο μεταλλικές μπάλες της προηγούμενης ερώτησης κυλούν πάνω σε ένα οριζόντιο τραπέζι με την ίδια ταχύτητα και όταν φτάσουν στην άκρη του τραπεζίου πέφτουν στο έδαφος. Σε αυτή την περίπτωση

(1) και οι δύο μπάλες χτυπούνε στο πάτωμα στην ίδια περίπου οριζόντια απόσταση από την άκρη (base) του τραπεζιού
(2) η βαρύτερη μπάλα χτυπά στο πάτωμα στην μισή περίπου οριζόντια απόσταση, από την άκρη του τραπεζιού, από αυτή στην οποία χτυπά η ελαφρύτερη μπάλα
(3) η ελαφρύτερη μπάλα χτυπά στο πάτωμα στην μισή περίπου οριζόντια απόσταση, από την άκρη του τραπεζιού, από αυτή στην οποία χτυπά η βαρύτερη μπάλα
(4) η βαρύτερη μπάλα χτυπά στο πάτωμα αρκετά πιο κοντά στην άκρη του τραπεζιού από ότι η ελαφρύτερη μπάλα, αλλά όχι απαραίτητα στη μισή οριζόντια απόσταση (σε σχέση με την ελαφρύτερη)
(5) η ελαφρύτερη μπάλα χτυπά στο πάτωμα αρκετά πιο κοντά στην άκρη του τραπεζιού από ότι η βαρύτερη μπάλα, αλλά όχι απαραίτητα στη μισή οριζόντια απόσταση

3.Μια πέτρα που πέφτει στο έδαφος από τη στέγη ενός μονώροφου κτιρίου

(1) αποκτά τη μέγιστη ταχύτητα της παρά πολύ σύντομα και μετά πέφτει με σταθερή ταχύτητα
(2) καθώς πέφτει επιταχύνεται γιατί η δύναμη βαρύτητας αυξάνει σημαντικά όσο η πέτρα πλησιάζει στο έδαφος.
(3) επιταχύνεται γιατί μια σχεδόν σταθερή δύναμη βαρύτητας ασκείται πάνω στην πέτρα καθόλη τη διάρκεια της πτώσης
(4) πέφτει εξαιτίας της φυσιολογικής τάσης όλων των σωμάτων να ηρεμούν πάνω στην επιφάνεια της Γης.
(5) πέφτει λόγω της συνδυασμένης δράσης της δύναμης της βαρύτητας, που σπρώχνει προς τα κάτω, και της δύναμης του αέρα, που σπρώχνει προς τα κάτω

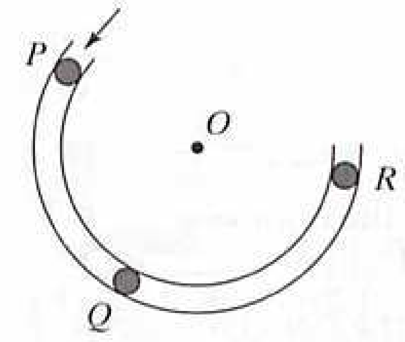
4.Ένα μεγάλο φορτηγό συγκρούεται μετωπικά με ένα μικρό αυτοκίνητο. Κατά τη διάρκεια της σύγκρουσης

(1) Το φορτηγό ασκεί μεγαλύτερης έντασης δύναμη στο αυτοκίνητο από ότι το αυτοκίνητο στο φορτηγό
(2) Το αυτοκίνητο ασκεί μεγαλύτερης έντασης δύναμη στο φορτηγό από ότι το φορτηγό στο αυτοκίνητο.

(3) Κανένα από τα δυο οχήματα δεν ασκεί δύναμη στο άλλο, απλά το αυτοκίνητο παραμορφώνεται γιατί πολύ απλά βρίσκεται στο δρόμο του φορτηγού
(4) Το φορτηγό ασκεί δύναμη στο αυτοκίνητο αλλά το αυτοκίνητο δεν ασκεί δύναμη στο φορτηγό
(5) Το φορτηγό ασκεί δύναμη στο αυτοκίνητο ίδιας έντασης με αυτή που ασκεί το αυτοκίνητο στο φορτηγό.

Χρησιμοποιήστε τις παρακάτω πληροφορίες και το σχήμα για να απαντήσετε τις επόμενες δυο ερωτήσεις (5 και 6)

Το παρακάτω σχήμα δείχνει ένα μεταλλικό κανάλι με αμελητέα τριβή, με σχήμα κυκλικού τόξου και κέντρο το σημείο O, μέσα στο οποίο κινείται μια μπίλια. Το κανάλι είναι σταθεροποιημένο πάνω στην επιφάνεια ενός οριζόντιου τραπεζίου που έχει αμελητέα τριβή. Στο παρακάτω σχήμα κοιτάτε το τραπέζι από πάνω. Οι δυνάμεις που ασκούνται λόγω του αέρα είναι αμελητέες. Μια μπίλια εκτοξεύεται με μεγάλη ταχύτητα μέσα στο κανάλι στο σημείο P και βγαίνει από το κανάλι στο σημείο R.



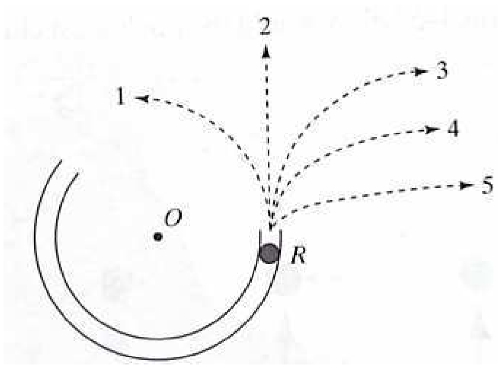
5.Θεωρείστε τις παρακάτω διαφορετικές δυνάμεις:

(A) Μια δύναμη προς τα κάτω λόγω βαρύτητας
(B) Μια δύναμη που ασκείται από το κανάλι με διεύθυνση από το Q προς το O
(C) Μια δύναμη με διεύθυνση αυτή της κίνησης
(D) Μια δύναμη με διεύθυνση από το O στο Q

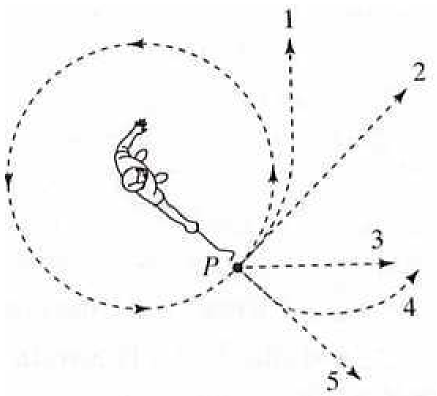
Ποια(ποιες) από τις παραπάνω δυνάμεις ασκείται (ασκούνται) στην μπίλια όταν αυτή φτάνει στο σημείο Q κινούμενη χωρίς τριβές μέσα στο κανάλι;

(1) Μόνο η (A)
(2) Η (A) και η (B)
(3) Η (A) και η (C)
(4) Οι (A), (B), και (C)
(5) Οι (A), (C), και (D)

6.Ποια από τις παρακάτω τροχιές 1-5 θα ακολουθήσει η μπίλια μόλις βγει από το κανάλι και ενώ κινείται χωρίς τριβή πάνω στην επιφάνεια του τραπεζίου;

