THESIS FOR DEGREE OF LICENTIATE OF PHILOSOPHY

Mathematical Reasoning in Physics Tests

Requirements, Relations, Dependence

Helena Johansson

CHALMERS | GÖTEBORGS UNIVERSITET

Department of Mathemtaical Sciences
Division of Mathematics
Chalmers University of technology and University of Gothenburg
Gothenburg, Sweden 2013

Mathematical Reasoning in Physics Tests
Requirements, Relations, Dependence

Helena Johansson
ISSN 1652-9715 NO 2013:18

Department of Mathematical Sciences
Division of Mathematics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Sweden
Telephone +46 (0)31 772 1000

Printed in Gothenburg, Sweden 2013

Abstract

By analysing and expanding upon mathematical reasoning requirements in physics tests, this licentiate thesis aims to contribute to the research studying how students' knowledge in mathematics influence their learning of physics. A sample of physics tests from the Swedish National Test Bank in Physics was used as data, together with information of upper secondary students' scores and grades on the tests. First it was decided whether the tasks in the tests required mathematical reasoning at all and if they did, that reasoning was characterised. Further, the relation between students' grades and mathematical reasoning requirements was examined. Another aim in this thesis is to try out if the Mantel-Haenszel procedure is an appropriate statistical method to answer questions about if there is a dependence between students' success on different physics tasks requiring different kinds of mathematical reasoning. The results show that 75% of the tasks in the physics tests require mathematical reasoning and that it is impossible to pass six out of eight tests without mathematical reasoning. It is also revealed that it is uncommon that a student gets a higher grade than Pass without solving tasks that require the student to come up with not already familiar solutions. It is concluded that the Mantel-Haenszel procedure is sensitive to the number of students each teacher accounts for. If there are not too few students, the procedure can be used. The outcome indicates that there is a dependence between success on tasks requiring different kinds of reasoning. It is more likely that a student manages to solve a task that requires the produce of new reasoning if the student has solved a task that is familiar from before.

Keywords: Mathematical reasoning, imitative reasoning, creative mathematical reasoning, physics tests, physics tasks, upper secondary school, Mantel-Haenszel procedure**.**

Acknowledgements

This licentiate thesis would not have been if it wasn't for my excellent advisors Mats Andersson and Jesper Boesen. Many thanks to both of you! And thank you Ann-Marie Pendrill for finding and organising this perfect combination of advisors in the beginning of my PhD studies. Thank you, Kerstin Wiklander for explaining and discussing the Mantel-Haenszel procedure and for your help with the statistical analysis. Further, I would like to thank Alan Schoenfeld and his functions group, as well as Andrea diSessa at Berkeley, University of California, for fruitful discussions and advise about my studies. My time as a visiting student researcher at UCB was partly funded by a scholarship from The Royal Swedish Academy of Science, thank you for this financial support. I would also like to thank my collegues in FLUM for positive support and fun times. Last but not least, thank you mum and dad for always supporting me, no matter what. To my beloved children; well, now you can get an idea of what I am been occupied with the last couple of years ;)

Contents

1 Introduction

There is a natural relation between the school subjects mathematics and physics, reflected both in mathematics education research and in physics education research. Independent of the two fields of research, mainly two strands about this relation could be found in the literature. The first strand has its focus on how the relation between the subjects is being manifested both in the curricula and in practice. The other one is more concerned with how knowledge in one of the subjects may contribute to the understanding of the other subject. Some of the discussions depart from the learning of mathematics and how the relation to physics could influence this learning, below called *physics in mathematics*. Other studies take a starting point in the learning of physics and discuss various aspects of the relation to mathematics, referred to below as *mathematics in physics*. Mathematics is used as the language of physics and knowledge in mathematics is a natural prerequisite for learning physics. It follows that some of the difficulties students encounter when learning physics likely relate to their ability to use mathematics, an experience shared by many physics teachers around the world.

By studying what kind of and to what extent mathematical reasoning is required to solve tasks in physics tests, the aim of this thesis is to contribute to the understanding of students' success, or lack of success, in learning physics. The tests chosen to be analysed come from the Swedish National Testbank in Physics. It is further examined if it is possible to say something about the relation between how students manage to solve different kinds of tasks, different with respect to requirements of mathematical reasoning. This licentiate thesis consists of two papers. The first paper describes a qualitative analysis of the mathematical reasoning requirements. The second paper is divided in two parts; whereas the first part analyse the relation between students' grades and which tasks they have solved with respect to reasoning category and the second part study if there is a relation between how students succeed on tasks requiring different kind of mathematical reasoning. By taking a starting point in physics and look at the mathematics required, this thesis best belongs to *mathematics in physics* research.

