UNIVERSITY OF JOENSUU
DEPARTMENT OF PHYSICS

DISSERTATIONS 41

High School Students' Conceptual Coherence of Qualitative Knowledge in the Case of the Force Concept

(slightly modified version)

Antti Savinainen

ACADEMIC DISSERTATION

To be presented, with permission of the Faculty of Science of the University of Joensuu, for public criticism in Auditorium M4 of the University, Yliopistokatu 7, Joensuu, on August 13th, 2004, at 12 noon.

JOENSUU 2004

Julkaisija Publisher	Joensuun yliopisto University of Joensuu
Toimittaja Editor	Timo Jääskeläinen, Ph.D., Professor
Ohjaajat Supervisors	Jouni Viiri, Ph.D., Ed.D., Docent University of Joensuu, Joensuu, Finland Ismo Koponen, Ph.D., Docent University of Helsinki, Helsinki, Finland Timo Jääskeläinen, Ph.D., Professor University of Joensuu, Joensuu, Finland
Esitarkastajat Reviewers	Maija Ahtee, Ph.D., Professor (Emerita) University of Jyväskylä, Jyväskylä, Finland David Meltzer, Ph.D., Assistant Professor Department of Physics and Astronomy, Iowa State University of Science and Technology, Ames, Iowa, USA
Vastaväittäjä Opponent	Daniel MacIsaac, Ph.D., Assistant Professor Department of Physics, State University of New York (SUNY) College at Buffalo, Buffalo NY, USA
Vaihto	Joensuun yliopiston kirjasto, vaihdot PL 107, 80101 JOENSUU Puh. 013-251 2677, telefax 013-251 2691 Email: vaihdot@joensuu.fi
Exchange	Joensuu University Library, exchanges P.O. Box 107, FIN-80101 JOENSUU Telefax +358 13 251 2691 Email: vaihdot@joensuu.fi
Myynti	Joensuun yliopiston kirjasto, julkaisujen myynti PL 107, 80101 JOENSUU Puh. 013-251 2652, 251 2677, telefax 013-251 2691 Email: joepub@joensuu.fi
Sale	Joensuu University Library, sale of publications P.O. Box 107, FIN-80101 JOENSUU Telefax +358 13 251 2691 Email: joepub@joensuu.fi

ISSN 1458-5332
ISBN 952-458-509-X

Joensuun yliopistopaino 2004

Antti Savinainen[*]; **High School Students' Conceptual Coherence of Qualitative Knowledge in the Case of the Force Concept –** University of Joensuu. Department of Physics. Dissertations 41, 2004 - 106 p.

Keywords: teaching of physics, force concept, conceptual coherence, the Force Concept Inventory

[*]Address: Kuopio Lyseo High School, Puijonkatu 18, Kuopio, FIN-70110, Finland

Abstract

This study consists of a theoretical and an empirical part. The theoretical research aims were to characterise students' conceptual coherence of qualitative knowledge in the case of the force concept, and how it can be evaluated. Students' conceptual coherence can be divided into three aspects: representational coherence, which is the ability to use multiple representations and move between them; contextual coherence, i.e. the ability to apply concepts in a variety of contexts (familiar and novel), and conceptual framework coherence, which addresses the relations - integration and differentiation - between relevant concepts. Certain groupings of the Force Concept Inventory (FCI), the Force and Motion Conceptual Evaluation (FMCE), and the Test for Understanding Graphs - Kinematics (TUG-K) questions were used to probe students' contextual and representational coherence of the force concept. Written extended response questions and interviews were also used in addition to multiple choice tests to provide complementary data.

The empirical part of this dissertation consists of designing a teaching approach (Interactive Conceptual Instruction (ICI)) and teaching sequences for kinematics and the force concept. The ICI approach involves several features or components: conceptual focus (concepts are introduced and rehearsed before quantitative problem solving), the use of multiple representations in varying contexts, classroom interactions (peer instruction), research-based materials, use of texts (reading before formal treatment), and concept maps. The teaching sequence for the force concept emphasises forces as interactions.

An empirical study was conducted to test the effectiveness of the ICI teaching. The study involved two pilot and two study groups in Kuopio Lyseo High School: Preparatory International Baccalaureate (Pre-IB) students (age 16; $n_{pilot} = 22$ and $n_{study} = 23$) and Finnish National Syllabus students (age 17; $n_{pilot} = 52$ and $n_{study} = 49$). The pilot groups followed the ICI approach without a focus on forces as interactions whereas the study groups followed the ICI approach with a focus on forces as interactions. The study groups were taught to think of forces as interactions through the systematic use of a modified version of the 'Symbolic Representation of Interactions', which provided a bridging representation to more abstract free-body diagrams. Otherwise, introductory mechanics was taught in a similar manner to the pilot and study groups (i.e., the same teacher - author AS - taught all the groups using the same textbooks, with generally similar exercises and activies, and the same ICI approach).

Average normalized gain (Hake gain) and effect size were used as measures of the practical significance of the overall FCI results. Hake gains for the pilot and study groups fall in the middle or upper end of the 'medium gain region' ($0.3 < (<g>) <0.7$): they were between 0.45 and 0.59. The effect sizes were well above the 'high boundary of 0.8': they were between 1.1. and 2.6. These indices show that the effect of both types of ICI teaching had practical significance at least as measured by the overall FCI results. The most impressive conceptual gains were made in Newton's first law in verbal representation, Newton's third law in verbal representation, and contact force in verbal representation. In almost all these cases Hake gains were above 0.50 and effect sizes above

1.1. The ICI teaching enhanced the contextual and representational coherence of the force concept in all the probed dimensions of the force concept for the pilot and study groups. In most dimensions the changes were also statistically significant ($p \leq 0.05$). In general, the most notable improvement in contextual and representational coherence occured in Newton's first law (all groups) and Newton's third law (the study groups) in verbal representation. In most groups, fewer students reached contextual coherence of Newton's first law in diagrammatic representation. It can also be concluded that Newton's second law proved to be harder for all groups than the first law.

The study groups had much better results in Newton's third law. More students' in the study groups exhibited contextual coherence in Newton's third law after teaching than in the pilot groups (the differences were statistically significant: $p \leq 0.023$). The differences were also practically significant: e.g. the effect size for the FCI questions addressing Newton's third law for the Pre-IB study group was extremely high (3.3). In other dimensions of the force concept the results are not conclusive: the Pre-IB study group did not do better than the Pre-IB pilot group in most of the dimensions and representations of the force, whereas the Finnish study group was better than the Finnish pilot group in the majority of the dimensions and representations of the force concept. Hence, it cannot be concluded that focusing on forces as interactions necessarily enhances students' conceptual coherence of the force concept in dimensions other than Newton's third law.

“*I don't know what's the matter with people: they don't learn by understanding;*
they learn by some other way – by rote, or something. Their knowledge is so fragile!
...So this kind of fragility is, in fact, fairly common, even with more learned people.”

R.P. Feynman (1991, 36-37)

Preface

This work was carried out during 2000-2004 mostly along with teaching duties in Kuopio Lyseo High School. However, I got a very good start when I worked as a Marie Curie Fellow in the University of Leeds (UK) for three months in 2001. The time was not long but it was quite intense. I had the pleasure of being supervised by Dr. Philip Scott, whose expertise helped me to clarify what I was actually trying to study. This collaboration was most fruitful: I wrote one conference article and two journal articles with Dr. Scott while in Leeds. I wish to express my deep gratitude to Dr. Scott and I am happy that our collaboration has continued after I came back to Finland. I also wish to thank Professor John Leach and Dr. Jenny Lewis for their support and kind hospitability.

I warmly thank my supervisor, Dr. Jouni Viiri, for his continuous support and guidance throughout. Dr. Viiri's help was 'multi-modal': he co-authored three articles, provided valuable advice in structuring this thesis, and sent me copies of numerous research articles. I feel very fortunate to have had him as my supervisor. I also wish to thank my other supervisor Dr. Ismo Koponen who encouraged me from the very beginning to work towards an article-based thesis. His constructive criticism has been most helpful in finalising this thesis. I especially benefited from Dr. Koponen's expertise when writing the historical outline of the force concept. I am indebted to Juhani Rautopuro, Lic.Ed., for his help with the SPSS statistical program. I thank Professor (Emerita) Maija Ahtee and Assistant Professor David Meltzer for very carefully reviewing this thesis. I also thank Professor Richard Hake for his help and encouraging comments on my first PER publication.

I wish to thank my principal, Leena Auvinen, for her positive attitude towards my research. I also thank my colleague Kauko Kauhanen for pleasant and good collaboration in teaching physics in Kuopio Lyseo High School. My editor and colleague Vivian Paganuzzi deserves special thanks for his careful editing of this thesis and his constructive criticism which greatly helped me to finalize the last chapter. I thank Professor Markku Kuittinen for providing valuable advice on preparing this thesis for printing. I also thank my former student Juhani Mykkänen for transcribing the interview data. Moreover, I am grateful to my students who have helped me to become a better teacher.

I thank the Finnish Cultural Foundation of Northern Savo and Kuopio Naturalists' Society (KLYY r.y.) for making it possible to take a leave of absence from teaching to write this thesis. I also thank the Physics Department of the University of Joensuu and *Matemaattisten aineiden opetustyön tukisäätiö* for making my conference trip to PERC 2003 (Madison, Wisconsin, USA) possible. I wish to thank my father, Toivo Savinainen, who gave me a substantial grant for the same conference trip. I take this opportunity to thank my parents, Tyyne and Toivo Savinainen, for giving me a living example of the value of hard work.

Finally, my warmest thanks belong to my family: this work would not have been possible at all without love and support from my wife Päivi and my daughter Ellamaria.

Kuopio, June 23, 2004

Antti Savinainen

List of Original Publications

This thesis is based on the articles referred to in the text by their Roman numerals. Some unpublished results are also presented.

I. Savinainen, A. and Scott, P. (2002a). The Force Concept Inventory: a tool for monitoring student learning. *Physics Education*, **37,** pp. 45–52.

II. Savinainen, A. and Scott, P. (2002b). Using the Force Concept Inventory to monitor student learning and to plan teaching. *Physics Education*, **37**, pp. 53-58.

III. Savinainen, A. and Viiri, J. (2003). Using the Force Concept Inventory to Characterise Students' Conceptual Coherence.
In L. Haapasalo and K. Sormunen (Eds.): Towards Meaningful Mathematics and Science Education, Proceeding on the IXX Symposium of Finnish Mathematics and Science Education Research Association. *Bulletin of Faculty of Education*, No 86, University of Joensuu, pp. 142-152.

IV. Savinainen, A. and Viiri, J. (2004). A Case Study Evaluating Students' Representational Coherence of Newton's First and Second Laws. In J. Marx, S. Franklin and K. Cummings (Eds.): *Proceedings of the Physics Education Research Conference,* Madison, Wisconsin. In press.

V. Savinainen, A., Scott, P. and Viiri, J. (2004). Using a bridging representation and social interactions to foster conceptual change: Designing and evaluating an instructional sequence for Newton's third law. Accepted for publication in *Science Education*.

Author's Contribution

Articles I and II

These two articles were written at the University of Leeds, UK, while on a Marie Curie Fellowship. I worked closely with Dr. Philip Scott who was my advisor in Leeds. I gathered and analysed the data and wrote most of the text in these two articles, while Dr. Scott provided guidance regarding the content and structure. He polished the style and wrote some parts of the text.

Articles III and IV

I gathered and analysed the data for these two conference articles and wrote most of the text. Dr. Jouni Viiri provided guidance regarding the content and structure of the articles. He also duplicated the interview data analysis to provide an investigator triangulation.

Article V

I gathered and analysed the data for this manuscript. I wrote several versions of the text, which were commented on and partially rewritten by Dr. Scott and Dr. Viiri. Their collaboration was so extensive that we decided on joint authorship, which is why our names are in alphabetical order.

Contents

Chapter 1

Introduction

1.1 Background to the Study

I work as a high school physics teacher, teaching both the International Baccalaureate and National Finnish syllabuses. This dissertation was initially motivated by the results that a group of students in a preliminary International Baccalaureate year (age 16) achieved in the Force Concept Inventory (FCI; Hestenes et al. 1992) several years ago. It was an eye-opener: the students had not learnt much in terms of conceptual understanding. The outcome was disappointing and made me wonder why only a few students scored well. My teaching then consisted of traditional lecturing and demonstrations. Students were asked questions involving conceptual understanding but only the most active students actually participated in the lessons; the majority were silent and spent their time writing down lecture notes. I realized that something very crucial was missing from the teaching.

This experience motivated me to seek for ways to improve my teaching. I had already done some research on teaching the force concept before (Savinainen 1994) so I knew where to look. My earlier encounter with Physics Education Research (PER) had taught me to emphasise conceptual understanding in teaching but obviously I had not been able to do it efficiently. However, I was not the only one: Hake's (1998a) results indicate that traditional lecturing does not significantly enhance conceptual understanding in basic mechanics whereas interactive-engagement teaching (IE) methods could offer much more in this respect.

I wanted to find an IE method which would fit into my personal teaching style and which would be easy to implement in a high school physics setting. Mazur's (1997) *Peer Instruction* fulfilled both of my criteria so I chose it as a starting point: by experimenting in my day-to-day teaching I tailored Peer Instruction and added several other components informed by PER into my teaching. This process was guided by the use of several well-validated conceptual inventories for monitoring the outcomes of the teaching approach in different domains of physics, and by students' feedback. The teaching approach I eventually developed was first systematically tested and documented in the context of thermal physics (Savinainen 2000b). This was used as a starting point for further refinement of the teaching approach.

The teaching approach developed aimed at enhancing students' conceptual understanding, which is a major goal in high school physics. But what is meant by 'conceptual understanding' in physics? In many PER articles it seems to refer to students' ability to answer *qualitative* questions addressing different aspects of physics concepts (in contrast to traditional quantitative questions which chiefly address the correct use of equations). There is no doubt that this is indeed a necessary character of conceptual understanding. I felt, however, that it should be possible to define more precisely what conceptual understanding means in the context of physics. This enquiry was inspired by Sabella's (1999) dissertation on the coherence of student knowledge and led to the present work, in which the main aims are the characterisation and evaluation of *students' conceptual coherence*[1]. Roughly speaking, the characterisation of students' conceptual coherence can be viewed as a clarification of what is meant by conceptual understanding. The characterisation of students' conceptual coherence is used in this study as a tool for evaluating the teaching approach (i.e., general principles and methods of teaching) together with teaching sequences (i.e., specific activities and the order at which the physical ideas are staged) in the case of the force concept. The teaching approach and teaching sequences together form the *instructional approach*.

The force concept was chosen for several reasons. Firstly, several research-based conceptual inventories have been developed on the force concept and related kinematics. Many of these instruments are used in this study to measure conceptual coherence and its development. Secondly, all preparatory International Baccalaureate (IB) students (as well as Finnish national syllabus students) start studying high school physics with mechanics. This made it possible to study the development of preparatory IB students' conceptual coherence from the very beginning. Thirdly, there is a lot of previous research in this domain, allowing some comparisons between different institutes and teaching approaches.

1.2 Overview of the Dissertation

This dissertation consists of a theoretical and an empirical part. The first theoretical research aim was to clarify what is meant by students' conceptual coherence of qualitative knowledge in physics, especially in the case of the force concept. Of course students' conceptual coherence needs to be somehow operationalized, and, the second theoretical research aim was to evaluate the degree of students' conceptual coherence using well-validated multiple-choice tests and interview questions. The empirical part of this dissertation consists of designing and evaluating the teaching approach and teaching sequences for the force concept and related kinematics. The research questions are presented at the end of this chapter (Chapter 1.3).

The most important research instrument used in this study is the FCI. Hence, it is crucial that its validity and reliability be discussed in detail: the development, structure, validation, and evaluation of the FCI are reviewed in Article I. It discusses the six dimensions of the force concept and the taxonomy of misconceptions probed by the FCI (Hestenes et al. 1992). In Chapter 2, the historical development of the force concept is outlined, followed by a presention of the contemporary versions of Newton's laws. This is done for several reasons. Firstly, the historical treatment facilitates a comparison between different forms of Newton's laws, especially in the case of the first law. Secondly, the historical perspective may help readers to appreciate the lengthy process of concept formation which was needed to formulate the Newtonian force concept. If it was very hard for Newton (see for instance Steinberg et al. 1990) and other great physicists to formulate the ideas, so it is hardly surprising that students encounter difficulties in learning the Newtonian view. Thirdly, sometimes students seem to hold views which resemble those presented in the history of physics (e.g. Boeha 1990), but this does not mean that students actually hold a systematic set of the ideas put forth by early scientists. Some of the most common specific difficulties ('misconceptions') that students have with the force concept are discussed in Chapter 2.3. A comparison of students' ideas with the historical ideas is also provided in that chapter.

[1] There are various notions of coherence in the literatures of various fields (Thagard 1992, 64). The notion of coherence in this study, however, is not derived from these.

A characterisation of students' conceptual coherence is presented in Article III and further elaborated in Chapter 3, which outlines earlier research on the consistency or coherence of students' ideas in physics, context dependency of learning, and the role of multiple representations in learning. The notion of the conceptual coherence of qualitative knowledge is then discussed in detail. It has three aspects: conceptual framework coherence, contextual coherence and representational coherence. Naturally the characterisation of conceptual coherence developed in this thesis has many links with the earlier research; these links are made explicit in Article III and more links are provided in Chapters 3.1 and 3.2. Article III also argues that the FCI can be used to analyse students' contextual coherence in the force concept. This and other instruments used in this study to measure the degree of students' contextual and representational coherence are presented in Chapter 3.3. Even though students' conceptual framework coherence is crucial in the characterisation of students' conceptual coherence, it is not directly evaluated in this study for the reasons explained in Chapter 3.3.

Article II outlines the components of the teaching approach (Interactive Conceptual Instruction, ICI) used in the empirical part of this study. The main components of the ICI are conceptual focus ('concepts first'), the use of multiple representations in varying contexts, classroom interactions (peer discussions), research-based materials, and the use of texts (reading before formal treatment) and concept maps. All these components have the potential to enhance students' conceptual coherence. The components of the ICI are all research-based but the combination of all of them has not been tested elsewhere. The theoretical background of the ICI approach is presented to some extent in Article V and more fully in Chapter 4, which discusses theories of conceptual change and social constructivism. Theories of conceptual change tend to focus on the individual learner, while social constructive views (sometimes also called Vygotskian or neo-Vygotskian theories) focus on social aspects of learning, especially on talk between teacher and students as well as talk between students (Leach & Scott 2003). Both individual and sociocultural views are useful in understanding learning (Leach & Scott 2002; Duit & Treagust 2003): they are applied in Chapter 5 as well as in Article V.

There were two pilot and two study groups in this study. The preparatory International Baccalaureate (Pre-IB) and Finnish National Syllabus pilot and study groups are described in Chapter 5.3. The pilot groups followed the same ICI approach as the study groups. There was, however, a significant difference in the teaching sequence used: the concept of force was introduced to the study groups using the idea of interactions, i.e. they followed the ICI approach with a focus on *forces as interactions*[2]. This focus was achieved using a certain diagrammatic representation providing a bridge, linking concrete physical situations and more abstract free-body diagrams: this is discussed in Article V. The details of the teaching sequences for kinematics and the force concept are presented in Chapters 5.2 and 5.3.

Articles I and II discuss the use of the normalized average gain (also knows as Hake gain) and effect size in analysing the change in pre- and post- FCI scores: they are used as indicators of *practical significance*. Chapter 6.1.2 further elaborates this discussion: for instance, the effect of possible 'hidden variables' in Hake gain is addressed. Other statistical methods applied in this study are discussed in Chapter 6.1.1: *p*-values are used as indicators of *statistical significance*. The measures of contextual and representational coherence are presented in Chapter 6.2 and 6.3. These measures are based on the instruments discussed in Chapter 3.3. Chapter 6 ends with a thorough discussion of the validity and reliability of the study.

[2] This notion signifies that forces arise from interaction between two objects and that this interaction is symmetrical. This is an essential element of Newton's third law.

The FCI results of the Pre-IB pilot group are presented and evaluated in Article II. The FCI is used to evaluate the contextual coherence of the Finnish study group in Article III. Article V provides a detailed analysis and comparison of the contextual coherence of the Pre-IB pilot and study groups in the case of Newton’s third law. Article IV describes a method for probing students’ representational coherence of Newton’s first and second laws. It also presents findings from five interviews with students in the preparatory International Baccalaureate study group. Chapter 7 makes use of the methods and results documented in the above-mentioned articles and provides a systematic comparison between the results of the pilot and study groups. It should be noted that both the pilot and study groups followed an interactive-engagement type of teaching but only the study groups focused on forces as interactions. Of course, it would have been interesting to compare the groups in this study with groups following a traditional course (i.e., lectures to passive students, ‘recipe-following’ laboratory sessions and algorithmic quantitative problem solving examinations (Hake 1998a)). Nevertheless, it is possible to make some comparisons with the traditional teaching: this is justified in Chapter 6.1.2 and 6.4.1. The answers to the research aims and research questions are provided in Chapter 7.3. Finally, Chapter 8 evaluates the study, discusses its limitations and reflects on the results from the point of view of conceptual change.

The original publications are included in an Appendix. Unnecessary duplication of the original publications will be avoided as much as possible. It is clear, however, that the flow of discussion necessarily demands some representation of the published material.

1.3 Research Aims and Research Questions

The research aims and research questions were formulated and focused in an iterative process in the course of the study. The theoretical research aims address the characterisation and evaluation of students' conceptual coherence. The research aims are also presented in the form of questions.

1. What does students' conceptual coherence entail?
2. How can students' conceptual coherence of the force concept be evaluated?

The empirical research questions address the Force Concept Inventory (FCI) as a measure of conceptual coherence and the evaluation of the two types of Interactive Conceptual Instruction (ICI) in terms of supporting conceptual gains and conceptual coherence of the force concept. 'The two types of the ICI' refer to the ICI teaching *without* and *with* the focus on forces as interactions.

3. a) What was the effect of the two types of ICI teaching on students' conceptual gains as measured by the FCI?
   b) How do the FCI results of the ICI groups compare with results in other institutions and instructional settings?

As pointed out in the previous overview of the dissertation this study focuses on students' contextual and representational coherence so the fourth research question is formulated thus:

4. What was the effect of the two types of ICI teaching on students' contextual and representational coherence of the force concept?

The most significant difference between the two types of the ICI teaching was in the focus on forces as interactions. The last research questions address possible differences between the two types of ICI teaching on the students' learning outcomes.

5. a) What was the effect of the focus on forces as interactions on students' contextual coherence regarding Newton's third law?
   b) What was the effect of the focus on forces as interactions on students' contextual and representational coherence in other dimensions of the force concept?

Chapter 2

The Newtonian Force Concept and Students' Conceptions

This chapter presents first a short historical overview of the force concept followed by a contemporary version of the Newtonian force concept. The historical outline is used to put the contemporary version in context, especially Newton's first law. The physical content of Newton's laws is summarised at the end of the discussion of each law. The validity of Newton's laws is briefly discussed since this is also a part of high school physics. Finally, the most common students' misconceptions regarding the force concept are presented. They are also compared with some ideas presented in the early phases of the historical development of the force concept. Sequira and Leite (1991a) argue that the teachers' knowledge about historical development of physical concepts can become a tool to anticipate students' difficulties in making their ideas more scientific.

2.1 A Historical Overview of the Force Concept

2.1.1 Force and Motion Before Newton

The work of Aristotle (384-322 B.C.) in physics was very influential for almost two thousand years. Aristotle (350 B.C.) presents his views on motion and force in *Metaphysics*. He categorised local motion as either natural or violent (Franklin 1978). He also had also two more categories of motion: alteration and celestial motion (Spielberg & Anderson 1995, 61). Natural motion was either 'up' or 'down'. Downward and upward motions were natural because the objects did not need to be pushed or pulled. Aristotle explained natural motion in terms of prime substances. For instance, a rock is of earth and hence it naturally moves toward the centre of the earth.

Aristotle's law of motion can be represented by (Franklin 1978):

$$\text{Velocity} = \frac{\text{Force}}{\text{Resistance}} \text{ or } V = \frac{kF}{R} \qquad (2.1)$$

By 'force' Aristotle referred to 'motive power'. According to Aristotle, velocity is directly proportional to force (k is the proportionality constant in equation (2.1)). The law implies that force is required to sustain motion: uniform force produces uniform motion. Aristotle recognised two kinds of forces: force inherent in matter and force as an emanation from substance (Jammer 1999, 35-36). The latter was the force of push and pull, which caused the motion in a second object. For Aristotle, rest and motion were essentially different things. Rest needed no explanation since it was a natural state of objects whereas motion was not (Viiri 1992, 17). It is also worth noting that Aristotle did not have any concept of acceleration since, in his view, change of the change was impossible (Lehti 1987, 282)

Aristotle had difficulties in explaining projectile motion which he regarded as violent motion (Franklin 1978). He proposed two alternative explanations: the medium provides the necessary force by rushing around to prevent the formation of a void ('nature abhors a void'), or the medium itself acquires a power to be a mover from the original projector. Several medieval critics noted the paradoxical use of the medium to both sustain and resist motion. Aristotle regarded motion in a void as impossible for two reasons: firstly, there is no medium to sustain the motion and secondly the absence of resistance due to medium (e.g. air resistance) would result in an infinite speed as implied by Aristotle's law of motion.

Aristotle's ideas of motion were commented on and criticised by several medieval scholars. For instance, Philoponus (late fifth and early sixth century A.D.) rejected Aristotle's law of motion and replaced it by $V = F - R$. Thus motion in a void, where the resistance R vanishes, becomes possible. His formulation implies that velocity in void is a measure of force since $V = F$. Philoponus also rejected Aristotle's explanations of projectile motion, suggesting instead that projectile motion is caused by a force impressed into the projectile by the projector (this idea was put forward before Philioponus by Hipparchus, who lived in the second century B.C.). The impressed force will not persist indefinitely and will gradually wear out even in a void; it will also be destroyed by the resistance due to medium. Using the idea of a self-expending impressed force, Philophonus rejected infinite motion in a void.

Buridan (1300-1358) introduced the impetus theory of motion. He regarded an impressed force as permanent unless acted on by resistances or other forces. Buridan also gave a quantitative definition of impetus: it is proportional to both the speed of the object and the quantity of matter (or mass) in the object. Buridan's impetus looks like the modern concept of momentum but Franklin (1978) argues that it would be a gross anachronism to equate 'impetus' and 'momentum'. Buridan applied his impetus theory to projectile motion and falling objects.

Galileo (1564-1642) arrived at the principle of inertia by examining motion in inclined planes. He noticed that motion down an inclined plane is accelerated and that motion upward is decelerated. Thus, he concluded that motion on a horizontal plane would be perpetual. Galileo stated in the *Dialogues Concerning Two New Sciences* (according to Franklin 1978):

> "Furthermore we may remark that any velocity once imparted to a moving body will be rigidly maintained as long as the external causes of acceleration or retardation are removed, a condition which is found only on horizontal planes;…"

Franklin (1978) notes that Galileo came very close to stating the inertial principle or Newton's first law of motion but he did not state it absolutely correctly, because he had previously defined horizontal as a surface equidistant from the centre of the earth. Galileo applied his ideas in projectile motion (Spielberg & Anderson 1995, 77-78). To Aristotle's question "Why do projectiles keep moving?", Galileo answered by pointing out that it is natural for a moving object to keep moving. He also realised that the effects of falling were independent of horizontal motion, whereas Aristotle had thought that motion cannot be divided.

Galileo was influenced by the tradition of impressed force. When explaining what happens when a stone is thrown upward he regarded the impressed force as an impetus that is gradually consumed by the opposing force of gravity (Jammer 1999, 100-101). However, he came close to the classical force concept by reducing the action of force to a gradual increase of velocity. This idea was possible only after he had assumed the principle of inertia. Thus Galileo prepared the basis for the formulation of Newton's first two laws of motion.

Kepler (1571-1630) sought for a quantitative definition of force (Jammer 1999, 81-92). From the Newtonian point of view, he was not successful in this quest. Nevertheless he introduced the idea of reciprocity into the concept of force: the moon is attracted by the earth as is the earth by the moon.

It implies that force does not belong to one single object; it contains a necessary relation to a second object, which is expressed in Newton's third law. Kepler did not, however, realise the equality of the two forces involved and their opposite directions.

2.1.2 The Force Concept in Newton's Principia

Newton's concept of force is historically and methodologically related to his study of gravitation (Jammer 1999, 116). Newton made a clear distinction between weight and mass, which he called a 'quantity of matter'. The notion of quantity of matter had already been conceived by Kepler, Gilbert and Galileo before Newton, but Newton was the first to explicitly recognise it as a basic concept in mechanics. This paved the way to the definition of momentum (Newton's 'quantity of motion') and force as determined by the change in momentum.

Newton's *Principia* was published in 1687. In it, the term 'force' (*vis* in Latin) appears for the first time in Definition III (Jammer 1999, 119):

> "The *vis insita*, or innate force of matter, is a power of resisting by which every body, as much as it lies, continues its present state, whether it be rest, or of moving uniformly forwards in a right line."

Definition III implies that inertia, in Newton's opinion, is a kind of force that is inherent in matter. This definition of force is not conceived as a cause of motion or acceleration. Jammer (1999, 120) explains this as a concession to pre-Galilean mechanics. Steinberg et al. (1990) argue that Newton's belief in the force of a moving body (impetus) hampered his development of mechanics from 1664 to 1685. Definition III shows that Newton did not completely abandon the belief. In contrast to 'innate force', Definition IV defines 'impressed force':

> "An impressed force is an action exerted upon a body, in order to change its state, either of rest, or of uniform motion in a right line"

Newton subscribed to a metaphysical principle of causality, so he perceived the change in motion as an effect and the impressed force as its cause.

In addition to the presented definitions, Newton had four definitions addresssing centripetal force. He presented his three axioms or laws of motion after the definitions (Jammer 1999, 123-124):

> "Law 1: Every body continues in its state of rest, or of uniform motion in a right line, unless it is compelled to change that state by force impressed upon it."
>
> "Law 2: The change of motion is proportional to the motive force impressed; and is made in the direction of the right line in which that force is impressed."
>
> "Law 3: To every action there is always opposed an equal reaction; or, the mutual actions of two bodies upon each other are always equal, and directed to contrary parts."

Newton credited the first two laws to Galileo and Huygens. The first law, the principle of inertia, can be interpreted in two ways: it can be either taken as either a qualitative definition of force or an empirical statement describing the motion of free bodies. The second law has also two possible interpretations: it can be understood as a quantitative definition of force or as a generalisation of empirical facts.

Newton's statement of the second law can be written in modern terms as $\sum \vec{F} \cdot \Delta t = \Delta \vec{p}$. Newton considered that this statement approaches $\sum \vec{F} = m\vec{a}$ as a limit when Δt approaches zero. It has been noted, however, that Euler was the first to present Newton's second law as the second derivative of position in 1747 (Lehti 1996, 124). Newton's formulation of the second law in terms of momentum reflects his early considerations of impact and the demands of the geometry used in Principia. (Westfall 1977, 152). This does not imply, however, that Newton would have just inferred his second law from the laws of impact. As Jammer (1999, 127) says, "it was a stroke of genius".

The first law as stated by Newton is just a special case or corollary of the second law (Taylor 1959, cited in Galili and Tseitlin 2003). Why then did Newton not regard the first law as a special case of the second law? He did not discard the first law even though he had discarded a number of other candidates for the status of fundamental principle. Steinberg et al. (1990) consider the hypothesis that "to do so [to discard the first law] would have obscured a conceptual issue which had been developmentally so important for him". Another perspective on the issue is given by Galili and Tseitlin (2003), who argue that Newton's original first law had two versions of complementary meaning. They argue that the *quantitative form*[3] of the first law is 'in a sense' an even more general statement than the second law, in which Newton further refined the first one.

The third law provides an important characteristic of force which is not present in the first two laws of motion: force is simultaneous action and reaction. Force as one side of a single interaction is clearly visible in the following passage (Newton 1962, 569):

> "It is not one action by which the Sun attracts Jupiter, and another by which Jupiter attracts the Sun;…but is one single intermediate action..."

This statement describes forces as interaction even though the term interaction is not explicitly used (Viiri 1995, 63).

Newton addressed the vector nature of force in Corollary I (Jammer 1999, 128):

> "A body acted on by two forces simultaneously, will describe the diagonal of a parallelogram in the same time as it would describe the sides by those forces separately."

Newton's derivation of the corollary was based on the kinematical composition of velocities. It tacitly assumes that the action of one force on a body does not depend of the action of another force. This assumption is by no means self-evident (Jammer 1999, 132).

It is interesting to note that his derivation of the parallelogram theorem of forces is not consistent with the second law since he explicitly spoke of a uniform motion instead of acceleration as resulting from a given force. As Jammer (1999, 130) points out, Newton could have reconciled this contradiction by considering the acceleration due to force as a series of successive increments of velocity.

[3] "Rapidity in states exchange of the body is in proportion to the applied force".

The concept of mass is briefly discussed next, since it an integral part of the Newtonian force concept.

2.1.3 Concepts of Mass

For Newton mass was the carrier of the *vis inertiae* (force of inertia), and the *quantitas of materiae* (quantity of matter) was proportional to it (Jammer 1997, 81-87). This concept of vis inertiae was widely used in the seventeenth and eighteenth centuries. There was, however, a notable exception: Euler's *Mechanica* provides a logical transition from Newton's concept of mass to the more modern abstract conception as a numerical coefficient which is characteristic of the individual physical body and determined by the ratio of (net) force to acceleration. Kant also criticised the concept of vis inertiae and paved the way for a more positivistic concept of mass.

Saint-Venant rejected the concept of 'quantity of matter' in his work published in 1851 and derived his definition of mass from the law of conservation of mass (Jammer 1997, 90). Then in 1867 Mach suggested a new kinematical definition of mass (Jammer 1997, 91-97). He considered two particles A and B interacting with each other but otherwise unaffected by all the other particles in the world. Experience shows that accelerations of the particles while interacting with each other are opposite in direction and their (negative inverse) ratio is a positive numerical constant (denoted by $m_{A/B}$) independent of the respective positions of the particles. Mach then considered a third particle C interacting with the other two particles separately. He showed that the numerical constants ($m_{A/B}, m_{A/C}$ and $m_{C/B}$) can each be represented as the ratio of two positive numbers. If one of these particles is chosen as the standard particle with its relative mass taken as unity (say, $m_A = 1$) then the remaining relative masses (m_B and m_C) can be called the 'masses' of the particles B and C.

Mach's operational definition of mass attracted some objections. Firstly, the mass ratio $m_{A/B}$ depends upon the system of reference: every observer in a non-inertial reference system arrives, in general, at a different value for the mass ratio. Secondly, it was questioned if Mach's definition of mass seemed to imply the existence of forces, since Mach assumed *interacting* particles. Mach defined the force concept in terms of the mass concept (Jammer 1999, 221): "The product of the mass and the acceleration induced in that body is called the moving force." Mach's approach is problematic in that his definitions of mass and force refer to an inertial reference frame but he does not consider whether the assumption of such a reference frame presupposes the concept of force and hence leads to a vicious circle (Jammer 1999, 240).

After Mach many attempts were made to formalize Newtonian mechanics into an axiomatic system (Jammer 1997, 111-121), but none using a precise explicit definition of mass has been very successful. Whitehead's remark encapsulates the issue (Jammer 1997, 120): "We obtain our knowledge of forces by having some theory about masses, and our knowledge about masses by having some theory about forces."