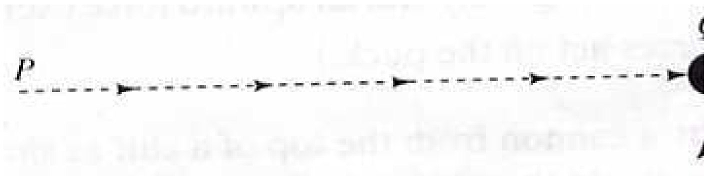


7. Μια μεταλλική μπάλα είναι δεμένη σε ένα σχοινί και στρέφεται σε κυκλική και οριζόντια τροχιά παράλληλη στο επίπεδο του χαρτιού όπως φαίνεται στο παρακάτω σχήμα.
   Όταν βρίσκεται στο σημείο P το σχοινί σπάει ξαφνικά κοντά στο σημείο που ήταν δεμένη η μπάλα. Ποια από τις τροχιές 1-5 περιγράφει ποιο πιστά την κίνηση της μπάλας αφού σπάσει το σχοινί, όπως την βλέπει ένας παρατηρητής που παρακολουθεί από πάνω την σκηνή,

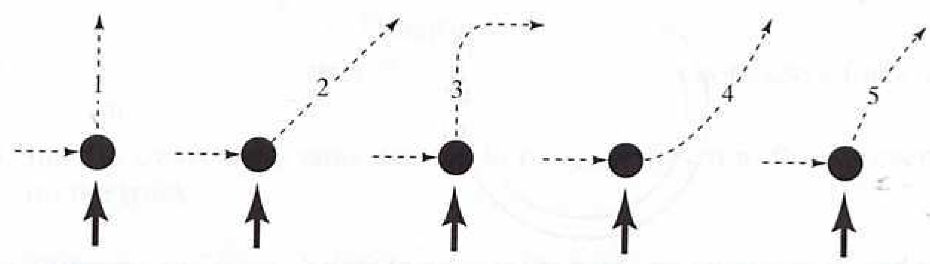


Χρησιμοποιήστε τις παρακάτω πληροφορίες και το σχήμα για να απαντήσετε τις επόμενες τέσσερις ερωτήσεις (8-11).

Το παρακάτω σχήμα δείχνει μια μπάλα του χόκεϊ καθώς ολισθαίνει με σταθερή ταχύτητα $\mathbf{υ_0}$ σε ευθεία τροχιά από το σημείο P προς το σημείο Q πάνω σε μια οριζόντια και χωρίς τριβές επιφάνεια. Οι δυνάμεις που ασκούνται από τον αέρα είναι αμελητέες. Στο σχήμα βλέπετε τη μπάλα από πάνω. Όταν η μπάλα φτάσει στο σημείο Q, δέχεται ένα πολύ γρήγορο κτύπημα κατά τη διεύθυνση του βέλους του σχήματος. Αν η μπάλα ήταν αρχικά ακίνητη στο σημείο P, τότε το παραπάνω κτύπημα θα την έθετε σε οριζόντια κίνηση με ταχύτητα $υ_κ$ κατά τη διεύθυνση του κτυπήματος.



8.Ποια από τις παρακάτω τροχιές 1-5 περιγράφει πιο πιστά την κίνηση της μπάλας αφού δεχτεί το χτύπημα;



9.Το μέτρο της ταχύτητας της μπάλας αμέσως μόλις δεχτεί το χτύπημα είναι

(1) Ίσο με την ταχύτητα $υ_0$ που είχε η μπάλα πριν δεχτεί το χτύπημα

(2) Ίσο με την παραπάνω ταχύτητα $υ_κ$ και ανεξάρτητο της ταχύτητας $υ_0$

(3) Ίσο με το αλγεβρικό άθροισμα των μέτρων των ταχυτήτων $υ_0$ και $υ_κ$.

(4) Μικρότερο και από την $υ_0$ και από την $υ_κ$.

(5) Μεγαλύτερο και από την $υ_0$ και από την $υ_κ$ αλλά μικρότερο από το αλγεβρικό άθροισμα των μέτρων των ταχυτήτων $υ_0$ και $υ_κ$.

10.Κατά μήκος της (χωρίς τριβές) τροχιάς που διαλέξατε στη ερώτηση 8, η ταχύτητα της μπάλας μετά που δέχτηκε το χτύπημα

(1) Είναι σταθερή

(2) Αυξάνεται συνεχώς

(3) Μειώνεται συνεχώς

(4) Αυξάνεται για λίγο και μετά μειώνεται

(5) Είναι σταθερό για λίγο και μετά μειώνεται

11. Κατά μήκος της (χωρίς τριβές) τροχιάς που διαλέξατε στη ερώτηση 8, η κύρια δύναμη (δυνάμεις) που ασκείται (ασκούνται) στην μπάλα μετά που δέχτηκε το κτύπημα είναι

(1) Η δύναμη βαρύτητας με διεύθυνση προς τα κάτω

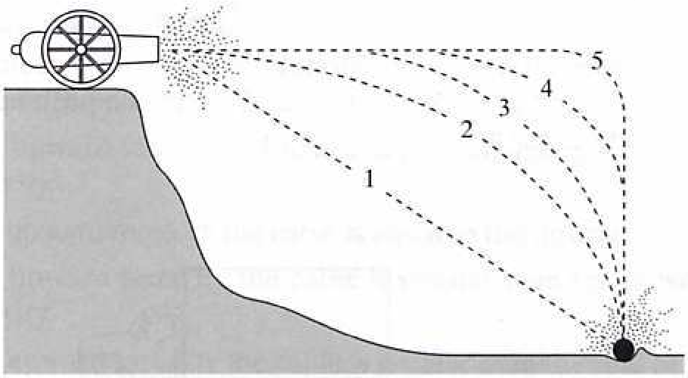
(2) Η δύναμη βαρύτητας με διεύθυνση προς τα κάτω, και μια οριζόντια δύναμη στη διεύθυνση της κίνησης

(3) Η δύναμη βαρύτητας με διεύθυνση προς τα κάτω, μια δύναμη προς τα πάνω που ασκείται από την επιφάνεια, και μια οριζόντια δύναμη στη διεύθυνση της κίνησης

(4) Η δύναμη βαρύτητας με διεύθυνση προς τα κάτω και μια δύναμη προς τα πάνω που ασκείται από την επιφάνεια

(5) Καμία. (Καμία δύναμη δεν ασκείται στη μπάλα)

12.Ένα κανόνι εκτοξεύει μια μπάλα από την κορυφή ενός γκρεμού όπως φαίνεται στο παρακάτω σχήμα. Ποια από τις τροχιές 1-5 περιγράφει με μεγαλύτερη ακρίβεια την τροχιά της μπάλας;



13.Ενα αγόρι πετάει μια μεταλλική μπάλα κατακόρυφα προς τα πάνω. Θεωρήστε την κίνηση της μπάλας αφού αυτή έχει φύγει από το χέρι του αγοριού και προτού αγγίξει το έδαφος, και υποθέστε ότι οι δυνάμεις που ασκούνται από τον αέρα είναι αμελητέες. Υπό αυτές τις συνθήκες, η δύναμη (δυνάμεις) που ασκείται (ασκούνται) στην μπάλα είναι

(1) Μια δύναμη με διεύθυνση κατακόρυφη προς τα κάτω λόγω βαρύτητας και μια δύναμη προς τα πάνω που συνεχώς μικραίνει.