2 Background

2.1 Physics in Mathematics

Blum and Niss (1991) noticed already in the 80's that the relation between the two subjects had become weakened in the mathematics education. The main reason is that new areas have developed, in which mathematics is important, and these areas can provide examples suitable for mathematical instructions instead of examples from physics. They agree on the necessity of the opening of mathematics instruction to other applicational areas, but at the same time they stress that it is of great value to keep a close contact between mathematics and physics in school. Examples from physics provide good representative cases for validating mathematical models. They discuss how a separation between the two subjects can lead to unnatural distances between the mathematical models and the real situation intended to model. The weakened relation between the school subjects mathematics and physics is also observed by Michelsen (1998), who describes how the separation of the two subjects has evolved in the Danish school. Doorman & Gravemeijer (2008) discuss the advantage of learning mathematical concepts through mathematical model building and how examples from physics are beneficial to symbolize the concepts. Hanna (2000) and Hanna and

Jahnke (2002) propose that it is advantageous to use arguments from physics in mathematical proofs to make them more explanatory. They refer to Polya (1954) and Winter (1978) and continue discussing the benefits of integrating physics in mathematics education while learning and dealing with mathematical proofs. The importance of using physics to facilitate students' learning of various mathematical concepts is also discussed by Marongelle (2004). Using events from physics can help students to understand different mathematical representations.

2.2 Mathematics in Physics

Tasar (2010) discusses how a closer relation between the school subjects, mathematics and physics, can contribute to the understanding of physics concepts. A closer relation might also prevent the assumption that students already understand the mathematical concepts needed in physics (ibid.). A closer relation, noticed by Basson (2002), might also decrease the amount of time physics teachers spend on redoing the mathematics students need in physics. The "redoing" is likely a consequence of e.g. that *"physics teachers claim that their students do not have the pre-requisite calculus knowledge to help them master physics"* (Cui, 2006, p.2). Michelsen (2005) discusses how interdisciplinary modelling activities can help students to understand how to use mathematics in physics and to see the links between the two subjects. Redish and Gupta (2009) emphasize the need to understand how mathematics is used in physics and also the cognitive components of expertise, in order to teach mathematics for physics more effectively to students. Basson (2002) mentions how learning problems in physics not only depends on the complexity of the subject, but also on improper mathematical knowledge. Bing (2008) discusses the importance of learning the language of mathematics when studying physics. Nguyen and Meltzer (2003) analyse students' knowledge of vectors and conclude that there is a gap between students' intuitive knowledge and how to apply their knowledge in a formal way, which can be an obstacle when learning physics.

A weaker relation between the subjects, mathematics and physics, in school is also observed more recently in Sweden in the Timms Advance 2008 report. A comparison between syllabuses for physics from different years revealed that the importance of mathematics in physics was more prominent ten years ago, than it is nowadays (Swedish National Agency for Education, 2009).

In a survey, Tuminaro (2002) classifies studies concerning students' use of mathematics in four categories according to their different approaches. *The observational approach* focuses on what students do when applying mathematics to physics problems and how they reason mathematically. Often there are no attempts to give any instructional implications. *The modelling approach* intends to describe the differences between experts and novices regarding their problem solving skills as well as to develop computer programs that can model the performance of the novices and the experts. Using results from these programs, one hopes to understand the learning process. The third approach is called *the mathematics knowledge structure approach*. Research placed in this category aims to explain the use of mathematics from cognitive structures of novices and experts. *General knowledge structure approach* is the last category, which includes research oriented towards an understanding of concepts in general (not only mathematical), using various kinds of cognitive structures. Tuminaro sees a hierarchical structure in the four approaches and compares this structure with a trend in cognitive psychology, towards a refined understanding of cognition. According to Tuminaro, the four approaches towards an understanding of how students use mathematics in physics do not reach the fully sophisticated level, as the trend in cognitive psychology. Tuminaro therefore suggests there still is a need for research about how the structure of students' knowledge coordinates when they draw conclusions about physics from mathematics. Since a part of the presented thesis focuses on what kind of mathematical reasoning that are required of students and not on students' use of mathematics, the thesis can be regarded as a complement to the studies categorised by Tuminaro.

2.3 Service subject

In some studies concerning the relation between mathematics and physics, the concept *service subject* emerges. It is though not always clear what is intended with this concept and therefore a very brief review of found definitions/descriptions follows. Howson (1988) describes *mathematics as a service subject* when mathematics is needed as a complement in other major subjects the students are studying e.g. physics. He stresses that this does not "imply some inferior form of mathematics or mathematics limited to particular fields" (ibid. p. 1). Blum & Niss (1991) observe that focusing on mathematics as a service subject and on co-operation between mathematics and other subjects has been treated separately in the education. They discuss different kinds of mathematical modelling and conclude that for physics situations, mathematics is primarily used to describe and explain the phenomena. This use is different from how mathematics is used in models for e.g. economic cases, in which norms are established by value judgements. Niss (1993) discusses different aspects of mathematics, one of which is mathematics as an applied science. In this shape, mathematics can serve as a service subject and provide help to understand phenomena in e.g. physics.