So far only *inertial* mass has been considered. Newton's law of gravitation involves the *gravitational* concept of mass, so definitions of mass in terms of weight are gravitational conceptions of mass. The law of gravitation addresses *active* gravitational mass (the mass of the central body, e.g. the Earth) and *passive* gravitational mass (the mass of the attracted body, e.g. a satellite revolving around the Earth). The proportionality of these two is a consequence of Newton's third law (Jammer 1997, 125-126), whereas the proportionality between inertial and (passive) gravitational masses is a purely empirical and accidental feature in classical physics. In Einstein's general relativity the proportionality between inertial and (passive) gravitational masses is a constitutive principle: it is one formulation of the equivalence principle (Jammer 1997, 203-204).

In the framework of relativity, mass and energy are identical (Jammer 1997, 188). Swartz and Miner (1998, 108) express this point nicely: "mass is energy is mass". They also point out that mass cannot be turned into energy since mass is energy.

Finally, it is interesting to note that Jammer (1997, 224) concludes that despite all the efforts no complete clarification of the concept of mass has been reached so far.

2.1.4 Concluding Remarks

The development of mathematical physics after Newton was essentially an attempt to explain physical phenomena in terms of mass points and their spatial relations (Jammer 1999, 229). This process of eliminating the force concept from mechanics was completed in the works of Mach, Kirchhoff, and Hertz (criticism of the force concept had already been started by the philosophers Berkeley and Hume). Newton's metaphysical idea of force as *causal* activity had no place in the domain of empirical measurements. This does not mean, however, that the concept of force was merely an illusion. As Jammer (1999, 242) points out, the force concept played a most constructive role in the advancement of physics and therefore fully justified its existence. He concludes that the modern treatment of classical mechanics admits the force concept as a methodological intermediate (Jammer 1999, 264). However, in the field theories of modern physics the notion of 'force' is treated only as an exchange of momentum and therefore replaced by the concept of 'interaction' between particles (Jammer 1999, V).

2.2 The Newtonian Concept of Force

In this presentation kinematics is presented before dynamics for two reasons. Firstly, it is the standard order in textbooks and secondly, it was the order in which the teaching in the empirical part proceeded (this decision is justified in Chapter 5.2).

Newton's laws are presented using Hestenes's (1998) modern formulation of mechanics, which takes into account the modifications and extensions that the Newtonian theory has undergone. Hestenes (1998,1) provides a formulation which allows "a smooth transition from pure particle mechanics to the classical theory of fields and particles". This aspect of Hestenes' formulation is utilised in the treatment of Newton's third law. The third law has a central role in the teaching of the force concept in this study (this is discussed in Chapter 5.2.2).

2.2.1 Underpinning Kinematics

Newton's laws are underpinned by kinematics: notions of particle, position, reference frame, velocity and acceleration must be developed first. Reference frame is merely mentioned in most high school courses on kinematics. Kinematics is usually taught in Finnish high schools before vectors are discussed in mathematics. Velocity is defined in terms of the rate of change in position, and acceleration in terms of the rate of change in velocity. Derivative gives a precise mathematical formulation for velocity and acceleration ($\vec{r}$ is a position vector):

$$\vec{v} = \frac{d\vec{r}}{dt} \tag{2.2}$$

$$\vec{a} = \frac{d\vec{v}}{dt} \tag{2.3}$$

In this study, kinematics was taught without the formal concept of derivative, which is itself a very complex concept with its own underpinnings of concepts of limit and continuity. Graphical techniques can be used to determine instantaneous rate of change (slope of a tangent, i.e. graphical derivation) and the average rate of change (slope of a secant). In addition graphical integration can be introduced before symbolic integration as a tool for concept formation and problem solving in kinematics.

Sometimes it is useful to resolve the acceleration vector into tangential and normal components (a_t and a_n, respectively):

$$\vec{a} = a_t\hat{e}_t + a_n\hat{e}_n \tag{2.4}$$

$$a_t = \frac{dv}{dt} \quad \text{and} \quad a_n = \frac{v^2}{r} \tag{2.5}$$

where $\hat{e}_t$ is the tangential unit vector, $\hat{e}_n$ is the normal unit vector, and v is the instantaneous magnitude of velocity (= speed).

The magnitude of tangential acceleration measures the rate at which speed changes and the magnitude of normal acceleration the rate at which direction of velocity changes. The physical meaning of the concept of acceleration can be summarised by stating all the possible cases when an object is accelerating:

- magnitude of velocity increases while the object is moving in a straight line
- magnitude of velocity decreases while the object is moving in a straight line
- magnitude of velocity is constant while the direction of velocity changes
- magnitude of velocity increases while the direction of velocity changes
- magnitude of velocity decreases while the direction of velocity changes

Hestenes and Wells (1992) argue that introductory physics should aim for at least a qualitative understanding of tangential and normal acceleration, even though they acknowledge that the concept of acceleration is too advanced for most high schools students. They claim that many physics teachers don't even understand it, and in fact a study of experts' understanding of the acceleration concept by Reif and Allen (1992) revealed that not even all professors of physics exhibit correct understanding.

Indeed, mastering kinematical concepts is not an easy task at all (for instance, see Trowbridge & McDermott 1980 and 1981; McDermott et al. 1987). A functional understanding entails clearly distinguishing between the concepts of position, velocity, change of velocity, and acceleration. In addition it demands the ability to make connections among the various kinematical concepts, their representations, and the motions of real objects (Rosenquist & McDermott 1987). This description of 'functional understanding' is close to the central issue of this dissertation, namely 'conceptual coherence', which is discussed in Chapter 3.2.

Students cannot be expected to master Newton's laws, especially the second law, before having a good grasp of kinematics. This does not mean, however, that students should fully master kinematics before studying dynamics. Studying the force concept allows returning to and reinvoking the kinematical concepts, i.e. spiralling back (Arons 1997, 10 and 45).

2.2.2 Newton's First Law of Motion

Modern classical mechanics defines the first law with reference to an inertial system or inertial reference frame. Hestenes (1998, 11) defines the first law or the law of inertia in the following way:

> "In an inertial system, every free particle has a constant velocity. A particle is said to be *free* if the total force on it vanishes." (Italics in the original)

This defines an inertial system implicitly by specifying a criterion which distinguishes it from noninertial reference frames (Hestenes 1998, 12). An inertial frame can be identified in principle by observing the motion of free particles. Since Newton's first law is needed to define what is meant by a free particle, it cannot be viewed just as a special case of Newton's second law. On the other hand the first law is not independent from Newton's other laws, because they are needed to define what is meant by 'free particle'; the definition of free particle necessarily involves the concept of total force (net force) in one form or another.

The presented formulation of the first law is not the same as the one given by Newton (see Chapter 2.1.2) since Newton did not have the notion of reference frame (Galili and Tseitlin 2003). The first law is stated in high school physics essentially in the same form as Newton stated it, Giancoli (1998, 79), for instance, states it thus:

> "Every body continues its state of rest or uniform speed in a straight line unless acted by a nonzero net force."

I do not know any high school physics text book which would *start* teaching Newton's laws by stating the first law as a definition of inertial reference frame. Many introductory physics text books at the university level also present the first law initially with no reference to inertial reference frames (e.g. Halliday et al. 2001, 73). Even though Newton's version of the first law might not be logically necessary (this claim was already addressed in Chapter 2.1.2), it could well be pedagogically very valuable; this point is elaborated in Chapter 5.2.2.

The aspects of the first law can be summarised in the following way:

- it is valid in an inertial reference frame (in advanced texts the first law is used to *define* an inertial reference frame)
- rest and constant velocity are equal (i.e. in either case there is no change in velocity and hence no acceleration)
- net force acting on the object is zero (no forces, or more commonly, all the forces cancel each other out)

2.2.3 Newton's Second Law of Motion

Hestenes (1998, 11) defines the second law thus:

> "The total force [net force] exerted on particle by other objects at any specified time can be represented by a vector $\vec{f}$ [4] such that $\vec{f} = m\vec{a}$, where *a* is the particle's acceleration and *m is a positive scalar constant* called the *mass* of the particle." (Italics in the original text)

Sometimes $\sum \vec{F} = m\vec{a}$ is considered to be a definition of force. Hestenes (1998, 12) emphatically rejects this notion and states that an explicit definition of force is impossible. The complete set of general laws is required to define (net) force implicitly; the equation $\sum \vec{F} = m\vec{a}$ represents only one characteristic of force. Mass in the equation can be interpreted as a measure of strength of a particle's response to a given net (or total) force. Hence, in Hestenes's formulation, the concept of force is used to define the concept of mass.

The second law is formulated in different ways. It can be expressed in the differential form, which emphasises the dynamic nature of the second law (i.e., the derivatives with respect to time as the rate of change with respect to time):

$$\sum \vec{F} = m\frac{d\vec{v}}{dt} = m\frac{d^2\vec{r}}{dt^2} \tag{2.6}$$

The second law can also be defined more generally in terms of the rate of change of linear momentum (e.g. Halliday et al. 2001, 177):

$$\sum \vec{F} = \frac{d\vec{p}}{dt} \tag{2.7}$$

where momentum is defined as $\vec{p} = m\vec{v}$. It is easy to show that equation (2.7) is equivalent to $\sum \vec{F} = m\vec{a}$ if mass is constant.

Newton formulated the second law in terms of impulse (Chapter 2.1.2). The net impulse $\vec{J}$ can be derived by integrating equation (2.7) over the interval Δt - from an initial time t_i to a final time t_f :

$$\vec{J} = \int_{t_i}^{t_f} \sum \vec{F}(t)\, dt \tag{2.8}$$

While the net force can be interpreted as an instantaneous measure of the strength of the interactions between the object and surroundings, the net impulse is a measure of the strength of the interactions between the object and surroundings during a time interval determined by the limits in the integral (Kurki-Suonio & Kurki-Suonio 1997, 184).

[4] The vector $\vec{f}$ denotes the total force or net force (this is elaborated in Chapter 2.2.5.)

Summarizing, the second law entails at least the following aspects (Chi et al. 1989; the last aspect is derived from Menigaux 1994):

- it applies to one body
- it involves all the forces acting on the body
- net force is the vector sum of all the forces
- magnitude of acceleration is directly proportional to net force
- direction of acceleration is the same as direction of net force
- acceleration is independent of the exact points where the forces are exerted on the body (the forces may or may not exert torque on the body)

For pedagogical reasons one might add one more aspect:

- there is no connection between net force and magnitude or direction of velocity (i.e. if only the net force acting on the object is known, *nothing* can be said about the direction or magnitude of velocity)

The additional aspect accords with Arons's (1997, 109) observation: "In order to understand what something is, one must also understand what it is *not*".

2.2.4 Newton's Third Law of Motion

Hestenes (1998, 11) defines the third law thus (the text in brackets by author AS):

> "To the force [$\vec{f}_{12}$] exerted by any object on a particle there corresponds an equal and opposite force [$\vec{f}_{21}$] exerted by the particle on that object."

For two interacting particles (Hestenes 1998, 13), the third law can be written:

$$\vec{f}_{12} = -\vec{f}_{21} \tag{2.9}$$

This relation is satisfied by Newton's gravitational force law and Coulomb's law, but it fails for direct magnetic interactions between charged particles. The terms in equation (2.9) can be rewritten using equation (2.7):

$$\frac{d\vec{p}_1}{dt} = \vec{f}_{12} \quad \text{and} \quad \frac{d\vec{p}_2}{dt} = \vec{f}_{21} \tag{2.10}$$

Hence, the third law can be rewritten:

$$\frac{d\vec{p}_1}{dt} = -\vec{f}_{21} \tag{2.11}$$

This equation can be interpreted as a law of momentum exchange. Hence, a failure of the third law would be a failure of the law of conservation of momentum. The law of conservation of momentum is regarded as more fundamental than Newton's laws because it holds in modern physics as well. Classical field theory can be used to explain magnetic interactions between charged particles by attributing momentum to the electromagnetic field. This saves the third law in magnetic interactions between charged particles if the ‘object’ in the third law is interpreted as a field.[5]

The third law can also be framed by stressing forces as interactions (Hellingman 1992):

> ”A force is one side of an interaction; the interaction takes place between two bodies, working equally strongly in the opposite directions.”

It should be noted that this formulation does not use the terms ‘action’ and ‘reaction’ forces, since they could imply for a student that ‘action’ comes before ‘reaction’. There is, however, another danger for a student in this definition. It uses the term ‘working’ which may be confusing since the concept of work has a definite meaning in classical mechanics. Arons (1997, 74) recommends replacing ‘working’ with ‘acting’.

Summarizing, the third law entails several aspects (Brown 1989):

- An object cannot experience a force in isolation and it cannot exert a force in isolation
- At all moments interaction is symmetrical i.e. two interacting objects exert the same magnitude of force on each other
- One implication of the above point is that neither force precedes the other force, i.e. ‘action’ does not come before ‘reaction’
- Forces arising from an interaction between two objects are always exactly opposite in direction

2.2.5 Newton’s Fourth Law of Motion

Usually only three laws of motion are presented. Hestenes (1998, 10-11) and Kurki-Suonio & Kurki-Suonio (1997, 80) argue that a superposition law needs to be separately stated. The fourth law can be formulated thus (Hestenes 1998,10):

> ”The total force $\vec{f}$ due to several objects acting simultaneously on a particle is equal to the vector sum of the forces $\vec{f}_k$ due to each object acting independently, that is $\vec{f} = \sum \vec{f}_k$.”

This law is already part of the second law, but formulating it independently emphasises its importance. It allows the lumping of a great many forces into a single force which can be analysed as a unit.

2.2.6 The Validity of Newton’s Laws of Motion

Newton’s laws of motion as presented here do not hold in noninertial reference frames. In fact, an inertial reference frame can be defined as one in which Newton’s first law holds. Newton’s second law can be extended to apply also in noninertial reference frames if an extra force - inertial force - due to noninertial effects is taken into account in the sum of forces. Inertial forces do not arise from interactions and hence they do not have the interaction ‘partner’ required by the third law (Giancoli 1998, 1051-52).

---

[5] Of course this is usually not an issue in high school physics. Nevertheless, the motivation for this extension of the third law in this presentation comes from a classroom situation: a student asked me once if the third law really is *always* valid.

Physics textbooks frequently warn against the fallacy of thinking that inertial forces are real: for instance, Giancoli (1998, 1052) makes the point that inertial forces are sometimes called *pseudo*forces or *fictitious* forces. However, it is a matter of convenience which reference frame is used for descriptions of phenomena. There is never a conflict between descriptions from different reference frames as long as they are not mixed up (Swartz & Miner 1998, 131). Moreover, in the framework of general relativity, gravitation (which is definitely considered to be real in the domain of classical mechanics) is viewed merely as an inertial force through the principle of equivalence; gravitation is fictitious to the same extent as an inertial force, such as a centrifugal force, is (Jammer 1999, 258). It can be questioned, whether it is pedagogically wise to introduce noninertial reference frames before a student can confidently apply Newton's laws in inertial reference frames.

Newtonian mechanics fails when the speed of an object becomes very high, i.e. at speeds approaching that of light (the discrepancy between the classical and relativistic predictions is not detectable at small speeds). Today classical mechanics is considered a limiting case of Einstein's special relativity. At speeds much lower than the speed of light, the relativistic formulas reduce to the classical ones (Giancoli 1998, 817).

Next we turn to the difficulties that students often have with the force concept. It is crucial that teachers are aware of these and can anticipate them in teaching (Viiri 1995, 159).

2.3 Students' Difficulties with the Force Concept

2.3.1 Students' Conceptions

There is a vast body of research showing that students have many ideas, both before and after teaching, which differ from the Newtonian framework regarding the force concept (see bibliographies in McDermott & Redish 1999; Duit 2004). Many terms have been used to describe students' (incorrect) ideas:

- preconception (e.g. Clement 1982)
- common sense conception (Halloun & Hestenes 1985)
- intuitive model (Thijs & Kuiper 1990)
- alternative conception or ideas (e.g. Sequira & Leite 1991a, b)
- misconception (e.g. Hestenes et al. 1992)
- p-prim (i.e. a knowledge structure that is smaller and more fragmentary than a physical concept; diSessa 1993)
- knowledge facet (i.e. individual pieces, or constructions of a few pieces, of knowledge and/or strategies of reasoning; Minstrell 2003)
- student view (i.e. student thinking [differing from the generally accepted understanding of a particular physical situation] about a limited aspect of particular area in physics; Thornton 1995)

Thornton (1995) points out that different terms imply different implicit or explicit models of human cognition. He also criticises the use of the term 'misconception' arguing that 'student thinking is not in general misconceived but often based on partial or incorrect information'. This may well be the case but in this study the term misconception means that there is a disparity between the student's idea and the Newtonian force concept, regardless of the origin of the student's idea.

There is no way of escaping students' misconceptions in teaching physics since every student has ideas before entering the classroom. These misconceptions are very hard to change (the process from initial (mis)conceptions to scientific conceptions, i.e. conceptual change, is discussed in Chapter 4). Hence, it is useful for the teacher to know the most common misconceptions in the target domain. As shown in the previous chapter, even Newton himself had misconceptions regarding the force concept. If the force concept was so difficult for intellectual giants like Galileo and Newton, it is hardly surprising that it is a problem for students today.

Table 2.1 presents a comparison between the Newtonian mechanics, students' common misconceptions and the history of science regarding the force concept. Sequira and Leite (1991a) claim that the history of science can provide help in anticipating students' misconceptions, and can give physics teachers some insights into how to deal with them. It should be noted, however, that the comparisons are made in terms of content, not in terms of frameworks, which are different and probably cannot be directly compared (Sequira and Leite 1991a).

Table 2.1: Newtonian mechanics, students' common misconceptions of the force concept, and their equivalent ideas in the history of science (derived from Sequira & Leite 1991a).

Newtonian mechanics	Students' misconception	History of science
Zero net force implies rest or constant velocity	Motion implies force in the same direction	Motion is maintained by the impetus (Buridan, 14th century)
Motion and rest are similar rule-governed stages (zero net force)	Motion and rest are different rule-governed stages: rest does not require an explanation, whereas motion does	Rest is a natural stage which does not require any explanation (Aristotle, 4th century B.C.)
Objects stop due to a net force opposite to motion [1]	Objects stop because they have used up all the force (i.e. force is seen as a property of an object)	Objects stop when the impetus vanishes (Buridan, 14th centuty)
Slowing down is due to a net force opposite to motion [2]	Slowing down is caused by the decrease of the force in the direction of motion [3]	A decrease in velocity is due to decrease in impetus (Buridan, 14th century)
Constant net force implies constant acceleration	Constant force implies constant speed [3]	Uniform force produces uniform motion (Aristotle, 4th century B.C.)
Net force is proportional to acceleration	Force is proportional to velocity [3]	Impetus is proportional to velocity (Buridan, 14th century)
Forces are due to interactions; force is a measure of the strength of an interaction between two objects	Objects have/acquire forces (i.e. force is seen as a property of an object)	Objects acquire and develop impetus (Buridan, 14th century)

[1] If the net force continues to act in the opposite direction to the original motion, the object does not stop: it has *instantaneously* zero velocity and changes its direction (e.g. an object thrown vertically up).
[2] More precisely: slowing down is due to the tangential component of a net force opposite to motion.
[3] The term net force is not used here because students may not understand the distinction between 'force' and 'net force'.

Table 2.1 is does not provide an exhaustive list of students' misconceptions regarding the force concept and kinematics. A thorough taxonomy of students' misconceptions in this field is provided by Halloun and Hestenes (1985) and Hestenes et al. (1992): the most common misconceptions in kinematics, for instance, relate to the vague concept of motion that students have. Concepts of distance, velocity, and acceleration are not well differentiated. Furthermore, average velocity (average acceleration) is not differentiated from instantaneous velocity (instantaneous acceleration).

The taxonomy of misconceptions does not imply that all of them should be explicitly dealt with in teaching. Hestenes et al. (1992) stated that some minor misconceptions tend to disappear spontaneously with the treatment of the most important misconceptions and the growth of Newtonian concepts. They suggest that the major misconceptions are the impetus concept of motion (an 'intrinsic force' that keeps things moving) and the dominance principle (the 'bigger', 'greater mass' or 'more active' exerts the greater force in a conflict). The dominance principle is discussed in more detail in Articles I and V.

Students' misconceptions have been ascribed to many sources in the research literature (see e.g. the summaries in Sequira & Leite 1991b and Viiri 1995, 71-75). At least two of these sources have immediate implications for teaching: teachers and textbooks sometimes present statements and concept definitions which are either scientifically incorrect or can lead students to the reinforcement of students' misconceptions (Sequira & Leite 1991b; Physics Textbook Review Committee 1998; Arons 1997, 73-74).

2.3.2 The Use of Language in Physics

Arons (1997, 73-74) warns that in everyday speech we tend to express things (e.g. saying that a force causes a body to "move") in a way which is inimical to the development of understanding of the force concept, and advises teachers to become sensitive to these usages, learn to avoid them, and divert students from their use. This is very good advice for any physics teacher. However, even when the use of language in physics is correct there is a danger of misunderstanding if the underlying assumptions are not made explicit. Sequira and Leite's (1991a) article can be used to illustrate the importance of language: some of their formulations of Newtonian ideas could in fact mislead students (This does not, of course, imply that the authors do not understand the force concept).

Sequira and Leite do not use the term 'net force' when referring to Newton's second law (the term 'net force' was added by author AS in Table 2.1). They state that 'constant force implies constant acceleration'. This is not generally true if the distinction between a force and the net force is not made: for instance, in order to move a wooden block on a horizontal table at constant *velocity*, a constant force along the motion must be exerted on the block (this force must equal the sum of the resistive forces; the net force is zero).

They also state that 'slow down motion is *caused* by negative acceleration' [emphasis added]. This statement makes no sense at all, since acceleration is the measure of the rate at which velocity changes, not a cause of changing velocity. What could be said is that the slowing down is caused by the (tangential component) of a net force acting in an opposite direction to motion.

To take a final example, Sequira and Leite state that “heavier objects fall with the same acceleration as lighter objects”[6]. This statement is strictly true only in vacuum where the air resistance vanishes. Perhaps the statement could be reformulated in the following way:

> *Heavier objects fall with the same acceleration as lighter objects through air as long as the air resistance exerted on them is negligible relative to the weight of the objects.*

The validity of ‘the same acceleration’ depends on the sensitivity of the measuring devices used. A legend has it that Galileo dropped two spheres, one of wood and one of iron, from the Leaning Tower of Pisa. The unbelieving spectators below observed that the two spheres hit the ground at the same moment (Gamow 1988). There is no physical reason why the legend could not be true, since the measuring devices in this case - human eyes - may not be sensitive enough to detect any difference in the moment of impact of the spheres.

It can be concluded that students’ misconception that ‘heavier objects fall faster than lighter objects’ is not necessarily a misconception, if air resistance plays a role. When asking students questions concerning this issue, enough contextual features or assumptions should be provided to allow them to decide whether air resistance is significant or not.

Summarizing, it can be asserted that the force concept is indeed very complex, with many dimensions (these are further elaborated in the next chapter). Hence it is no wonder that students have difficulties with it[7].

[6] In general relativity, this is a consequence of the equivalence principle.
[7] One might add that teachers too may have difficulties with the force concept.

Chapter 3

Students' Conceptual Coherence in High School Mechanics

Chapters 3.1 and 3.2 are based on Articles III and IV. These chapters discuss in detail what is meant in this study by 'conceptual coherence of qualitative knowledge'.

3.1 Earlier Research on Students' Conceptual Coherence

There is a body of research showing that students' views are not 'consistent' or 'coherent' after an introductory course in mechanics (see the references in the following presentation). Most of the studies do not, however, provide a detailed definition of what is meant by consistency or coherence. Generally speaking, lack of coherence or consistency seems to mean that students respond differently to different types of tasks involving the same concepts. The following discussion mainly focuses on the research in the domain of physics.

3.1.1 Consistency of Students' Ideas

Clough and Driver (1986) investigated the consistency with which students used ideas in different contexts. The tasks were posed in contexts which were familiar to the students interviewed. Clough and Driver found that generally students did not use their alternative (incorrect) ideas as consistently as their scientifically correct ideas.

Finegold and Gorsky (1991), who investigated consistency in students' concept of force, found that many students did not understand, or had great difficulty in applying, Newton's laws. They found that no alternative framework was consistently used by students. A Newtonian framework was consistently used across different tasks by a few students. The same conclusions were drawn by Halloun and Hestenes (1985) from their study of over 4000 college students' concept of force.

Reif (1987) investigated physics students following traditional teaching, and concluded that 'they [students] rely on various special knowledge elements stored in memory, try to achieve one of these, and apply it without much subsequent reasoning'. Reif also points out that many students are unable to identify particular instances of general laws and the important elements in a system. McDermott (1993) concluded that 'a coherent framework is not typically an outcome of traditional instruction'. These conclusions are well supported by Hake's (1998a; 2002) more recent large survey, which strongly suggests that *traditional* courses (i.e. passive-student lectures, 'recipe-following' laboratory sessions and algorithmic quantitative problem-solving examinations) fail to convey much basic conceptual understanding of Newtonian mechanics to the average student.

It is possible, however, that sometimes students' views may be consistent but flawed from the point of view of science. Many researchers have interpreted research data as evidence of theory-like alternative (i.e. scientifically incorrect) explanatory frameworks which appear to be consistent or coherent and applicable across a range of phenomena (e.g. Vosniadou 1992) Interestingly, it seems there are data both for and against the notion that students have stable coherent conceptual frameworks. Taber (2000) suggests that neither view is likely to be exclusively right. He presents a case study which demonstrates that an individual learner can simultaneously have *several* alternative stable and coherent explanatory schemes - multiple frameworks - that are applied to the same concept area. The debate on the nature of alternative frameworks is not, however, the issue in this study, which explicitly concentrates on investigating the degree of students' conceptual coherence with respect to the Newtonian framework.

All the cited studies on the consistency of students' ideas address different contexts. The context dependence of learning and multiple representations are discussed next in greater detail, since they are central in defining conceptual coherence in this study.

3.1.2 Context and the Context Dependence of Learning

The word 'context' has a variety of meanings in the research literature (Finkelstein 2001). Context can have a micro meaning referring to how a specific problem is represented (e.g. verbal, pictorial or symbolic representation) or the setting of a problem (e.g. an inclined plane). Macro meanings of context address the macrocultural influences of various disciplines in western culture and their implications for student learning. Finkelstein proposes three levels of context:

- task formation: the particular form a task takes
- the situation in which such action takes place: e.g. working alone or with other students
- idioculture: the broader setting that creates the circumstances of the situation (e.g. working with other students could take place in a high school class)

Bao (2002) also proposes three major categories of context factors (the first two categories resemble Finkelstein's levels of context):

- content-based context factors: the actual scenarios and specific features of context scenarios employed in or related to the learning of a particular piece of knowledge
- learning environment context factors: specific educational settings and features of such settings used in teaching and learning
- student-teacher internal cognitive status: the students' and instructors' general views and attitudes on the learning and teaching of a particular context and the background knowledge of the content area.

While the possible usefulness of different levels or categories of context is acknowledged, this study analyses context dependence from the point of view of task formation in physics. Furthermore, representations involved in tasks are discussed separately from context in the next chapter.

Many contextual features in task formation can have an effect on students' responses. A student may show correct understanding in some exercise involving e.g. the force concept but fail to apply this in other contexts (Steinberg & Sabella 1997). These investigators argued that 'different contexts and presentations can trigger different responses from a given student, even if the underlying physics is identical'. Even varying the magnitudes of the quantities involved can trigger different responses in the same context (Mildenhall & Williams 2001). Clough and Driver (1986) noticed that the inclusion of scientific terminology in questions can generate a different distribution in the pattern of response.

Students may respond to physically irrelevant contextual features of the question, such as the type of object in motion or the direction of the motion (Palmer 1997). Another example is provided by Bao et al. (2002), who identified four contextual features that students use in their reasoning regarding Newton's Third Law: velocity, mass, pushing and acceleration. For instance, 'pushing' can imply for a students that the object that 'pushes' exerts a larger force. So the students recognise that both objects exert force on each other but they fail to appreciate the fact that the forces arising from an interaction are always symmetrical. Furthermore students may use combinations of different contextual features in their reasoning and may think that they have different levels of significance for specific questions. Even slight changes in context can make a difference.

Bao and Redish (2001) state that strong context dependence in student responses is very common, especially when students are just beginning to learn new material. Students are not sure of the conditions under which the rules they have learned apply, and tend to use the rules either too broadly or too narrowly. These conclusions are in good agreement with research on experts and novices (see for instance Chi et al. 1981).

3.1.3 Multiple Representations in Learning

Multiple representations (such as texts, pictures, diagrams, graphs, or mathematical) have many functions in learning (Ainsworth 1999). First, multiple representations can complement each other because they differ either in the information each expresses or in the processes each supports. A single representation may be insufficient to carry all the information about the domain or be too complicated for learners to interpret if it does so. Multiple representations can also encourage students to use more than one strategy to solve a problem. A second function of multiple representations is to help students develop a better understanding of a domain by using one representation to constrain their interpretation of a second one. For instance, graphs can be used to constrain the interpretation of equations. Thirdly, multiple representations can support the construction of deeper understanding when students integrate information from more than one representation.

The combination of representations that both complement and constrain each other can have synergetic effects, since they enable students to deal with the material from different perspectives and with different strategies (Seufert 2003). However, this synergy does not emerge easily. Even though introducing multiple representations in teaching has great potential benefits, it can also jeopardise the learning process due to an increased cognitive load (de Jong et al. 1998, 34). There are a number of cognitive tasks that students have to perform to cope successfully with multiple representations: students must (Ainsworth et al. 1998, 123-125):

- learn the format and operators of each representation
- understand the relation between the representation and the domain it represents
- understand how the representations relate to each other

To help students to perform the cognitive tasks, Ainsworth et al. (1998, 131) recommend that a new representation should be supported by a familiar complementary one.

Given the cognitive tasks required, it is not surprising that many studies show that students rarely use multiple representations effectively (van Someren et al. 1998) and that they have difficulty moving across or connecting multiple representations (Kozma 2003). And yet these are the very skills required for constructing coherent knowledge structures: moving within a representation (e.g.

moving from one kinematics graph to another) and between different representations (e.g. from verbal description of motion to graphical representation) (Seufert 2003).

The importance of multiple representations has also been realized in physics education (Larkin & Simon 1987; Hestenes 1996; Van Heuvelen & Zou 2001; Meltzer 2002a; 2003). Hestenes (1996) argues that students' ability to understand physics depends on the representational tools at their disposal. Van Heuvelen and Zou (2001) give several reasons why multiple representations are useful in physics education: they

- foster students' understanding of physics problems since, as visual aids, they automatically enhance human perceptual reasoning (Larkin & Simon 1987)
- build a bridge between verbal and mathematical representations
- help students develop images that give mathematical symbols meaning

Van Heuvelen and Zou further argue that an important goal of physics education is to help students to learn to construct multiple representations of physical processes, and to learn to move in any direction between these representations. This notion is supported by Gardner (1991, 13), who says that "Genuine understanding is most likely to emerge...if people possess a number of different ways of representing knowledge of a concept or skill and move readily back and forth among these forms of knowing...".

However, Meltzer's (2002a, 2003) preliminary results in physics suggest that there are possible discrepancies in student learning abilities when using oral and written representations compared with diagrammatic and mathematical representations. His results also suggest that certain representations may pose particular learning difficulties in physics. For instance, after the introduction of microcomputer-based laboratory tools in an inquiry-based elementary physics course, Meltzer et al. (1997) found that students' ability to give correct responses to questions involving Newtonian dynamics in graphical representation seemed to have significantly increased. However, no corresponding improvement was found when the questions were posed in the form of ordinary language.

3.2 Conceptual Coherence of Qualitative Knowledge

The conceptual coherence of students' qualitative knowledge can be divided into three aspects: representational, contextual and conceptual framework coherence (Figure 3.1). Each aspect is defined in terms of skills they entail. Naturally there is some overlap between the aspects of conceptual coherence.

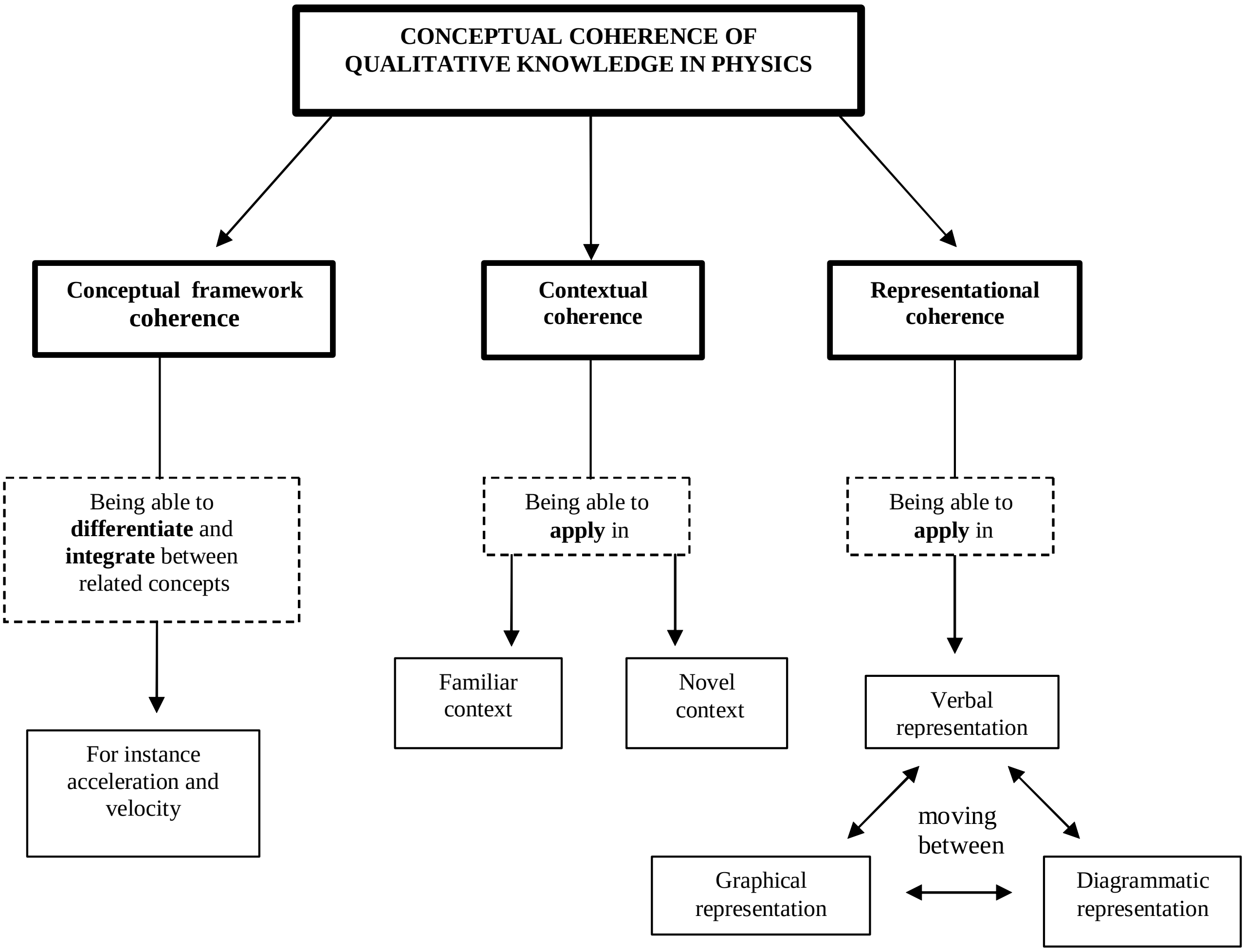


Figure 3.1: Dimensions of the conceptual coherence of qualitative knowledge in physics.