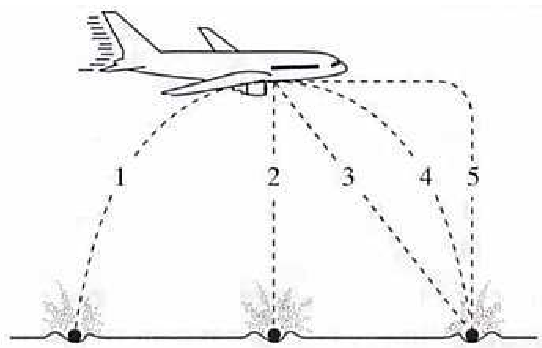
(2) Μια δύναμη προς τα πάνω που μικραίνει συνεχώς από τη στιγμή που η μπάλα φεύγει από το χέρι του αγοριού και μέχρι να φτάσει στο ψηλότερο σημείο της τροχιάς της. Καθώς η μπάλα κατεβαίνει, υπάρχει μια δύναμη προς τα κάτω λόγω βαρύτητας που συνεχώς αυξάνεται καθώς η μπάλα πλησιάζει στη Γη.

(3) Μια σχεδόν σταθερή δύναμη λόγω βαρύτητας με διεύθυνση προς τα κάτω και μια δύναμη προς τα πάνω που συνεχώς μικραίνει μέχρι η μπάλα να φτάσει στο ψηλότερο σημείο της τροχιάς της. Καθώς η μπάλα κατεβαίνει, υπάρχει μόνο μια σχεδόν σταθερή δύναμη προς τα κάτω λόγω βαρύτητας.

(4) Μόνο μια σχεδόν σταθερή δύναμη βαρύτητας προς τα κάτω

(5) Κανένα από τα παραπάνω. Η μπάλα πέφτει πίσω στο έδαφος εξαιτίας της φυσιολογικής τάσης των σωμάτων να έρχονται σε ηρεμία στην επιφάνεια της Γης.

14.Μια μπάλα του μπόουλινγκ (bowling) πέφτει κατά λάθος από την πόρτα αποσκευών ενός αεροπλάνου που πετάει κατά μήκος μιας οριζόντιας διεύθυνσης. Ποια από τις τροχιές 1-5 περιγράφει με μεγαλύτερη ακρίβεια την τροχιά της μπάλας αφού αυτή φύγει από το αεροπλάνο, όπως την βλέπει ένας άνθρωπος που στέκεται στο έδαφος και παρατηρεί το αεροπλάνο όπως στο παρακάτω σχήμα



Χρησιμοποιήστε τις παρακάτω πληροφορίες και το σχήμα για να απαντήσετε τις επόμενες δυο ερωτήσεις (15 και 16)

Η μηχανή ενός μεγάλου φορτηγού σταματάει να δουλεύει ενώ αυτό κινείται πάνω στο δρόμο. Για να βγει από τη μέση του δρόμου το σπρώχνει ένα μικρό αυτοκίνητο όπως φαίνεται στο παρακάτω σχήμα.

15.Ενώ το αυτοκίνητο, σπρώχνει το φορτηγό, επιταχύνει μέχρι να φτάσει τη μέγιστη ταχύτητα του

(1) Η δύναμη με την οποία το αυτοκίνητο σπρώχνει το φορτηγό είναι ίση κατά μέτρο με τη δύναμη με την οποία το φορτηγό σπρώχνει το αυτοκίνητο

(2) Η δύναμη με την οποία το αυτοκίνητο σπρώχνει το φορτηγό είναι μικρότερη από τη δύναμη με την οποία το φορτηγό σπρώχνει το αυτοκίνητο

(3) Η δύναμη με την οποία το αυτοκίνητο σπρώχνει το φορτηγό είναι μεγαλύτερη από τη δύναμη με την οποία το φορτηγό σπρώχνει πίσω το αυτοκίνητο

(4) Η μηχανή του αυτοκίνητου δουλεύει και άρα το αυτοκίνητο σπρώχνει το φορτηγό, ενώ η μηχανή του φορτηγού δεν δουλεύει και άρα το φορτηγό δεν μπορεί να σπρώξει το αυτοκίνητο. Το φορτηγό σπρώχνεται προς τα εμπρός διότι βρίσκεται στο δρόμο του αυτοκίνητου.

(5) Ούτε το αυτοκίνητο Ούτε το φορτηγό ασκούν δυνάμεις το ένα στο άλλο. Απλώς το φορτηγό σπρώχνεται προς τα μπρος διότι βρίσκεται στο δρόμο του αυτοκίνητου

16.Όταν το αυτοκίνητο φτάσει τη μέγιστη και σταθερή ταχύτητα με την οποία επιθυμεί να σπρώχνει το φορτηγό ο οδηγός του,

(1) Η δύναμη με την οποία το αυτοκίνητο σπρώχνει το φορτηγό είναι ίση κατά μέτρο με τη δύναμη με την οποία το φορτηγό σπρώχνει πίσω το αυτοκίνητο

(2) Η δύναμη με την οποία το αυτοκίνητο σπρώχνει το φορτηγό είναι μικρότερη από τη δύναμη με την οποία το φορτηγό σπρώχνει πίσω το αυτοκίνητο

(3) Η δύναμη με την οποία το αυτοκίνητο σπρώχνει το φορτηγό είναι μεγαλύτερη από τη δύναμη με την οποία το φορτηγό σπρώχνει πίσω το αυτοκίνητο

(4) Η μηχανή του αυτοκίνητου δουλεύει και άρα το αυτοκίνητο σπρώχνει το φορτηγό, ενώ η μηχανή του φορτηγού δεν δουλεύει και άρα το φορτηγό δεν μπορεί να σπρώξει το αυτοκίνητο. Το φορτηγό σπρώχνεται προς τα μπροστά διότι βρίσκεται στο δρόμο του αυτοκίνητου.

(5) Ούτε το αυτοκίνητο ούτε το φορτηγό ασκούν δυνάμεις το ένα στο άλλο. Απλώς το φορτηγό σπρώχνεται προς τα μπρος διότι βρίσκεται στο δρόμο του αυτοκίνητου

17.Μια καμπίνα ασανσέρ ανεβαίνει με σταθερή ταχύτητα καθώς τραβιέται προς τα πάνω με ένα ατσάλινο καλώδιο όπως φαίνεται στο παρακάτω σχήμα. Όλες οι τριβές είναι αμελητέες. Σε αυτή την περίπτωση, οι δυνάμεις που ασκούνται στο ασανσέρ είναι τέτοιες ώστε

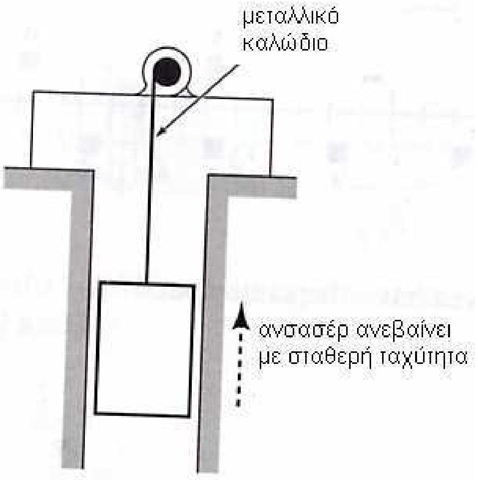
(1) Η δύναμη προς τα πάνω που ασκείται στο ασανσέρ από το καλώδιο είναι μεγαλύτερη από τη δύναμη προς τα κάτω που ασκείται στο ασανσέρ λόγω βαρύτητας.