3 The Swedish upper secondary school

3.1 The national curriculum

The upper secondary school in Sweden is governed by the state through the curriculum, the programme objectives and the syllabuses. In the curriculum are laid down the fundamental values that are to permeate the school's activities and also the goals and guidelines that are to be applied. The syllabuses, on the other hand, detail the aims and objectives of each specific course. They also indicate what knowledge and skills students must have acquired on completion of the various courses.

In the curriculum it is stated that the school should aim to ensure that students acquire good knowledge in the various courses that together constitute their study programme and that they can use this knowledge as a tool for example; to *formulate and test assumptions* and to *solve practical problems and work tasks*. It is the responsibility of the school to ensure that students, after they have finished school, can formulate, analyse and solve mathematical problems of importance for vocational and everyday life (Lpf94, 2006, p. 10 - 12). Upper secondary school in Sweden is divided into *different national programmes*; *different specially designed programmes* and *programmes provided at independent schools*. A special designed programme could be considered similar to a national programme, and programmes at an independent school could be approved as one of the national programme. Two of the national programmes, the *Natural Science Programme* (NV) and the *Technology Programme* (TE), are oriented towards science and include higher courses in mathematics and courses in physics. About 12% of all students in the upper secondary school in Sweden attend the Natural Science Programme or the Technology Programme (Swedish National Agency for Education, 2011).

According to the programme objectives (Swedish National Agency for Education, 2001), NV aims at developing the ability to use mathematics in the natural science and in other areas. It is also stated in the programme objective for NV that in order to develop concepts, students need an understanding of the inter-relationships within and between subjects. The importance of information technology (IT) in for example mathematics and science is outlined in the programme objective for TE. Therefore, one responsibility for TE is to give the students opportunity to attain familiarity with using computers as a tool and to use IT for learning and communication. The different courses in each program are chosen to fulfil the aims in the different programme objectives. Courses in a school subject are labelled with capital letters, starting with A for the first course and B for the succeeding course and so on. For all students in NV, Mathematics A to D and Physics A are compulsory courses. For students in TE, Mathematics A to C and Physics A are compulsory. In each of the programmes students can choose between different branches. NV has three branches and TE has five. For the branch *Natural Science* for NV (NVNA), Physics B is compulsory and for the branch *Mathematics and Computer Science* (NVMD), Mathematics E is compulsory. Both Physics B and Mathematics E must be offered as optional courses to all students in NV regardless their choice of branch. None of the branches for TE includes requirements of more courses in mathematics or physics, but Physics B and Mathematics D to E must be provided the students as optional (ibid.).

3.2 Syllabuses

Mathematics is one of the core subjects In Swedish upper secondary school, together with e.g. English, religion and social science, and Mathematics A is compulsory to all students. This importance of mathematics is expressed in the syllabuses for mathematics –a core subject– as e.g. "The school in its teaching of mathematics should aim to ensure that pupils: develop confidence in their own ability to … use mathematics in different situations, …, develop their ability with the help of mathematics to solve … problems of importance in their chosen study orientation" (Swedish national Agency for Education, 2001, p.112). In addition to core subjects there are programme-specific subjects, as for example physics for NV and TE. According to the syllabus in physics, some of the aims are to: "*develop [students'] ability to quantitatively and qualitatively describe, analyse and interpret the phenomena and processes of physics in everyday reality, nature, society and vocational life", …," develop [students'] ability with the help of modern technical aids to compile and analyse data, as well as simulate the phenomena and processes of physics"* (Swedish National Agency for Education, 2000c).

Explicitly, mathematics is important when making quantitative descriptions and implicitly, when analysing data, although the analysing part is mentioned in relation to technical aids. In the syllabuses for the various courses Physics A and Physics B, mathematics is mentioned more explicitly. In Physics A, the students should *be able to make simple calculations using physical models* (Swedish National Agency for Education, 2000a). In Physics B there is more than one aim that includes mathematics. The student should *be able to handle physical problems mathematically*. They should also *be able to make calculations in nuclear physics using the concepts of atomic masses and binding energy* (Swedish National Agency for Education, 2000b). Physics B has Physics A as a prerequisite and the students should attain a deeper understanding for some of the physical concepts when studying Physics B. It is also explicated that there are higher demands on the mathematical processing in Physics B (Swedish National Agency for Education, 2000c). Besides the aims are also the requirements for the grades in each course stated in the different syllabuses. The final grades students are awarded after the courses depend on the achieved level of proficiency (Swedish

National Agency for Education, 2000a, 2000b). The grades vary between *Not Pass* (IG), *Pass* (G), *Pass with distinction* (VG) and *Pass with special distinction* (MVG). The descriptions of the different grade levels are quite vague and the intention is that the syllabuses should be processed and interpreted locally at the schools.