1) Representational coherence

Representational coherence entails the unification of multiple representations based on correct conceptual understanding (Bao 2001). The student is able to use multiple representations correctly and move between representations. Verbal (written and oral), diagrammatic (free-body diagrams, other vector representations, motion maps, path diagrams) and graphical (graphs, e.g. velocity against time) representations are efficient tools in analysing physical situations (Van Heuvelen 1991). Mathematical representation is naturally also very important but it is not within the scope of this study. This characterisation of representational coherence is closely linked with expertise which can be viewed as the possession and coordinated use of multiple representations of the same domain (de Jong et al. 1998).

Meltzer (2002a) illustrates the significance of representation in assessing student understanding in physics. He gives examples of test items addressing Newton's Third Law in the context of gravitational forces between the Earth and other heavenly objects. Similar questions were framed in verbal and vector diagram representations. The proportion of correct responses was halved when vector diagram representation was used. Meltzer concludes that the question 'was measuring not only students' knowledge of Newton's third law of motion and law of gravitation, but also (in part) students' understanding of vector diagrams'. While I agree with the conclusion I would also infer that the students lacked representational coherence in their understanding of Newton's third law.

2) Contextual coherence

The student can apply a concept (e.g. acceleration) or a physical law (e.g. Newton's Laws) in a variety of familiar and novel contexts. The context here refers to the circumstantial features in which a task is posed. Contextual coherence cannot be evaluated in isolation, since the student must use some representation to express his/her understanding in given situations. The effect of contextual factors can be probed if a representation is kept the same while the context changes.

Even slight changes in a general context or contextual features within the same general context can make a difference for students who lack contextual coherence. For instance, Schecker and Gerdes (1999) found that when certain Force Concept Inventory (FCI) questions were posed in slightly different general contexts, i.e. a golf ball was replaced by a soccer ball or a steel ball thrown upward was replaced by a vertical pistol shot, student responses were significantly different. The FCI also provides an example of varying contextual features in the same general context of a car pushing a truck: in one question the car is pushing the truck with increasing speed, whereas in the next question the car continues pushing the truck at constant speed. This change in contextual features (acceleration vs. constant speed) is irrelevant from the point of view of Newton's third law but it proved to be a crucial difference for many students as described in Article II.

3) Conceptual framework coherence

This aspect addresses relations between concepts and overlaps the other aspects to some extent. In order to apply a concept in a variety of contexts, the student must relate (integrate) a concept to other concepts. The student also needs to differentiate that concept from related concepts (McDermott 1993). It is worth noting that Rosenquist and McDermott (1987) use the term 'functional understanding' in the context of kinematics in a way which is very close to the characterisations of conceptual framework and representational coherence in this study (see Chapter 3.2)

Evaluation of conceptual framework coherence is possible if the given tasks demand the use of many related concepts at the same time. This can be done in many levels of the hierarchical structure in mechanics. For instance, answering questions on Newton's second law demands framework coherence, since Newton's second law includes the concept of acceleration. Furthermore, acceleration is underpinned by the concept of velocity. The force concept is central also in higher levels of the hierarchy: it is involved in the concepts of work and momentum, for example.

Conceptual understanding in a certain representation and context implies that the student has reached at least some degree of conceptual framework coherence. A student may have achieved framework coherence in some representation and context and still fail in other representations and contexts. Hence, the framework coherence is a necessary but not sufficient condition for representational and contextual coherence.

Figure 3.2 presents example questions derived from Hake (2002) to illustrates how the aspects of conceptual coherence can be used in categorising a question.

A student in a lab holds a brick of weight W in her outstretched horizontal palm and lifts the brick vertically upward at a constant speed. All the following questions refer to the situation of the brick moving vertically upward at a constant speed.

1. The magnitude of the force on the brick by the student's hand is:
   - A. constant in time and zero.
   - B. constant in time, greater than zero, but less than W.
   - C. constant in time and W.
   - D. constant in time and greater than W.
   - E. decreasing in time but always greater than W.

2. Draw a free-body diagram showing all the forces acting on the brick.

3. a) Graph velocity against time.
   b) Graph position against time.
   c) Graph acceleration against time.

Figure 3.2: Conceptual questions derived from Hake (2002) to illustrate analysis in terms of conceptual coherence.

The questions in Figure 3.2 are framed in one context for all three parts. Question 1 involves application of Newton's first law in verbal representation whereas question 2 involves moving between representations (from verbal to diagrammatic representation), as does question 3 a (from verbal to graphical representation). Questions 3 b and 3 c address moving within graphical representation. In addition to representational coherence questions 3 b and 3 c also address conceptual framework coherence, since the successful performance demands integration of position, velocity and acceleration.

Contextual coherence could be addressed by framing another set of similar questions in different contexts: for example, in terms of a metal ball moving vertically upward at constant speed on a hydraulic platform (Steinberg & Sabella 1997). Of course, this change in context is irrelevant from the point of view of physics. Contextual coherence could also be addressed by changing a contextual feature within the original context, e.g. reversing the direction of the motion of the brick.

3.3 Instruments for Measuring Students' Conceptual Coherence of the Force Concept

This study aimed at measuring students' contextual and representational coherence in the case of Newton's laws and related kinematics. No attempt was made to measure conceptual framework coherence directly. Of course, conceptual framework coherence underpins both contextual and representational coherence since it is not possible to exhibit good understanding of the force concept without well-differentiated and integrated relations between the key concepts. It would be possible to evaluate students' conceptual framework coherence separately from other aspects of conceptual coherence if the representation and contextual factors were kept the same. However, this would demand the lengthy process of designing and validating a new set of questions. It was decided to make use of the existing questions which were already validated.

3.3.1 The Force Concept Inventory

Since the Force Concept Inventory (FCI) is central to this dissertation, it is discussed in more detail than the other instruments used to evaluate students' conceptual coherence. Article I also discusses the history, validation, and critique of the FCI.

Characterisation of the FCI

The original version of the Force Concept Inventory (FCI) was published in 1992 (Hestenes et al. 1992). A revised version was developed and placed on the web in 1995 (Halloun et al. 1995) and later appeared in Mazur's book (Mazur 1997). This revised version has 30 items whereas the original FCI had 29. All the FCI questions have five possible responses. In every question, four of the responses are distracters which representing common student misconceptions. One of the authors of the 1995 version of the FCI (Hake 1998a) claims that it has 'fewer ambiguities and a smaller likelihood of false positives' than the earlier version. Since the earlier version has been shown to be relatively free from the tendency towards false positives (correct answers for incorrect reasons), it is very reasonable to assume that this is also the case with the 1995 version.

The FCI and Conceptual Coherence

Hestenes and Halloun (1995) have argued that the entire FCI test should be used for the purposes of evaluating courses and teaching. They argue that 'the total FCI score is the most reliable single index of student understanding, because it measures coherence across all dimensions of the Newtonian force concept'. Single FCI items cannot be used to make reliable conclusions but several items addressing the same dimension of the force concept can provide valuable information about specific learning difficulties that students may have (Article II). It may well be the case that the total score is the best single measure of a student's overall conceptual coherence of the force concept, but I believe that a more detailed analysis in terms of aspects of conceptual coherence is possible.

Hestenes et al. (1992) classified the FCI questions in terms of the six dimensions of the force concept:

- Kinematics
- Newton's First Law
- Newton's Second Law
- Newton's Third Law
- The Superposition Principle
- Kinds of Forces: solid and fluid contact forces (combined together as contact forces in this study), gravitational forces

In addition to the classification by Hestenes et al. (1992) we use the categories of representational coherence in classifying the FCI questions. Table 3.1 presents the classification in terms of the dimensions of the force concept and representation for the 1995 version of the FCI (Halloun et al. 1995). Some questions were classified into two dimensions by Hestenes et al. (1992). These questions were carefully considered in order to decide the most appropriate dimension. Question 27 (about slowing down due to friction) is classified in the Newton's Second Law dimension in Table 3.1 whereas Hestenes et al. classified it as Kinds of Forces (solid contact). Two solid contact questions (5 and 18) have a dynamic situation and hence at least implicitly address Newton's second law. In addition, question 8 demands quite complex reasoning at first in terms of Newton's second law, then in terms of the vector nature of velocity and finally in terms of Newton's first law. To answer this question correctly with correct reasoning demands conceptual framework coherence. It is clear that the classifications in Table 3.1 are not mutually exclusive.

Table 3.1: The classification of FCI questions in terms of dimensions and representations of the force concept. This classification can be used to measure students' contextual coherence of the force concept.

Kinematics	**Newton's First Law**		**Newton's Second Law**	**Newton's Third Law**	**Kinds of Forces**	
					Gravitation	**Contact**
Diagram	Verbal	Diagram	Verbal	Verbal	Verbal	Verbal
12, 14 19, 20	10, 17 24, 25	6, 7 8, 23	22, 26, 27	4, 15, 16, 28	1, 2 3, 13	5, 11 18, 29, 30

The superposition question (9) and diagrammatic Newton's Second Law question (21) are not included, since one question in those domains is not enough to allow the evaluation of conceptual coherence. Questions 15 and 16 (Newton's Third Law) and also questions 26 and 27 (Newton's Second Law) have the same general context but the contextual features (states of the systems) are different: for instance, in question 15 the velocity of the car pushing the truck increases while in question 16 the velocity is constant. This change of contextual feature, which is irrelevant from the point of Newton's third law, was crucial for many students in a previous study (Article II).

Several questions in the FCI are framed in the same contexts but fall into different categories of dimension and representation of the force concept. From the point of view of the classification used a clear majority of the FCI questions have different contexts. Hence, the FCI results can provide information on contextual coherence within verbal and diagrammatic representations in different dimensions of the force concept. Interestingly, critics of the FCI lend support to this conclusion when they argue that determinations of students' understanding of the force concept and of students' familiarity with the context are inextricably tied together in the case of the FCI (Huffman

& Heller 1995). I would add that students' contextual coherence implies correct responses even in *novel* contexts.

3.3.2 The Test of Understanding Graphs - Kinematics

Beichner (1994) has developed a multiple-choice test for evaluating the understanding of graphs in kinematics: the Test of Understanding Graphs - Kinematics (TUG-K). Answer distracters represent typical incorrect student responses. Most test questions are framed with as little contextual information as possible. Hence, the test is not suitable for the evaluation of students' contextual coherence. Beichner strove to ensure that only kinematics graph interpretation skills were measured: for instance, no questions regarding the components of the velocity of a ball tossed in the air were designed, since they would also test knowledge of projectile motion.

9. An object starts from rest and undergoes a positive, constant acceleration for ten seconds. It then continues on with a constant velocity. Which of the following graphs correctly describes this situation?

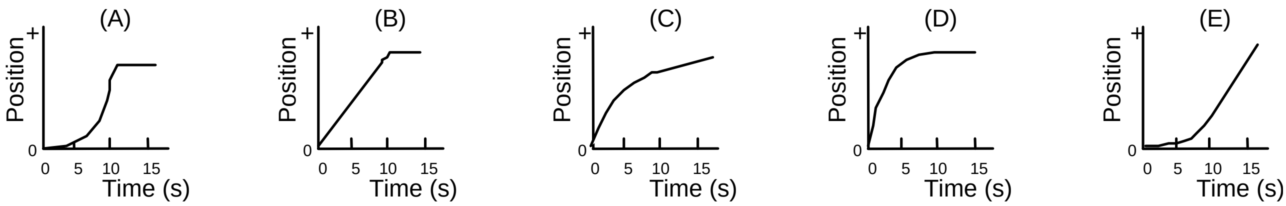


Figure 3.3: A sample question from the TUG-K test (Beichner 1994).

The TUG-K has 21 items. Figure 3.3 presents a sample item from the test. The questions which address moving between representations (Questions 3, 8, 9, 12, 19, 21) can be used for evaluating students' representational coherence in kinematics. Naturally these questions involve conceptual framework coherence as well.

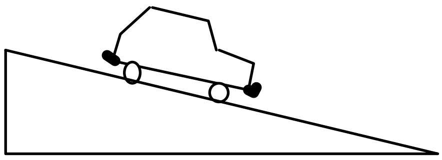
3.3.3 The Force and Motion Conceptual Evaluation

Thornton and Sokoloff (1998) have developed a research-based, multiple-choice assessment of student conceptual understanding of Newton's laws of motion and kinematics: the Force and Motion Conceptual Evaluation (FMCE). It consists of 43 questions (actually, a new version of the FMCE test has 47 questions but the last four questions address work and energy which are not part of this study). The questions involve verbal and graphical representations and they are formed into groups, and each group of questions share the same general context and a story line to introduce the questions. Contextual features are varied within the same general contexts in the groups. Hence the FMCE can be used to evaluate contextual coherence within verbal and graphical representations. The FMCE also has questions which demand moving between representations.

It is worth noting that the FMCE differs from the FCI in that the FMCE has far fewer general contexts than the FCI. Another major difference between them is that the FCI does not address graphical representation. Thornton and his collaborators have found, however, that there is a very strong correlation (r = 0.8) between the FCI and FMCE results, but the FMCE appears to be harder than the FCI for low scoring students, i.e. their scores on the FMCE are significantly lower than their FCI scores (Redish 2003, 104). The strong correlation between them suggests that both tests seem to measure the same target domain, namely the Newtonian concept of force and related kinematics.

Figure 3.4 shows sample questions from the FMCE, and Table 3.2 presents the categorisation of the FMCE questions used to measure students' contextual and representational coherence.

Questions 8-10 refer to a toy car which is given a quick push so that it rolls up an inclined ramp. After it is released, it rolls up, reaches its highest point and rolls back down again. *Friction is so small it can be ignored*.



Use one of the following choices (**A** through **G**) to indicate the **net force** acting on the car for each of the cases described below. Answer choice **J** if you think that none is correct.

A Net **constant** force **down** ramp

B Net **increasing** force **down** ramp

C Net **decreasing** force **down** ramp

D Net force zero

E Net **constant** force **up** ramp

F Net **increasing** force **up** ramp

G Net **decreasing** force **up** ramp

______8. The car is moving up the ramp after it is released.

______9. The car is at its highest point.

______10. The car is moving down the ramp.

Figure 3.4: Sample questions from the FMCE test (Thornton & Sokoloff 1998).

Table 3.2: FCME questions used to measure students' contextual and representational coherence.

Dimension	Representation	Context	Questions
Newton's First law	Verbal [1]	Sled	2, 5
	Verbal to Graphical [2]	Car	14, 15, 17, 21
Newton's Second Law	Verbal [1]	Sled	1, 3, 4, 6, 7
		Car ramp	8, 9, 10
		Coin toss	11, 12, 13
	Verbal to Graphical [2]	Car	16, 18, 19, 20
Newton's Third Law	Verbal [1]	Car Collision	30, 31, 32, 33, 34
		Car Push	35, 36, 37, 38
		Student pushing	39
Kinematics	Verbal [1]	Coin toss	27, 28, 29
	Verbal to Graphical [2]	Car	22, 23, 24, 25, 26 40, 41, 42, 43

[1] Evaluates contextual coherence.
[2] Evaluates representational coherence.

3.3.4 Survey on Newton's Third Law

Bao et al. (2002) developed the Survey on Newton's Third Law to study the effects of contextual (or physical, as they call them) features. Since each multiple-choice question in it measures students' reasoning related to a single contextual feature of the third law, the Survey addresses contextual coherence in the case of Newton's third law, using verbal representation with some pictures to clarify the situations. The Survey is discussed in Article V.

3.3.5 Interview Questions

A set of interview questions were used to measure students' representational coherence of Newton's first and second laws and contextual coherence of Newton's third law. They are presented and analysed in Articles IV and V. Another set of interview questions was designed to measure students' contextual coherence of Newton's laws within verbal representation (Figures 3.5a, 3.5b and 3.5c).

1. A bug hits the windshield of a car driving at 80 km/h along a highway.

a) Is the magnitude of the force exerted on the bug by the car larger than, smaller than, or equal to the magnitude of the force on the car by the bug? Describe your reasoning in reaching your answer. Consider three instants of time:

(i) just when the collision starts
(ii) in the middle of the collision
(iii) just before the collision ends.

b) As a result of this collision, is the acceleration of the bug larger than, smaller than, or equal to the acceleration of the car? Describe your reasoning in reaching your answer.

Figure 3.5 a: Interview question on contextual coherence of Newton's third law (derived from Reif 1995a).

2. Two crates, A and B, are in the elevator (crate A on the top of crate B). The mass of crate A is greater than the mass of crate B.

a) The elevator moves upward at constant speed.

(i) How does the acceleration of crate A compare to that of crate B? Explain.
(ii) Draw and label separate free-body diagrams for the crates.[1]
(iii) Rank the forces on the crates according to magnitude, from largest to smallest. Explain your reasoning.
(iv) Consider the direction and magnitude of the net force acting on crate A.
Consider the direction and magnitude of the net force acting on crate B.
Compare the magnitudes of the net forces.

b) As the elevator approaches its destination, its speed decreases while it continues to move downward. The same questions were asked as in case a).

[1] This is the only question addressing diagrammatic representation. Hence there are not enough questions for the evaluation of students' contextual coherence in diagrammatic representation

Figure 3.5 b: Interview question on contextual coherence of Newton's first and second laws (derived from McDermott et al. 1998).

3. A man is pushing two crates in contact with each other in the World's Strongest Man competition. The bigger crate has mass of 140 kg and the smaller crate has mass of 70 kg. The mass of crate A is greater than the mass of crate B. Consider the following situations.

a) The crates do not move.
(i) Compare the forces that 140 kg and 70 kg boxes exert on each other.
(ii) Compare the forces that the man and 140 kg box exert on each other.
(iii) Compare the net forces acting on the crates.

b) The crates are moving at constant velocity.
(i) Compare the forces that 140 kg and 70 kg boxes exert on each other.
(ii) Compare the forces that the man and 140 kg box exert on each other.
(iii) Compare the net forces acting on the crates.

c) The crates are moving at constantly increasing velocity.
(i) Compare the forces that 140 kg and 70 kg boxes exert on each other.
(ii) Compare the forces that the man and 140 kg box exert on each other.
(iii) Compare the net forces acting on the crates.

Figure 3.5 c: Interview question on contextual coherence of Newton's laws (derived from Brown 1989).

Verbal representations of Newton's second and third laws are addressed in three different contexts, whereas Newton's first law is addressed in two contexts. All the cases involving zero net force are classified under Newton's first law. The questions have three different contexts with varying contextual features (e.g. constant velocity vs. changing velocity). The classification of the interview questions in terms of Newton's laws is presented in Table 3.3. All the questions involve only verbal representation. Hence, students' contextual coherence can be evaluated within verbal representation.

Table 3.3: The dimensions of the force concept addressed in the interview questions (Figures 3.5 a, b, c).

Newton's First Law	Newton's Second Law	Newton's Third Law
2 (a) (iv) 3 (a) (iii) 3 (b) (iii)	1 (b) 2 (b) (iv) 3 (c) (iii)	1 (a) (i), (ii), (iii) 2 (a) (iii); 2 b) (iii); 3 (a) (i), (ii) 3 (b) (i), (ii); 3 c) (i), (ii)

3.3.6 Overview of the Research Instruments

The research instruments and the evaluated aspects of conceptual coherence are presented in Figure 3.6. As explained above, the conceptual framework coherence is not directly measured but it is implicitly addressed in students' contextual and representational coherence. The instruments overlap with each other, since it is not possible to evaluate students' representational coherence without a context or students' contextual coherence without a representation. When students' representational coherence is evaluated the context should be kept as constant as possible. Students' contextual coherence can be evaluated within one representation, so for example students' contextual coherence in Newton's first law within verbal representation means that all questions and students' answers regarding Newton's first law are provided using verbal representation. Only the contexts and contextual features are varied.

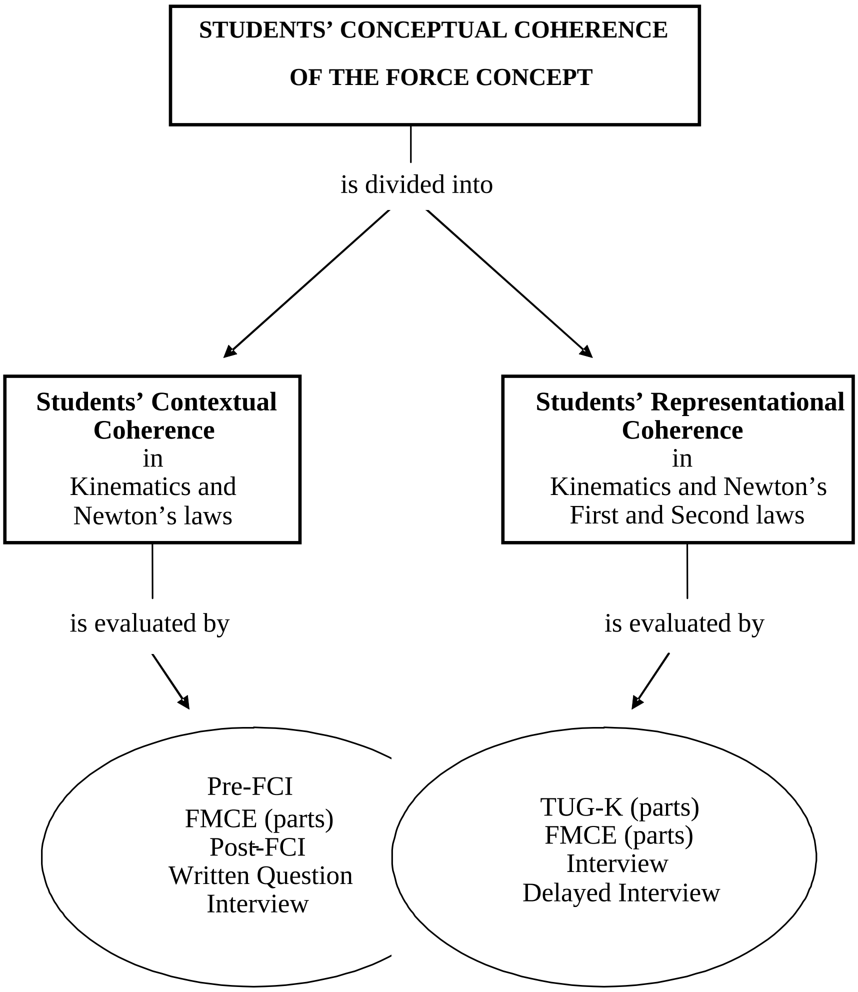


Figure 3.6: The research instruments and their relations to the evaluated aspects of conceptual coherence.

Chapter 4

Teaching and Learning Physics

This chapter discusses theories of learning physics, addressing both individual and sociocultural perspectives. No attempt is made to provide an exhaustive description of all possible views in this field: rather, a theoretical background for the adopted teaching approach in this study is presented. Article V provides an example of how these views can be used in designing and evaluating a teaching sequence in the case of Newton's third law.

4.1 Learning as Conceptual Change

4.1.1 Theories of Conceptual Change

Conceptual change is a research agenda which evolved from the study of students' alternative conceptions (Tyson et al. 1997; Schnotz et al. 1999). It is related to contructivistic theories of learning in which a metaphor of construction is central: individuals build their knowledge in a process involving the learners' activity and creativity, not just through repeating and memorisation (Tynjälä 1999, 37-38). This view emphasises that it is not possible to acquire knowledge directly through observations and experiences: these always demand interpretation on the basis of earlier ideas and experiences (i.e., observation is 'theory-laden').

A range of terms has been used to describe degrees or kinds of conceptual change, for instance:

- assimilation and accommodation (e.g. Posner et al. 1982)
- weak and strong restructuring (Carey 1985)
- conceptual capture and conceptual exchange (Hewson & Hewson 1992)
- differentiation and reconceptualization (Dykstra et al. 1992)
- enrichment and revision (Vosniadou 1994)
- shifts across parallel categories within a major ontological tree and shifts from one major ontological tree to another (e.g. matter to process) (Chi et al. 1994)

The various descriptions of conceptual change imply that there are ”big” and ”small” changes (Tyson et al. 1997). Changes can occur in the conceptual structure involving the simple *addition* of knowledge, or some kind of *revision* of the existing conceptual structure. The latter type of change is divided by most theorists into weak and strong revision. Tyson et al. (1997) compare the language used by various researchers to describe degrees or kinds of conceptual change, and claim that there is common ground between the various perspectives. Naturally, researchers also differ on the nature of conceptual change. For instance, one debate concerns whether conceptual change is revolutionary (a sudden shift from one theory to another) or evolutionary (a gradual adjustment process). It has also been argued that theories of conceptual change have suffered from inexplicitness and imprecision in terms of what constitutes a concept and what actually changes in conceptual change (diSessa & Sherin 1998).

Conceptual change does not happen easily - students' misconceptions are very resistant to change. Hewson and Thorley (1989), on the basis of the work by Posner et al. (1982), identify four conditions that need to be satisfied for a student to experience conceptual change: *dissatisfaction* with existing conceptions, *intelligibility* of the new competing conception, *plausibility* (students also need to believe the new conception), and *fruitfulness* (the new conception needs to be valuable in a pragmatic sense). Redish (1994) claims this is a corollary of the following principle: 'It is very difficult to change an established mental model substantially'. Motivational beliefs and the roles of individual students in a classroom learning community can also facilitate or hinder conceptual change (Pintrich et al. 1993).

4.1.2 Strategies of Teaching for Conceptual Change

Scott et al. (1991) identify two main groupings of strategies to promote conceptual change. The first consists of strategies based on cognitive conflict and the resolution of conflicting perspectives. In some of these strategies, the conflict must be recognised by the student in the early stages of teaching; in other strategies an alternative (scientific) 'way of looking' is introduced first and the conflict is highlighted later. An example of a successful conflict-based strategy for teaching the force concept is discussed in Chapter 4.1.3: it is also used to illustrate the process of conceptual change.

It is not surprising that conflict-based strategies have been criticised by some science educators. It is clear that the success of any conflict-based strategy depends upon students' willingness and ability to recognise and resolve the conflict (Scott et al. 1991). Sometimes a conflict from the teacher's point of view is not a conflict for a student (e.g. Gunstone & Watts 1985; Roth et al. 1997). Dreyfus et al. (1990) point out that even meaningful conflicts do not always ensure the construction of the required knowledge. This observation gets strong support from Chinn and Brewer (1993; 1998), who provide a taxonomy of possible responses to anomalous data (i.e., 'conflicts'); most ways of responding do not result in an alteration of the current theory. Limon (2001) identifies three kinds of problems which may explain why the cognitive conflict strategy is not as successful as is often expected:

- making the cognitive conflict meaningful for students
- theoretical problems related to conceptual change (e.g., more refined methodological tools are needed to take account of students' prior knowledge)
- problems in implementing instructional strategies developed to promote conceptual change

Dreyfus et. al. (1990) found that bright and successful students welcomed cognitive conflicts whereas unsuccessful students ended up, as a result of the conflict-based teaching, with diminished self-confidence and negative attitudes towards school tasks. However, it is not clear to what extent this might happen to 'weaker' students with any teaching strategy.

The second grouping of strategies build on students' existing ideas and extend them to a new domain. In these strategies conflicts may occur but they are not seen as being essential for promoting learning, and may even be avoided (Scott et al. 1991). Less emphasis is placed in these strategies on the students' role in reorganising their knowledge and more on the design of appropriate interventions by the teachers than the former grouping. Use is made of analogies and 'anchoring examples', which draws upon students' intuitive knowledge (e.g. Brown & Clement 1989; Brown 1992; Camp & Clement 1994). Brown and Clement (1989) describe four steps in the bridging strategy:

1) A target question is used to make explicit students' misconceptions relating to the topic under consideration (e.g. forces acting on a book on a table).
2) An analogous case, an anchoring example, is suggested by the instructor (e.g. a hand pushing down a spring).
3) The instructor asks students to make an explicit comparison between the anchor and target cases in order to establish the analogy relation.
4) If students do not accept the analogy, the instructor attempts to find an intermediate bridging analogy between the target and anchor (e.g. a book on top of a spring and on top of a noticeably flexible board).

Clement (1987) has reported significant gains in high school students' understanding of the concept of force using the bridging strategy. It is noted, however, that the use of analogies has its dangers as well: for example, uncritical use of analogies may itself generate misconceptions (Treagust et al. 1996 and references therein).

It is argued in Article V that the notion of *bridging representation* is useful in linking concrete physical situations and more abstract free-body diagrams. A bridging representation has close links with the concept of the bridging analogy discussed here. The role of a bridging representation in fostering conceptual change is discussed and evaluated in detail in Article V.

4.1.3 An Example of Conceptual Change in the Case of Force and Motion

An example of conceptual change in the context of the force concept in the case of college students is provided by Dykstra et al. (1992). Their approach can be considered conflict-based. Figure 4.1 presents a successive series of conceptual changes in the concepts of force and motion based on observations made in an introductory course on mechanics. It serves at the same time as an example on a teaching approach used to induce conceptual change.

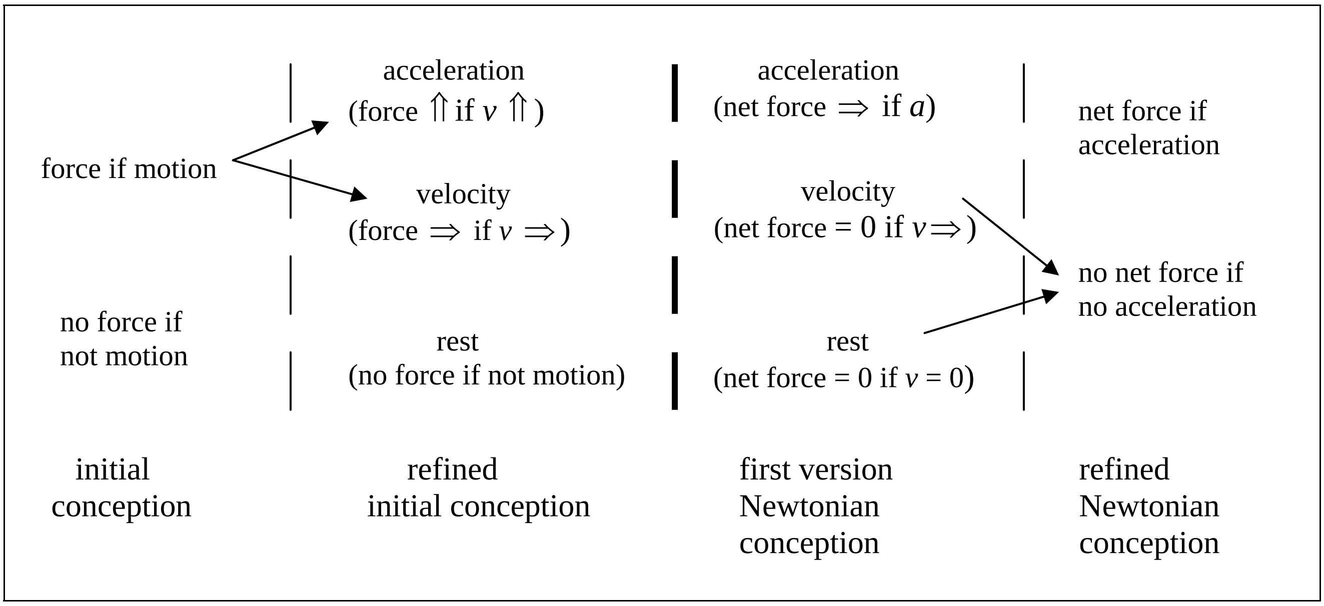


Figure 4.1: A series of conceptual changes regarding force and motion. The bold, dashed line at the centre of the figure indicates a substantial conceptual change (i.e. strong revision) and the regular, vertical, dashed lines on either side indicate less substantial conceptual refinements (i.e. weak revision). ("$\Uparrow$" = increases; "$\Rightarrow$" = remains constant). (Dykstra et al. 1992).

In Figure 4.1 'initial conception' refers to the situation at the beginning of teaching. The process of conceptual change in this particular example starts with teaching the various interrelationships between various quantities used to describe motion: distance, velocity and acceleration as a function of time (at the beginning, no reference is made to forces). For this the students had microcomputer-based laboratory equipment (MBL) with graphing possibilities and instruction materials from the Tools for Scientific Thinking Project (Thornton & Sokoloff 1990) available to them. The differentiation of different motions (e.g. constant velocity, acceleration) provides a basis for enriching their current conceptions about the causes of motion. This phase is presented in Figure 4.1 as 'refined initial conception'. Students still think that maintaining constant velocity requires a constant force and that changing velocity requires changing force. They have not yet discriminated 'applied force' and 'net force' from each other.

The next phase is based on creating a surprise due to the mismatch between students' predictions and observed motion. To achieve this, Dykstra et al. use a demonstration of motion in which a constant net force is applied to an object and the resulting motion is recorded using the MBL equipment. It is worth noting that as a result of familiarity with the MBL equipment, the students have started to trust it as an extension of their senses. Before performing the demonstration students make a prediction about the motion. They are surprised to find out that a constant acceleration, not constant velocity, is the result (some students may predict that there will be acceleration *and* increasing (net) force). Dykstra et al. argue that a substantial conceptual change (depicted as the bold, dashed line in Figure 4.1) necessarily requires *cognitive disequilibration* which is not the same as contradiction in a logical sense. It is a consequence of the surprise produced when an expected event does not take place. As indicated earlier, Hewson and Thorley (1989) call this requirement a dissatisfaction with existing conceptions.

At this point, discussion between students is encouraged, in which students reach the new phase of the 'first version Newtonian conception': zero net force implies constant velocity and a constant net force implies constant acceleration. Students also shift their thinking from applied force to net force. The final step in this example of conceptual change is understanding rest as a particular state of constant velocity. The 'refined Newtonian conception' also entails differentiating between rest and zero instantaneous velocity.

The role of the teacher in the teaching sequence is to ask leading, even provocative questions but not to provide the answers. The development of ideas and convincing the class is left as much as possible to the class. Peers try together to make sense out of the disequilibrating experiences. The teacher makes sure that before experiments and demonstrations are carried out, students' beliefs and predictions are first examined by the individual and the group. After students have *invented* the idea - e.g. Newton's second law - the teacher summarises the result and quotes it formally.

On the basis of their empirical data, Dykstra et al. identified three types of conceptual change, which are compared here with the framework suggested by Tyson et al. (1997) (briefly described in Chapter 4.1.1):

1. *Differentiation*: new concepts emerge from existing, more general ones (e.g. motion is differentiated into velocity and acceleration in the transition from initial to refined initial conception). Tyson et al. (1997) classify differentiation as addition of knowledge.

2. *Class extension*: existing concepts that are considered to be different are found to be cases of one subsuming concept (e.g. rest becomes a special case of constant velocity in the transition from initial Newtonian conception into the refined Newtonian conception). Tyson et al. (1997) see this as an example of weak revision.

3. *Reconceptualization*: a significant change in the nature of and relationship between concepts occurs (e.g. the change from the refined initial conception into initial Newtonian conception in which 'force implies motion' changes into 'force implies acceleration). This corresponds to strong revision in the scheme proposed by Tyson et. al. (1997).