(2) Η δύναμη προς τα πάνω που ασκείται στο ασανσέρ από το καλώδιο είναι ίση με τη δύναμη προς τα κάτω που ασκείται στο ασανσέρ λόγω βαρύτητας.

(3) Η δύναμη προς τα πάνω που ασκείται στο ασανσέρ από το καλώδιο είναι μικρότερη από τη δύναμη προς τα κάτω που ασκείται στο ασανσέρ λόγω βαρύτητας

(4) Η δύναμη προς τα πάνω που ασκείται στο ασανσέρ από το καλώδιο είναι μεγαλύτερη από το άθροισμα της δύναμης προς τα κάτω που ασκείται στο ασανσέρ λόγω βαρύτητας και της δύναμης προς τα κάτω που ασκείται στο ασανσέρ από τον αέρα.

(5) Τίποτα από τα παραπάνω. (Το ασανσέρ ανεβαίνει διότι το καλώδιο όλο και κονταίνει, όχι γιατί ασκείται στο ασανσέρ μια δύναμη προς τα πάνω από το καλώδιο).



18. Το παρακάτω σχήμα δείχνει ένα αγόρι να κάνει κούνια ξεκινώντας από μια θέση ψηλότερη από το σημείο P. Θεωρείστε τις παρακάτω διαφορετικές δυνάμεις:

(A) Μια δύναμη προς τα κάτω λόγω βαρύτητας

(B) Μια δύναμη που ασκείται από το σχοινί με διεύθυνση από το P προς το O.

(C) Μια δύναμη παράλληλη στη διεύθυνση κίνησης του αγοριού

(D) Μια δύναμη με διεύθυνση από το O προς το P

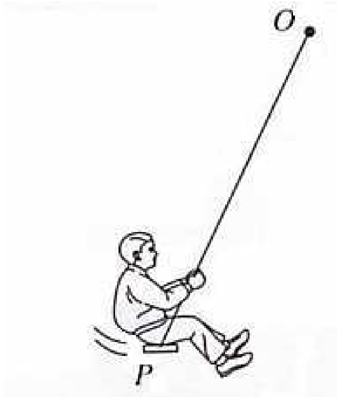
Ποια(ποιες) από τις παραπάνω δυνάμεις ασκείται (ασκούνται) στο αγόρι όταν αυτό βρίσκεται στη θέση P;

(1) Μόνο η A

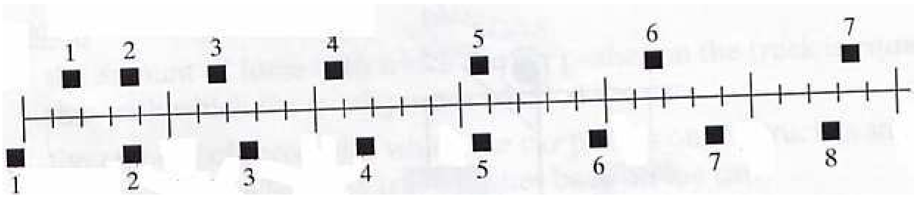
(2) Η A και η B

(3) Η A και η C

(4) Οι A, B, και C



19. Οι θέσεις δυο σωμάτων σε διαδοχικές χρονικές στιγμές που απέχουν μεταξύ τους κατά 0.20 s παριστάνονται από τα αριθμημένα τετράγωνα στο παρακάτω σχήμα. Τα δυο σώματα κινούνται προς τα δεξιά.



Έχουν κάποια στιγμή τα δυο σώματα την ίδια ταχύτητα;

(1) Όχι

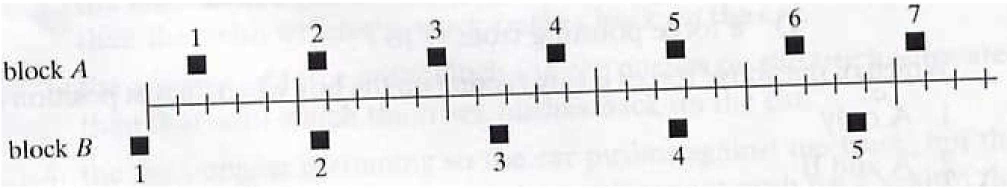
(2) Ναι, τη χρονική στιγμή 2

(3) Ναι, τη χρονική στιγμή 5

(4) Ναι, τις χρονικές στιγμές 2 και 5

(5) Ναι, κάποια στιγμή ανάμεσα στις χρονικές στιγμές 3 και 4

20. Οι θέσεις δυο σωμάτων σε διαδοχικές χρονικές στιγμές που απέχουν μεταξύ τους κατά 0.20 s παριστάνονται από τα αριθμημένα τετράγωνα στο παρακάτω σχήμα. Τα δυο σώματα κινούνται προς τα δεξιά.



Οι επιταχύνσεις των δυο σωμάτων συνδέονται ως εξής:

(1) Η επιτάχυνση του Α είναι μεγαλύτερη από την επιτάχυνση του Β

(2) Η επιτάχυνση του Α είναι ίση με την επιτάχυνση του Β. Και οι δυο επιταχύνσεις είναι μεγαλύτερες του μηδενός

(3) Η επιτάχυνση του Β είναι μεγαλύτερη από την επιτάχυνση του Α

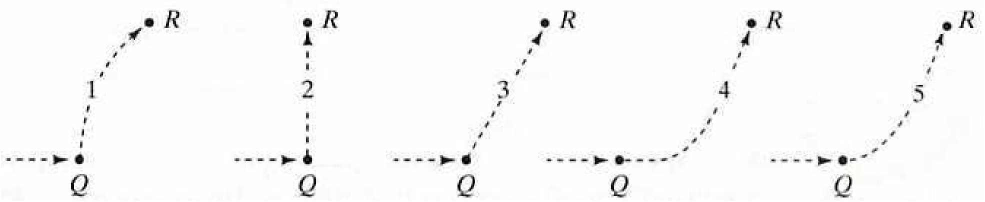
(4) Η επιτάχυνση του Α είναι ίση με την επιτάχυνση του Β. Και οι δυο επιταχύνσεις είναι μηδέν.

(5) Δεν δίνονται αρκετές πληροφορίες για να απαντηθεί η ερώτηση

Χρησιμοποιήστε τις παρακάτω πληροφορίες και το σχήμα για να απαντήσετε τις επόμενες τέσσερις ερωτήσεις (21 ως 24)

Ένα διαστημόπλοιο κινείται πλάγια στο διάστημα από το σημείο P στο σημείο Q όπως δείχνει το παρακάτω σχήμα. Το διαστημόπλοιο δεν δέχεται εξωτερικές δυνάμεις. Όταν φτάσει στο σημείο Q παίρνει μπροστά η μηχανή του διαστημόπλοιου και προκαλεί μια σταθερή ώθηση (δηλαδή δύναμη στο διαστημόπλοιο) με διεύθυνση κάθετη στην ευθεία PQ. Αυτή η σταθερή ώθηση ασκείται μέχρι το διαστημόπλοιο να φτάσει στο σημείο R στο διάστημα.