3.3 Physics tests from the National Test Bank

To accomplish equivalent assessment in physics, assessment supports are provided by the Swedish National Agency of Education. One of these supports is the National Test Bank in Physics. In this respect the tests can be considered as a governmental concretisation of the syllabuses for physics. The character and the design of the tasks in tests stress what is covered in the taught curriculum as well as teachers' interpretation of the syllabuses, and by extension what students focus on (Ministry of Education and Research, 2001; Swedish National Agency for Education, 2003). The test material in the National Test Bank for Physics is developed by the Department of Applied Educational Science at Umeå University, who has had this commission since shortly after the new national curriculum was implemented in 1994. Most of the material provided by the test bank is not open to public, only to teachers in physics at upper secondary school. In total there are so far 847 tasks to choose from besides 16 complete tests for each of the courses Physics A and Physics B. The first tests are from 1998 and the latest is from spring 2011. In addition, there are five tests for each course open for students (or anyone interested) to practice on. In this way the students get an idea of what the tests looks like and what is required when taking a test. (Department of Applied Educational Science, 2011). There might be a difference between the national tests and teacher made tests; this is not investigated in this thesis.

The provided tests are constituted of two parts; the first part consists of tasks for which a short answer is enough as a solution and the second part consists of tasks that require more analysing answers. For the last ten years, the final task in the tests is an aspect-task. The task should be easy to start with, but it should also include challenge to more proficient students. These aspect-tasks are corrected according to an assessment matrix, with scores according to achieved qualitative level for the different aspects, e.g. the use of concepts and models, the use of physics reasoning, and the accounting for the answer. The first three years, 1998-2000, it was an experimental part included in the tests; this part is not included in the analysis in this thesis. A part of the assessment support is that scoring rubrics are provided to the teachers with each test. The guidance in these rubrics has changed some over the years. In the more recent rubrics are e.g. more examples of acceptable answers outlined. Furthermore, the criteria for the highest grade were not explicated in the scoring rubrics for the earliest tests.

As opposed to national tests in for example mathematics, the teachers are not obligated to use the tests from the National Test Bank. However a majority of all registered teachers uses the provided physics tests as a final exam in the end of the physics courses (Swedish National Agency for Education, 2005). It is important to stress that National tests, although students must take them in for example mathematics, are not high-stake test. The final grade in a course is not solitarily dependent on the achievement on the national test. In fact, teachers are not allowed to grade a course only on a single test, they have to account for all the various aspects the student has shown his/her knowledge in during the entire course. After a test from the Test Bank is used, the teachers are intended to report back students' results on the test to the Test Bank.

This thorough description of the National Test Bank tests' purpose and their influence on the physics education hopefully clarifies and motivates the choice to use these tests as an indicator of what are required mathematically from upper secondary students while studying physics.

4 Conceptual framework

4.1 Mathematical problem solving

The conceptual framework used in this thesis is related to the various phases of problem solving (Lithner, 2008). Problem solving is used in various contexts with different meanings. Solving mathematical problems can include everything between finding answers to already familiar tasks and trying to proof new theorems. In this thesis *problem* implies when an individual does not have easy access to a solution algorithm (Schoenfeld, 1985). The term *task* on the other hand comprises most work students are involved in during class and while doing homework (Lithner, 2008), which in this thesis narrows down to the work students do while taking a physics test. Different advantages of working with mathematical problem solving in school are that students' ability to reason mathematically improves, their problem solving skills develop and they become more prepared for life outside school, compared to not working with problem solving (Lesh & Zawojewski, 2007, Schoenfeld, 1985; Wyndham et al., 2000;). Learning mathematics through problem solving can also help students to develop their mathematical thinking and their skills in reason mathematically in other areas than pure mathematics, for example physics, (Blum & Niss, 1991).

4.2 Mathematical reasoning

The impact of mathematical reasoning *on mathematical learning* has been discussed and studied from multiple perspectives. Schoenfeld (1992), for example, points out that a focus on rote mechanical skills leads to bad performance in problem solving. Lesh & Zawojeskij (2007) discuss how emphasising on low-level skills does not give the students the abilities needed for mathematical modelling or problem solving, neither to draw upon interdisciplinary knowledge. Lithner (2008) refers to his studies of how rote thinking is a main factor behind learning difficulties in mathematics. The definition of mathematical reasoning and the conceptual framework that is used for the analyses in this thesis are developed by Lithner (2008) through his empirical studies of how students are engaging in various kinds of mathematical activities. As a result, *reasoning* as ”the line of thought adopted to produce assertions and reach conclusions in task solving” was defined (p. 257 ibid.). Reasoning is considered as a product of all reasoning sequences required to reach an answer (ibid.). Each sequence includes a choice that defines the next sequence and the reason is the justification for the choice made (Ball & Bass, 2003).