The interesting point to be made here is that Dykstra et al. clearly see conceptual change as a process in which students' views are gradually transformed. This suggests that conceptual change in the case of force and motion is evolutionary rather than revolutionary. Thornton's (1995) data supports this view. He used a phenomenological description (i.e. one based on empirical data instead of drawn from a particular model of cognition) of conceptual dynamics, i.e. the process by which students' views are transformed during instruction. Thornton's (1995) idea of a transitional state between students' initial state and the Newtonian view is utilised in the characterisation of students' conceptual coherence later in this thesis (also in Article III). The transitional state can also be identified in Dykstra et al.'s (1992) example of conceptual change: the 'initial refined conception' and 'first version Newtonian conception' phases can be interpreted to constitute different degrees of the transitional state.

4.2 The Role of Classroom Interactions in Learning

4.2.1 The Social Constructivist Perspective

Leach and Scott (2003) refer to the theories of conceptual change as individual views of learning since they portray science learning fundamentally in terms of changes in individuals' 'mental structure'. These theories recognise the social nature of formal learning but do not go much further than that, whereas a sociocultural view (also referred to in the literature as Vygotskian or neo-Vygotskian views) of learning portrays learning and meaning-making as originating in social interactions between individuals, or as individuals interacting with books or other sources. Both individual and sociocultural views are useful in understanding learning: these are incorporated in a social constructivist perspective on learning (Leach & Scott 2002). Some aspects of this view are discussed here, since the teaching approach in this study emphasizes peer discussions and teacher's talk (Articles II and V).

Language and other semiotic mechanisms (such as mathematical symbols, diagrams, gesture, stance) provide the means for ideas to be talked through and communicated on the social or intermental plane (Leach & Scott 2002). Scott and Jewitt (2003) provide an example in the context of magnetic fields of how one teacher uses talk, visual communication, and demonstration to move from the phenomenon to the scientific theory. Their analytical approach focuses upon a range of communicative resources (gesture, movement, image, talk, etc.). Multiple representations, discussed in Chapter 3.1.3, form a subset of these communicative resources.

The process in which individuals appropriate and become able to use for themselves (on the intramental plane) conceptual tools first encountered on the social plane is called internalisation (Leach & Scott 2003). This relates the sociocultural view with the individual view: an individual has to come to a personal understanding of the ideas encountered in the social plane, reorganise and reconstruct the talk and activities of the social plane.

Teaching sequences involve three key features from the social constructivist point of view proposed by Leach and Scott (2002):

1) Staging the scientific story

This concerns the way in which science is made available on the social plane of the classroom. This process is interactive involving both teacher and students. The teacher presents new ideas, talks through ideas with the whole class, and discusses ideas with individuals and groups of students. The whole aim of the staging process, which involves talk, other semiotic modes and various activities (such as experiments and demonstrations), is to make the scientific story intelligible and plausible (these criteria for conceptual change were addressed in Chapter 4.1.1) to students.

The classroom discourse can have 'authoritative' and 'dialogic' functions (Scott 1998). The teacher uses the authoritative discourse to convey information and make new ideas available to the students. This mode of discourse is almost exclusively used in a traditional, lecture-based teaching. Dialogic discourse allows opportunities for the exploration of meanings. It means that the teacher asks for, and discusses, students' opinions, or that the students discuss ideas with each other (peer discussion) (Leach & Scott 2002). Scott (1998) argues that learning in the classroom is enhanced through a balance between authoritative and dialogic discourse. This calls for an appropriate 'rhythm' to the discourse.

2) Supporting student internalisation

Teachers' interventions support student internalisation of the scientific story throughout the teaching sequence. It is central to the teacher's role to monitor students' understandings and respond to those understandings in terms of how they relate to the physical ideas being taught. This can be achieved through whole class questioning and discussion, small group activities, or individual writing activities. Research-based exercises and well-validated multiple-choice tests can also be used for this purpose. Teachers' support for student internalisation is greatly enhanced if the teacher has a detailed knowledge and understanding of the conceptual terrain of the subject area, including the canonical physics knowledge and student misconceptions (Article II).

3) Handing over responsibility to the students

This feature involves providing opportunities for students to 'try out' and practise the new ideas for themselves, to make the new ideas 'their own'. At the beginning this may demand the teacher's support and guidance but gradually the teacher hands over responsibility to the students when their competence and confidence has increased.

Leach and Scott (2002) note that the presented conceptualisation of the teaching sequence is not typical in the research literature, where teaching sequences tend to be conceptualised in terms of activities, with no reference to the talk which surrounds them. Of course, there are exceptions as well: for instance, Dykstra et al. (1992) give a fairly good account on the role of teacher's and students' talk in the activities, as described in Chapter 4.1.3. Furthermore, Leach and Scott (2002) state that 'it would certainly come as no surprise if different teachers achieved very different outcomes in student learning with comparable groups of students, by following the same sequence of activities without any attempt to stage those activities in the same way'. Hake (1998b) lends support to this view by concluding that effective teaching methods appear to be a necessary but not sufficient condition for improved learning gains: they also need to be implemented successfully in the classroom.

4.2.2 Small Group Discussions

Small group discussions play an important role in helping students to explore meanings in the framework of social constructivism. This role is examined in more detail in this chapter.

Peers can assist each other in seeing how new knowledge has meaning (Jones & Carter 1998). The mere process of verbalising one's thoughts may affect cognitive growth: it helps students to realise their own conceptions. A peer may be able to assist a confused student by rewording the teacher's explanation. Peers can use metaphors and analogies to promote conceptual understanding by bridging a new concept to a familiar concept in the same way as the teacher may do, as discussed in Chapter 4.1.2 (bridging analogies). Student-generated analogies can be equally or more powerful than those employed by the teacher. Students can also offer other students examples from their own common ”everyday” experiences.

Peer instruction can be valuable in creating cognitive conflict and conceptual change: it can help students to develop plausible new concepts which are useful in different contexts (Jones & Carter 1998). This was utilised in, for instance, the study described in Chapter 4.1.3 (Dykstra et al. 1992). Peer discussions make it possible for students to express their representations and beliefs. This helps students to increase their metaconceptual awareness (Mason 1998; Vosniadou et al. 2001) i.e., to become aware of their explanatory frameworks and presuppositions. As mentioned earlier, motivational beliefs and the roles of individual students in a classroom learning community can also facilitate or hinder conceptual change (Pintrich et al. 1993). Mason (1998) argues that in peer discussions the authority for learning and knowing is shared between teacher and students: this enhances the intrinsic motivation for learning.

It should be noted, however, that just providing students the opportunity to work collaboratively does not ensure that learning takes place (Jones & Carter 1998). Within a sociocultural context there are many factors that can affect the learning in a group setting. One factor is the group size: on one hand, in the pair both students have an opportunity to participate in the discussion; on the other hand, fewer ideas may be generated in the pair than in small groups of more than two students (Jones & Carter 1998). Furthermore, pairing two students operating at low cognitive levels (i.e., ‘weak’ students) is not fruitful, as there are virtually no cognitive resources available to the pair. Alexopolou and Driver (1996) found that students in pairs seemed to face difficulties in negotiating their views and dealing with their disagreements. Students in fours handled these difficulties more easily and recognised that different people can have different views, which fostered the negotiation of meanings. Gunstone et al. (1999) argue that there is evidence that three may be the optimal group size for balancing the constraints of too small and too large groups. Heller and Hollabaugh (1992) found that in addition to group size the gender and ability composition of groups, seating arrangement, role assignment, textbook use, and group, as well as individual testing, contribute to the problem-solving performance of cooperative groups.

It is crucial that students in small groups are willing to work together. Consequently, Alexopolou and Driver suggest that there may be advantages in using self-selecting groups rather than groupings on predetermined factors (e.g., ability levels). Another important requirement is that students recognise and accept the worth of peer discussions (Gunstone et al. 1999). This is strongly emphasised by Crouch and Mazur (2001), who have used Peer Instruction for more than ten years at Harvard University (Peer Instruction is briefly described in Article II). They have noticed that students can be initially skeptical about Peer Instruction since it requires students to be significantly more actively involved and independent in learning than does a traditional lecture format. For instance, when Mazur changed his method of instruction from traditional lecture to Peer Instruction, a student asked him: “Professor Mazur, when are we going to do some *real* physics?” (italics in original; Mazur 1998). Consequently, Crouch and Mazur (2001) argue that proper student motivation is essential and they recommend several ways to motivate students: for example, explaining the reasons for teaching the ”new way”, and grading students on conceptual understanding in exams.

Chapter 5

Implementation of Teaching

The detailed discussion of the teaching approach - Interactive Conceptual Instruction (ICI) - in this study can be found in Articles II and V, and in Savinainen (2001a). Chapter 5.1 provides additional details and makes links to the ideas discussed in Chapters 3 and 4. These theoretical considerations (i.e., students' conceptual coherence, conceptual change and the social constructivist perspective on learning physics) are used to explain and theorise the instructional approach developed. Chapters 5.2 and 5.3 outline the teaching sequences for teaching kinematics and the force concept, with references to the ideas discussed in Chapter 2.

It is worth noting that the teaching approach (ICI) provides principles and methods which can be used to teach any domain of physics: its application in teaching thermal physics, for instance, is discussed in Savinainen (2000b). The teaching sequence presents specific activities and the order in which the physical ideas are staged in teaching of the force concept. The teaching approach and teaching sequence together describe how the teaching was implemented: this is referred to as the instructional approach. In this way, Leach and Scott's (2002) advice on presenting both the activities and how they are staged is taken into account.

5.1 Teaching Approach: Interactive Conceptual Instruction

The teaching approach was developed to promote conceptual coherence and was based on the premise that developing an understanding of physics requires an interactive process in which there is opportunity for ideas to be thought through, and talked through, between teacher and students. In other words, the process should be ongoing teaching and learning dialogues, as discussed in Chapter 4.2. Both 'authoritative' and 'dialogic' discourse was used in 'staging the scientific story'.

Interactive Conceptual Instruction entails several features or components which overlap with each other to some extent (Article II):

1) conceptual focus and 'concepts first' (concepts are introduced and rehearsed before quantitative problem solving as recommended by Van Heuvelen (1991))
2) use of multiple representations in varying contexts
3) classroom interactions (peer discussions)
4) research-based materials
5) use of texts (reading before formal treatment) and concept maps

The components of the ICI approach address all the aspects of conceptual coherence (Chapter 3.2). The second and fourth components were designed to address the representational and the contextual coherence in the dimensions of the force concept. While conceptual framework coherence was at least implicitly addressed in all the components, the concept maps were designed especially to help the students to see 'the big picture' and to foster framework coherence, i.e. differentiation and integration of related concepts.

Peer discussions were initially inspired by Mazur's (1997) Peer Instruction, which presents students with conceptual questions in a multiple-choice format. In the ICI approach, only a few multiple-choice questions were used: instead, points arising from demonstrations and laboratory experiments were used as a basis for oral conceptual exercises, which can be flexibly tailored to match the teaching situation at hand. In addition to oral conceptual questions designed by the teacher, research based exercises were used as a basis for peer discussions: for instance, Ranking Task Exercises (O'Kuma et al. 2000), CUPs (Mills et al. 1999) and the Rollerblade video (Etkina 1998).

The students discussed conceptual questions mostly in pairs; only a few students preferred groups of three or four. The grouping was self-selective, following the suggestion of Alexopolou and Driver (1996). Some quiet students needed help in finding a partner. The teacher (author AS) monitored the pair discussions and intervened if students were not well engaged with given tasks (this happened very rarely). There were a few students, however, who refused to work in pairs or small groups. The teacher tried to persuade them to participate but did not force them to do so since these students seemed to be satisfactorily engaged with the given tasks by themselves.

Students were encouraged to take part in peer discussion by the teacher at the beginning of teaching. It was argued that knowing physics means being able to 'talk physics' as well as being able to use other representations in analysing physical situations. As advocated by Crouch and Mazur (2001), it was made clear that conceptual understanding is required in the exams. However, it was pointed out that there is no risk in terms of grades when students express their ideas in pair discussions and whole-class discussions (the emphasis in the teaching was on the pair discussion). Indeed, students were told that "a mistake is our friend since it shows us what we don't master *yet*". They were also assured that answering various conceptual instruments (described in Chapter 3.3) would have a minor or no effect on their marks: the students still seemed to take the tests seriously. A deliberate attempt was made to create an atmosphere in which the students would feel safe to explore their understandings.

To support students internalisation (see Chapter 4.2.1) the teacher monitored their progress by asking them to raise their hands if they thought the answer and reasoning given by a pair for a given task was correct. If only few students responded correctly, the teacher continued to explain the topic and devised a new conceptual question or task. This strategy is a direct application of Mazur's (1997) Peer Instruction. It also offers a flexible way to implement the social constructivist principle of 'handing over responsibility to the students'.

5.2 Teaching Sequences

The teaching sequence used to teach the Pre-IB study group the force concept (the groups are described in Chapter 5.3) is the result of the author's experimenting over several years[8]. The main idea is to take forces as interactions – essentially, Newton's third law - as an entry point to the concept of force and then use the balancing metaphor in introducing Newton's first and second laws. The teaching of Newton's first and second laws relied heavily on concepts of velocity, change in velocity and acceleration: this is why kinematics had to precede the force concept (kinematics was introduced before dynamics in Chapter 2 as well). Trying to introduce forces along with kinematics would be too much for students to handle (Dykstra et al. 1992).

[8] Teaching sequences for the Pre-IB pilot group are presented in Tables 5.1 and 5.2. The instructional approach for the Pre-IB pilot group is also discussed in Savinainen (2001a) and Article II.

5.2.1 Kinematics Sequence

At first the students in the Pre-IB study group were introduced to kinematics graphs through a simple walking experiment. The teacher (author AS) walked the same distance three times with different velocities. Students measured the time at equal distances: these measurements were represented in a position against time graph. They had already had some experience with graphing, since graphs and straight lines are part of the lower secondary school mathematics. Students then discussed in pairs what might be the difference between the lines and how these differences might relate to the motions they observed. After the differences were resolved on the qualitative level, the slopes of the lines were determined. It was emphasized that the students must take the unit of the slope into account. Once the slopes with units were determined, a new quantity (average velocity) was defined in terms of the slopes. Average speed was also defined and its difference from average velocity was elaborated using several examples, using peer discussions.

The same walking data were next represented in a velocity against time graph (i.e., in a (t,v) coordinate system). The students had to figure out by themselves what the outcome was and then compare it with their neighbors' answers. The teacher encouraged negotiation if any differences arose. The students were asked to figure out what would be the common factor between different lines; they were reminded to relate their conclusion to the observed motions. Using students' responses, the physical meaning of the area in the (t, v) - coordinate system was deduced. This was followed by suitable exercises.

The students were then asked to explain what happens to an object, when the graph in the (t, s) - coordinate system is not a straight line. The teacher then imitated the motion which was represented by the graph. A graphical method was introduced for determining average and instantaneous velocities. These were also expressed in words as 'average and instantaneous rates at which position changes with respect to time'. The mathematical formulation of instantaneous velocity (discussed in Chapter 2.2.1) was not given, since the students had not done any calculus in mathematics.

At this point the students had an idea of what change in velocity means. This idea was quantified by representing uniform acceleration in the (t, v) coordinate system. The average acceleration was defined in words as 'the average rate at which velocity changes with respect to time', and also algebraically. Students were asked to deduce the physical meaning of the area in the (t,v) coordinate system with the help of the unit consideration. The students were given plenty of opportunities to practice the ideas that were produced. At the end of the kinematics treatment the students analyzed a ticker timer data from a dropping experiment using both the (t, s) and the (t, v) - coordinate systems. This analysis allowed several conclusions to be drawn:

- uniform acceleration is represented by a parabola in the (t, s) coordinate system whereas in a the (t, v) coordinate system it is represented by a straight line
- acceleration of the free fall is called 'acceleration due to gravity' (actually an object in the ticker timer experiment is not freely falling, but it is a good enough approximation)
- the ticker timer paper offered a new representation for uniform and non-uniform motion (successive dots have the same time separation, whereas the distance between them may change)
- changing velocities were also represented using arrows, i.e. a vector notation

Care was taken that the concepts developed had empirical meanings[9]. Either the teacher or students themselves produced motions corresponding to graphs (by walking, running, or moving hands) or represented observed motions graphically. Of course more efficient ways to do this exist: for instance, Dykstra et al. (1992) utilized microcomputer-based lab (MBL) equipment.

The treatment of kinematics graphs was finished by giving the students the TUG-K test. This was followed by deriving the kinematics equations using the graphs which the students were already familiar with. Falling motion and horizontal projectile motion were then discussed from the point of view of kinematics. Algebraic representations were accompanied with graphical representations throughout the teaching to promote representational coherence.

As mentioned earlier the above presentation concerns the Pre-IB study group. The Pre-IB pilot group followed a very similar teaching sequence in kinematics. For a comparison, an example lesson on introducing acceleration to the Pre-IB pilot group is presented in Table 5.1.

Table 5.1: An example lesson for the pre-IB pilot group on introducing acceleration (Savinainen 2001a). The students read the chapter on acceleration in the textbook before the lesson.

Activity	Time	Comments
Checking homework on velocity (problem solving and conceptual exercises from the book)	10 min	Some problems are presented on a white board. Conceptual homework exercises are first discussed in pairs.
Teacher explains the idea of change in velocity.	5 min	It is explained that magnitude or direction (or both) of velocity can change.
Short demonstrations in which the teacher moves in different ways. Students discuss in pairs if the teacher is accelerating or not.	10 min	Demonstrations are used as a source of peer discussion questions. The questions are posed orally.
Teacher explains uniform acceleration.	3 min	A falling marker is used as an example.
Students are asked to present uniform motion and uniform acceleration using velocity against time graphs. The graphs are constructed and discussed in pairs.	7 min	Uniform motion and the graphical determination of average velocity were introduced using data measured by the students.
Velocity against time graphs are presented on the whiteboard. Students explain and execute the motions in pairs.	10 min	Graphical exercises will be continued in the next lesson in which students empirically determine acceleration of a falling object using ticker timer tapes.
Conceptual and graphical homework exercises and/or reading tasks are set for the next lesson.		No calculations are carried out in the introduction lesson. Calculations will follow the qualitative treatment.

[9] This emphasis was inspired by the perceptional approach to teaching physics (see for instance Kurki-Suonio & Kurki-Suonio (1994)). The presented teaching sequence, however, is not based on the perceptional approach.

5.2.2 The Force Concept Sequence

In many textbooks Newton's third law is introduced after Newton's first and second laws (e.g. Giancoli 1998). This does not, however, give a central role to forces as interactions. The 'symbolic representation of interactions' (SRI diagram) developed by Jiménez and Perales (2001) was adopted in this study in order to permit a strong emphasis on forces as interactions throughout the teaching. The use of SRI diagrams as a *bridging representation* to free-body diagrams in the study courses is described in Article V.

The concept of force was introduced to the students at first in the context of contact interaction. They were asked to press a table with their thumbs and then observe what happened to the thumb. This touching was characterised as an interaction between the thumb and the table. The students were also asked to press their books and notebooks to find out if they too were deformed. This made it easier for them to believe that also table and even a wall do deform in an interaction with another object. Next the students were asked to press the table softly and hard and then observe if there were any differences in the deformations of the thumb. This simple activity helped them realize that the strength of an interaction can vary. At this point 'force' was defined as a measure of the strength of an interaction[10]. An ordinary scale and a spring balance were introduced as tools to measure the strength of an interaction, i.e. forces. A contest between the teacher and students was announced: who could exert a greater force on the scale? It was pointed out that the units shown by the scale are not units of force, whereas a spring balance shows force properly in Newtons.

Next the SRI diagram was introduced as a tool to represent interactions: two-headed arrows are used to show an interaction between two objects. It was emphasised that both objects participate in the interaction and that the interaction is symmetrical: as the force is the measure of interaction, the same amount of force is necessarily exerted on both objects. This was the meaning given to the two-headed arrows.

Student were asked if touching is the only way to interact, and they proposed magnets, probably because almost all of them had played with magnets. Gravitational interaction was also identified, with the help of the teacher. Hence, two broad categories of interactions were used: contact and distance interactions (these were denoted by 'c' and 'd', respectively in the SRI diagrams as shown in Figure 5.1).

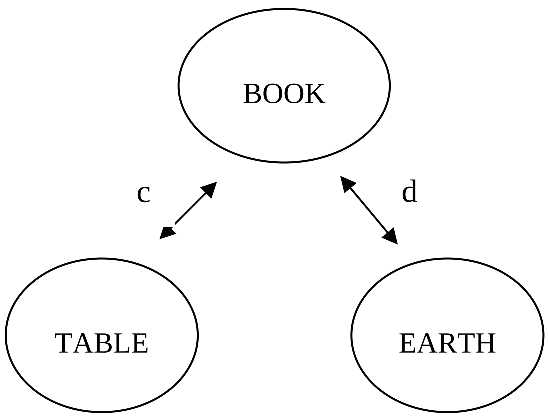


Figure 5.1: A SRI diagram for a book on a table. Contact and distance interactions are denoted by 'c' and 'd', respectively.

[10] Force is not the only measure of the strength of an interaction in mechanics: torque, work, and energy are also interaction descriptors (Reif and Heller 1982; Kurki-Suonio 2000).

After treating Newton's third law via the SRI diagrams, the idea of balancing was used to teach the free-body diagram and Newton's first and second laws. The students were already familiar with a scale as a measure of the strength of the contact interaction. A student volunteer was asked to come forward and step on the scale, and the reading was recorded. At this point a free body diagram was introduced as a tool for representing forces acting on an object (the students had already had some experience with vectors in the context of kinematics). The connection between the mass (strictly speaking *gravitational* mass as discussed in Chapter 2.1.3) and weight was introduced. This made it possible to express the normal force and gravitational force in Newtons in connection with the free-body diagram. Then the students were asked to decide whether the vector representing the normal force should be less than, the same as, or greater than the gravitational force. This discussion was used to formulate the idea of balancing forces. A sign convention allowed the sum of forces to be deduced: the balanced forces imply that the sum of the forces is zero, i.e. the net force is zero.

Then the teacher pushed down and pulled up the student standing on the scale, and the readings were recorded. The students were asked to identify the interactions in those cases and use the idea of balancing when constructing corresponding free-body diagrams. They were encouraged to draw a free-body diagram of their own and then compare and discuss it with their partners. It was stressed that if the forces on the object balance (zero net force), then the object is at rest or moving at constant velocity and hence there is no acceleration. The students were given opportunities to practice the use of SRI diagrams and corresponding free-body diagrams.

Next the teacher demonstrated jumping and hitting the floor 'forcefully'. First the students were asked whether the teacher was accelerating *during* the hit and if so what was the direction of the acceleration. After this issue had been resolved (in the peer discussions, of course) the students were asked whether the scale reading during the hit was smaller than, the same as, or greater than when the teacher just stands on the scale. Then they were asked whether the forces on the teacher balanced each other, i.e. cancelled, during the hit. This discussion led to the idea of non-balanced forces: the unbalanced force was called the (non-zero) net force or the sum of forces. Finally, the idea of net force was linked with acceleration in the same way as zero net force had earlier been linked with zero acceleration. This was summarised by stating that non-zero net force implies non-zero acceleration, and in addition, the acceleration and the net force have the same direction.

At this point the students were told that net force and acceleration are related to each other, but they did not have the formula yet. The connection between the quantities was made clear through an experiment with a linear air-track and a glider attached to a string. The string ran over a pulley, and small weights were hung on the string. Several weights were used to accelerate the glider (the mass of the weights was much smaller than that of the glider). Acceleration was determined using photogates and relevant kinematics formulas. The weight of the mass was taken as a net force acting on the glider[11]. Then the accelerations were graphed against the net forces, resulting in a straight line which passed close to the origin (the small intercept value was interpreted as a sign of small systematic error). Hence it was deduced that the net force is directly proportional to acceleration. The proportional coefficient - the slope of the line - was determined and its value was in the same order as the mass of the glider. This analysis justified (or more precisely: made acceptable to the students) Newton's second law: $\sum \vec{F} = m\vec{a}$.

[11] Of course, the net force on the glider is due to the tension force exerted by the string but at this point in teaching this would have been an unnecessary complication for the analysis. However, using the mass of the attached weight as the net force is a reasonable approximation if the masses of the attached weights are small relative to the mass of the glider.

After the quantitative form of the second law had been introduced the teaching proceeded to the traditional end-of-chapter calculations addressing the force concept. These calculations were often complemented with graphical and verbal representations to promote representational coherence.

5.2.3 Reflections on the Teaching Sequences

The use of a scale and the idea of balancing as a springboard to Newton's first and second laws can be seen as building on students' existing ideas and extending them to new domains (see Chapter 4.1.2). The balancing metaphor in the context of Newton's laws has been criticised (Physics Textbook Review Committee 1998), but it is argued here that it serves as a fruitful starting point especially for students who are not familiar with the vector concept (Pre-IB groups had not had any vectors in their mathematics lessons).

After the students had started applying Newton's second law the first law could be taught as effectively a special case of the second law (see Chapters 2.1.2 and 2.2.2). However, the students continued using the first law in their reasoning, probably because of the powerful balancing metaphor. In the first exposure to Newton's laws there is no point in introducing Newton's first law as a definition of an inertial reference frame, as discussed in Chapter 2.2.2.

The concept of mass was introduced in the context of weighing, since this was familiar to the students before any physics teaching. The conceptual difference between inertial and gravitational mass (see Chapter 2.1.3) was not mentioned: it was assumed that the same concept of mass was involved both in both weighing and Newton's second law. It is worth noting, however, that the inertial and gravitational masses - as well as inertial and non-inertial reference frames - are discussed later in the International Baccalaureate syllabus (as part of the 'Relativity Option').

The order in which Newton's laws were presented can be compared with that reported by Dykstra et al. (1992). After treating kinematics, they start with experiments and demonstrations utilising MBL. Students together invent Newton's second law and the teacher summarises the result and quotes it formally. Students figure out Newton's first law after the second law. In this teaching sequence Newton's third law is dealt with later on. Hence, the order in which Newton's laws are treated is exactly the opposite to the order used in this study, where Newton's third law was introduced first via forces as interactions. This order was inspired by the earlier failure in teaching the third law to the Pre-IB pilot group (see Article II). The use of peer discussion is a common factor in the presented sequences and those used by Dykstra et al. (1992), but the role of the teacher as a communicator of the 'scientific story' was somewhat more prominent in this study.

Algebraic calculations (problem solving) were practiced only after extensive treatment of motion in terms of graphical and verbal representations according to the 'concepts first' principle. The students were often asked to read the relevant section of the textbook addressing new topics prior to the lessons, in the spirit of Mazur's (1997) Peer Instruction. While many of the exercises used were invented by the teacher, research-based conceptual exercises, such as the Ranking Task Exercises (O'Kuma et al. 2000) and CUPs (Mills et al. 1999), were also used frequently.

Students' misconceptions (discussed in Chapter 2.3) were implicitly addressed in the conceptual exercises used [12]. Both research-based exercises and exercises designed by the teacher were used for this purpose. The students had plenty of opportunities to compare their explanations with those provided by the teacher. This most likely induced cognitive conflicts (see Chapter 4.1.2), but they were not highlighted by the teacher. There were, however, occasions when students' misconceptions were addressed in teaching a topic: for instance, when projectile motion was taught, the students were told that 'many people think there must be a force along the direction of the motion'. The students were then invited to analyse the interactions in the case of a projectile fired with initial velocity at some angle to the horizontal.

5.3 Course Descriptions and Data Gathering

5.3.1 Preparatory International Baccalaureate Groups

The preparatory International Baccalaureate study group (denoted Pre-IB study group) consisted of Finnish students (age 16, n = 23) following a preparatory International Baccalaureate programme in the school year 2002-2003. A similar pilot group (denoted as Pre-IB pilot group; n = 22) was involved in the pilot study (Article II and Savinainen 2001a) in the school year 2000-2001. Both the Pre-IB study and pilot group students had encountered mechanics in their lower secondary school studies and were taught by the same teacher (author AS). All teaching took place through the medium of English. The level of the students' English language proficiency was good enough to start studying in English: this had been tested in the entrance exams of the IB program. Furthermore, the students did not complain at any point of teaching that they did not understand English well enough.

The Pre-IB pilot and study groups had available to them an American physics textbook based on algebra and trigonometry (Giancoli 1998). Both groups followed the same ICI approach and they had virtually the same kinematics sequence. The main difference in the teaching was in the emphasis on forces as interactions: the pilot group did not have the SRI diagram as their representational tool and hence did not have as thorough teaching on the third law as the study group. Partially due to the introduction of this new representational tool, slightly more time was devoted to teaching of the force concept with the study group than with the pilot group. In addition, the pace of teaching was slower for the Pre-IB study group than for the Pre-IB pilot group for the reasons discussed in Chapter 8.2. Consequently, the pilot group covered more topics in physics during their pre-IB year. In other respects both groups did generally similar exercises (for instance, the same end-of-chapter exercises from Giancoli (1998) were used).

Concept maps designed by the present author were used. Those on the force concept are presented in Figures 5.2 and 5.3. The pilot group was given almost the same concept map as the study group: it addressed but did not emphasise forces as interactions. Both concept maps were given to the students and discussed at the revision phase before the force tests. A concept map on kinematics (not shown here) was also used for revision before the TUG-K test.

---

[12] Chapter 2.3. compared students' misconceptions with conceptions in the history of physics. Historical considerations were not addressed in teaching but the author believes that that the knowledge of the historical development of physical concepts helped in anticipating students' difficulties in modifying their ideas in a scientific direction.

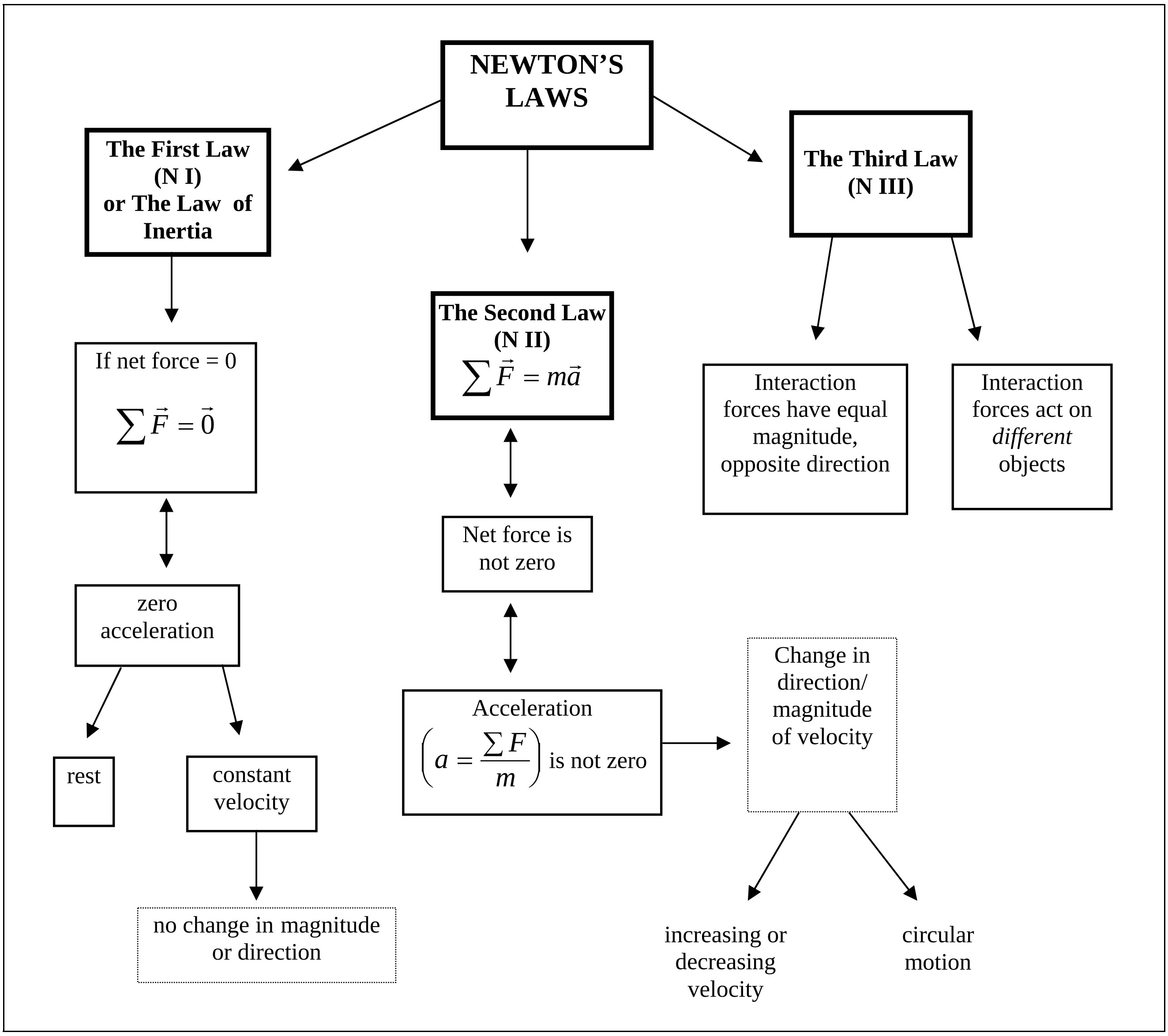


Figure 5.2: Newton's laws and their characterisations.

The two headed arrows in Figure 5.2 emphasise that the implication works both ways: for instance, if net force is zero, it implies zero acceleration and vice versa.

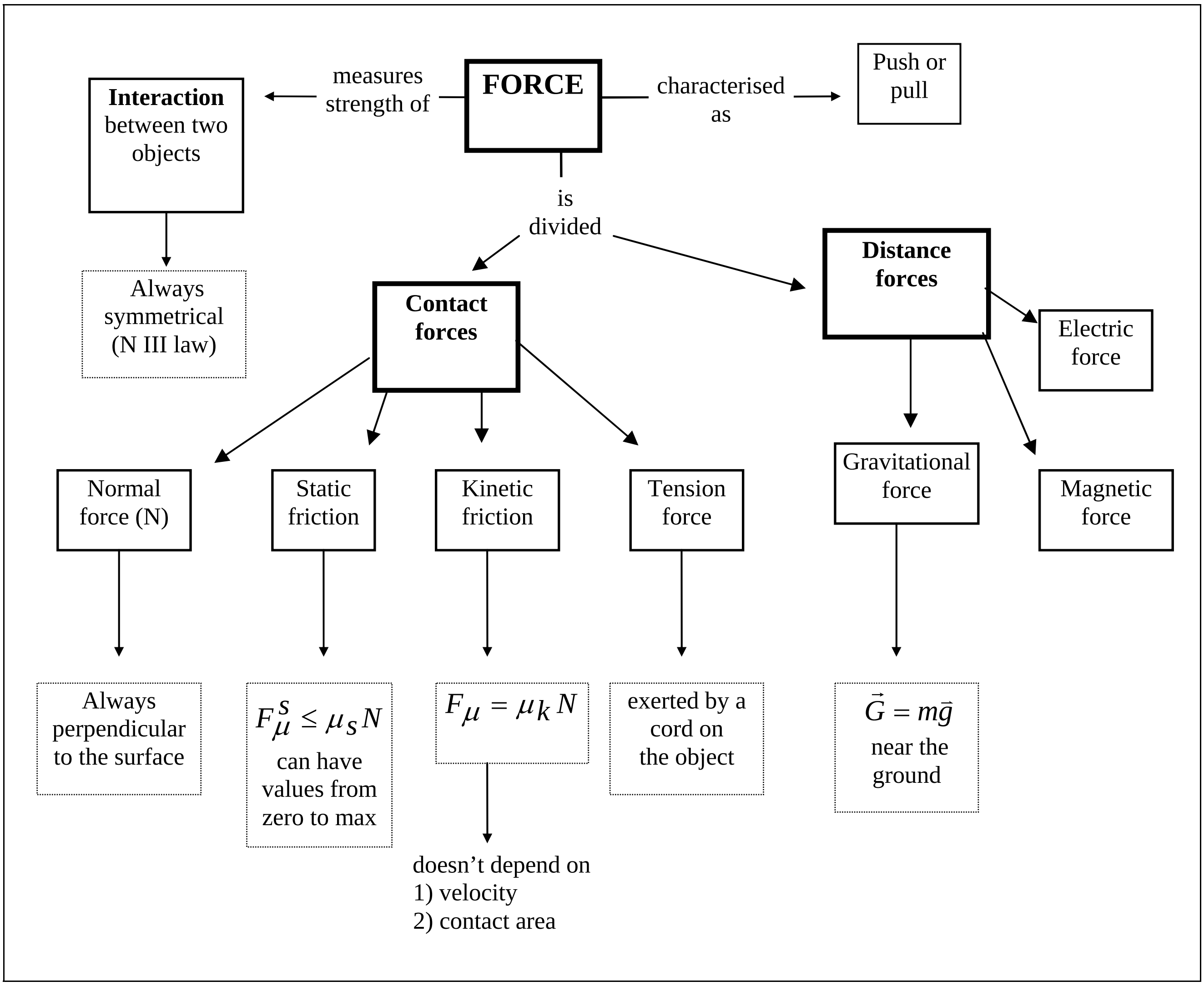


Figure 5.3: The force concept and types of force.