21. Ποια από τις τροχιές 1-5 περιγράφει καλύτερα την τροχιά του διαστημόπλοιου ανάμεσα στα σημεία Q και R;



22. Ενώ το διαστημόπλοιο κινείται από το σημείο Q στο σημείο R η ταχύτητα του

(1) Παραμένει σταθερή

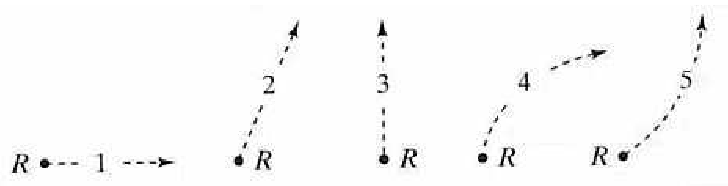
(2) Αυξάνεται συνεχώς

(3) Μειώνεται συνεχώς

(4) Αυξάνεται για λίγο και μετά παραμένει σταθερή

(5) Είναι σταθερή για λίγο και μετά μειώνεται.

23. Στο σημείο R το διαστημόπλοιο σβήνει τη μηχανή του και αμέσως η ώθηση που δέχεται μηδενίζεται. Ποια από τις παρακάτω τροχιές 1-5 θα ακολουθήσει το διαστημόπλοιο μετά το σημείο R.



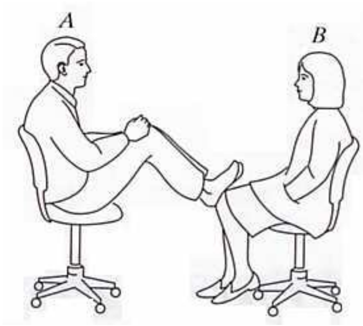
24. Μετά το σημείο R, η ταχύτητα του διαστημόπλοιου
(1) Είναι σταθερή
(2) Αυξάνεται συνεχώς
(3) Μειώνεται συνεχώς
(4) Αυξάνεται για λίγο και μετά μένει σταθερή
(5) Είναι σταθερή για λίγο και μετά μειώνεται

25. Μια γυναίκα ασκεί μια σταθερή οριζόντια δύναμη σε ένα μεγάλο κουτί. Ως αποτέλεσμα, το κουτί κινείται κατά μήκος του οριζόντιου δαπέδου με σταθερή ταχύτητα $υ_0$. Η σταθερή οριζόντια δύναμη που ασκεί η γυναίκα
(1) Έχει μέτρο ίσο με αυτό του βάρους του κουτιού
(2) Είναι μεγαλύτερη από το βάρος του κουτιού
(3) Έχει το ίδιο μέτρο με τη συνολική δύναμη που αντιστέκεται στην κίνηση του κουτιού
(4) Είναι μεγαλύτερη από τη συνολική δύναμη που αντιστέκεται στην κίνηση του κουτιού
(5) Είναι μεγαλύτερη είτε από το βάρος του κουτιού είτε από τη συνολική δύναμη που αντιστέκεται στην κίνηση

26. Αν η γυναίκα της παραπάνω ερώτησης διπλασιάσει τη σταθερή οριζόντια δύναμη που ασκεί στο κουτί για να το σπρώξει πάνω στο ίδιο οριζόντιο δάπεδο, τότε το κουτί κινείται
(1) με σταθερή ταχύτητα διπλάσια της $υ_0$ της προηγούμενης ερώτησης
(2) με σταθερή ταχύτητα μεγαλύτερη αλλά όχι απαραίτητα διπλάσια της $υ_0$ της προηγούμενης ερώτησης
(3) για ένα χρονικό διάστημα με ταχύτητα σταθερή και μεγαλύτερη της $υ_0$ της προηγούμενης ερώτησης, και μετά με ταχύτητα που αυξάνει διαρκώς
(4) για ένα χρονικό διάστημα με ταχύτητα που αυξάνει και μετά με σταθερή ταχύτητα
(5) Με ταχύτητα που αυξάνει συνεχώς

27. Αν η γυναίκα της ερώτησης 25 ξαφνικά σταματήσει να ασκεί την οριζόντια δύναμη στο κουτί, τότε αυτό
(1) Θα σταματήσει αμέσως
(2) Θα συνεχίσει να κινείται με σταθερή ταχύτητα για ένα χρονικό διάστημα και μετά θα επιβραδύνει μέχρι να σταματήσει
(3) Θα αρχίσει αμέσως να επιβραδύνει μέχρι να σταματήσει
(4) Θα συνεχίσει να κινείται με σταθερή ταχύτητα
(5) Θα αυξήσει την ταχύτητα του για ένα χρονικό διάστημα και μετά θα αρχίσει να επιβραδύνει μέχρι να σταματήσει

28. Στο παρακάτω σχήμα, ο φοιτητής Α έχει μάζα 75 kg και η φοιτήτρια Β έχει μάζα 57 kg. Κάθονται σε όμοιες καρέκλες γραφείου αντικριστά. Ο φοιτητής Α βάζει τα πόδια του πάνω στα γόνατα της φοιτήτριας Β, όπως φαίνεται στο σχήμα. Ο φοιτητής Α ξαφνικά κλωτσάει με τα πόδια του προς τα έξω κάνοντας και τις δυο καρέκλες να κινηθούν.



Κατά τη διάρκεια της κλωτσιάς και ενώ οι φοιτητές αγγίζουν ο ένας τον άλλο

(1) Κανένας από τους δυο φοιτητές δεν ασκεί δύναμη στον άλλο
(2) Ο φοιτητής Α ασκεί δύναμη στη φοιτήτρια Β, αλλά η Β δεν ασκεί δύναμη στον Α
(3) Και οι δυο φοιτητές ασκούν δύναμη ο ένας στον άλλο αλλά η Β ασκεί μεγαλύτερη δύναμη
(4) Και οι δυο φοιτητές ασκούν δύναμη ο ένας στον άλλο αλλά ο Α ασκεί μεγαλύτερη δύναμη
(5) Και οι δυο φοιτητές ασκούν την ίδια δύναμη ο ένας στον άλλο

29. Μια άδεια καρεκλά γραφείου βρίσκεται ακίνητη πάνω στο δάπεδο. Θεωρείστε τις παρακάτω δυνάμεις:

(A) Μια δύναμη προς τα κάτω λόγω βαρύτητας
(B) Μια δύναμη προς τα πάνω που ασκείται από το δάπεδο
(C) Μια συνιστάμενη δύναμη προς τα κάτω που ασκείται από τον αέρα

Ποια (ποιες) από τις παραπάνω δυνάμεις ασκείται (ασκούνται) πάνω στην καρέκλα ;