Just as problem solving, mathematical reasoning is a term that is used with different meanings in various contexts (Yackel & Hanna, 2003). For some scholars, mathematical reasoning is used as a synonym for a strict mathematical proof (e.g. Duval, 2002; Harel, 2006); others talk about pre-axiomatic reasoning e.g. Leng (2010). The NCTM (2000) distinguishes between mathematical reasoning and mathematical proofs when setting the standards for school mathematics. Ball and Bass (2003) equate mathematical reasoning with a mathematical ability every student need in order to understand mathematics. In this thesis, to be considered as *mathematical reasoning* the justifications for the different reasoning sequences should be anchored in mathematical properties and mathematical reasoning is used as an extension of a strict mathematical proof (Lithner, 2008). When reasoning, one starts with an object, a fundamental entity that can be a function; an expression; a diagram etc. To this object, a transformation is done and another object is acquired. A series of transformations performed to an object is called a procedure (ibid.). The mathematical properties of an object are of different relevancy in different contexts e.g. what kind of task one is trying to solve. This leads to a distinction between *surface* properties and *intrinsic* properties, where the former ones have little relevance in the actual context and the latter ones are central and have to be regarded. How the student makes and motivates the choices in the reasoning sequences is dependent on what resources he/she has access to. Schoenfeld (1985) defines the term resources as the tools; e.g. mathematical knowledge; the student has access to when solving a task. The justification for a choice does not have to be mathematical correct, but it has to be a plausible argument. This means that there is some logic to why a guess would be more reasonable, form a mathematical point of view, than another guess (Polya, 1954). Depending on whether the reasoning is based on surface properties, *superficial reasoning*, or intrinsic properties, *intrinsic reasoning*, the framework distinguishes between *imitative* and *creative mathematical founded reasoning*.

One example, described in Bergqvist, Lithner and Sumpter (2008), of when only surface properties are considered, is a student who tries to solve a max-min problem: "Find the largest and the smallest values of the function $y = 7 + 3x - x^2$ on the interval [-1, 5]". This task can be solved with a straightforward solution procedure: One first uses that the function is differentiable on the whole interval to find all possible extreme points in the interval, (i.e. solve $f'(x) = 0$). If there are extreme points, the values at these points are calculated and compared with the values at the endpoints. In the situation described, the student does not remember the whole procedure, but reacts on the words largest and smallest and starts differentiating the function and solves $f'(x) = 0$. This calculation only gives one value and the answer demands two. Instead of considering intrinsic mathematical properties, the student seeks a method that will provide two values and instead solves the second degree equation $7 + 3x - x^2 = 0$. Two points are now obtained and the function values at these points are accepted as the solution by the student. Although the student gives these values as an answer, it is with some hesitation because the method used did not involve any differentiation, something remembered by the student to be related to a max-min problem.

4.3 Creative mathematical founded reasoning

Creativity is another term that is used in various contexts and without an unequivocal definition, just as problem solving and mathematical reasoning are. There are though mainly two different use of the term: one where creativity is seen as a thinking process which is divergent and overcomes fixation; and another one, where creativity is used when the result is a product that is ascribed great importance to a group of people (Haylock, 1997). Regardless of context, there are two main components that can be crystalized when discussing creativity; these are the usefulness and novelty (Franken, 2002; Niu & Sternberg, 2006).

When creativity is discussed in a mathematical context, it has often been an ability ascribed to experts (Silver, 1997). A quantitative study by Kim (2005) shows a nominal correlation between students' creativity and their scores on IQ-tests, a result supporting the view of not ascribing creativity only to experts or "genius". In a study by Schoenfeld (1985), where he compares novices' problem solving abilities with experts', he concludes that professional mathematicians succeed because of their different way from students of tackling a mathematical problem. These are abilities that can be developed and improved by the students (ibid.). Silver (1997) makes a similar conclusion in his paper when he discusses the value for educators in mathematics of changing their view of creativity from professional mathematicians' skills, to a mathematical activity every student can improve in school. Sriraman (2009, p.15) makes a definition of mathematical creativity "as the process that results in unusual and insightful solutions to a given problem, irrespective of the level of complexity".

In the framework used in this thesis, the creativity perspective from Haylock (1997) and Silver (1997) is adopted. That means that creativity is seen as a thinking process that is novel, flexible and fluent. The flexibility indicates that the students have overcome fixation behaviours at some level. The two types of fixation that are intended are content universe fixation, which limits the range of elements that are seen as useful; and algorithmic fixation, which concerns the repeated use of an algorithm once successful (Haylock, 1997).

Creative mathematical founded reasoning (CR)[1] fulfils all of the following criteria. (Lithner, 2008, p.266)

i. Novelty. A new (to the reasoner) reasoning sequence is created or a forgotten one is recreated.
ii. Plausibility. There are arguments supporting the strategy choice and/or strategy implementation motivating why the conclusions are true or plausible.
iii. Mathematical foundation. The arguments are anchored in intrinsic mathematical properties of the components involved in the reasoning.