The outline of the course for the Pre-IB pilot group is presented in Table 5.2 and for the Pre-IB study group in Table 5.3. The total duration of lessons for Pre-IB pilot group on kinematics and the force concept was approximately 30 hours. The total duration of lessons for the Pre-IB study group on kinematics and the force concept was approximately 33 hours, after which the teaching addressed work, energy, momentum, and wave motion (see Table 5.4). The force concept was naturally involved in the later topics as well: for instance, the conservation of momentum was derived using Newton's second and third laws. The pilot group also continued studying the same topics as the study group after the force concept.

Table 5.2: Content of the course for the Pre-IB pilot group on the force concept, with comments. Unfortunately, no exact time allocation for the lessons was recorded.

Content	Comment
Pre-FCI	at the beginning of the course
Velocity, change in velocity, acceleration	utilising multiple representations; walking experiment; ticker timer practical; concept map on kinematics
TUG-K test	after completing graphs in kinematics
Falling motion	experiment on height of a projectile, end-of-chapter exercises (Giancoli 1998)
Introduction of the force concept	forces due to interactions (merely mentioned), push or pull
Newton's laws[1] and free-body diagrams	net force and its connection to accelereration introduced first (Newton's second law qualitatively); then Newton's first law as a special case of the second law
Newton's second law	quantitative form of the second law was given by the teacher; using verbal and diagrammatic representations
Exercises on kinematics and forces	moving between verbal and graphical representations; ConcepTests from Peer Instruction (Mazur 1997)
Kinetic and static friction	practical with a block and spring balance
Problem solving on forces	end-of-chapter exercises from Giancoli (1998)
Vector concept introduced (the students had had no vectors in mathematics)	velocity and force vectors; resolving vectors into components; inclined plane problems; horizontal projectile motion
Revision exercises on the force concept	qualitative exercises addressing Newton's laws and free-body diagrams (McDermott et al. 1998; O'Kuma et al. 2000); concept maps (almost the same as in Figures 5.2 and 5.3)
FMCE test	as a part of course test 1
Circular motion	analysing components of acceleration and net force
Applying the force concept and related kinematics	Rollerblade video (Etkina 1998); end-of-chapter exercises (Giancoli 1998)
Relative velocity	motions analysed from different reference frames (making use of Etkina 1998)
Newton's law of gravitation	(the study group did not study the law of gravitation before the post-FCI)
Post-FCI	end of teaching of the force concept
Written question on Newton's third law	as a part of the course test 2

[1] Newton's laws were given and explained by the teacher in a more direct manner than with the Pre-IB study group (for instance, no experiment was carried out to justify the quantitative form of Newton's second law). Of course, peer discussions were used when students practised applying Newton's laws.

Table 5.3: Content of the course for the Pre-IB study group on the force concept, with comments.

Content	Comment	Time (h)
Pre-FCI	at the beginning of the course	0.75
Measurement, error and uncertainty	introduction to physics	1.5
Velocity, change in velocity, acceleration	utilising multiple representations; walking experiment; ticker timer practical; concept map on kinematics	8.5
TUG-K test	five students selected for interviews	0.5
Falling motion, vector concept introduced (the students had had no vectors in mathematics); horizontal projectile motion	experiment on height and range of a projectile, problem solving	4.5
Forces as interactions	SRI diagrams introduced before free-body diagrams; linear air-track practical on Newton's II law; CUPs (Mills et al. 1999);	6
Revision exercises on kinematics and forces	moving between multiple representations; ConcepTests from Peer Instruction (Mazur 1997) concept maps (see Figures 5.2 and 5.3)	1.5
FMCE test	as a part of course test 1	(0.75)
Practice on applying Newton's laws	explicit use of SRI diagrams was discontinued; end-of-chapter exercises (Giancoli 1998) and the Ranking Task Exercises (O'Kuma et al. 2000)	2
Kinetic and static friction	practical with a block and spring balance; free-body diagram exercises with multiple representations (Court 1999); inclined plane problems	4
Circular motion	analysing components of acceleration and net force	3
Relative velocity; vector nature of velocity	motions analysed from different reference frames	0.75
Applying the force concept and related kinematics	Rollerblade video (Etkina 1998); exercises using multiple representations in many contexts	1.5
Summary on kinematics and Newton's laws	students asked to write a summary of their own as a homework	-
Post-FCI	end of teaching of the force concept	0.75
Interview questions	five students interviewed	($\approx$0.75 per student)
Written question on Newton's third law	as a part of the course test 2	

Table 5.4: Content of the course for the Pre-IB study group after finishing with the force concept, with comments (the Pre-IB pilot group covered the same topics).

Content	Comment	Time (h)
Work	introduced as transfer of energy via forces	2.5
Mechanical energy, conservation of energy	conceptual discussions and problem solving	5
Power	practical on work and power	1.5
Revision exercises on work and energy	problem solving	2
Written question on Newton's third law	as a part of course test 2	(0.25)
Linear momentum and impulse	general formulation of the second law	1
Conservation of linear momentum	practical using a linear air-track (practical report)	3.5
Writing practical reports	advice on how to write a practical report	1
Springs	practical determination of the spring constant	2
Waves	no explicit references to mechanics	(16)
Survey on Newton's Third Law	as a part of course test 3	(< 0.5)
Delayed interview questions	the same five students interviewed after 1.5 months after physics instruction	(0.5)

A Finnish translation of the 1995 version of the Force Concept Inventory (Halloun et al. 1995; Koponen et al. 2000) was used as a pre-test because students' competence in English was not sufficiently good at the beginning of the programme. After the pre-test, all the instruments and questions were given in English. It should be noted that the students probably could not have adequately answered the Finnish version of the FCI *after* teaching since they had learned all the material in English.

Five students were chosen for the interview on the basis of their success in the early part of the course: they represented the top, middle and bottom of the study group (see Article IV). A semi-structured interview was used in all interviews. The students were asked 'to think aloud' while answering the interview questions. The students were free to change their answers during the interviews if they so wished. Some students spontaneously wrote down formulas or diagrams. The interviews were conducted by author AS, using neutral questions to clarify the students' views according to the advice given by White and Gunstone (1992).

Table 5.5: The research instruments administered to the Pre-IB study group.

Research Instrument	Timing
Pre-FCI (in Finnish)	at the beginning of the teaching
TUG-K	a month after the pre-FCI
FMCE	two months after the pre-FCI
Post-FCI (in English)	after completing the course on the force concept (a month after the FMCE)
Interview questions	five students interviewed a week after the post-FCI
Written question on Newton's third law	a month after the post-FCI (as a part of the course test 2)
Survey on Newton's Third Law	three months after the post-FCI
Delayed interview questions	the same five students interviewed 4 months after the first interview and 1.5 months after physics instruction

The students were allowed to see the correct answers (except in the case of the pre-FCI) after the administration of the instruments but all the material regarding the instruments and answers were collected. Administration of the research instruments is presented in chronological order in Table 5.5 (also shown in Tables 5.3 and 5.4). The research instruments were administered to the Pre-IB study group over a period of six months. The Pre-IB pilot group students answered the pre-FCI, TUG-K, FMCE, post-FCI, and essentially the same written question on Newton's third law as the study group. They were not interviewed and did not answer the Survey on Newton's Third Law.

5.3.2 Finnish National Syllabus Groups

Two groups of Finnish national syllabus students (the Finnish study group; combined n = 49, age 17) followed a 27-hour course on mechanics in autumn 2001 using the same principles (the ICI approach and focusing on forces as interactions) as the Pre-IB study group. Two similar Finnish national syllabus groups (the Finnish pilot group; combined n = 52) followed the ICI teaching with no focus on forces as interactions in autumn 2000 in the same way as the Pre-IB pilot group. It should be noted that the Finnish pilot and study groups are used to provide complementary data for the Pre-IB groups. Hence, no detailed teaching sequences for the Finnish groups are described here.

A Finnish physics textbook based on algebra and trigonometry (Lehto & Luoma 1999) was used with the Finnish groups. The level of treatment was more demanding in these groups than in the Pre-IB groups because the students in the Finnish groups had already finished three more-or-less traditional introductory courses (not taught by author AS) addressing the force concept and related kinematics before the ICI teaching. For instance, the Finnish groups had more practise in problem solving in complex situations. Many of the research-based exercises used with the Pre-IB groups were also used with the Finnish groups.

The Finnish translation of the 1995 version of the FCI was used (Halloun et al. 1995; Koponen et al. 2000). It was administered prior to, and on completion of, the teaching programme for both Finnish national syllabus groups. Six students from the Finnish study group were chosen for interviews on the basis of their performance in the pre-FCI, representing the top, middle and bottom of the groups. The Finnish pilot group students answered the pre- and post-FCI but they were not interviewed. Administration of the research instruments to the Finnish study group is presented in chronological order in Table 5.6. The Finnish pilot group answered only pre- and post-FCI.

Table 5.6: The research instruments administered to the Finnish study group.

Research Instrument	Timing
Pre-FCI	at the beginning of the teaching
Post-FCI	after completing the course on the force concept
Interview questions	six students interviewed at the end of the course

The results of the Finnish study group are presented and evaluated from the point of view of conceptual coherence in Article III. They are not, however, compared with the FCI results of the Finnish pilot group: this is done in Chapter 7.2. Research methods and research design are discussed and evaluated in the next chapter.

Chapter 6

Research Methods

This chapter discusses methods of testing statistical and practical significance. Firstly, tests of inferential statistics used in the study to test statistical significance are presented. Secondly, Hake gain and effect size are discussed as indicators of practical significance and as an overall measure of the relative success of the courses. These indices were applied to pre- and post-FCI total scores as well as to the dimensions and representations of the force concept addressed by the FCI. Thirdly, the classification of the degrees of coherence is introduced. This provides complementary information to Hake gain and effect size on changes in students' contextual and representational coherence. Fourthly, the validity and reliability of the research methods used are evaluated at the end of this chapter.

6.1 Statistical Methods

6.1.1 Tests of Statistical Significance

Statistical significance provides evidence that an event did not happen by chance (McLean & Ernest 1998). The *p*-value is often used for this purpose: it is the smallest significance level (α) at which the null hypothesis is rejected (Mann 1995, 450). The level of significance α chosen in this study was 0.05. It is very important to note that the *p*-value tells nothing about the magnitude of significance (Nix & Jackson 1998). *P*-value depends on sample size: even small differences in population parameters can be significant if the sample size is large enough.

The difference between the means of two independent samples can be tested using the *t*-test if the following assumptions are met (Ranta et al. 1999, 185-186):

1) samples are random
2) samples are independent
3) measurements taken using an interval scale
4) populations involved are approximately normal (this is important especially when the sample sizes are small)

It is advisable to use a non-parametric test when the sample sizes are small and the estimated variances differ strongly (Ranta et al. 1999, 192). A non-parametric alternative for the *t*-test can be used if the assumptions of the *t*-test are violated: the Mann-Whitney U-test can be applied when two random samples are independent and when the random variable is continuous (Johnson 1996, 699-700). It has almost the same power as the *t*-test (Ranta et al. 1999, 195). The Mann-Whitney test was used in this study to compare the statistical significance of FCI results between the groups.

The data in this study which will be analysed using the Mann-Whitney U-test have two limitations. Fistly, the test score data are not really continuous. This is not a serious limitation: the Mann-Whitney U-test is often used to analyse exam or test score data (for instance, Johnson 1996, 700-701). The second limitation is more serious: the school setting in this study did not allow random samples from one population of students. Random sampling guarantees that the units of the samples are independent from each other. This limitation is addressed in Chapters 6.4 and 8.2.

The Chi-square test can be used to test independence between two variables arranged in a contingency table. It has the following restrictions regarding its use (Key 1997):

1) data must be in frequency form (nominal data).

2) the sample must be representative (random)

3) the individual observations must be independent of each other

4) sample size must be adequate.

   a) In a 2 x 2 table, the chi-square test should not be used if *n* is less than 20.

   b) In a larger table, no expected value should be less than 1, and not more than 20% of the variables can have expected values of less than 5.

5) the distribution basis must be decided on before the data are collected.

6) the sum of the observed frequencies must equal the sum of the expected frequencies.

The data in this study satisfy the above requirements, except randomness of the samples.

McNemar's test is a non-parametric test assessing the significance of the difference between two dependent samples when the variable of interest is a dichotomy. It is used primarily in pre-post studies to test for an experimental effect (Garson 2004). It is the only test available for two dependent samples when the nominal scale is used (Ranta et al. 1999, 221).

6.1.2 Hake gain and Effect Size

As mentioned above, the *p*-value tells something about the statistical significance but does not tell the magnitude of the significance: this calls for different measures. Robinson and Levin (1997, 23) have expressed the relationship between statistical and practical significance in the following way:

"First convince us that a finding is *not due to chance*, and only then, assess how *impressive* it is" (italics in the original)

The statistical tests discussed above are used to demonstrate that findings are not due to chance, while Hake gain and effect size are used to assess how impressive the findings are.

Hake (1998a) introduced an average normalized gain in the context of the FCI. It is the ratio of the actual average gain to the maximum possible average gain:

$$< g > = \frac{< S_{post} > - < S_{pre} >}{100\% - < S_{pre} >} \tag{6.1}$$

where $< S_{post} >$ and $< S_{pre} >$ are the final (post) and initial (pre) class averages. The average gain provides a way to compare classes with very different FCI scores since it normalizes scores based on how well the class performed before the instruction. Hake gain is practically independent of the students' initial scores since there was virtually no correlation between the gain and pre-FCI averages in Hake's (1998a) large survey study on high school, college, and university students. The individual student data in this study also support this conclusion: the correlation coefficient (*r*) between individual student *g* values with individual pretest scores was 0.035 (n = 146).

In his survey Hake identified two broad types of instruction: 'Traditional/Conventional' and 'Interactive–Engagement'. The traditional courses relied primarily on passive-student lectures, 'recipe-following' laboratory sessions and algorithmic quantitative problem-solving examinations. Hake defined Interactive Engagement methods 'as those designed 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'. The average normalized gain (<<g>>) in the 14 traditional survey courses observed (2048 students) was 0.23 $\pm$ 0.04sd (Hake 1998a). The 48 interactive-engagement (IE) survey courses yielded <<g>> = 0.48 $\pm$ 0.14sd.

On the basis of these data Hake (1998a) classified introductory mechanics courses into the following regions:

1) 'High-g' courses: (<g>) > 0.7;
2) 'Medium-g' courses: 0.7 > (<g>) > 0.3;
3) 'Low-g' courses: (<g>) < 0.3.

No course from the study reached the High-g region (the best <g> was 0.69). All the traditional courses fell into the Low-g region. The findings imply that gains in *traditional* high school, college and university physics courses were largely independent of the teachers' experience and academic background, a result which is consistent with the findings outlined earlier in the context of the Mechanics Diagnostic Test (Halloun & Hestenes 1985) and the original FCI (Hestenes et al. 1992).

It is also possible to determine the normalized gain g_{ave} for each single student in the class and consequently the average of the single-student gains (Hake 2001):

$$g_{\text{ave}} = \left(\tfrac{1}{N}\right)\sum_{i=1}^{N} g_i = \left(\tfrac{1}{N}\right)\sum_{i=1}^{N} \frac{(\text{post}_i - \text{pre}_i)}{(100\% - \text{pre}_i)} \tag{6.2}$$

According to Hake (2001), when the number of students taking the test is greater than about 20, g_{ave} is usually within 5% of <g>.

Hake gain provides background data for evaluating the general success of the courses in this study with respect to courses in other institutions. However, this should be done with caution, as demonstrated by Meltzer (2002b), who studied students' conceptual understanding in the domain of electricity. The students in his study followed non-traditional, interactive engagement (IE) instruction. He found a significant correlation between Hake gain and mathematical skill and concluded that 'the observed correlations might imply that widely diverse populations taught with identical instructional methods might manifest different normalized learning gains [Hake gains]'. If it is assumed that mechanics courses would also show similar correlation, Meltzer estimated that the variation in Hake gain ascribable to mathematics preparation would be ± 0.07 for <g> $\approx$ 0.45 (a typical value for mechanics courses employing IE methods). He concludes that even though there is uncertainty in comparing small differences in <g>, large differences are probably due to instructional method.

Effect size is a family of indices that measure the magnitude of a treatment. It is strongly recommended by many psychologists and biologists as the preferred alternative to the usual *t*-tests and *p*-values associated with null-hypothesis testing (Hake 2002 and references within).

Effect size can be defined as follows (there are other definitions):

$$d = \frac{<Post\_FCI> - <Pre\_FCI>}{\sigma_{pooled}} \tag{6.3}$$

$$\sigma_{pooled} = \sqrt{\frac{\left(\sigma_{pre}{}^{2} + \sigma_{post}{}^{2}\right)}{2}} \tag{6.4}$$

where $<Post\_FCI>$ and $<Pre\_FCI>$ are the final (post) and initial (pre) class averages; σ_{post} and σ_{pre} are the final and initial standard deviations. Cohen (1988) defined effect sizes as "small, d = 0.2; medium, d = 0.5; and large, d = 0.8". Both normalized gain and effect size are reported because they do not provide identical information on practical significance: the effect size depends on standard deviation of the sample whereas Hake gain does not. For instance, it is possible to have an insignificant Hake gain (say, less than 0.30) and at the same time the effect size could be very impressive if the standard deviation were very small (Meltzer 2002c).

6.2 Measures of Students' Contextual Coherence

Tests of statistical and practical significance provide valuable information on the effect of teaching regarding the force concept and its dimensions. A different view of the same data can be achieved when the notions of students' contextual and representational coherence are applied (Articles III, IV and V). The test items in various instruments can be classified in terms of aspects of conceptual coherence along different dimensions of the force concept (Chapter 3.3). Moreover, it is possible to categorise the achievement levels of students' conceptual coherence into three classes: 'no coherence', 'partial coherence' and 'coherence'. Of course, the choice of the limits between the categories of achievement levels is somewhat arbitrary. After the data had been categorised, the chi-square and McNemar's tests were used (Chapter 7) to compare the coherence results between the pilot and study groups.

It should be noted that the data gathered do not allow the evaluation of students' contextual or representational coherence along all the dimensions of the force concept.

Newton's Laws

Newton's first law is addressed in the following multiple-choice questions:

- FCI: questions 10, 17, 24 and 25 (verbal representation)
- FCI: questions 6, 7, 8 and 23 (diagrammatic representation)
- FMCE: questions 2 and 5 (verbal representation)

The results regarding *Newton's first law in verbal representation* in the multiple-choice tests were classified into three levels of achievement:

I. 'No contextual coherence': zero or one out of four verbal FCI questions correct; zero out of two verbal FMCE questions correct
II. 'Partial contextual coherence': two or three out of four verbal FCI questions correct; one out of two verbal FMCE questions correct
III. 'Contextual coherence': all four verbal FCI questions correct; both verbal FMCE questions correct.

These quantitative criteria for coherence refer to a specific test and a specific representation, simultaneously. Hence, the criteria were applied to the FCI and FMCE results separately.

The results regarding *Newton's first law in diagrammatic representation* in the multiple-choice tests were classified into three levels of achievement:

I. 'No contextual coherence': zero or one out of four diagrammatic FCI questions correct
II. 'Partial contextual coherence': two or three out of four diagrammatic FCI questions correct
III. 'Contextual coherence': all four diagrammatic FCI questions correct

Newton's second law is addressed in the following multiple-choice questions in verbal representation:

- FCI: questions 22, 26, 27
- FMCE: questions 1, 3, 4, 6, 7 (sled context); questions 8, 9, 10 (car ramp context); questions 11, 12, 13 (coin toss context)

The results of *Newton's second law in verbal representation* were also classified into three levels of achievement:

I. 'No contextual coherence': zero or one out of three verbal FCI questions correct; 0-3 out of eleven verbal FMCE questions correct
II. 'Partial contextual coherence': two out of three verbal FCI questions correct; 4-9 out of eleven verbal FMCE questions correct
III. 'Contextual coherence': all four verbal FCI questions correct; only one verbal FMCE incorrect or all eleven FMCE questions correct (one lapse could be due to carefulessness).

Again, the above criteria were applied to the FCI and FMCE results separately.

Multiple-choice questions require only the identification of correct answers. Interview questions were used to evaluate students' reasoning in more complex situations. Interview responses of six students from the Finnish study group regarding *Newton's laws in verbal representation* (see Chapter 3.3.5) were classified using a corresponding scheme of achievement levels in contextual coherence:

I. 'No contextual coherence': zero or one answer correct.
II. 'Partial contextual coherence': two answers with explanations correct and at least one answer or explanation incorrect.
III. 'Contextual coherence': all answers and explanations correct.

Pre-IB study group students' contextual coherence in the case of *Newton's third law in verbal representation* was investigated using the following criteria:

1. all four FCI questions addressing the third law are correctly answered
2. at least 9 out of 10 FMCE questions on the third law are correctly answered (one lapse was allowed since it could be due to carelessness)
3. the written question has the correct answer with correct reasoning
4. at least 15 out of 16 Survey questions are correctly answered

The above criteria were applied to the different instruments separately. Pre-IB students' contextual coherence in the case of Newton's third law is discussed in detail in Article V.

It is noted that our classification resembles Thornton's (1995) three-fold classification (Student View, Transitional State and the Physicist View) to describe conceptual dynamics, i.e. the process by which students' views are transformed during instruction. Thornton gathered his data using the FMCE test instead of the FCI.

6.3 Measures of Student's Representational Coherence

Students' representational coherence was classified in the same manner as students' contextual coherence. Article IV provides an example of the analysis of representational coherence in the case of Newton's first and second laws. It should be noted that representational coherence here refers to the ability to move between representations.

Kinematics

Representational coherence is addressed in kinematics in the following multiple-choice questions:

- TUG-K: questions 3, 8, 9, 12, 19, 21
- FMCE: questions 22, 23, 24, 25, 26 (car acceleration) and questions 40, 41, 42, 43 (car velocity)

The results regarding kinematics in the TUG-K and FMCE were classified into three levels of achievement in the same way as in the case of contextual coherence:

I. 'No representational coherence': 0-3 questions correct
II. 'Partial representational coherence': more than three questions correct and at least two questions incorrect
III. 'Representational coherence': only one question incorrect or all questions correct

Newton's First and Second Laws

The representational coherence of Newton's first and second laws is addressed in the following questions:

- FMCE: Newton's first law in questions 14, 15, 17 and 21 and Newton's second law in questions 16, 18, 19 and 20
- Interview and delayed interview questions (Chapter 3.3.5) are analysed in Article IV

The results of Newton's first and second laws in the FMCE were classified in the following way:

I. 'No representational coherence': zero or one question correct
II. 'Partial representational coherence': more than one question correct and at least one question incorrect
III. 'Representational coherence': all questions correct

6.4 Research Design and Validity Issues

6.4.1 Research Design

The research design for the Pre-IB study and pilot groups is presented in Table 6.1.

Table 6.1: Research design for the Pre-IB groups.
O = measurement, X = experimental treatment.

	Pretest (FCI)	ICI teaching	Tests during the course	ICI teaching	Post-test (FCI)	Interviews and other tests
Study group	O_1	X^1	O_2, O_3	X^1	O_4	O_5, O_6, O_7
Pilot group	O_1	X^2	O_2, O_3	X^2	O_4	O_5

[1] ICI teaching with the focus on forces as interactions.
[2] ICI teaching with no focus on forces as interactions.

The research design has features of a pre-experimental design: there was no real control group following traditional teaching instead of the ICI. Cohen et al. (2000, 213, 217) criticise the pre-experimental design by pointing out that the effect of extraneous variables cannot be ruled out and these variables can offer plausible hypotheses explaining a possible difference between pre and post-test. It should be noted, however, that in this study multiple research instruments were used (see Chapter 3.3) to monitor students' developing conceptual coherence during and after the ICI teaching. Hence concurrent validity (i.e. agreement of the different data collecting instruments) is sought through methodological triangulation (Cohen et al. 2000, 112-115). Furthermore, it is unlikely that the 'effect of extraneous variables' would be in the same direction for many groups over a period of almost three years.

Although the research design adopted does not permit direct comparisons of learning gains between matched control and experimental classes subjected to 'new' and 'traditional' teaching, it is argued here that comparisons with FCI data gathered in research studies in Finland and elsewhere allow an evaluation of the relative success of the ICI teaching approach. Jauhiainen et al. (2001) sampled the post-FCI results of almost 400 upper secondary school students from different parts of Finland. They found that the distributions of the alternatives chosen by the Finnish students were largely the same as the alternatives chosen by a sample of over 2100 American students. Their data lends convincing support for the same conclusion drawn by Viiri (1995) in the case of Finnish engineering students. Jauhiainen et al. (2001) conclude that their results concerning understanding of the force concept in Finnish upper secondary schools can be generalised and can be used as a reference when evaluating the effectiveness of teaching experiments. Savinainen (2001a) also argues that comparisons between Finnish and American FCI results is possible. It is acknowledged that there are no published Finnish 'baseline' data on Hake gain regarding the FCI. Nevertheless, in the light of data by Jauhiainen et al. (2001) and Viiri (1995) it is extremely unlikely that the results of typical Finnish introductory courses on mechanics would differ dramatically from those of the traditional courses in Hake's (1998a) large survey in the USA.

The research design also has an element of a quasi-experimental design. The groups were not randomly sampled but the Pre-IB pilot group acts as a control group since the teaching those students received did not have a focus on forces as interactions. A 'true' experimental design which entails statistical sampling of students into control (traditional teaching) and experiment (the ICI teaching) groups would have been problematic in this study for several reasons. Firstly, a rigorous sampling of students into control and experiment groups would be very hard to implement in a Finnish high school system, where students choose their courses themselves, i.e. they are not assigned to a given course by the administration. Secondly, an experimental approach in which the author would teach the control group using conventional methods would be ethically difficult to justify, since the author is convinced that the teaching approach developed is better than the lecturing employed earlier. Using other teachers to teach control groups would have been problematic as well since it would have introduced possible differences in the teachers' expertise (Leach & Scott 2002).

The research design of the Finnish group is presented in Table 6.2. It resembles the design for the Pre-IB groups with one exception: the Finnish groups answered only pre- and post-FCI. The Finnish groups act as comparison groups for the Pre-IB groups, providing additional data on the effects of the ICI teaching.

Table 6.2: Research design for the Finnish groups.
O = measurement, X = experimental treatment.

	Pretest (FCI)	ICI teaching	Post-test (FCI)	Interview
Study group	O_1	X^1	O_2	O_3
Pilot group	O_1	X^2	O_2	-

[1] ICI teaching with the focus on forces as interactions.
[2] ICI teaching with no focus on forces as interactions.

6.4.2 Reliability and Validity of the Research Instruments

Reliability addresses consistency and replicability over time, over instruments and over groups of respondents. There are many forms of validity but roughly speaking it refers to a demonstration that a particular instrument measures what it purports to measure. Reliability can be seen as a necessary but not sufficient precondition of validity (Cohen et al. 2000, 105-133). Reliability and certain forms of validity regarding the research instruments used in this study are addressed next.

Multiple-choice tests

A summary of the validation process and the reliability for the FCI is provided in Article II. Additional support for the pre-post FCI design is provided by Henderson (2002), who concluded that giving the pre-FCI does not affect the post-FCI results. Rebello and Zollman (2004) compared students' responses to four FCI questions and equivalent open-ended questions. Their results indicated a good agreement between the percentages of correct responses in each of the two formats. This finding shows that the FCI does not adversely affect performance as measured by the number of correct answers. However, these researchers also found that a significant percentage of the open-ended responses fall into categories that are not included in the FCI multiple choices.

A serious concern concerning the validity of the FCI was raised by Huffman and Heller (1995). They claim on the basis of a (rotated) factor analysis that the FCI does not measure students' coherence of the force concept but more likely 'bits and pieces of student knowledge'. Factor analysis of their data revealed only few significant factors even among so called 'confirmed Newtonian thinkers' who scored over 85% on the FCI. Halloun and Hestenes (1996) answered their criticism by pointing out that the use of factor analysis has been severely criticised in the case of dichotomous variables[13] (Mislevy 1986). Halloun and Hestenes (1996) also show using hypothetical data on near-Newtonian thinkers that the correlation coefficient used by Huffman and Heller (1995) can produce very misleading results. Hence, Halloun and Hestenes (1996) provide evidence that the statistics used for the analysis of the FCI data by Huffman and Heller (1995) are indeed invalid.

The reliability and validity of the TUG-K was clearly established by Beichner (1994). Thornton and Sokoloff (1998) also provide evidence for the validity of the FMCE.

Interviews

All the interview questions used were derived from earlier research. This is why no specific attempts were made to validate the interview questions, since they had already gone through a refinement process. It is admitted, however, that most of the interview questions were not used in exactly the same form as the original sources. Changes in the context of the questions (e.g. an elevator was replaced by a rocket) were introduced, though the structure and wording remained very much the same. In addition, new additions (e.g. graphs) to the interview questions were designed by the author. In all interviews care was taken to make sure that students understood the question and what they were expected to do.

A semi-structured interview was used. All the students interviewed were asked the same questions (in almost all cases) in the same order but the students were free to come back to their answers and

[13] The FCI data as such are not dichotomous but running a factor analysis demands that the data must be dichotomized (i.e., answers to the FCI items are classified as right or wrong).

change them if they so wished. The interview questions were highly structured: this helps to ensure reliability (Cohen et al. 2000, 121). In this study the researcher interviewed his own students. The students were assured that their answers would not have any effect whatsoever on their grades. The students were paid a small amount of money (5 euros) for participating: the author would otherwise have felt that obligating the students to participate in the research would have been unfair.

Investigator triangulation (using more than one observer) was used in analysing the interview data in order to ensure more valid and reliable data analysis (Cohen et al. 2000, 114). The principal supervisor (Jouni Viiri) analysed the interview data independently. In general, his analysis was in very good agreement with that of the author AS. All discrepancies were thoroughly negotiated.

It can be argued the researchers inevitably have some influence on interviewees and consequently on the data (Cohen et al. 2000, 121). Scherr and Wittmann (2002) also discuss how research agendas dictate which students' statements are considered to constitute data. This affects both the way the interviewer asks specifying questions such as 'what do you mean by this?' and the analysis of interview data. In this study, the research agenda addressed only conceptual knowledge of physical mechanism, not for instance beliefs about knowledge.

Schoultz et al. (2001) argue that children's responses in interview studies should be regarded as situated and as dependent on the tools available as resources for reasoning. They make a compelling case by showing how children's understanding of astronomical concepts was dramatically different from the reported literature when a globe was introduced as a tool for thinking. Their results are in good agreement with the view of context-dependent learning as discussed in Chapter 3.1.2. The children in the study by Schoulz et al. (2001) demonstrated good conceptual understanding regarding the earth as a planet in space. There is no guarantee, however, that they would have demonstrated the same understanding in a more abstract, verbal framework used in other studies in the field. In the interviews in this study, the contexts in which the questions were posed were made as explicit as possible by simple demonstrations where possible.

Internal and External Validity

Internal validity addresses the question whether the experimental treatment (in this case the ICI teaching) is really connected with the outcomes. External validity addresses the question, to what population or settings can the demonstrated effects be generalised? Cohen et al. (2000, 126-128) list several threats to internal and external validity, some of which - the most relevant in this study - are addressed in Tables 6.3 and 6.4.

Table 6.3: Some threats to internal validity (Cohen et al. 2000, 126-127) and comments regarding this study.

Threat to internal validity	*Comment*
History: Events other than the experimental treatments occur during the time between the pretest and post-test	It is not likely that students would learn the force concept independently of teaching; it is also unlikely that test-question leakage would have taken place since all test material was always gathered after the tests.
Maturation: Students change in a variety of ways between observations	The changes in students' command of the force concept was monitored during and after teaching; anything other than the desired 'maturation' (i.e. learning) is not likely.
Statistical regression: Students scoring highest in a pretest are likely to score relatively lower on a post-test; those scoring lowest on a pretest are likely to score relatively higher on a post-test	This effect has never been reported in the research literature on the FCI. On the contrary, in Hake's (1998a) large survey there was a positive correlation (+0.55) between the pre-FCI and post-FCI scores.
Testing: Pretest can sensitize students to the purposes of the experiment and consequently produce higher post-test scores	Henderson (2002) found that giving the FCI as a pretest does not affect the post-test. Besides, the students in this study did not know beforehand that there would be a post-test.
Instrumentation: Unreliable test instruments can introduce serious error into experiments	The reliability and validity of the research instruments used is already discussed at the beginning of this chapter.
Selection: Bias maybe introduced if there are differences in the selection of students for experimental and comparison groups	Students were not randomly sampled. Any differences between the groups are identified and evaluated in Chapters 7 and 8.
Instrument reactivity: The effects that the instruments of the study exert on the students in the study	This effect, if it happened in this study, was the same for all groups.

Table 6.4: Some threats to external validity (Cohen et al. 2000, 127-128) and comments regarding this study.

Threat to external validity	*Comment*
Failure to describe independent variables explicitly: If this is not done, future replications of the experimental conditions are impossible	Relevant factors related to the teaching approach are carefully described. It is admitted, however, that the exact replication of the course(s) is not possible even by the author himself. Also factors related to the author's expertise as a teacher cannot be addressed within this study.
Lack of representativeness of available and target populations	All the available students in Kuopio Lyseo High School participated in this study. There may be some but not great variation between students from year to year. Kuopio Lyseo High School attracts very good students so the students in this study do not constitute a random sample of all students in the Kuopio area, or in Finland. Hence direct generalisation of results to the whole student population in Finland is questionable. Nevertheless Hake gain provides a measure which can be used - with some care as discussed in Chapter 6.1.2 - for comparisons between different institutes in Finland and elsewhere.
Hawthorne effect: students perform better because they realise that they are part of an experiment	Hawthorne effect would be expected to diminish when the treatment (i.e. ICI teaching) is applied as a regular long-term routine (three years in this study) to relatively large numbers of subjects (Hake 2002).
Sensitization/reactivity to experimental conditions: the same argument as with internal threats to validity (see Table 6.3).	See the comments for *Testing* and *Instrument reactivity* in Table 6.3.
Invalidity or unreliability of instruments	See the comments for *Instrumentation* in Table 6.3.
Ecological validity: problems in relating experimental conditions to everyday life	Not a problem in this study since the research was carried out as a part of day-to-day teaching

One threat to internal and external validity not mentioned in Tables 6.3 and 6.4 could be that the teacher was simultaneously a researcher and knew what test questions the students had to answer: 'teaching to the test' would indeed invalidate any results. In a sense, there *was* teaching to the test since the very aim of the teaching was to give students some understanding of the force concept. Conceptual understanding was much more emphasized than in the traditional approach, which focuses mainly on problem solving. There was also a deliberate attempt to avoid exercises and contexts identical to the test items. However, some overlap is bound to happen: for instance, it is hard to teach projectile motion without discussing thrown objects, which are addressed in some FCI questions.