(1) Μόνο η Α
(2) Η Α και η Β
(3) Η Β και η C
(4) Οι Α, Β, και C
(5) Καμία από τις παραπάνω δυνάμεις (αφού η καρέκλα είναι ακίνητη δεν ασκούνται δυνάμεις πάνω της)

30. Παρολο τον πολύ δυνατό αέρα, μια τενίστρια καταφέρνει να κτυπήσει τη μπάλα του τένις με τη ρακέτα της έτσι ώστε η μπάλα να περάσει πάνω από το δίχτυ και να προσγειωθεί στο γήπεδο της αντίπαλου της. Θεωρείστε τις παρακάτω δυνάμεις

(A) Μια δύναμη προς τα κάτω λόγω βαρύτητας
(B) Μια δύναμη λόγω του «κτυπήματος»
(C) Μια δύναμη που ασκείτε από τον αέρα

Ποια (ποιες) από τις παραπάνω δυνάμεις ασκείται (ασκούνται) πάνω στη μπάλα του τένις αφού έφυγε από τη ρακέτα και πριν προσγειωθεί στο έδαφος

(1) Μόνο η Α
(2) Οι Α και Β
(3) Οι Α και C
(4) Οι Β και C
(5) Οι Α, Β, και C

APPENDIX B

The open-ended questionnaire (Part A)

PART A:

Here are some of your fellow-students' statements on issues concerning mechanics. Do you agree or disagree with these statements? Try to explain to them <u>why</u> you agree or disagree. In case you do not know, tick the option 'I do not know' and explain why you cannot decide (e.g. you do not know the theory, you do not understand the statement…)

1. Georgia asserts that*:' If I release simultaneously two metallic balls from the roof of a building and one weights twice as much as the other, then the heavier ball needs half the time to reach the ground than the lighter ball.'*

□ I AGREE □I DISAGREE □I DO NOT KNOW

Explain why...

2. John says that: '*If the same above balls roll on a horizontal table with the same velocity, reach the edge of the table and fall, then the heavier ball will hit the ground at approximately half the horizontal distance, from the base of the table, as compared to the lighter ball.'*

□ I AGREE □I DISAGREE □I DO NOT KNOW

Explain why...

3. Manos asserts that: *'If a big lorry is involved in a head-on collision with a small car, then during the collision the lorry exerts force on the car which has the same magnitude as the one that the car exerts on the lorry.'*

□ I AGREE □I DISAGREE □I DO NOT KNOW

Explain why...

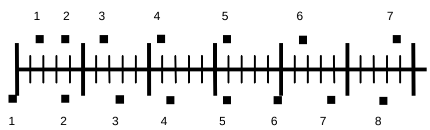
4. The engine of a big truck stops working while the truck is moving. A small car pushes the truck out of the road. Maria says that*: 'While the car pushes the truck and the car accelerates till it reaches its maximum velocity, the force by which the car pushes the truck is bigger than the force that the truck exerts on the car.'*

□ I AGREE □I DISAGREE □I DO NOT KNOW

Explain why...

5. **5.** Vaggelis observes the following figure which illustrates the positions of two bodies in successive time intervals of 0.20 s. The two bodies are moving towards the right. He says that: *'The two bodies have the same velocity at positions 2 and 5.'*

□ I AGREE □I DISAGREE □I DO NOT KNOW



Explain why...

6. Jack says*: 'If a person stands on an elevator that accelerates, then it is not correct to say that N=W (where N is the normal force from the elevator floor, and W is the person's weight).'*

□ I AGREE □I DISAGREE □I DO NOT KNOW

Explain why...

APPENDIX C

The open-ended questionnaire (Part B)

PART B:

Here are some of your classmates' arguments. Are you convinced by them or not? Try to explain to them what is/are exactly the element/s of the argument that persuaded or did not persuade you. If you do not know, tick the option 'I do not know' and explain why you cannot decide (e.g. you do not know the theory, you do not understand the statement…)

7.	Some students discuss in the classroom the possibility that the rotational speed of the earth changes, if the ice in the North Pole melts and the level of the sea rises. Tom says: *'I believe that it can happen. Ice floats because of Pascal's principle. The density of ice is less than the density of water and it floats because of the buoyant force. If ice melts, then the level of the sea will rise. And because the angular momentum of the earth must be conserved, then inevitably the rotational speed of the earth must change.'* □He convinced me □ He did not convince me □I do not know Explain why….
8.	Some students discuss in the classroom about a man who is standing on a scale inside an elevator. The question under discussion is the reading of the scale when the elevator is going up or down. Nik says that: *'The reading of the scale is determined by the normal force from the elevator floor. Given that the normal force from the elevator floor equals the person's weight and that the person's weight does not change, then the normal force from the elevator floor does not change. Therefore the reading of the scale does not change.'* □He convinced me □ He did not convince me □I do not know Explain why…
9.	Evi says that: *'If during the pre-historic period the ice in the North Pole melted and the earth flooded, then the dinosaurs would have been extinct. Dinosaurs are extinct. Therefore, in the pre-historic period the ice in the North Pole melted and the earth flooded.'* □She convinced me □ She did not convince me □I do not know Explain why…
10.	Some months ago, thousands of people received an e-mail, urging them to jump at a specific date and time. The person who sent this e-mail asserted that in that way it is possible to make the earth change its orbit around the sun. You are present in the classroom as the students discuss on whether something like this is possible, even if the whole population jumps at exactly the same time. At some point, a famous Physics professor from Harvard University who is visiting our University and is present in the discussion says: *'Yes, it may happen!'* She gives no further explanation. □She convinced me □ She did not convince me □I do not know Explain why…

REFERENCES

ALAO, S. & GUTHRIE, J., 1999. Predicting conceptual understanding with cognitive and motivational variables. *The Journal of Educational Research*, 92 (4), 243-54.

BELL, P. & LINN, M., 2000. Scientific arguments as learning artifacts: designing for learning from the web with KIE. *International Journal of Science Education*, 22(8), 797-817.

BREWER, W. F., CHINN, C. A. & SAMARAPUNGAVAN, A., 2000. Explanations in scientists and children. In F.C. KEIL & R. A. WILSON, eds., *Explanation and cognition,* Cambridge, MA: The MIT Press, 279-98.

BOOTH, K. M. & JAMES, B. W., 2001. Interactive learning in a higher education Lenel 1 mechanics module. . *International Journal of Science Education*, 23(9), 955-67.

BUNCICK, M., C. & HORGAN, D.D, 2001. Using demonstrations as a contextual road map: enhancing course continuity and promoting active engagement in introductory college physics. *International Journal of Science Education*, 23 (12), 1237-55.

CHAIMALA, F., 2004. *The purposes of physics education: An analysis of the Greek national curriculum and textbooks*. Unpublished masters' dissertation. University of Southampton, Southampton, UK.

CHAN, C., BURTIS, J. & BEREITER, C., 1997. Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction*, 15, 1-40.

CHANG, W., 2005. The rewards and challenges of teaching innovation in university physics: 4 years' reflection. *International Journal of Science Education*, 27(4), 407-25.

CHINN, C. A. & BREWER, W. F., 1993. The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science education. *Review of Educational Research,* 63 (1), 1-49.

CHIU, M., GUO, C. & TREAGUST, D.F., 2007. Assessing students' conceptual understanding in science : An introduction about a national project in Taiwan. *International Journal of Science Education,* 29 (4), 379-90.