4.4 Imitative reasoning

The other kind of reasoning used is imitative reasoning (IR). The difference between imitative and creative mathematical reasoning, is that there is no flexibility in the thinking process. There are no new reasoning sequences created and the arguments, that motivate the chosen solution method, could be anchored in surface mathematical properties. The reasoner just uses a solution procedure that seems to fit that kind of task. Imitative reasoning is distinguished into *memorised reasoning* and *algorithmic reasoning*. When it is enough just to recall an answer to be able to solve a task, this is regarded as memorised reasoning, for example the proof of a theorem.

Memorised reasoning (MR) fulfils the following conditions (Lithner, 2008, p. 258)

i. The strategy choice is founded on recalling a complete answer.
ii. The strategy implementation consists only of writing it down.

If some kind of calculations is required to solve the task, there is often no use in remembering an answer. Instead it is more suitable to recall an algorithm. Algorithm is here used in a wide sense and refers to all the procedures and rules that are needed to reach the conclusion to a specific type of tasks, not only the calculations.

[1] Henceforward called mathematical reasoning

Algorithmic reasoning (AR) fulfils the following conditions (Lithner, 2008, p.259)

i. The strategy choice is to recall a solution algorithm, which if it is followed step by step will give the right answer without any demands of novelty.
ii. The remaining parts of the strategy implementation are trivial for the reasoner and just a careless mistake can obstruct the reaching of an answer.

AR is subdivided into three different categories, depending on how the proper algorithm is argued for. The categories are *familiar algorithmic reasoning, delimiting algorithmic reasoning* and *guided algorithmic reasoning*. In this thesis/licentiate, only the categories familiar algorithmic reasoning and guided algorithmic reasoning are used during the analysis of the tests.

Familiar AR (FAR) fulfils the following conditions (p. 262 ibid.)

i. The reason for the strategy choice is that the task is identified as belonging to a familiar class of tasks which are known to be solved by a specific algorithm.
ii. The algorithm is implemented.

If the reasoner does not recall any algorithm or is not able to delimit any from the known ones, there can be a need for guidance from an external source to perform the reasoning. The guidance can either be text-guided, e.g. when following an example in the text book that look similar on the surface, or person-guided when for instance the teacher tells every step in the reasoning sequence that has to be made to fulfil the reasoning, without discussing any intrinsic-based mathematical arguments for the choices.

Text- guided AR (GAR) fulfils (ibid. p.263)

i. The strategy choice concerns identifying surface similarities between the task and an example, definition, theorem, rule or some other situation in a text source.
ii. The algorithm is implemented without verificative argumentation.

4.5 Local and global creative mathematical reasoning

Lithner (2008) introduces a refinement of the category CR into local CR (LCR) and global CR (GCR) that captures some significant differences between tasks categorised as LCR and GCR. This differentiation has been more elaborated by other scholars that have used the framework e.g. Boesen, Lithner and Palm (2010) and Palm, Boesen and Lithner (2011). The difference between LCR and GCR is that in LCR, the reasoning is mainly MR or AR but contains a minor step that requires CR. If instead there is a need for CR in several steps, it is called GCR, even when some parts contain AR and/or MR. Important to stress is that as soon as CR is involved there has to be some understanding of the intrinsic mathematical properties in the task.

4.6 Non-mathematical reasoning

In the application of the framework for the analyses described in this thesis, an additional category is defined. This category consists of those tasks that can be solved by only using physics knowledge; and this category is called *non-mathematical reasoning* (NMR). Physics knowledge is here referred to as relations and facts that are discussed in the physics courses and not in the courses for mathematics, according to the syllabuses and textbooks, e.g. that angle of incidence equals angle of reflection. In the same way, a solution that requires mathematical reasoning refers to mathematics taught in courses at upper secondary school or assumed already to be known by the students according to the curricula.

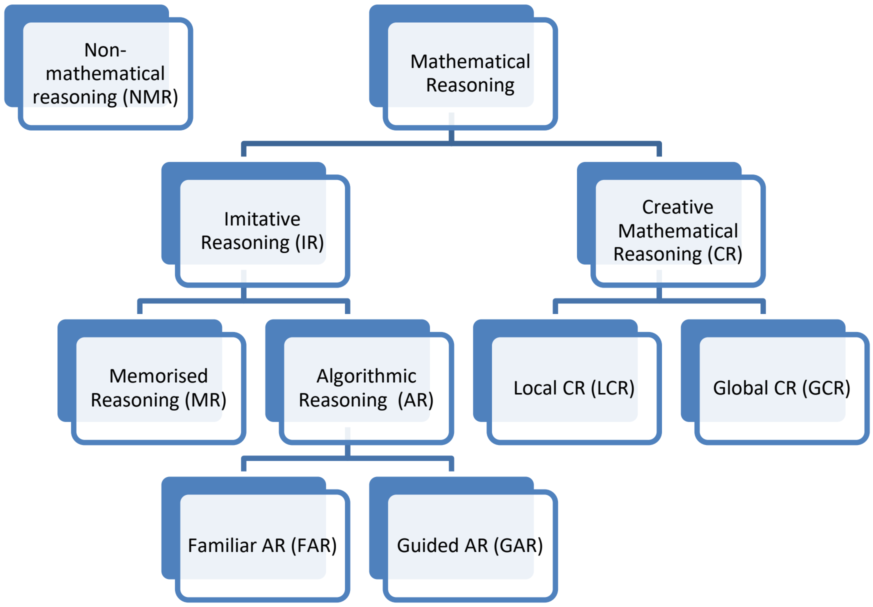


Figure 1. Overview of the conceptual framework.