Chapter 7

Results

This chapter presents the results achieved when ICI teaching was used with and without focusing on forces as interactions. Some of the results are included in the publications (see Articles II, III, IV and V): they are not repeated here in detail. The purpose of this chapter is to bring together all the relevant data needed to answer the research questions.

The preparatory International Baccalaureate (Pre-IB) and Finnish study groups were taught using ICI with a focus on interactions, whereas the Pre-IB and Finnish pilot groups were taught using ICI but with no focus on interactions. The results are presented in a way which makes comparisons between the pilot and study groups possible. First the FCI data were analysed using average normalized gain (Hake gain), effect size (indicative of practical significance), and the Mann-Whitney U-test (test of statistical significance). Then the same FCI data and other complementary data were analysed from the point of view of students' contextual and representational coherence. This could well be called methodological triangulation (Cohen et al. 2000, 113).

7.1 Preparatory International Baccalaureate Groups

7.1.1 Force Concept Inventory Results

The FCI results, Hake gains, and effect sizes are shown in Table 7.1. The results of the Pre-IB pilot group are discussed in Article II.

Table 7.1: Pre- and post-FCI results, average normalized gains (<g>, Hake 1998a), and effect sizes (Cohen's d). Standard deviations are shown in parentheses.

Pre-IB Groups	N	Pre-test % (S. Dev.)	Post-test % (S. Dev.)	<g>	Effect size
Pilot Group	22	28 (14)	69 (17)	0.57	2.6
Study Group	23	31 (12)	64 (19)	0.48	2.2

The post-FCI averages can first be compared with the Finnish baseline post-test average (58%, in Jauhiainen et al. 2001; Jauhiainen 2003). It should be noted, however, that the students in the study by Jauhiainen et al. (2001) had had more courses in physics than the Pre-IB students in this study. The difference between the combined Pre-IB group (n = 45; $< S_{post,\mathrm{Pr}e-IB} > = 67\%$) and the Finnish baseline group (n = 386; $< S_{post,baseline} > = 58\%$ %) was statistically significant (p = 0.010 in the Mann-Whittney U-test for two independent samples). The non-parametric Mann-Whitney U-test of two independent samples was used to test statistical difference between the FCI averages. A non-parametric test was used rather than the *t*-test, because the data did not follow the normal distribution. The Mann-Whitney test was applied to the data sets gathered in this study and those provided by Jauhiainen (2003) consisting of all individual post-test scores. All statistical testing was carried out using the SPSS for Windows (Version 10) program.

Interestingly, the pilot group which followed ICI teaching without a focus on forces as interactions had a higher post-test average than the study group. The difference was not, however, statistically significant (*p*-values for the pre- and post-FCI results were 0.224 and 0.322, respectively).

Hake gains are in the medium gain region ($0.3 < (\langle g \rangle) < 0.7$; Hake 1998a) indicating that the ICI teaching was relatively successful for both groups. Effect sizes for both Pre-IB groups were well above the 'large' boundary of 0.8 (Cohen 1988). Effect sizes support the conclusion based on Hake gain of effective treatment (i.e. teaching).

Hake gains for each student in the class (Hake 2001) were also determined. They were used to compare statistical significance between the averages of the single-student gains. Again, the Mann-Whitney U-test was utilised: there was no statistically significant difference between the averages of the single student gains ($g_{ave,pilot} = 0.582$ and $g_{ave,study} = 0.494$; $p = 0.199$). It can be seen that the averages of the single student gains are very close to Hake gains. This result is consistent with Hake (2001).

Hake gains and effect sizes can also be used to evaluate changes in different dimensions and representations of the force concept as revealed by the pre- and post-FCI (Table 7.2; see also Table 3.1, which identifies the questions in each category).

Table 7.2: Average normalized gains (<g>, Hake 1998a), and effect sizes with respect to the dimension and representation of the force concept for the Pre-IB pilot and study group. The values were calculated on the basis of the pre- and post-FCI.

Dimension and representation (No. of Questions)	Pilot Group <g>	Study Group <g>	Pilot Group Effect Size	Study Group Effect Size
Kinematics diagrammatic (4)	0.48	0.35	1.1	0.81
Newton I verbal (4)	0.69	0.63	1.9	2.0
Newton I diagrammatic (4)	0.30	0.37	0.71	0.71
Newton II verbal (3)	0.37	0.30	0.86	0.71
N III verbal (4)	0.67	0.91	1.7	3.3
Gravitational verbal (4)	0.75	0.48	1.9	1.2
Contact force verbal (5)	0.83	0.51	3.0	1.8

It can be seen that significant improvement took place along all the dimensions and representations for both groups. There are, however, some differences between the groups. The pilot group clearly improved more in the gravitational and contact force dimensions (statistical significance will be detailed later on). This could be partially due to the fact that the pilot group had somewhat more teaching regarding the gravitational force: the study group did not encounter Newton's law of universal gravitation. On the other hand the FCI did not explicitly require it. The study group showed much greater improvement in Newton's third law than the pilot group. Both Hake gain (0.91; 'high gain region') and effect size (3.3) indicate that the ICI course with a focus on forces as interactions was exceptionally effective for teaching Newton's third law in verbal representation at least at the level of identification of correct answers. There were two gains in the pilot group (Gravitational and Contact forces) and one gain in the study group (Newton's third law) which were above the 'high gain region' (≥ 0.7) .

Comparisons between the combined Pre-IB groups (n = 45) and the Finnish baseline group (Jauhiainen et. al. 2001; Jauhiainen 2003; n = 386) were also made in terms of post-FCI averages across different dimensions and representations of the force concept. The Mann-Whitney test revealed that the combined Pre-IB group had significantly better post-FCI average in the following dimensions in verbal representation:

- Newton's first law ($p = 0.012$)
- Newton's third law $(p < 0.001)$
- Gravitational force $(p < 0.001)$
- Contact forces $(p < 0.001)$

In other dimensions and representations there were no statistically significant differences.

Next, the results regarding Newton's laws are analysed from the point of view of contextual coherence and representational coherence. It could be expected that the dimensions and representations of the force concept, which had showed good improvement, would also exhibit significant contextual coherence.

7.1.2. Pre-IB Students' Contextual Coherence

Newton's First and Second Law

The results of the Pre-IB pilot group in terms of contextual coherence (see Chapter 6.2) of Newton's first law are presented in Figure 7.1. It should be noted that the FMCE was given as a part of the course test 1 about 8 weeks after the pre-FCI and four weeks before the post-FCI.

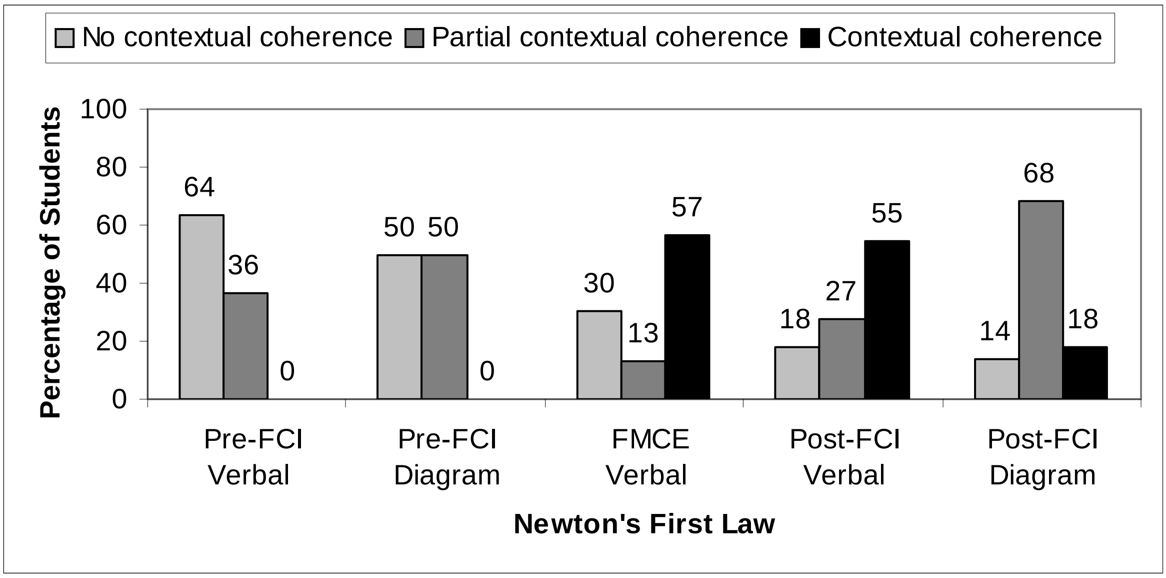


Figure 7.1: Achievement levels of contextual coherence in Newton's First Law for the Pre-IB pilot group. The FMCE test was answered between the pre- and post-FCI tests.

There is significant improvement in Newton's first law in verbal representation for the pilot group. At the beginning of the teaching no students exhibited contextual coherence in Newton's first law in verbal representation, whereas more than half the students did so at the end of the course. About the same percentage of students showed contextual coherence in the FMCE test which addressed the first law in verbal representation in only two questions. There was some progress from the 'no contextual coherence' to the 'partial contextual coherence' category. One cannot conclude, however, that the course was successful in teaching Newton's first law since less than 20% of the students correctly answered all the questions regarding the first law in diagrammatic representation.

The results of the Pre-IB study group in terms of contextual coherence of Newton's first law are presented in Figure 7.2.

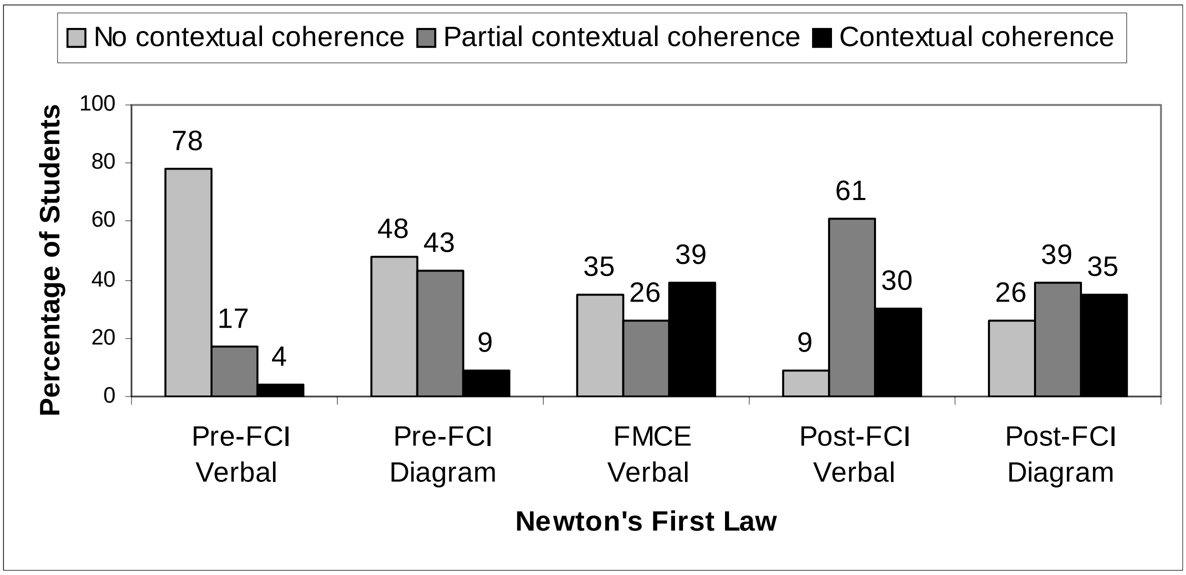


Figure 7.2: Achievement levels of contextual coherence in Newton's First Law for the Pre-IB study group. The FMCE test was answered between the pre- and post-FCI tests.

The results of the study group also show that the students' improved in the first law in verbal representation even though the percentage of students showing contextual coherence at the end of the teaching was lower than in the pilot group. On the other hand, more students in the study group mastered the first law in diagrammatic representation but this is not statistically significant (see Table 7.5).

It should be noted that these differences in the distributions of contextual coherence could not have been directly predicted from Hake's average normalised gains and effect sizes, which were very similar for the pilot and study groups in the case of Newton's first law (both in verbal and diagrammatic representations).

The results of the Pre-IB pilot group in terms of contextual coherence (see Chapter 6.2) of Newton's second law are presented in Figure 7.3.

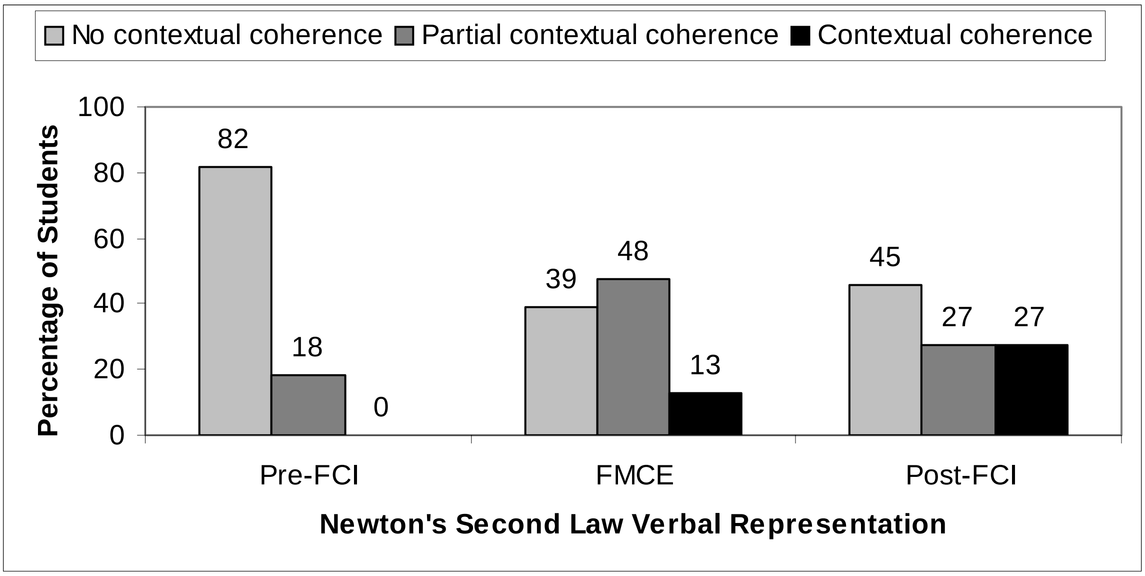


Figure 7.3: Achievement levels of contextual coherence in Newton's Second Law for the Pre-IB pilot group. The FMCE test was answered between the pre- and post-FCI tests.

Clearly, the second law was much harder for the students than the first law. At the end of the teaching, fewer than 30% of the pilot group students mastered the second law in verbal representation. The evaluation of contextual coherence in the case of diagrammatic representation of the second law was not possible, since only one question in the FCI addresses it.

The results of the Pre-IB study group in terms of contextual coherence of Newton's second law are presented in Figure 7.4.

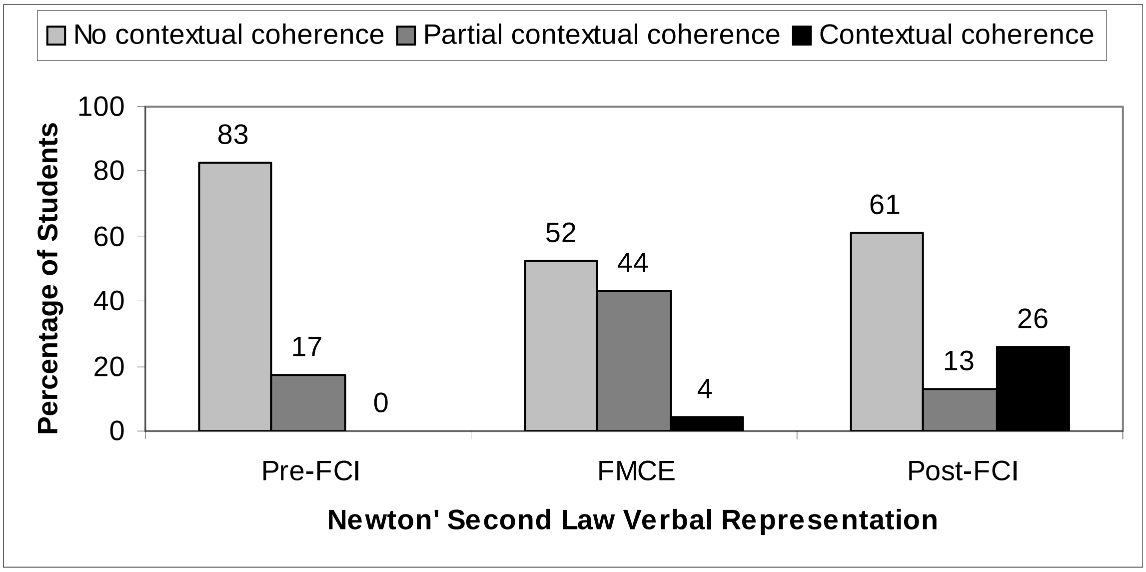


Figure 7.4: Achievement levels of contextual coherence in Newton's Second Law for the Pre-IB study group. The FMCE test was answered between the pre- and post-FCI tests.

Again, the pilot group seemed to have had slightly better results (the issue of statistical significance of the difference is addressed below) in the FMCE and the post-FCI. However, the percentage of the study group students exhibiting contextual coherence of the second law in verbal representation (in the post-FCI) is about the same for both groups.

Newton's Third Law

This is discussed in detail in Article V. The third law results are presented in Figure 7.5.

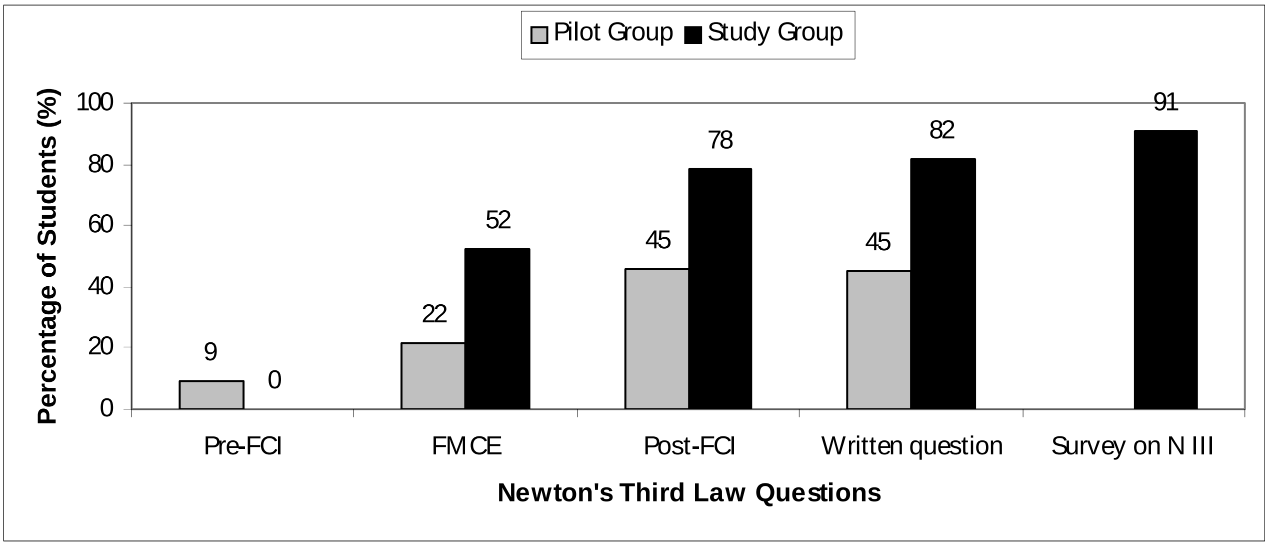


Figure 7.5: Percentage of students who exhibited contextual coherence in the questions addressing Newton's third law for the pilot and study groups. The tests are in chronological order.

The results indicate that almost all the Pre-IB study students reached contextual coherence in Newton's third law at least at the level of the identification of the correct answer. In this respect the focus on forces as interactions in the ICI teaching was successful. The interview data of five students from the study group indicated, however, that the level of identification of correct answers addressed by the FCI is not necessarily sufficient for success in more open, complex situations (see Article V).

Contextual Coherence and Statistical Significance

In the above analysis it is not clear whether the observed differences are statistically significant or not. The main interest is in students who showed contextual coherence in various dimensions of the force concept. These students could be called 'experts' in a limited sense, since they showed expertise across different contexts in one dimension of the force concept at a time. To obtain enough data to fulfil the requirements of the Chi-square and McNemar's tests (see Chapter 6.1), three achievement levels were combined into two:

- non-experts: 'no contextual coherence' or 'partial contextual coherence'
- experts: 'contextual coherence'

McNemar's test can be used to determine whether there were statistically significant changes within groups when the nominal scale is used. An example of McNemar's test data is presented in Table 7.3: the first case (Newton's first law, verbal representation) is statistically significant, whereas the second case (Newton's first law, diagrammatic representation) is not. *P*-values of 0.05 or less are considered to be statistically significant.

Table 7.3: Example of McNemar's test data regarding Newton's first law for the Pre-IB pilot group (using pre- and post-FCI; see Table 3.1). The change from non-experts to experts is statistically significant in the case of verbal representation $(p < 0.001)$ but not in the case of diagrammatic representation (p = 0.063). The data set for diagrammatic representation is in italics.

		After (Post-FCI)	
		Non-experts	Experts
Before (Pre-FCI)	Experts	0	0
	Non-experts	10	12
	Experts	*1*	*1*
	Non-experts	*15*	*6*

The *p*-values for the changes within the Pre-IB groups are shown in Table 7.4. The FCI data are analysed in this manner to allow comparisons with the analysis using Hake gain and effect size (Table 7.2).

Table 7.4: Statistical significance of the changes in the percentages of experts of the pre- and post-FCI results for the Pre-IB groups, using McNemar's test. *P*-values of 0.05 or less (underlined) are considered to be statistically significant.

Dimension and representation (No. of Questions)	**Pilot Group (Pre vs. post FCI) *p*-value**	**Study Group (Pre vs. post FCI) *p*-value**
Kinematics diagrammatic (4)	<u>0.031</u>	<u>0.031</u>
Newton I verbal (4)	<u>< 0.001</u>	<u>0.016</u>
Newton I diagrammatic (4)	0.063	0.063
Newton II verbal (3)	<u>0.016</u>	<u>0.016</u>
N III verbal (4)	<u>0.004</u>	<u>< 0.001</u>
Gravitational verbal (4)	<u>0.002</u>	<u>0.016</u>
Contact force verbal (5)	<u>< 0.001</u>	<u>0.031</u>

The changes in the percentages of experts in almost all dimensions of the force concept were statistically significant. The changes in diagrammatic representation of Newton's first law were not statistically significant in terms of percentages of experts. The corresponding Hake gains in Newton's first law (diagrammatic reprsentation) for the pilot and study groups were 0.30 and 0.37, respectively. The corresponding effect sizes were 0.71. These indices indicate that the results in terms of averages had moderate practical significance. This is an example of a case when an examination of the percentage of experts provides more detailed information on students' success: significant changes took place chiefly from 'no coherence' to 'partial coherence' achievement levels, whereas the change in the number of experts was not statistically significant (though, it was probably practically significant).

The Chi-square test can be used to test the differences between groups, in other words to answer the question: were the differences between percentages of experts in the pilot and study groups statistically significant? There were no statistically significant differences in the pre-FCI results. In fact, the percentages of experts were virtually the same (very low) in both groups. This could have been anticipated directly from the pre-FCI averages which were about 30% (see Table 7.1). There were, however, some differences between the post-FCI results (Table 7.5).

Table 7.5: Statistical significance of the differences in the percentage of experts between post-FCI results of the Pre-IB groups, using the Chi-square test. *P*-values of 0.05 or less (underlined) are considered to be statistically significant.

Dimension and representation (No. of Questions)	**Pilot Group vs. Study Group *p*-value**
Kinematics diagrammatic (4)	0.666
Newton I verbal (4)	0.102
Newton I diagrammatic (4)	0.208
Newton II verbal (3)	0.928
N III verbal (4)	<u>0.023</u>
Gravitational verbal (4)	<u>0.051</u>[1]
Contact force verbal (5)	<u>0.011</u>

[1]This is considered to be statistically significant.

There were statistically significant differences in Newton's third law, gravitational and contact force dimensions. Almost three quarters (71%) of the study group were expert students in Newton's third law compared with less than half (45%) of the pilot group. The pilot group was better than the study group in gravitational force (55% vs. 29% were experts, respectively) and in contact force (59% vs. 30% were experts, respectively).

7.1.3 Pre-IB Students' Representational Coherence

Multiple Choice Data

Representational coherence was probed by the TUG-K and the FMCE tests (see Table 3.2 and Chapter 6.3). Figures 7.6 and 7.7 present the results for the Pre-IB pilot and study groups, respectively.

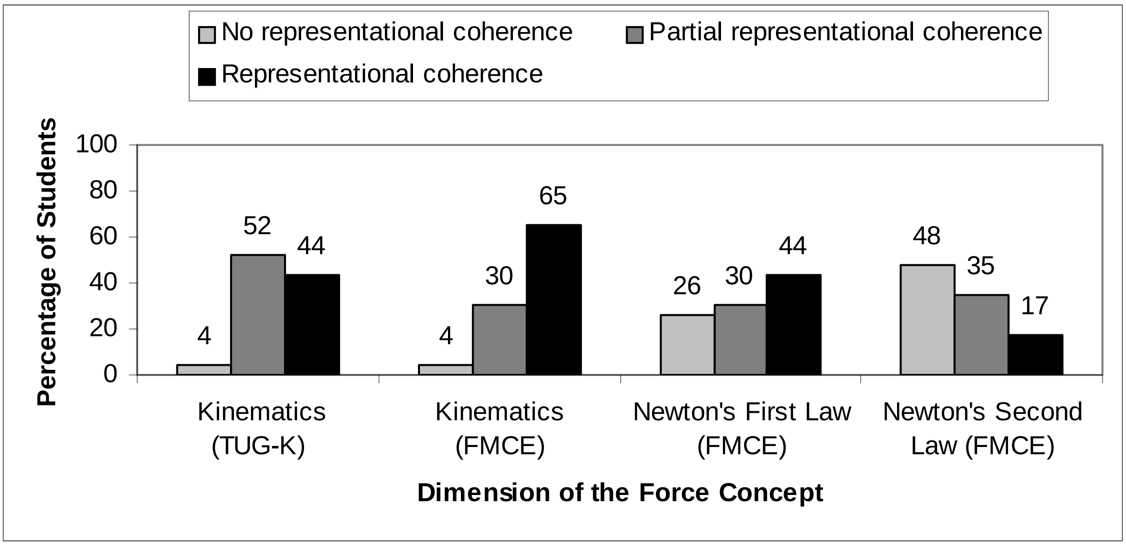


Figure 7.6: Achievement levels of representational coherence in kinematics and Newton's first and second laws for the Pre-IB pilot group.

As can be seen in Figure 7.6, the representational coherence of the pilot group developed positively in kinematics between the TUG-K and FMCE tests. The students exhibited better representational coherence in Newton's first law than in Newton's second law (according to McNemar's test the difference was statistically significant: p = 0.016). In the FMCE test, students' ability to apply Newton's first law in moving between verbal and graphical representations is somewhat lower than the ability to apply Newton's first law in verbal representation (compare Figure 7.6 with Figure 7.1). The second law was equally hard for the students both in verbal and graphical representation (compare Figure 7.6 with Figure 7.3).

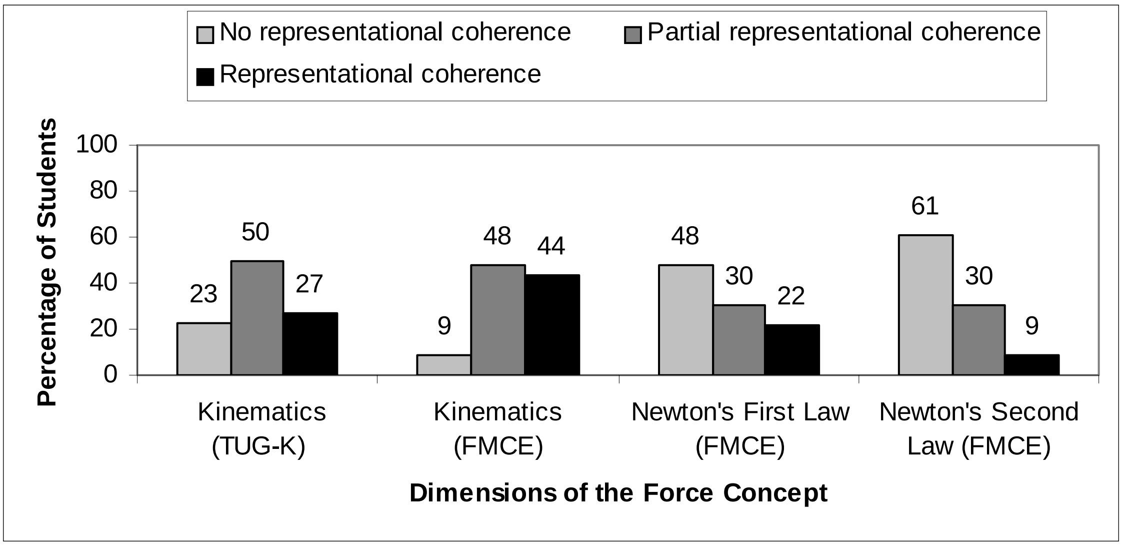


Figure 7.7: Achievement levels of representational coherence in kinematics and Newton's first and second laws for the Pre-IB study group.

The same conclusions can be drawn in the case of the study group as well. It should be noted, however, that the pilot group exhibited better representational coherence than the study group in all the areas addressed by the TUG-K and the FMCE tests.

Interview Data

The interview data of the Pre-IB study group is discussed in Article IV. Five students were interviewed a week after the post-FCI and again four months after the first interview (delayed interview). The students exhibited representational coherence of Newton's first law in both interviews but did not do as well in the case of Newton's second law. These results are in good accordance with the results of the multiple-choice tests. The interview data also revealed that the second law was easier for the students in verbal representation than in graphical or diagrammatic representation. This suggests that students' understanding is not just context dependent, but also representation dependent. Viiri and Savinainen (2004) discuss the same interview data from the point of view of conceptual change.

7.2 Finnish National Syllabus Groups

7.2.1 Force Concept Inventory Results

The FCI results, Hake gains, and effect sizes for the Finnish pilot and study groups are shown in Table 7.6. It can be seen that both the pre- and post-FCI averages of pilot groups A and B differ significantly from each other (*p*-values 0.000 and 0.017 in the Mann-Whitney test, respectively). In the Finnish high school system students choose freely which courses they attend from the available possibilities (in this case the same course on mechanics was offered twice). The statistical differences between pilot groups A and B imply that course A attracted, on average, better students than course B. There were no statistical differences in the pre-FCI or post-FCI results for the study groups A and B (*p*-values 0.123 and 0.565 in the Mann-Whitney test, respectively).

Table 7.6: Pre- and post-FCI results, average normalized gains (<g>, Hake 1998a), and effect sizes (Cohen's d). Standard deviations are shown in parentheses.

Finnish Groups	N	Pre-test % (S. Dev.)	Post-test % (S. Dev.)	<g>	Effect size
Pilot Group A	32	64 (19)	81 (14)	0.45	0.96
Pilot Group B	20	44 (13)	68 (17)	0.44	1.6
Pilot Group (combined)	52	56 (20)	76 (17)	0.45	1.1
Study Group A	27	61 (15)	84 (20)	0.58	1.3
Study Group B	22	53 (17)	81 (13)	0.61	1.9
Study Group (combined)	49	57 (19)	82 (14)	0.59	1.5

The pilot and study groups were combined into single pilot and study groups, because

- the pilot and study groups had the same pre-course curriculum history
- the pilot groups followed the same ICI teaching and the study groups followed the ICI teaching with a focus on forces as interactions
- all the groups were taught by the same teacher (author AS)
- the pilot groups as well as the study groups included all students studying physics in that age group (second year students, i.e. aged 17) in Kuopio Lyseo High School

In the following discussion the terms ‘pilot group’ and ‘study group’ refer to the combined groups unless otherwise stated.

Both groups had followed more or less traditional introductory courses addressing the force concept and related kinematics before being taught using ICI. This is reflected in the relatively high pre-test averages, which are very close to the reported post-teaching average (58%) in Finland (Jauhiainen et al. 2001).

The Mann-Whitney U-test revealed that there was no statistically significant difference between the pre-FCI results of the pilot and study groups ($p = 0.535$). However, the difference between the post-FCI averages between the groups was statistically significant ($p = 0.051$). Hake gain and effect size also indicate that the ICI teaching with a focus on forces as interactions was more successful in the case of the Finnish study group than in the case of the Finnish pilot group. This conclusion is reinforced by an examination of the averages of the single student gains: there was a statistically significant difference ($g_{ave,pilot} = 0.414$ and $g_{ave,study} = 0.625$; $p = 0.003$).

Table 7.7: Average normalized gains (<g>, Hake 1998a), and effect sizes (Cohen’s d) with respect to the dimension and representation of the force concept for the Finnish pilot and study group. The values were calculated on the basis of the pre- and post-FCI.

Dimension and representation (No. of Questions)	Pilot Group <g>	Study Group <g>	Pilot Group Effect Size	Study Group Effect Size
Kinematics diagrammatic (4)	0.34	0.50	0.54	0.79
Newton I verbal (4)	0.61	0.85	0.87	1.4
Newton I diagrammatic (4)	0.51	0.59	0.66	0.96
Newton II verbal (3)	0.43	0.40	0.78	0.68
N III verbal (4)	0.28	0.95	0.23	1.3
Gravitational verbal (4)	0.34	0.55	0.53	0.89
Contact force verbal (5)	0.60	0.67	1.3	1.2

Table 7.7 presents Hake’s average normalized gains and effect sizes for the Finnish pilot and study groups. The study group had greater indices of improvement in all dimensions except Newton’s second law: the difference is especially dramatic in the case of Newton’s third law. There were no ‘high region’ gains (≥ 0.7) in the pilot group, whereas the study group had two gains (Newton’s first and third laws, and verbal representations) above the high gain region. Incidentally, the only

'low region' gain (< 0.30) was Newton's third law in the pilot group. This was mainly due to poor results in pilot group B (n = 20).

Pilot group A (n = 32) had almost the same post-test average as the study group. In general, Hake gains for pilot group A were only slightly smaller than for the study group except in two cases: pilot group A had a smaller gain (0.66) in Newton's third law and larger gain in Newton's second law (0.55) than the study group. It is noteworthy that pilot group B actually had a negative gain in Newton's third law even though the students followed almost identical teaching sequences as those in pilot group A. This suggests that teaching is not the only factor that affects students' conceptual gains.