CLARK, D. B. & SAMPSON, V., 2006. Personally-seeded discussions to scaffold online argumentation. . *International Journal of Science Education,* 29 (3), 253-77

COHEN, E. G., 1994. Restructuring the classroom: conditions for productive small groups. *Review of Educational Research*, 64, 1-35.

COHEN, L., MANION, L. & MORRISON, K., 2000. *Research methods in education*. 5th ed. London and New York: Routledge Falmer.

CROUCH, C. & MAZUR, E., 2001. Peer Instruction: Ten years of experience and results. *American Journal f Physics*, 69(9), 970-76.

DANCY, M. H., 2000. *Investigating animations for assessment with an animated version of the FCI.* Doctoral dissertation, North Carolina State University.

DREYFUS, A., JUNGWIRTH, E. & ELIOVITCH, R., 1990. Applying the cognitive conflict strategy for conceptual change – some implications, difficulties and problems. *Science Education*, 74, 555-569.

DRIVER, R., ASOKO, H., LEACH, J., MORTIMER, E. & SCOTT, P., 1994. Constructing scientific knowledge in the classroom. *Educational Researcher*, 23, 5-12.

DRIVER, R., NEWTON, P. & OSBORNE, J., 2000. Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84, 287-312.

DUIT, R., 1995. The constructivist view: A fashionable and fruitful paradigm for science education research and practice. In L. STEFFE & J. GALE, eds. *Constructivism in education*. Hillsdale, NJ: Lawrence Erlbaum, 272 – 285.

DUIT, R., 1999. Conceptual change approaches in science education. In W. SCNOTZ, S. VOSNIADOU & M. CARRETERO, eds. *New perspectives on conceptual change*, Amsterdam, NL: Pergamon, 263-82.

DUIT, R. & TREAGUST, D.F., 2003. Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education*, 25(6), 671-88.

DUSCHL, R. A. & OSBORNE, J., 2002. Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, 38, 39-72.

EICHINGER, D.C., ANDERSON, C.W., PALINCSAR, A.S & DAVID, Y.M., 1991. An illustration of the roles of content knowledge, scientific argument and social norm in collaborative problem solving. Paper presented at AERA, Chicago April.

ERDURAN, S., SIMON, S. & OSBORNE, J., 2004. TAPping into argumentation: developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, 88 (6), 915-33.

ERDURAN, S., 2008. Methodological foundations in the study of argumentation in science classrooms. In: S. ERDURAN & M. P. JIMENEZ-ALEIXANDRE, eds. *Argumentation in Science Education: Perspectives from classroom-based research.* Springer.

EURYDICE, THE INFORMATION NETWORK ON EDUCATION IN EUROPE, 2006. *Science teaching in schools in Europe: Policies and Research*. Brussels: Eurydice.

EYLON, B. & LINN, M. C., 1998. Learning and instruction: An examination of four research perspectives in science education. *Review of Educational Research*, 58 (3), 251-301.

FAGEN, A. P., CROUCH, C. H. & MAZUR, E., 2002. Peer instruction: Results from a range of classrooms. *The physics teacher*, 40, 206-9.

FLICK, U., 2002. *An introduction to qualitative research*. 2nd ed. London: Sage.

GAUTREAU, R. & NOVEMSKY, L., 1997. Concepts first- A small group approach to physics learning. *American Journal of Physics*, 65(5), 418-28.

GILBERT, J. K., BOULTER, C. & RUTHERFORD, M., 1998. Models of explanations, Part 1: Horses for courses? *International Journal of Science Education*, 20 (1), 83-97.

GOODLAD, J., 1994. Curriculum as a field of study. In: T.HUSEN & T.N.POSTLETHWAITE, eds. *The international Encyclopedia of Education*. Oxford: Pergamon Press, 1262-1267.

GOW, L. & KEMBER, D., 1990. Does higher education promote independent learning? *Higher Education*, 19 (3), 307-22.

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

HELLER, P. & HOLLABAUGH, M., 1992. Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics*, 60(7), 637-44.

HESTENES, D., WELLS, M. & SWACKHAMER, G., 1992. Force concept inventory. *The Physics Teacher*, 30, 141-58.

HESTENES, D. & HALLOUN, I., 1995. Interpreting the Force Concept Inventory. *The Physic Teacher*, 33, 502-6.

HOWE, K.R., 1998. Against the quantitative-qualitative incompatibility thesis, or, Dogmas die hard. *Educational Researcher*, 17, 10-16.

HUFFMAN, D. & HELLER, P., 1995. What does the force concept inventory actually measure? *The Physics Teacher*, 33, 138-43.

JIMENEXEX-ALEIXANDRE, M. P., RODRIGUEZ, A.B. & DUSCHL, R., 2000. “Doing the lesson” or “doing science”: Argument in high school genetics. *Science Education*, 84, 757-92.

JOHNSON , R. B. & TURNER, L. A., 2003. Data collection strategies in mixed methods research. In A. TASHAKKORI & C. TEDDIE, Eds. *Handbook of mixed methods in social and behavioral research.* Thousand Oaks, CA: Sage

JOHNSON, R., B. & CHRISTENSEN, L. B., 2004. *Educational research: quantitative, qualitative and mixed approaches*. Boston, MA: Allyn and Bacon.

JOHNSON, R. B. & ONWUEGBUZIE, A. J., 2004. Mixed-methods research: a research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.

KELLY, G. J. & TAKAO, A., 2002. Epistemic levels in argument. An analysis of University Oceanography Students’ use of evidence in writing. *Science Education*, 86, 314-42.

KOLSTO, S. D. & RATCLIFFE, M., 2008. Social aspects of Argumentation. In: S. ERDURAN & M. P. JIMENEZ-ALEIXANDRE, eds. *Argumentation in Science Education: Perspectives from classroom-based research.* Springer.

KOSLOWSKI, B., 1996. *Theory and evidence: the development of scientific reasoning*. Cambridge, MA: MIT Press.

KUHN, D., 1991. *The skills of argument.* Cambridge: Cambridge University Press.

KUHN, D., SHAW, V., & FELTON, M., 1997. Effects of dyadic interactions on reasoning. *Cognition and Instruction*, 15(3), 287-315.

LAWS, P. W., 1997. Millikan lecture 1996: Promoting active learning based on physics education research in introductory physics courses. *American Journal of Physics*, 65(1), 14-21.

LAWSON, A. E., 2003. The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International Journal of Science Education*, 25 (11), 1387-408.

LEMKE, J.L., 1990. *Talking science: Language, learning and values*. Norwood, NJ: Ablex.

LENSEN, M. S. & FINLEY, J., 1995. Teaching evolution using historical arguments in a conceptual change strategy. *Science Education*, 79 (2), 147-166.

LIGHT, G. & COX, R., 2001. *Learning and teaching in Higher Education: the reflective professional.* London: Sage Publications.

LIMON, M. & CARRETERO, M., 1997. Conceptual change and anomalous data: a case study in the domain of natural sciences. *European Journal of Psychology of Education*, 12 (2), 213-230.

LIMON, M., 2001. On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction*, 11, 357-80.

LIMON, M. & MASON, L., 2002. *Reconsidering conceptual change: Issues in theory and practice.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

LINCOLN, Y. S. & GUBA, E. G., 1985. *Naturalistic Inquiry*. Beverly Hills: Sage Publications.