5 Research related to the framework

5.1 Rote learning, procedural knowledge and imitative reasoning

Although imitative reasoning, AR and MR, refers to the kind of knowledge that is learned by heart/rote and that studies have shown that rote learning contributes to learning difficulties, this study does not imply that students should not learn algorithms. Sfard (1991) discuss how operational and structural knowledge are complementary, and that one of them cannot exist without the other. A three stage hierarchical model is presented and it is stated that before a concept is fully understood, the student has to learn processes/operations that are related to the concept (ibid.). I.e. it is necessary to learn some algorithms in order to achieve a deeper mathematical understanding, but it is not enough. In the paper of Gray and Tall (1994), the dichotomy between procedures and concepts is discussed and they introduce a new word “procept”, referring to both the concept and the process that are represented by the same symbol. Although there is an agreement that procedural knowledge is important, it is not enough when students learn mathematics (Baroody, Feil & Johnson, 2007; Gray & Tall, 1994; Sfard, 1991; Star, 2007). Further there is an argumentation about whether deep procedural knowledge can exist without involvement of conceptual knowledge (Baroody, Feil & Johnson 2007; Star, 2005, 2007). To be successful in mathematics it is necessary for the students to do proceptual thinking, which includes the use of procedures. But as Grey and Tall (1994) stress, the proceptual thinking is also flexible i.e. it includes the capacity to view the symbols as a procedure or a mental object depending on the situation. The definitions of the different subcategories of imitative reasoning accounted for above include no such thing as flexibility. On the contrary, the reasoning could be very fixated.

5.2 Physics reasoning

Since mathematical reasoning in physics tasks is the focus in this thesis, it seems natural to include a brief review of how scholars discuss reasoning in physics; and descriptions of the most commonly used concepts. diSessa (1993) uses the term *p-primes* to describe people's sense of physical mechanism. P-primes are described as small knowledge structures that in some cases are self-explanatory i.e. things happen because that is the way they are. The p-primes originate from the students' experiences of the real world. Through learning, appropriate p-primes are activated in relevant situations and new ones can be generated. The function of a p-prime as self-explanatory, also change during learning, as it must be consistent with the physics laws. P-primes are neither wrong nor right in themselves, in some circumstances they are correct and in others not. In this respect, p-primes can be used both when studying reasoning about unproblematic situations and problematic situations. This is different from using the concept misconception, then only wrongly understood situations can be analysed (ibid.).

Bliss (2008) accounts for studies, conducted by Joan Bliss, Jon Ogborn and others from a period of twenty years, about how students use *common sense reasoning* to explain/describe physical phenomena. Common sense reasoning refers to when students use experiences from everyday life in their reasoning and is explained as "It is the type of reasoning we use to make sense of what is happening around us, or what may have happened, or what will happen" (ibid. p.126). One of the results of the studies in Bliss (2008) was that concrete physical schemes are developed through the interaction with real world experience. These schemes are combined to mental models and used when one is trying to understand or predict different physical events i.e. reasoning about physics.

Another concept used for reasoning about physics situations is *qualitative reasoning* or *qualitative physics* (Forbus, 1981, 2004). This concept is mainly used for an area of artificial intelligence (AI) that is modelling the world, (from a scientific perspective), using the intuitive notions of human mental models instead of mathematical models. The origin of qualitative reasoning is peoples intuitively reasoning about the physical world, i.e. their common sense reasoning (Klenk et al 2005). Qualitative reasoning seems also being used by physicists when first trying to understand a problem and later when interpreting quantitative results (Forbus, 2004).

Wittmann (2002) introduce the concept *pattern of association* when discussing reasoning in physics. This refers to the linked set of reasoning resources brought by a student to some specific situation. Some of the resources can be described by diSessa's (1993) concept p-primes. How these resources are organised when explaining a physics situation is what distinguishes novices from experts, not the existence of the resources (ibid.).

6 Aims and research questions

The literature accounted for above shows that there are several studies in educational research about the relations between the school subjects mathematics and physics. The importance of mathematics and the effects the use of mathematics may have when learning physics is a common share. The review also shows that how students reason mathematically affect their learning in mathematics. If students only look for superficial properties when they are solving a mathematical task, i.e. using imitative reasoning, it is more likely they end up not understanding the underlying mathematical concepts. Focusing on surface properties is a kind of rote-mechanical procedure that contributes to poorer performance in mathematical problem solving. At the same time skills in mathematical problem solving is considered to have a positive effect on the abilities to reason mathematically in other areas than mathematics.