The difference between the Finnish baseline group (n = 386) and the combined Finnish group in this study (n = 101) was statistically significant (Mann-Whitney U-test, $p < 0.001$) in all the dimensions and representations of the force concept. It should be noted that the combined Finnish group and the Finnish baseline group students had had approximately the same amount of studies in physics.

7.2.2 Finnish National Syllabus Students' Contextual Coherence

Multiple-choice Data

The results of the Finnish pilot and study groups in terms of contextual coherence of Newton's first law are presented in Figures 7.8 and 7.9, respectively. The results of the Finnish study group are discussed in more detail in Article III.

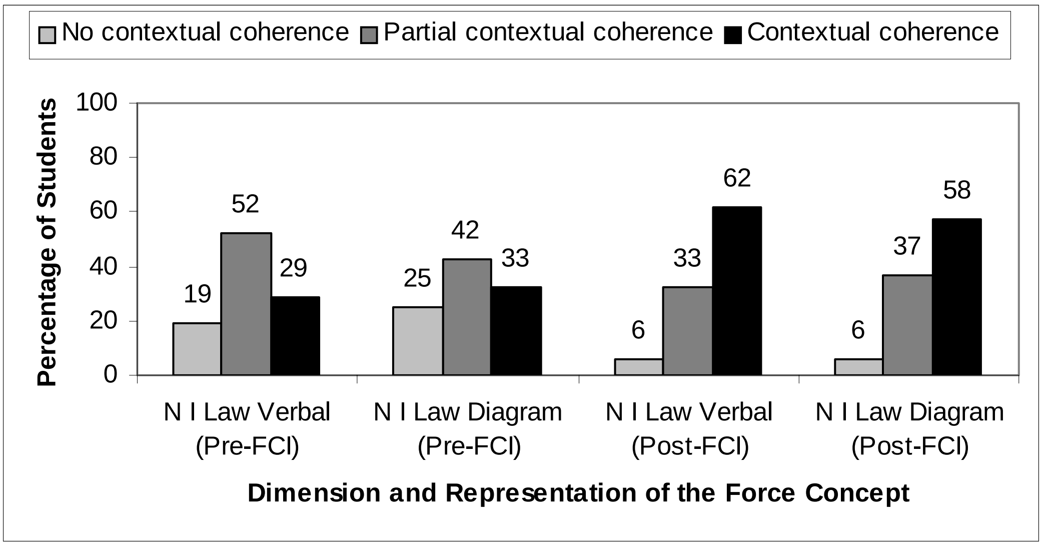


Figure 7.8: Contextual coherence of Newton's First Law in the pre- and post-FCI for the Finnish pilot group.

There is a clear improvement in the contextual coherence of Newton's first law both in verbal and diagrammatic representation (Figure 7.8). The pilot group students seemed to master Newton's first law about equally well in verbal and diagrammatic representations.

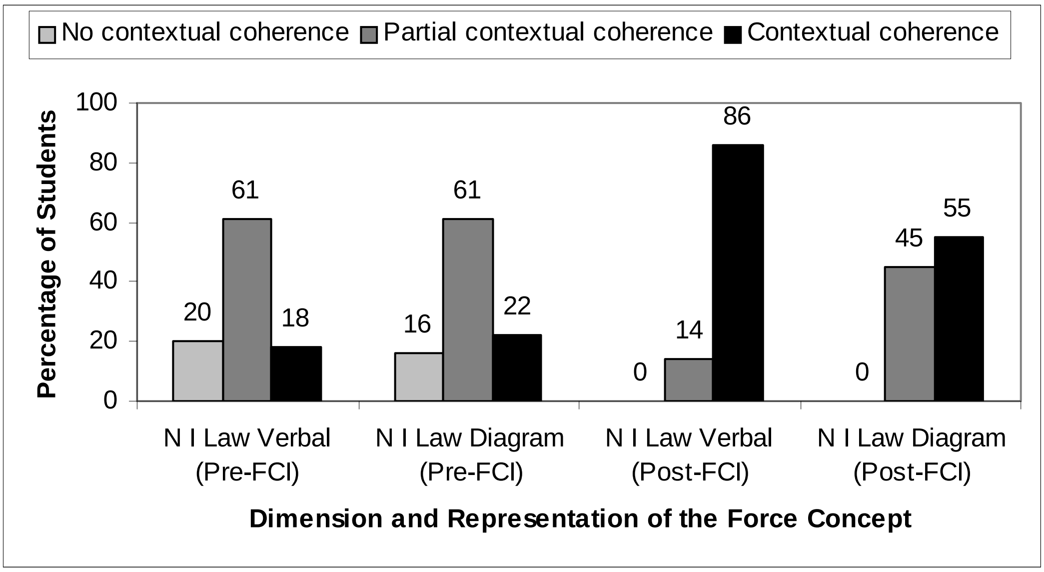


Figure 7.9: Contextual coherence of Newton's First Law in the pre- and post-FCI for the Finnish study group.

The same positive development is evident in the case of the study group as well (Figure 7.9). The Finnish study group exhibited a greater degree of contextual coherence of Newton's first law in verbal representation than the Finnish pilot group. The corresponding post-FCI averages were also greater in the study group: 85% of for the pilot group, and 94% for the study group. The difference is statistically significant (*p*-value = 0.011, again using the Mann-Whitney U-test). It should be noted that even though the post-FCI average of the pilot group in Newton's first law verbal representation was very high (85%) only 62% of the students exhibited contextual coherence.

The results of the Finnish pilot and study groups in terms of contextual coherence of Newton's second and third laws are presented in Figures 7.10 and 7.11, respectively.

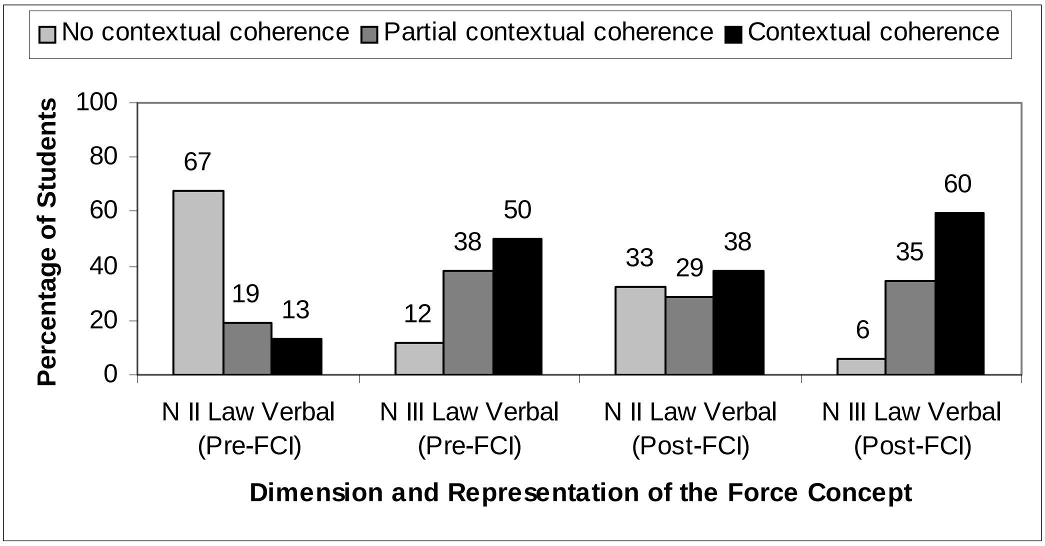


Figure 7.10: Contextual coherence of Newton's Second and Third Law in the pre- and post-FCI for the Finish pilot group.

There was a moderate improvement in the contextual coherence of Newton's second law in verbal representation. Students' contextual coherence of the third law, however, changed very little.

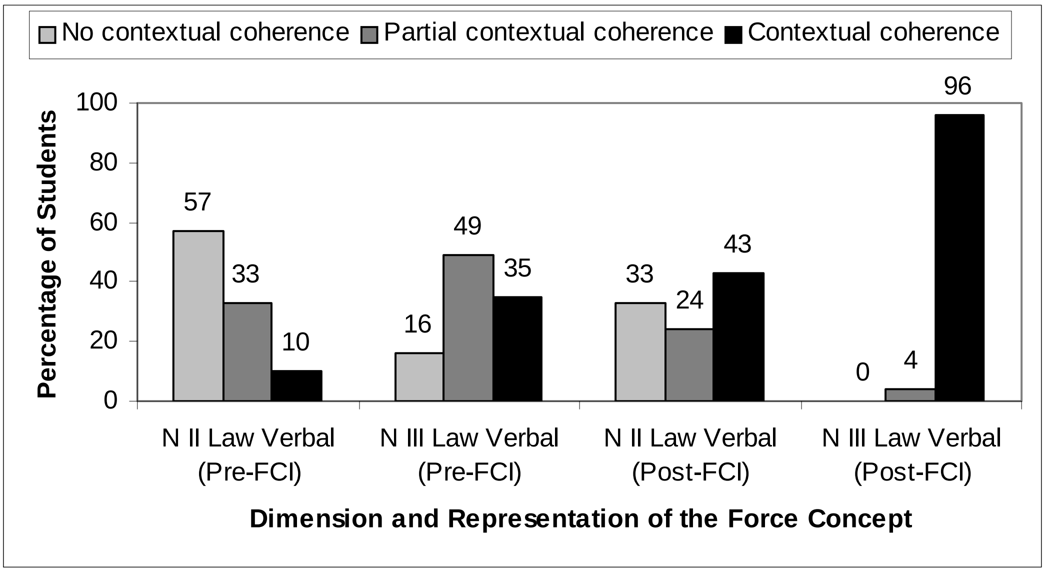


Figure 7.11: Contextual coherence of Newton's Second and Third Law in the pre- and post-FCI for the Finish study group.

The development of the contextual coherence of the second law is very similar also in the case of the study group. There is a dramatic difference, however, in the contextual coherence of Newton's third law: almost all the students in the study group exhibited contextual coherence at the level of identification of correct answers at the end of the teaching.

Contextual Coherence and Statistical Significance

Statistical comparisons of the percentage of experts in the Finnish group were carried out in the same manner as in the Pre-IB groups. McNemar's test showed that changes in all dimensions of the force concept from pre-FCI to post-FCI were statistically significant ($p \leq 0.05$), except for one case: there was no statistically significant change in Newton's third law in the Finnish pilot group ($p = 0.151$; percentage of experts changed from 50% to 60%).

The Chi-square test showed that there were no statistically significant differences in the pre-FCI results between the Finnish pilot and study groups (all *p*-values were above 0.10). Table 7.8 shows the *p*-values for the post-FCI results in different dimensions and representations of the force concept.

Table 7.8: Statistical significance of the differences between the post-FCI results of the Finnish groups, using the Chi-square test. *P*-values of 0.05 or less (underlined) are considered to be statistically significant.

Dimension and representation (No. of Questions)	**Pilot Group vs. Study Group *p*-value**
Kinematics diagrammatic (4)	0.143
Newton I verbal (4)	<u>0.006</u>
Newton I diagrammatic (4)	0.793
Newton II verbal (3)	0.653
N III verbal (4)	<u>0.000</u>
Gravitational verbal (4)	<u>0.008</u>
Contact force verbal (5)	0.061

The percentage of experts in the study group was significantly greater than in the pilot groups in Newton's first law, verbal representation (86% vs. 62%), Newton's third law (96% vs. 60%, respectively) and gravitational force (57% vs. 31%, respectively). Both Hake gains and effect sizes (Table 7.7) show that there were also differences in terms of practical significance. These differences are especially notable in the case of Newton's third law.

Interview Data

Six students in the Finnish study group were chosen for the interview on the basis of the pre-FCI results. They represented the top, middle and bottom of the groups. The FCI and interview results regarding Newton's laws are presented in Figures 7.12 and 7.13. The students (labeled from 1 to 6) are shown in the boxes and the numbers of students moving between achievement levels are indicated by arrows. Students 1 - 3 were interviewed before and students 4 - 6 after the post-FCI, which is why no arrows are shown between post-FCI and interview. The analysed interview questions addressed only verbal representation.

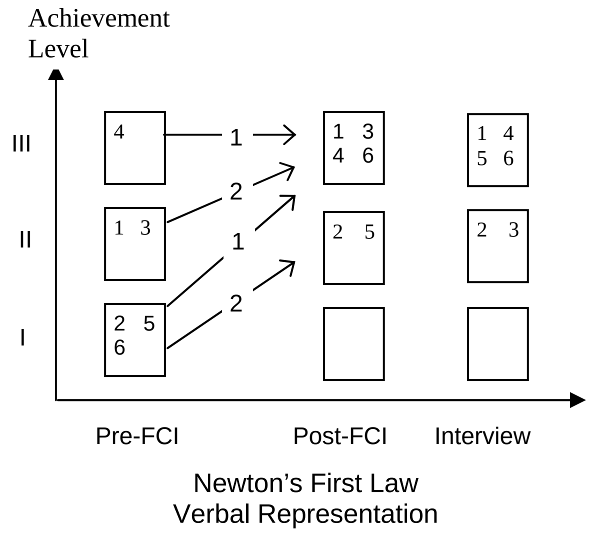


Figure 7.12: Newton’s first law: achievement levels of contextual coherence in verbal representation. Achievement level I means ‘no contextual coherence’, level II ‘partial contextual coherence’ and level III ‘contextual coherence’. The students were drawn from the Finnish study group.

There was a clear improvement in Newton’s first law (Figure 7.12). Both the FCI and interview questions indicate that students had reached a good level of contextual coherence in verbal representation of the first law.

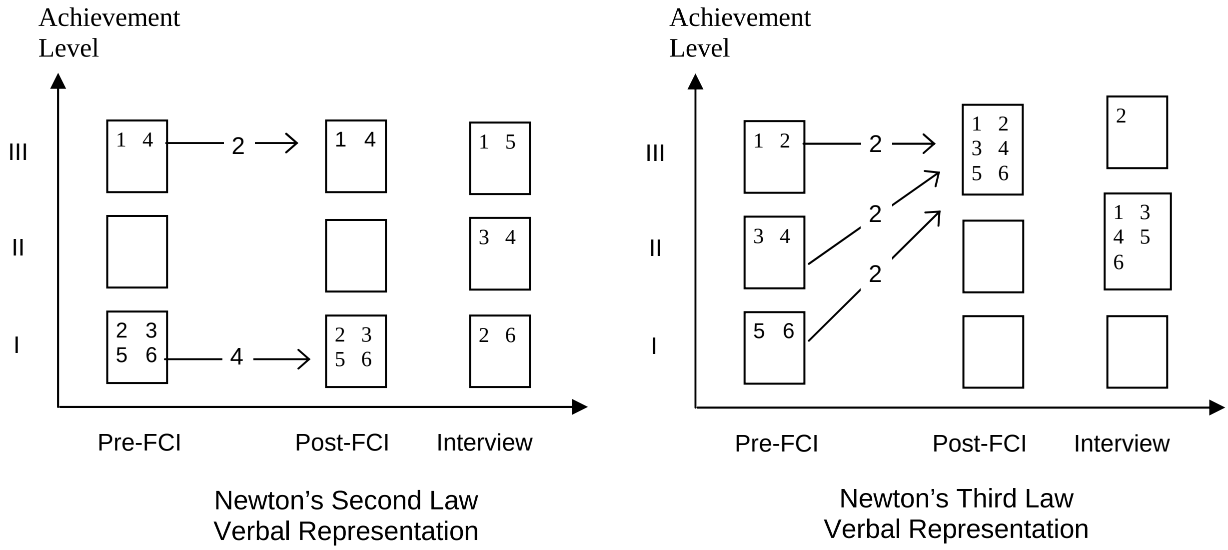


Figure 7.13: Newton’s second and third laws: achievement levels of contextual coherence in verbal representation. Achievement level I means ‘no contextual coherence’, level II ‘partial contextual coherence’ and level III ‘contextual coherence’. The students were drawn from the Finnish study group.

The FCI results show that in terms of contextual coherence the teaching of Newton’s second law was ineffective for the students (Figure 7.13). The students did do better in the interviews than in the FCI. This could be partially due to the fact that the interview questions addressed the second law only in one dimension, whereas one FCI question (22) required the application of Newton’s second law in a 2-dimensional case. It can be seen that student 5 exhibited basic understanding of Newton's Second Law in the interview (i.e., if an object is accelerating, there must be a net force acting in the direction of acceleration) but this level of understanding was not enough to answer the corresponding FCI questions correctly.

One student (student 4), who exhibited perfect understanding of Newton's second law in the FCI, had some difficulty with the idea of net force acting on the sub-systems of the system. This is why the views he expressed in the interview were not classified as 'contextual coherence' in the second law. Otherwise he answered the interview questions very well. No questions in the FCI address net forces acting on sub-systems. This is not surprising, since one test or interview cannot exhaustively assess all aspects of the force concept.

All the students identified the correct answers in the post-FCI regarding Newton's third law. The first interview question, which involves identification of the magnitudes of the interaction pair forces in a collision, is analogous to the FCI question 4. It was very well answered: only one students gave an incorrect answer. However, the students were not as successful in more complex questions. Contextual coherence of the third law at the level of identification of correct answers demonstrated in the FCI did not guarantee mastery in the situations which addressed many related concepts at the same time. This result lends support to the conclusion drawn in the case of Pre-IB study group (Article V).

The interview results revealed the same trend as the FCI results: Newton's first law was easier for the students than Newton's second law. Those who scored highly in the FCI exhibited very good (but not perfect) understanding in the interviews as well. The number of interviewed students was quite small but their interview results support the conclusion that the FCI appears to be a good measure of contextual coherence of the force concept at least at the level of identification of correct answers of correct anwers.

7.3 Summary of the Results

Research aims

1. What does students' conceptual coherence entail?

This research aim is analysed in detail in Chapter 3.2. As a brief summary, conceptual coherence can be divided into three aspects:

- representational coherence, which addresses students' ability to use multiple representations (such as verbal, diagrammatic, graphical) and move between them[14]
- contextual coherence, which addresses students' ability to apply a concept or a physical principle in a variety of familiar and novel situations, i.e. the dependency of students' understanding from contextual features which are irrelevant from the physicist's point of view
- conceptual framework coherence. which addresses students ability to differentiate and integrate related concepts in a certain domain of physics (e.g. in the case of the Newtonian force concept)

[14] It should be noted that only the students' ability to move between multiple representations was explicitly tested in this study.

2. How can students' conceptual coherence of the force concept be evaluated?

Certain groupings of the FCI and the FMCE questions can be used to probe contextual coherence, as can written and interview questions addressing several contexts (or different contextual features within the same general context) in a certain representation of the same dimension of the force concept. A student exhibits contextual coherence in a certain dimension and representation of the force concept if she/he answers most of the questions addressing that dimension and representation correctly (see Chapters 3.3 and 6.2).

Certain groupings of the TUG-K and FMCE questions can be used to probe students' representational coherence in kinematics. Interview questions can be used to probe students' representational coherence in other dimensions of the force concept within a certain context. A student exhibits representational coherence if she/he correctly answers most of the questions addressing that dimension of the force concept using multiple representations (see Chapters 3.3 and 6.3).

Conceptual framework coherence was not evaluated in this study. It should be noted that the framework coherence is a necessary but not sufficient condition for representational and contextual coherence.

Research questions

3. a) What was the effect of the two types of ICI teaching on students' conceptual gains as measured by the FCI?

Hake gains in the FCI for all ICI groups fall in the middle or upper end of the 'medium gain region' (0.3 < (<g>) <0.7): they were between 0.45 and 0.59. The effect sizes were well above the 'high boundary of 0.8': they were between 1.1. and 2.6. These indices show that the effect of the two types of ICI teaching had practical significance at least as measured by the overall FCI results.

Hake gains were at least 0.3 (the lower limit of the 'medium gain region') and effect sizes were above 0.5 ('medium') for all the dimensions of the force concept (except in one case: see Table 7.7). The most impressive conceptual gains were made in Newton's first law in verbal representation, Newton's third law in verbal representation and contact force in verbal representation. In almost all these cases, Hake gains were above 0.50 and effect sizes above 1.1.

Furthermore, it can be concluded that ICI teaching is useful for groups with very different initial levels, since it yielded equally good results in terms of Hake gains when the students' initial level was very poor (pre-FCI average about 30%) and quite high (pre-FCI average nearly 60%).

b) How do the FCI results of the ICI groups compare with results in other institutions and instructional settings?

Although the research design adopted does not permit direct comparisons in learning gains between matched control and experimental classes subjected to 'new' and 'traditional' teaching, relative success of ICI teaching is well justified, as shown above. The Hake gain for the ICI groups was

0.52 ± 0.07sd [15]. This compares well with results reported in the USA in different institutions where 48 interactive- engagement (IE) survey courses yielded <<g>> = 0.48 $\pm$ 0.14sd (Hake 1998a). It should also be noted that three out of four Finnish subgroups (Table 7.6 in Chapter 7.2.1) had results of over 80% in the post-FCI. No high school and only one college in Hake's (1998b) survey exceeded 80% in the post-FCI. The pre-FCI results of the Finnish groups were also very high by U.S. standards (Hake 1998b). This may be explained by the fact that the students had already had instruction in mechanics before the ICI course.

There are no Finnish data on Hake gains, but comparisons in terms of post-FCI results with the Finnish baseline (Jauhiainen et al. 2001; Jauhiainen 2003) indicate that both the combined Pre-IB group and the combined Finnish group had better post-FCI averages; the differences were statistically significant ($p = 0.010$ and $p < 0.001$, respectively).

4. *What was the effect of the two types of ICI teaching on students' contextual and representational coherence of the force concept*?

The ICI teaching enhanced contextual and representational coherence of the force concept in all the probed dimensions of the force concept for the pilot and study groups. In most dimensions the changes were also statistically significant ($p \leq 0.05$). In general, the most notable improvement took place in Newton's first law (all groups) and Newton's third law (the study groups) in verbal representation.

In most groups, fewer students reached contextual coherence of Newton's first law in diagrammatic representation. It can also be concluded that Newton's second law proved to be harder for all groups than the first law. This is not surprising since the first law can be viewed, at least in the high school level, as a special case of the second law.

5. a) *What was the effect of the focus on forces as interactions on students' contextual coherence regarding Newton's third law*?

The effect of emphasising forces as interactions resulted in much better results on Newton's third law. More students in the study groups exhibited contextual coherence on Newton's third law after teaching than in the pilot groups (the differences were statistically significant: $p \leq 0.023$). It is clear that the differences were also practically significant: for example, the effect size for the FCI questions addressing Newton's third law for the Pre-IB study group was extremely high (3.3).

b) *What was the effect of the focus on forces as interactions on students' contextual and representational coherence in other dimensions of the force concept*?

In other dimensions of the force concept the results are not conclusive: the Pre-IB study group did not do better than the Pre-IB pilot group in most dimensions and representations of the force concept, whereas the Finnish study group was better than the Finnish pilot group in most dimensions and representations of the force concept. Hence, it cannot be concluded that focusing on forces as interactions would necessarily enhance students' conceptual coherence of the force concept in dimensions other than Newton's third law.

[15] To be precise, there are not enough data (n = 4) to justify the use of standard deviation (Swartz & Miner 1998, 62). However, it is nevertheless reported here to allow a more direct comparison with Hake's (1998) results which are reported in terms of standard deviations.

Chapter 8

Discussion

8.1 The Force Concept Inventory

The FCI data gathered in this study throw light onto the debate regarding the FCI as a measure of coherence in students' force concept. The Finnish study group students' post-FCI average was very high (82%) and 53% of these students scored over 85% (the limit for the 'confirmed Newtonian thinkers' as defined by Hestenes and Halloun (1995)). The analysis of the same post-FCI data from the point of view of contextual coherence revealed that most students exhibited contextual coherence in Newton's first and third laws in verbal representation. About half of the students reached contextual coherence in other dimensions and representations of the force concept. These results indicate that high FCI scores and significant coherence in students' force concept are closely related to each other, which suggests that the FCI does not just measure 'bits and pieces of student knowledge', as Huffman and Heller (1995) state. Their conclusion was based on a type of factor analysis which is not well suited to FCI data as pointed out by Halloun and Hestenes (1996) (see Chapter 6.4.2). But despite their criticism, physics education research articles seem to continue employing "simple" factor analysis (e.g. Singh & Rosengrant 2003; Engelhardt & Beichner 2004). It is not argued here, however, that factor analysis is useless in analysing multiple-choice data. The usefulness of a more sophisticated factor analysis in the context of the FCI will be addressed in future research.

Clearly, the FCI has merits: its reliability and validity have been well confirmed (an overview is provided in Article I). It should be kept in mind, however, that the FCI, like any multiple-choice test, also has limitations: for instance, it can directly test only the ability to choose a correct answer from amongst attractive distractors. It is quite a different task to demonstrate understanding in a more open situation. The interviews conducted in this study provide some insights to the relationship between students' performance on the FCI and in situations, where they need to produce the answers themselves. On one hand, the interviews in this study suggest that if a student does well in the FCI, he/she generally demonstrates good command of the Newtonian view in the interview situation as well. Hence, this lends further support for the validity of the FCI. On the other hand, the interview results suggest that the identification of a correct answer in the FCI is not necessarily enough for good performance in a more open and complex situation. This should not come as a surprise if the FCI is really interpreted as a minimum competence test in the domain of the force concept: this finding is in good agreement with that of Hestenes and Wells (1992), who concluded that '*a good score on the* [*Force Concept*] *Inventory is a necessary but not a sufficient condition for a good score on the Baseline* or on other problem-solving tests on mechanics' [italics by Hestenes and Wells]. The present study did not address problem solving, but the interview results suggest that their conclusion is also valid in the case of complex qualitative questions. It should be noted, however, that in this study only five or six students per group were interviewed. This means that great caution must be exercised when generalising.

The FCI data in the present study revealed that in most groups, fewer students mastered Newton's first law in diagrammatic representation than in verbal representation. Furthermore, the interview data (Article IV) indicate that Newton's second law was easier for the students in verbal representation than in diagrammatic or graphical representations. These findings suggest that students' understanding is not just context dependent, but also representation dependent. Hence, the results of this study lend support to Meltzer's (2003) findings. However, it should be noted that the FCI cannot be used for the *direct* evaluation of representational coherence, for two reasons: firstly, most questions are framed in different contexts (probing of representational coherence requires that the effects of contextual factors should be minimized), and secondly, the FCI does not involve graphical representation.

The wide use of the FCI in different institutions at different levels of instruction provides a lot of data for comparison. Moreover, these comparisons are faciliated by the use of Hake gain (with some caution, as discussed in Chapter 6.1.2), which does not depend on the initial scores. It is worth noting that Hake gain is calculated in the PER literature only from average pre-post test scores: hence it does not reveal what dimensions of the force concept were hardest or easiest for the students. However, Hake gain appears to be a useful measure to analyse changes along different dimensions and representations as well as the total pre-post averages: the analysis of the FCI in terms of the dimensions and representations of the force concept can provide more detailed information, as shown in this study.

8.2 Evaluation of the Study Design and the Teaching Sequences

The study design made use of comparison or pilot groups, which were assumed to be approximately equivalent to the study groups. This assumption was tested by comparing the pre-FCI results of the study groups and pilot groups: the pre-FCI results were very close to each other and there were no statistically significant differences. It was not possible, however, to randomize the students into pilot and study groups since the groups were taught in separate years. This means that it is possible that the groups were not really equivalent despite the very similar pre-FCI results. In fact, it was thought almost from the beginning of the teaching that the Pre-IB study group might be somewhat weaker than the Pre-IB pilot group, on the basis of observing how well and how fast students managed the given tasks. Consequently, the pace was somewhat slower in the Pre-IB study group than in the Pre-IB pilot group[16]. This raises an interesting question: is it possible for a teacher to make reliable comparisons between groups just by observing them while teaching? Many experienced teachers would probably believe that it is. Some evidence to support the notion that the Pre-IB study group might be somewhat weaker that the Pre-IB pilot group was gained from the FCI results: the Pre-IB study group had a lower post-FCI results and smaller Hake gain. It should be noted, however, that these differences were not great enough to be statistically significant.

The students could not be randomized, but, many other factors were kept as constant as possible to allow comparisons between the groups: the pilot and study groups had the same teacher and the same textbook, they followed the same teaching approach, and they had generally similar exercises and activities. It should be noted, however, that there were some differences in the teaching sequences (compare Tables 5.2 and 5.3). The main difference concerns the force sequence, whereas the kinematics teaching sequence was virtually the same for both Pre-IB groups[17]. For the pilot groups, forces were briefly characterized as interactions along with Newton's third law but this did not really act as an entry point to the force concept, since they had no efficient tool to make

[16] It is interesting to note that despite the slower pace the Pre-IB group did not do better than the Pre-IB pilot group in most dimensions and representations of the force concept.

[17] This does not mean, however, that the kinematics sequences were *identical* since the staging of the same activities might have varied. This is because skilful physics teaching calls for careful listening to what students say and observing what they do, and then acting according to the feedback.

interactions visible. Instead, the concepts of acceleration and net force (Newton's second law) were taken as an effective starting point. Newton's first law was then identified as a special case of the second law. Newton's third law was addressed throughout the teaching: for instance, this was often done when constructing free-body diagrams.

The study groups started studying forces with Newton's third law and had a new representational tool - the SRI diagram or the Bridging Representation - at their disposal. This emphasis on forces as interactions with the study groups significantly improved students' contextual coherence in the third law: the study groups had better results than the pilot groups, with the difference being significant both in terms of statistical and practical significance. This implies that introducing forces as interaction by utilizing the bridging representation is very helpful in teaching Newton's third law. The results of the Finnish study group also strongly support this conclusion. Of course, the introduction and use of a new representation (the SRI diagram in this case) take some extra time, perhaps one hour altogether[18]. In view of the results, this may not be a bad investement.

On the other hand, focusing on forces as interactions did not produce conclusive differences in other dimensions than Newton's third law: the Finnish study group did generally better than the Finnish pilot group, whereas the Pre-IB pilot group did generally better than the Pre-IB study group in other dimensions than the third law. This suggests that the teaching order might not be as crucial a factor in learning as is the teaching approach and representational tools at students' disposal. The author's belief is, however, that the teaching sequence used with the study groups (see Chapters 5.2 and 5.3) is better than the one used with the pilot groups, since it has a clearer line of development. One could also argue that teaching forces as interactions might pave the way for those students who will continue studying physics at university, since the concept of interaction is more fundamental than the concept of force in modern field theories.

8.3 The ICI Approach

It is worth reflecting on possible reasons why the ICI approach and the teaching sequences designed were relatively successful in teaching the force concept. Firstly, the ICI had a strong conceptual focus via the "concepts first" -principle: students were given enough time to create meanings for the concepts studied before solving problem. This focus was emphasised by using exercises which were research based or at least informed by PER (e.g. Ranking Task Exercises). The second reason is related to this one: the teaching activities and exercises were designed to address the most common difficulties that students have with the force concept (e.g. force as an innate property of an object, impetus). This design was informed by the conceptual change research tradition. It is worth noting that conflict-based strategies were not explicitly used: instead, the teaching activities built on students' existing ideas and extended them to a new domain. For instance, the notion of interaction was developed using students' sensory experiences as a starting point (i.e., pressing their thumbs on different objects and realising that both the thumb and the object are deformed). Apparently, stressing the symmetric nature of interaction convinced the students that inanimate objects can also exert forces. This is by no means self-evident to students as demonstrated by Minstrell (1982). The impetus conception was also addressed by resorting to the notion of interaction: before being asked whether there is a force along the direction of motion of, say a cannonball, the students were asked to identify the interactions.

[18] Actually, there were no significant differences between the Finnish groups in terms of the total instructional time or topic coverage.

Thirdly, the students were given plenty of opportunities to talk and think about physical situations in multiple contexts and representations. Careful attention was devoted to creating both contextual and representational coherence: the students were asked to analyse different physical situations addressing the same physical principles using verbal, graphical, and diagrammatic representations. They were also encouraged to make explicit comparisons between the representations and to compare their answers with those of other students. In terms of the sociocultural view, peer discussions supported students' 'internalisation of the scientific story'. Furthermore, peer discussions help students to increase their metaconceptual awareness , i,e. to become aware of their existing ideas and also question them (Mason 1998; Vosniadou et al. 2001).

Fourthly, the author has been interested in improving his teaching through physics education research for ten years (Savinainen 1994, 1999, 2000a, 2000b, 2001a, 2001b). It is very likely that the author's effectiveness in engaging with students' thinking and staging the teaching sequences has improved in this process (Leach & Scott 2002). The skilful use of interactive-engagement methods is clearly a necessary but not sufficient condition for a significant improvement over traditional methods (Hake 1998b). As Reif (1995b) puts it, "even the best instructional materials and methods are useless if students do not actually engage in the recommended learning activities". Clearly, motivational factors play a role in conceptual change (Pintrich et al. 1993): students need to be cognitively engaged in order to have conceptual change in the first place. The teaching approach relied heavily on peer discussions as ways of exploring meanings, so it was vital that students recognised and accepted the value of peer discussions (Mazur 1997; Gunstone et al. 1999).

Student motivation was enhanced in the ICI approach through several means. Firstly, the aims of the teaching were carefully explained to the students, making use of a language metaphor: to know physics is to be able to 'speak' physics fluently using many alternative ways of expressing physical ideas (multiple representations) in many different situations (various contexts). Secondly, the students knew that they were supposed to demonstrate conceptual understanding in addition to problem solving skills in the exams. Thirdly, the students were further motivated during the courses when they noticed that they could correctly answer questions which had initially been designed for introductory physics courses at the university level (e.g. at Harvard University).

Might the ICI approach be useful for other teachers than the author? This question can be partially answered by pointing out that the components of ICI have been field tested in many institutes and by many teachers in the USA. On the other hand, this particular combination of the components forming the ICI approach has probably been used so far only by the present author. Experiences from another research-based teaching approach provide some insights here. The Modeling Method was very successful in the case of one expert teacher, but the first Modeling Workshop was not successful in enhancing teachers' effectiveness: the FCI scores of each teacher's class before and after the first workshop were almost identical (Wells et al. 1995). Encouragingly, the Modeling Workshops have subsequently been successful, especially in the case of teachers taking two full summer workshops instead of a single four-week workshop (Hestenes et al. 2000). However, the wider dissemination of the ICI approach might not be easy, since developing the necessary expertise in any innovative teaching approach takes a lot of time and effort.

8.4 Reflections on Conceptual Change

Different perspectives can be used in theorizing about students' conceptual change: for instance, Chi et al's (1994) ontological model of conceptual change offers one prominent perspective. Initially, students' force concept in this study belonged to the ontological category of 'matter', since they perceived forces to be an innate property of objects (e.g. the heavier an object is and the faster it moves, the more 'forceful' it is). Conceptual change occurred when the students reassigned the force concept to the ontological category of 'processes', i.e. when they started thinking of forces as properties of interactions between objects. This ontological shift was supported in teaching by developing the idea of interaction from the very beginning of the instruction. This change would be called 'reconceptualization' in Dykstra et al.'s (1992) framework: furthermore, it would most likely be qualified as a 'big' change in various descriptions of conceptual change (Tyson et al. 1997). Hence, teaching was very successful in bringing about conceptual change in this dimension of the force concept.