MALONEY, J. & SIMON, S., 2006. Mapping students' discussions of evidence in science to assess collaboration and argumentation. *International Journal of Science Education*, 28(15), 1817-41.

McNEILL, K. L.& KRAJCIK, J., 2008. Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching,* 45 (1), 53-78.

McROBBIE, C., & TOBIN, K., 1997. A social constructivism perspective on learning environments. *International Journal of Science Education*, *19*(2), 193–208.

MARTIN, M., 1972. *Concepts of science education: A philosophical analysis*. London: Scott, Foresman.

MARTUNNEN, M., 1992. Commenting on written arguments as a part of argumentation skills-comparison between students engaged in traditional vs. on-lone study. *Scandinavian Journal of Educational Research*, 36 (4), 289-302.

MASON, L., 2000. Role of anomalous data and epistemological beliefs in middle school students' theory change about two controversial topics. *European Journal of Psychology of Education*, *15 (2),* 329-346.

MAZUR, E, 1997. *Peer Instruction: A users' manual*. Upper Saddle River, NJ, Prentice Hall.

MELTZER, D. E. & MANIVANNAN, K., 2002. Transforming the lecture-hall environment: The fully interactive physics lecture. American Journal of Physics, 70 (6), 639-54.

MILLAR, R. & OSBORNE, J.F, 1998. *Beyond 2000: Science education for the future.* London: King's College.

MILLS, D., McKITTRICK, B., MULHALL, P. & FETERIS, F., 1999. CUP: cooperative learning that works. *Physics Education*, 34(1), 11-5.

MURPHY, J. P., 1990. Pragmatism: from Pierce to Davidson. Boulder, CO: Westview.

NEWTON, P., DRIVER, R. & OSBORNE, J., 1999. The place of argumentation in the pedagogy of school science. International Journal of Science Education, 21(5), 553-576.

NIESWANDT, M., 2007. Student affect and conceptual understanding in learning chemistry. *Journal of Research in Science Teaching*, 44 (7), 908-37.

NORRIS, S. P., GUILBERT, S. M., SMITH, M. L., HAKIMELAHI, M. & PHILLIPS, L. M., 2005. A theoretical framework for narrative explanation in science. *Science Education*, 89, 535-63.

NOVAK, J., 1977. *A theory of education*. Ithaca, New York: Cornell University Press.

NRC (National Research Council), 1995. *National Science Education Standards.* Washington: National Academic Press.

OSBORNE, J., 1997. Practical alternatives. *School Science Review*, 78, 61-6.

OSBORNE, J. F., 1998. Beyond constructivism. *Science Education*, 80, 1, 53-82.

OSBORNE, J., ERDURAN, S. & SIMON, S., 2004. Enhancing the quality of argument in school science. *Journal of Research in Science Teaching*, 41, 994-1020.

PINTRICH, P. R. & GARCIA, T., 1991. Students' goal orientation and self-regulation in the college classroom. In M. MAEHR & P. R. PINTRICH, eds. *Advances in motivation and achievement: motivation, enhancing environment, vol.6.* Greenwich, CT: JAI Press.

PINTRICH, P.R., MARX, R.W. & BOYLE, R.A, 1993. Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63, 167-99.

POSNER, G. J., STRIKE, K.A., HEWSON, P.W. & GERTZOG, W.A., 1982. Accommodation of a scientific conception: Towards a theory of conceptual change. *Science Education*, 66, 211-27.

SAMIULLAH, M., 1995. Effect of in-class student-student interaction on the learning of physics in a college physics course. *American Journal of Physics*, 63(10), 944-50.

SANDOVAL, W. A. & REISER, B. J., 2004. Explanation-driven inquiry: integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88, 345-72.

SANDOVAL, W. A. & MILLWOOD, K. A., 2005. The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, 23 (1), 23-55.

SAVINAINEN, A. & SCOTT, P., 2002. The FCI: a tool for monitoring student learning. *Physics Education*, 37, 45-52.

SAVINAINEN, A. & VIIRI, J., 2008. The force concept inventory as a measure of students' conceptual coherence. *International Journal of Science and Mathematics Education*, 6(4), 719-40.

SEYMOUR, E., 2001. Tracking the process of change in US undergraduate education in science, mathematics, engineering, and technology. *Science Education*, 86, 79-105.

SHARMA, M., D., MILLAR, R. & SETH, S., 1999. Workshop tutorials: accommodating student-centered learning in large first year university physics courses. *International Journal of Science Education,* 21(8), 839-53.

SIMON, S., OSBORNE, J. & ERDURAN, S., 2003. Systemic teacher development to enhance the use of argumentation in school science activities. In J. WALLACCE & J. LOUGHARAN, Eds. *Leadership and professional development in science education*. London: Routledge Falmer, 198-217.

SOLOMON, J., 1986. Children's explanations. *Oxford Review of Education*, 12 (1), 41-51.

SUPPE, F., 1998. The structure of a scientific paper. *Philosophy of science*, 65, 381-405.

THACKER, B., EUNSOOK, K., TREFZ, K. & LEA, S., M., 1994. Comparing problem solving performance of physics students in inquiry-based and traditional introductory physics courses. *American Journal of Physics*, 62(7), 627-33.

TOULMIN, S., 1958. *The uses of argument*. Cambridge: Cambridge University Press.

TYSON, L. T., VENVILLE, G. J., HARRISON, A.G. & TREAGUST, D. F., 1997. A multidimensional framework for interpreting conceptual change events in the classroom. *Science Education,* 81, 387-404.

VAN DEN AKKER, J., 1998. The science curriculum: Between ideals and outcomes. In: B. J. FRASER AND K. G. TOBIN, eds. *International handbook of science education.* Dordrecht, Boston, London: Kluwer Academic Publishers, 421-47.

VENVILLE, G. J. & TREAGUST, D.F., 1998. Exploring conceptual change in genetics using a multidimensional interpretive framework. *Journal of Research in Science Teaching*, 35, 1031 – 55.

VON AUFSCHNAITER, C., 2003. Interactive processes between university students: Structures of interactions and related cognitive development. *Research in Science Education*, 33, 341-74.

VOSNIADOU, S., 1994. Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4, 45-69.

VOSNIADOU, S., 2004. Extending the conceptual change approach to mathematics learning and teaching. *Learning and Instruction*, 14, 445-51.

VOSNIADOU, S., IOANNIDES, C., DIMITRAKOLOULOU, A. & PAPADEMETRIOU, E., 2001. Designing learning environments to promote conceptual change in science. *Learning and Instruction*, 11, 381-419.

WALTON, D. N., 1997. *Appeal to expert opinion: arguments from authority*. Pennsylvania: The Pennsylvania State University Press.

ZEIDLER, D. L., 1997. The central role of fallacious thinking in science education. *Science Education*, 81(4), 483-96.

ZOHAR, A. & NEMET, F., 2002. Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39 (1), 35-62.

ZOLLER, U., BEN-CHAIM, D., PENTIMALLi, R., & BORSESE, A., 2000. The disposition towards critical thinking of high school and university science students: An inter–intra Israeli–Italian study. *International Journal of Science Education*, 22, 571–582.