'Small' conceptual changes also took place in this study. Many students initially do not consider that 'at rest' is a special case of constant velocity: to see this requires class extension (Dykstra et al. 1992) or weak revision of existing conceptual structure (Tyson et al. 1997). The good progress made by students in applying Newton's first law in this study suggests that class extension or weak revision of their conceptual structure occurred. However, the teaching sequence was not very successful in bringing about conceptual change in all the dimensions of the force concept: only a relatively small percentage of the students achieved contextual coherence in Newton's second law. This can be interpreted as meaning that only 'peripheral conceptual change' (Chinn & Brewer 1993) took place in this respect. This would not be very surprising since the second law and also the acceleration concept requires a well-developed vector concept; it is worth noting that the pre-IB students had not had any instruction on vectors in their mathematics lessons. Moreover, the concept of acceleration can be difficult even for physics professors (Reif & Allen 1992).

In this study the change in students' contextual and representational coherence was used as a measure of conceptual change. These measures were applied at different stages of instruction. The analysis suggests that the development of the two types of coherence is a gradual process (see also Article V and Viiri & Savinainen (2004)). Hence, this result lends support to the view that conceptual change is evolutionary rather than revolutionary (see e.g. Vosniadou & Ioannides, 1998). It means that students' learning progresses from initial, scientifically more or less incorrect views via some intermediate or 'transitional states' to scientific views (Thornton 1995). The results also support Harrison et al.'s (1999) conclusion that students' conceptual change requires time and explicit attention to developing the concepts.

Despite the relative success of the ICI approach and the teaching sequences designed, it is clear that the force concept poses a great challenge for students. This is not surprising when one considers that there are at least six dimensions involved in the Newtonian force concept (Hestenes et al. 1992): its mastery demands expert-like conceptual coherence, which entails differentiation and integration of related concepts as well as the ability to apply all the relevant representational tools in a variety of contexts. It is also worth noting that the development of the force concept was a slow and difficult process, as discussed in Chapter 2. However, this and other studies in the field of PER show that it is possible to significantly foster students' conceptual change in this domain. Furthermore, the author can testify that becoming familiar with PER (and with science education in general) and doing PER also constitutes an efficient way to foster conceptual change in different domains of physics in the teacher as well.

References

Ainsworth, S.E. (1999). The functions of multiple representations. *Computer & Education*, **33**(2/3), 131-152.

Ainsworth, S.E., Bibby, P. and Wood, D. (1998). Analysing the costs and benefits of multiple representations learning environments. In van Someren et al. (Eds.) (1998).

Alexopoulou, E. and Driver. R. (1996). Small group discussions in physics: Peer interaction modes in pairs and fours. *Journal of Research in Science Teaching*, **33**, 1099-114.

Aristotle (350 B.C.). *Metaphysics*. An English translation by W.D. Ross is online at <http://classics.mit.edu/Aristotle/metaphysics.html>. Accessed on 22.4.2004.

Arons, A. (1997). *Teaching introductory physics*. USA: John Wiley & Sons.

Bao, L. (2001). Understanding Students' Representations in Learning Physics: Theories and Assessment Methods. *AAPT Announcer* **31** (2), 127. Online at <http://www.physics.ohio-state.edu/~lbao/talks.htm>. Accessed on 22.4.2004.

Bao, L. (2002). Explicit and implicit states of model mixing in conceptual development. Submitted to *PERS of American Journal of Physics*, in editorial process.
Online at <http://www.physics.ohio-state.edu/~lbao/publications.htm>. Accessed on 22.4.2004.

Bao, L. and Redish, E.F. (2001). Model Analysis: Assessing the Dynamics of Student Learning. Submitted to *Cognition and Instruction*. Online at <http://www.physics.ohio-state.edu/~lbao/papers.htm>. Accessed on 22.4.2004.

Bao, L. , Zollman, D., Hogg, K. and Redish, E.F. (2002). Model analysis of fine structures of student models: an example with Newton's third law. *PERS of American Journal of Physics*, **70**, 766-778.
Online at <http://www.physics.ohio-state.edu/~lbao/papers.htm>. Accessed on 22.4.2004.

Beichner, B. (1994). Testing Student Interpretation of Kinematics Graphs. *American Journal of Physics*, **62**, 750-762. Online at <http://www2.ncsu.edu/ncsu/pams/physics/Physics_Ed/Authors/Beichner.html>. Accessed on 22.4.2004.

Boeha, B. (1990). Aristotle, alive and well in Papua New Guinea science classrooms. *Physics Education*, **25**, 280-283.

Brown, D.E. (1989). Students’ concept of force: the importance of understanding Newton’s third law. *Physics Education*, **24**, 353- 358.

Brown, D.E. (1992). Using examples to remediate misconcpetions in physics: factors influencing conceptual change. *Journal of Research in Science Teaching*, **29**, 17-34.

Brown, D.E. and Clement, J. (1989). Overcoming misconceptions by analogical reasoning: abstract transfer versus explanatory model construction. *Instructional Science*, **18**, 237-261.

Camp, C. W. and Clement, J.J. (1994). *Preconceptions in mechanics. Lessons dealing with students’ conceptual difficulties*. USA: Kendall/Hunt Publishing Company.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Chi, M., Bassok, M., Lewis, M. Reimann, P. and Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science*, **13**, 145-182.

Chi, M., Feltovich, P. and Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science,* **5**, 121-152.

Chi, M., Slotta, J. and de Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, **4**, 27-43.

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

Chinn, C.A. and Brewer, W.F. (1998). An empirical taxonomy of responses to anomalous data in science, *Journal of Research in Science Teaching*, **35**, 623-654.

Clement, J. (1982). Students’ preconceptions in introductory mechanics. *American Journal of Physics*, **50**, 66-71.

Clement, J. (1987). Overcoming students’ misconceptions in physics: The role of anchoring intuitions and analogical validity. In J. D. Novak (Ed.). Proceedings of the Second International Seminar, Misconceptions and educational strategies in science and mathematics, Vol. III, Ithaca, New York: Cornell University.

Clough, E and Driver, E. (1986). A study of consistency in the use of students’ conceptual framework across different task contexts. *Science Education*, **70**(4), 473-496.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

Cohen, L., Manion, L. and Morrison, K. (2000). *Research Methods in Education*. Great Britain: Routledge Falmer.

Court, J.E. (1999). Free-body diagrams revisited—I. *The Physics Teacher*, **37**, 427-433.

Crouch, C.H. and Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American Journal of Physics*, **69**, 970 -977. Online at <http://mazur.deas.harvard.edu/publications.php>. Accessed on 22.4.2004.

de Jong, T., Ainsworth, S., Dobson, M., van der Hulst, A., Levonen, J., Reimann, P., Sime, J-A., van Someren, M., Spada, H. and Swaak, J. (1998). Acquiring knowledge in science and mathematics: the use of multiple representations in technology-based learning environments. In van Someren et al. (Eds.) (1998).

diSessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, **10**, 105-225.

diSessa, A. and Sherin, B.L. (1998). What changes in conceptual change? *International Journal of Science Education*, **20**, 1155-1191.

Dreyfus, A., Jungwirth, E. and Eliowitch, R. (1990). Applying the 'cognitive conflict' strategy for conceptual change - some implications, difficulties and problems. *Science Education*, **74**(5), 555-569.

Duit, R. (2004). Bibliography - STCSE Students' and teachers' conceptions and science education. Online at <http://www.ipn.uni-kiel.de/aktuell/stcse/stcse.html>. Accessed on 22.4.2004.

Duit, R. and Treagust, D.F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education*, **25**, 671-688.

Dykstra, D.I., Boyle, F.C. and Monarch, I.A. (1992). Studying conceptual change in learning physics. *Science Education*, **76**(6), 615-652.

Engelhardt, P. and Beichner, R. (2004). Students understanding of direct current resistive electrical circuits. *American Journal of Physics,* **72**, *98-115.*

Etkina, E. (1998). Physics on rollerblades. *The Physics Teacher*, **36**, 24-27.

Feynman, R.P. (1991). *"Surely You're Joking, Mr. Feynman! ": Adventures of a Curious Character by Richard P. Feynman as told to Ralph Leighton*. USA: Quality Paperback Book Club, by arrangement with W.W. Norton & Company, Inc., Edition copyright by Book-of-the-Month Club, Inc..

Finegold, M. and Gorsky, P. (1991). Students' concepts of force as applied to related systems: A search for consistency. *Inernational Journal of Science Education*, **13**(1), 97-113.

Finkelstein, N. (2001). Context in the context of physics and learning, *Proceedings of the 2001 Physics Education Research Conference*. Rochester, NY, 1-4. Online at <http://spot.colorado.edu/~finkelsn/research.html>. Accessed on 22.4.2004.

Franklin, A. (1978). Inertia in the Middle Ages. *The Physics Teacher*, **16**, 201-108.

Galili, I. and Tseitlin, M. (2003). Newton's first law: text, translations, interpretations and physics education. *Science & Education*, **12**, 45-73.

Gamow, G. (1988). *The great physicists from Galileo to Einstein*. New York: Dover Publications.

Gardner, H. (1991). *The unschooled mind*. New York: Basic Books.

Garson, G.D. (2004). *PA 765 Statnotes: An Online Textbook*. Online at <http://www2.chass.ncsu.edu/garson/pa765/statnote.htm>. Accessed on 14.6.2004.

Giancoli, D. (1998). *Physics - Principles with applications* (5th edition). USA: Prentice Hall International.

Gunstone, R., McKittrick, B. and Mullhall, P. (1999). Structured cognitive discussions in senior high school physics: Student and teacher perceptions. *Research in Science Education*, **29**(4), 527-546.

Gunstone, R. and Watts, M. (1985). Force and motion, In R. Driver, E. Guesne, and A. Tiberghien (Eds.). *Children's ideas in science*, Open University Press, Milton Keynes, 85-104.

Hake, R. (1998a). Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, **66**, 64- 74. Online at <http://www.physics.indiana.edu/~sdi/welcome.html#z44 >. Accessed on 22.4.2004.

Hake, R. (1998b). Interactive-engagement methods in introductory mechanics courses. Online at <http://www.physics.indiana.edu/~sdi/welcome.html#z44 >. Accessed on 22.4.2004.

Hake, R. (2001). Suggestions for Administering and Reporting Pre/Post Diagnostic Tests, 5/18/01. Online at <http://www.physics.indiana.edu/~hake>. Accessed on 22.4.2004.

Hake, R. (2002). Lessons from the physics education reform effort. *Conservation Ecology* 5(2), 28. Online at <http://www.consecol.org/vol5/iss2/art28>. Accessed on 22.4.2004.

Halliday, D., Resnick, R. and Walker, J. (2001). *Fundamentals of Physics* (6th edition). USA: John Wiley & Sons.

Halloun, I., Hake, R., Mosca, E. and Hestenes, D. (1995). Force Concept Inventory (Revised 1995). Password protected at <http://modeling.la.asu.edu/modeling.html>. Accessed on 22.4.2004.

Halloun, I. and Hestenes, D. (1985). Common sense concepts about motion. *American Journal of Physics*, **53**, 1056-1065. Online at < http://modeling.la.asu.edu/R&E/Research.html>. Accessed on 22.4.2004.

Halloun, I. and Hestenes, D. (1996). The search for conceptual coherence in FCI data, Working paper. Online at < http://modeling.la.asu.edu/R&E/Research.html>. Accessed on 22.4.2004.

Harrison, A., Grayson, D. and Treagust, D. (1999). Investigating a Grade 11 Students' Evolving Conceptions of Heat and Temperature. *Journal of Research in Science Teaching*, **36**, 55-87.

Heller, P. and Hollabaugh, M. (1992). Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics*, **60**(7), 637-644.

Hellingman, C. (1992). Newton's third law revisited. *Physics Education*, **27**, 112-115.

Henderson, C. (2002). Common Concerns About the Force Concept Inventory. *The Physics Teacher*, **40**, 542-547.

Hestenes, D. (1996). Modeling method for physics teachers. *Proceedings of the International Conference on Undergraduate Physics Education (College Park).* Online at <http://modeling.la.asu.edu/modeling.html>. Accessed on 22.4.2004.

Hestenes, D. (1998). *Foundations of mechanics* (Originally published as Chapter 9 in the first edition of *New Foundations for Classical Mechanics*, 1986). Online at <http://modelingnts.la.asu.edu/pdf/Foundations.pdf>. Accessed on 22.4.2004.

Hestenes, D. and Halloun, I. (1995). Interpreting the FCI. *The Physics Teacher*, **33**, 502-506. Online at < http://modeling.la.asu.edu/R&E/Research.html>. Accessed on 22.4.2004.

Hestenes, D., Wells, M. and Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher*, **30**, 141-158. Online at < http://modeling.la.asu.edu/R&E/Research.html>. Accessed on 22.4.2004.

Hestenes, D. and Wells, M. (1992). A Mechanics Baseline Test. *The Physics Teacher*, **30**, 159-166. Online at < http://modeling.la.asu.edu/R&E/Research.html>. Accessed on 22.4.2004.

Hestenes, D. et al. (2000). Findings of the Modeling Workshop Project (1994-00). Online at <http://modeling.la.asu.edu/R&E/Research.html>. Accessed on 22.4.2004.

Hewson, P.W. and Hewson, M.G. (1992). The status of students' conceptions. In R. Duit, F. Goldberg, and H. Niedderer (Eds.). *Research in physics learning: Theoretical issues and empirical studies*, IPN, Kiel, Germany, 59-73.

Hewson, P.W. and Thorley, N.R. (1989). The conditions of conceptual change in a classroom [special issue]. *International Journal of Science Education*, **11**, 541-553.

Huffman, D. and Heller, P. (1995). What does the Force Concept Inventory Actually Measure? *The Physics Teacher*, **33**, 138-143.

Jammer, M. (1997). *Concepts of mass in classical and modern physics*. Mineola, New York: Dover Publications.

Jammer, M. (1999). *Concepts of force*. Mineola, New York: Dover Publications.

Jauhiainen, J., Koponen, I. and Lavonen, J. (2001). The Force Concept Inventory in diagnosing the conceptual understanding of Newtonian mechanics in Finnish upper secondary schools. In M. Ahtee, O. Björkvist, E. Pehkonen and V. Vatanen (Eds.): *Research on mathematics and science education - From beliefs to cognition, from problem solving to understanding*. Institute for Educational Research, University of Jyväskylä, 101-114.

Jauhiainen, J. (2003). Personal correspondences through electronic mail on 25th and 26th June.

Jiménez, J.D. and Perales, F.J. (2001). Graphic representation of force in secondary education: analysis and alternative educational proposals. *Physics Education*, **36**, 227 - 235.

Johnson, R. (1996). *Elementary statistics* (6th edition). USA: Duxbury Press.

Jones, M.G. and Carter, G. (1998). Small groups and shared constructions. In J. Mintzes, J. Wandersee and J. Novak (Eds*.). Teaching science for understanding - A human constructivist view*. Academic Press, Cornell University, Ithaca, New York, 261-279.

Key, J.P. (1997). Research Design in Occupational Education: Module S7 – Chi Square, Oklahoma State University, USA. Online at <http://www.okstate.edu/ag/agedcm4h/academic/aged5980a/5980/newpage28.htm>. Accessed on 22.4.2004.

Koponen, I., Jauhiainen, J. and Lavonen, J. (2000). A Finnish translation of the 1995 version of the Force Concept Inventory, available upon request. Department of Physics, University of Helsinki.

Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction*, **13**, 205-226.

Kurki-Suonio, K. (2000). Modernin fysiikan perushahmojen synty (The birth of the primary perceptions in modern physics). *Dimensio* 2, 32-36. Online at <http://didactical.physics.helsinki.fi/didfys/artikkeleita.htm>. Accessed on 22.4.2004.

Kurki-Suonio, K. and Kurki-Suonio, R. (1994). The concept of force in the perpectional approach. In H. Silfverberg and K. Seinelä (ed.) *Ainedidaktiikan teorian ja käytännön kohtaaminen* (*Theory and practice in didactics*). Reports from the Department of Teacher Education in Tampere. University of Tampere. Online at <http://didactical.physics.helsinki.fi/tutkimus.htm>. Accessed on 22.4.2004.

Kurki-Suonio, K. and Kurki-Suonio, R. (1997). *Vuorovaikuttavat kappaleet - mekaniikan perusteet* (*Interacting objects - foundations of mechanics*). Helsinki: Limes ry.

Larkin, J. and Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words? *Cognitive Science*, **11**, 65-99.

Leach, J. and Scott, P. (2002). Designing and evaluating science teaching sequences: An approach drawing upon the concept of learning demand and a social constructive perspective on learning. *Studies in Science Education*, **38**, 115-142.

Leach, J. and Scott, P. (2003). Individual and sociocultural views of learning in science education. *Science & Education*, **12**, 91-113.

Lehti, R. (1987). Nopeuden käsite ja sen laskutekniset probleemit aristoteelisessa fysiikassa (The concept of velocity and its calculation problems in Aristotelian physics). *Arkhimedes*, **37**, 94 - 112.

Lehti, R. (1996). *Tähtiä ja ihmisiä (Stars and men)*. Jyväskylä: URSA.

Lehto, H. and Luoma, T. (1999). *Fysiikka 3. Mekaniikka, lämpö ja energia (Physics 3. Mechanics and thermal Physics)*. Jyväskylä: Kirjayhtymä.

Limon, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction*, **11**, 357-380.

Mann, P.S. (1995). *Introductory statistics* (2nd edition). USA: John Wiley & Sons.

Mason, L. (1998). Sharing cognition to construct scientific knowledge in school context: The role of oral and written discourse. *Instructional Science*, **26**, 359-389.

Mazur, E. (1997). *Peer Instruction: A User's Manual*.USA: Prentice Hall.
'Professionals' may obtain free copies of the book (a) from Prentice Hall campus representatives or (b) by downloading the Adobe Acrobat portable document file at <http://galileo.harvard.edu/> Accessed on 22.4.2004.

Mazur, E. (1998). Moving the mountain: impediments to change. *Proceedings of Third Annual NISE Forum Washington DC*, 91-93. Online at <http://mazur.deas.harvard.edu:16080/library.php>. Accessed on 22.4.2004.

McDermott, L. (1993). Guest Comment: How we teach and how students learn – A mismatch? *American Journal of Physics,* **61**(4), 295-298.

McDermott, L.C. and Redish, E.F. (1999). Resource Letter: PER-1: Physics Education Research. *American Journal of Physics,* **67**(9), 755-767. Online at <http://www.phys.washington.edu/groups/peg/pubs.html>. Accessed on 22.4.2004.

McDermott, L.C., Rosenquist, M. and van Zee, E. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics,* **55**, 503-13.

McDermott, L.C. , Shaffer, P. and the PERG at the University of Washington (1998). *Tutorials for Introductory Physics - Workbook, Preliminary Edition*. USA: Prentice Hall.

McLean, J.E. and Ernest, J.M. (1998). The Role of Statistical Significance Testing in Educational Research. *Research in the Schools,* 5, 15-22. Online at <http://www.personal.psu.edu/users/d/m/dmr/sigtest/sigtest.htm>. Accessed on 22.4.2004.

Meltzer, D. (2002a). Student learning of physics concepts: Efficacy of verbal and written forms of expression in comparison to other representational modes. Paper presented at the *Conference on Ontological, Epistemological, Linguistic and Pedagogical Considerations of, Language and Science Literacy: Empowering Research and Informing Instruction*, University of Victoria, Victoria, British Columbia, Canada. Online at <http://www.physics.iastate.edu/per/articles/>. Accessed on 22.4.2004.

Meltzer, D. (2002b). The relationship between mathematics preparation and conceptual learning gains in physics: a possible "hidden variable" in diagnostic pretest scores. *American Journal of Physics*, **70**, 1259-1268. Online at < http://www.physics.iastate.edu/per/articles/>. Accessed on 22.4.2004.

Meltzer, D. (2002c). Issues related to data analysis and quantitative methods in PER, In S. Franklin, K. Cummings, and J. Marx (Eds.): *Proceedings of the 2002 Physics Education Research Conference [Boise, Idaho, August 7-8, 2002],* New York, PERC Publishing, 21-24. Online at <http://www.physics.iastate.edu/per/articles/>. Accessed on 22.4.2004.

Meltzer, D. (2003). Relation between students' problem-solving performance and representational mode, submitted for publication. Online at <http://www.physics.iastate.edu/per/articles/>. Accessed on 22.4.2004.

Meltzer, D., Manivannan, K. and Tassara, T. (1997). Effectiveness of a microcomputer-based laboratory environment for teaching force and motion in an elementary physics course. *AAPT* Announcer, **27,** 122.

Menigaux, J. (1994). Students' reasoning in solid mechanics. *Physics Education*, 29, 242-246.

Mildenhall, P. and Williams, J. (2001). Instability in students' use of intuitive and Newtonian models to predict motion: the critical effect of parameters involved. *International Journal of Science Education*, **23**, 643-660.

Mills, D., McKittrick, B., Mulhall, P., and Feteris, S. (1999). CUP: cooperative learning that works. *Physics Education*, **34**, 1-5. Examples of CUP's are online at <http://www.education.monash.edu.au/centres/scienceMTE/teachingLearningPhysics.html/>. Accessed on 22.4.2004.

Minstrell, J. (2003). What are facets? Online at <http://www.talariainc.com/facet/pWhatAre.htm>. Accessed on 22.4.2004.

Mislevy, R.J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, **11**(1), 3-31.

Newton, I. (1962). *Principia*. California: University Press. (Original in 1686).

Nix, T.W. and Barnette, J.J. (1998). The Data Analysis Dilemma: Ban or Abandon. A Review of Null Hypothesis Significance Testing. *Research in the Schools*, **5**, 3-14. Online at <http://www.personal.psu.edu/users/d/m/dmr/sigtest/sigtest.htm>. Accessed on 22.4.2004.

O'Kuma, T., Maloney, D. and Hieggelke, C. (2000). *Ranking task exercises in physics*. USA: Prentice Hall.

Palmer, D. (1997). The effect of context on students' reasoning about forces. *International Journal of Science Education*, **6**, 681-696.

Physics Textbook Review Committee (1998). Quibbles, misunderstandings, and egregious mistakes. *The Physics Teacher*, **37**, 297-305.

Pintrich, P.R., Marx, R.W. and 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**, 167-200.

Posner, G., Strike, K., Hewson, P., and Gerzog, W. (1982). Accomodation of a scientific conception: toward a theory of conceptual change. *Science Education*, **63**(2), 167-199.

Ranta, E., Rita, H. and Kouki, J. (1999). *Biometria – tilastotiedettä ekologeille (Biometry – statistics for ecologists).* Helsinki: Yliopistopaino.

Rebello, N.S and Zollman, D.A. (2004). The effect of distracters on student performance on the force concept inventory. *American Journal of Physics*, **72**, 116-125. Online at <http://www.phys.ksu.edu/~srebello/mid-tenure/reprints/reprints.htm>. Accessed on 15.6.2004.

Redish, E.F. (1994). The implications of cognitive studies for teaching physics. *American Journal of Physics*, **62**(6), 796-803.

Redish, E.F. (2003). *Teaching physics with the physics suite*. USA: John Wiley & Sons.

Reif, F. (1987). Instructional design, cognition, and technology: Application to the teaching of scientific concepts. *Journal of Research in Science Teaching*, **24**, 309-324.

Reif, F. (1995a). *Understanding basic mechanics - Workbook*. New York: John Wiley & Sons.

Reif, F. (1995b). Millikan Lecture 1994: Understanding and teaching important scientific thought processes. *American Journal of Physics*, **63**, 17-32.

Reif, F. and Allen, S. (1992). Cognition for interpreting scientific concepts: A study of acceleration. *Cognition and Instruction*, **9**(1), 1-44.

Reif, F. and Heller, J. (1982). Knowledge structure and problem solving in physics. *Educational Psychologist*, **17**, 102-127.

Robinson, D.G. and Levin, J.R. (1997). Reflections on statistical and substantive significance, with a slice of replication. *Educational Researcher*, **26**(5), 21-26.

Rosenquist, M. and McDermott, L. (1987). A conceptual approach to teaching kinematics. *American Journal of Physics*, **55**, 407-415.

Roth, W.-M., McRobbie, C. and Boutonné, S. (1997). The local production of order in traditional science laboratories: a phenomenological analysis. *Learning and Instruction*, **7**, 107-136.

Sabella, M. (1999). Using the context of physics problem solving to evaluate the coherence of student knowledge. Dissertation, University of Maryland.
Online at <http://webs.csu.edu/~msabella/dissertation/>. Accessed on 23.4.2004.

Savinainen, A. (1994). *Käsityksiä voimasta - Konstruktivistisen opetuksen vaikutus 7. luokkalaisten voimakäsitysten muutokseen*. (*Conceptions of the force concept - Constructivist teaching and conceptual change in the 7th grade*). MSc Thesis in Education. University of Joensuu 1994. Published in Finnish.

Savinainen, A. (1999). Laatufysiikkaa - Kuinka opettaa opiskelijoita ymmärtämään fysiikkaa? (Quality physics - How to teach students to understand physics?). *Dimensio* 4, 40-44. Published in Finnish. Online at <http://kotisivu.mtv3.fi/physics/>. Accessed on 22.4.2004.

Savinainen, A. (2000a). Muistiinpanot fysiikassa - tehokasta opiskelua vai ajanhukkaa? (Notetaking in physics - Effective learning or a waste of time?). *Dimensio* 3, 16-18. Published in Finnish. Online at <http://kotisivu.mtv3.fi/physics/>. Accesed on 22.4.2004.

Savinainen, A. (2000b). *Sovellettu Peer-menetelma lämpöopin käsitteiden opetuksessa*, (*Applied peer method in teaching the concepts of thermal physics*). Licenciate Thesis, University of Joensuu, Department of Physics. Published in Finnish.

Savinainen, A. (2001a). An evaluation of interactive teaching methods in mechanics: Using the Force Concept Inventory to monitor student learning. In M. Ahtee, O. Björkqvist, E. .Pehkonen and V. Vatanen (Eds.). *Research on mathematics and science education - From beliefs to cognition, from problem solving to understanding*, Institute for Educational Research. University of Jyväskylä, 115-132. Online at <http://kotisivu.mtv3.fi/physics/>. Accessed on 22.4.2004.

Savinainen, A. (2001b). Peer-menetelmä fysiikan käsitteiden opetuksessa, (Collaborative peer instruction in teaching the concepts of physics). *Dimensio* 1, 28-32. Published in Finnish. Online at <http://kotisivu.mtv3.fi/physics/>. Accessed on 22.4.2004.

Schecker, H. and Gerdes, J. (1999). Messung von Konzeptualisierungsfähigkeit in der Mechanik: Zur Aussagekraft des FCI. *Zeitschrift für Didaktik der Naturwissenschaften*, **5**(1), 75-89.

Scherr, R. and Wittmann, M. (2002). The challenge of listening: The effect of researcher agenda on data collection and interpretation, In S. Franklin, K. Cummings and J. Marx (Eds.): *Proceedings of the 2002 Physics Education Research Conference [Boise, Idaho, August 7-8, 2002]*, New York, PERC Publishing. Online at <http://perlnet.umephy.maine.edu/research/papers.htm>. Accessed on 22.4.2004.

Scott, P. (1998). Teacher talk and meaning making in science classrooms: a review of studies from a Vygotskian perspective. *Studies in Science Education*, **32**, 45-80.

Scott, P., Asoko, H.M. and Driver, R.H. (1991). Teaching strategies for conceptual change: A review of strategies. In R. Duit, F. Goldberg and H. Niedderer (Eds.). *Research in physics learning: Theoretical issues and empirical studies*, IPN, Kiel, Germany, 310-329.

Scott, P. and Jewitt, C. (2003). Talk, action and visual communication in teaching and learning science. *School Science Review*, **84**(308), 117- 124.

Sequira, M. and Leite, L. (1991a). Alternative conceptions and history of science in physics teacher education. *Science Education*, **75**, 45-56.

Sequira, M. and Leite, L. (1991b). Alternative ideas in mechanics: where they come from? In M. Ahtee, V. Meisalo and H. Saarikko (Eds.): *Proceedings of the international symposium on the evaluation of physics education*, Department of Teacher Education, University of Helsinki. Research Report 96, 187 - 201.

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, **13**, 227-237.

Schnotz, W., Vosniadou, S. and Carretero, M. (Eds.) (1999). *New Perspectives on Conceptual Change*. Amsterdam: Pergamon

Shoultz, J., Säljö, R. and Wyndham, J. (2001). Heavenly talk: discourse, artifacts, and children's understanding of elementary astronomy. *Human Development*, **44**, 103-118.

Singh, C. and Rosengrant, D. (2003). Multiple-choice test of energy and momentum concepts. *PERS of American Journal of Physics,* **71**(6), 607-617.

Spielberg, N. and Anderson, N. (1995). *Seven ideas that shook the universe.* New York: John Wiley & Sons.

Steinberg, M., Brown, D. and Clement, J. (1990). Genius is not immune to persistent misconceptions: conceptual difficulties impeding Isaac Newton and contemporary physics students. *International Journal of Science Education*, **3**, 265-273.

Steinberg, R. and Sabella, M. (1997). Performance on multiple-choice diagnostics and complementary exam problems. *The Physics Teacher*, **35**, 150-155.

Swartz, C.E. and Miner, T.D. (1998). *Teaching introductory physics - a sourcebook*. USA: AIP Press, Springler-Verlag.

Taber, K.S. (2000). Multiple frameworks? Evidence of manifold conceptions in individual cognitive structure. *International Journal of Science Education*, **22**, 399-417.

Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.

Thijs, G.D. and Kuiper, J. (1990). Use of intuitive models of force among secondary school students as found in a cross-cultural study. In N. Bleichrodt and P.J.D. Drenth (Eds.): *Contemporary issues in cross-cultural psychology*, Amsterdam, Swets & Zeitlinger, 309-321.

Thornton, R.K. (1995). Conceptual dynamics: Changing students views of force and motion. In C. Tarsitani, C. Bernandini and M. Vincentini (Eds.). *Thinking Physics for Teaching*., London: Plenum, 157-183.

Thornton, R.K. and Sokoloff, D. (1990). Learning motion concepts using real-time microcomputer-based laboratory tools. *American Journal of Physics*, **58**(9), 858-66.

Thornton, R.K. and Sokoloff, D.R. (1998). Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics*, **66**, 338-351.

Trowbridge, D. and McDermott, L. (1980). Investigation of student understanding of the concept of velocity in one dimension. *American Journal of Physics,* **48**, 1020-28.

Trowbridge, D. and McDermott, L. (1981). Investigation of student understanding of the concept of acceleration in one dimension. *American Journal of Physics*, **49**, 242-53.

Treagust, D.F., Harrison, A.G. and Venville, G.J. (1996). Using analogical teaching approach to engender conceptual change. *International Journal of Science Education,* **18**, 213-229.

Tynjälä, P. (1999). *Oppiminen tiedon rakentamisena. Konstruktivistisen oppimiskäsityksen perusteita (Learning as construction of knowledge. Introduction to the constructivist view of learning).* Tampere: Kirjayhtymä Oy.

Tyson, L., Venville, G., Harrison, A. and Treagust, D. (1997). A multidimensional framework for interpreting conceptual change events in the classroom. *Science Education*, **81**, 387-404.

Van Heuvelen, A. (1991). Overview, Case Study Physics. *American Journal of Physics*, **59**(10), 898-907.

Van Heuvelen, A. and Zou, X. (2001). Multiple representations of work-energy processes. *American Journal of Physics*, **69**(2), 184- 194.

van Someren, M.W., Reimann, P., Boshuisen, H.P.A.and de Jong, T. (Eds.) (1998). *Learning with multiple representations*. Oxford: Elsevier.

Viiri, J. (1992). *Aristoteleen liikeoppia (Aristotle's theory of motion)*. University of Joensuus, Department of Physics, Selected Papers 2.

Viiri, J. (1995). *Voimakäsitteen opettaminen ja oppiminen insinöörikoulutuksen fysiikan kurssissa* (*Teaching and learning of the force concept in a course of physics for engineering students*). University of Joensuu, Joensuu.

Viiri, J. and Savinainen, A. (2004). The development of students' representational coherence. In S. Vosniadou, C. Stathopoulou, X. Vamvakoussi and N. Mamalougos (Eds.): *Abstracts, 4th European Symposium Conceptual Change: Philosophical, Historical, Psychological, and Educational Approaches,* European Association for Research on Learning and Instruction, May 19-23, Delphi, Greece, pp. 123-126. Online at <http://kotisivu.mtv3.fi/physics/>. Accessed on 22.4.2004.

Vosniadou, S. (1992). Knowledge acquisition and conceptual change. *Applied Psychology: An International Review*, **41**, 347-357.

Vosniadou, S. (1994). Capturing and modelling the process of conceptual change [special issue], *Learning and Instruction*, **4**, 45-69.

Vosniadou, S. and Ioannides, C. (1998). From conceptual development to science education: a psychological point of view. *International Journal of Science Education*, **20**, 1213-1230.

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A. and Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, **11**, 381-419.

Wells, M., Hestenes, D. and Swackhamer, G. (1995). A Modeling Method for high school physics teachers. *American Journal of Physics*, **63**, 606-619. Online at <http://modeling.la.asu.edu/modeling-HS.html>. Accessed on 23.4.2004.

Westfall, R. (1977). *The construction of modern science. Mechanism and mechanics*. Cambridge: Cambridge University Press.

White, R. and Gunstone, R. (1992). *Probing understanding*. London: Falmer Press.

Corrections (*in italics*) to Antti Savinainen's Printed Dissertation

Abstract, the first page, the last paragraph:
- "In all these cases Hake gains were..." to "In *almost* all these cases Hake gains were...".

Abstract, the second page, the first paragraph and page 89, the third paragraph:
- "In general, the most notable... in Newton's first and third laws in verbal representation." to "In general, the most notable…in Newton's first *law (all groups) and Newton's third law* (*the studygroups*) in verbal representation."

Motto page: The second citation mark should be at the end of the Feynman citation.

Page 3, the fourth paragraph:
- "Chapter 6.1.1 further elaborates…" to "Chapter *6.1.2* further elaborates…"
- "…are discussed in Chapter 6.1.2:…" to "…are discussed in Chapter *6.1.1*:…"

Page 10, the third paragraph:
- "…(Jammer 1997, 90-97)" to "…(Jammer 1997, *91*-97)"
- "…the particles after interacting with…" to "…the particles *while* interacting with…"
- "…(m_B and m_B) can be called the 'masses' of the particles A and B." to "…(m_B and m_C) can be called the 'masses' of the particles *B* and *C*."

Page 13, the sixth paragraph:
- "…at the university level also present the first law with no reference…" to "….at the university level also present the first law *initially* with no reference…"

Page 32, Table 3.2, the first row:
- "Sledge" to "*Sled*"

Page 44, the first paragraph:
- "Chapters 5.2 outlines the teaching sequences…" to "Chapters *5.2 and 5.3 outline* the teaching sequences…"

Page 47:
- Table 5.1.: The frame of the table should be *bolded*.
- Remark 9: "The presented teaching sequence presented,…" to "The presented teaching sequence ~~*presented*~~,…"

Page 48, the third paragraph and page 49, the second paragraph:
- "modified SRI diagram…" to "~~*modified*~~ SRI diagram…"

Page 74, the third paragraph:
- "…more students in the pilot group mastered" to "…more students in the *study* group mastered"

Article V, page 9: "BOOK" in the Figure 2 to "*BLOCK*"

Article V, References:
- p. 21: "Ainsworth S.E" to "Ainsworth S.E."
- p. 24: "Meltzer, D. (2002)….Cummings, C. " to "Meltzer, D. (2002)….Cummings, *K*. "
- p. 26: "Vosniadou, S. & Ioannides, C. (1998). ….Internal Journal of Science Ecucation…" to "Vosniadou, S. & Ioannides, C. (1998). ….*International* Journal of Science Education…"