From the previous discussion it is likely to assume that students will not understand the underlying mathematical concepts and the mathematical models the concepts are used to illustrate if they are focusing on imitative reasoning when solving physics tasks. Because the mathematical models express the physics used for describing what and why events happen in the world around us, a lack of understanding of the mathematics likely affects the understanding of the physics and then also students' learning of physics. The conducted studies in this thesis are based on the assumption that students' ability to reason mathematically when solving tasks in physics has an impact *on their learning of physics*, as it has on their learning of mathematics.

It seems that the mathematical reasoning required by students when they are solving physics tasks is not as well researched as the reasoning students use in physics classes. By analysing the mathematical reasoning demands students are confronted when taking physics tests, as well as the relation between types of tasks student solves and their grades, this thesis attempts to contribute to research of how mathematical reasoning may affect students' learning of physics. The aim is specified in the following research questions:

- What is the character of the mathematical reasoning that is required of students in upper secondary school to solve the tasks in physics tests from the Swedish national test bank?
- Is it possible for a student to get one of the higher grades without using creative mathematical reasoning and if it is how common is it?
- Does a student's success on tasks categorised as requiring creative mathematical reasoning depend on the student's success on tasks requiring imitative reasoning?

As mentioned in the introduction, this thesis consists of two papers. Paper I deals with the first of the questions and Paper II with the other two. To answer the last question an appropriate method is needed. Therefore, Paper II also serves as a pilot, to see if the Mantel-Haenszel procedure is a suitable quantitative method to answer questions about the dependence between students' success on the different kinds of tasks. If the model works out well it could be used for further analyses of relation between IR and CR on mathematics tests that have been categorised according to Lithner's (2008) framework in previous studies.

According to the above discussion about similarities between IR – procedural knowledge and CR – conceptual knowledge, the answer to the third question may also shed some light on the relation between procedural and conceptual knowledge.

7 Methods

For the analysis in Paper I, ten tests were randomly chosen from the available ones in the National Test Bank, five from each course Physics A and Physics B respective. Results from eight of these ten tests were used for the analysis in Paper II.

7.1 Categorisation of Mathematical Reasoning requirements

The objects of study were the mathematical reasoning required from students to solve the tasks in the physics tests from the National Test Bank. No students with their actual solutions were included in the study. Required reasoning refers to what kind of reasoning that is sufficient to solve a task, and the chosen framework gives the possibility to determine this. Whether the solution to a task requires creative mathematical reasoning or imitative reasoning depends on whether the solution requires some novelty. In order for a task to be categorised as requiring algorithmic reasoning or memorised reasoning, the student should be able to recognize the type of task. This in turn depends on the education history of the solver. According to studies of how the education in physics and mathematics are organised, a major part of the learning activities seem to be controlled by the textbooks in respective subject (Engström, 2011; Swedish National Agency for Education, 2003 & 2009; Swedish Schools Inspectorate, 2010; Ministry of Education and Research, 2001). A description of the references' respective findings can be found in Paper I. The learning history of an average student is in this thesis therefore reduced to the content in the textbooks. Both textbooks in mathematics and physics were considered. Since students are allowed to use a physics handbook during a physics test, the access to formulas and definitions in this handbook also has to be taken into account when analysing the tasks. The textbooks and the handbook were chosen among the books commonly used in the physics courses in upper secondary school. Even if not all students in the Swedish upper secondary school are using the books above, it is a reasonable assumption of the learning situation for a conceivable student. The following textbooks were used in the analysis of the mathematical reasoning requirements; "Ergo Fysik A" and "Ergo Fysik B" (Pålsgård, Kvist & Nilsson, 2005a, 2005b); "Matematik 3000 Kurs A och B" and "Matematik 3000 Kurs C och D" (Björk & Brolin, 2001, 206) as well as "Tabeller och formler för NV- och TE- programmen" (Ekbom et al., 2004).

The procedure for analysing the tasks is given by the chosen framework and an analysis sheet was used to structure the procedure. The steps comprised in the procedure are outlined below and are used earlier in e.g. Palm et al. (2011).

I. Analysis of the assessment task – Answers and solutions
   a) Identification of the answers (for MR) or algorithms (for AR)
   b) Identification of the mathematical subject area
   c) Identification of the real life event

II. Analysis of the assessment task – Task variables
   1. Assignment
   2. Explicit information about the situation
   3. Representation
   4. Other key features

III. Analysis of the textbooks and handbook – Answers and solutions
   a) In exercises and examples
   b) In the theory text

IV. Argumentation for the requirement of reasoning

Below follows a thoroughly description of the steps in the procedure

I. Analysis of the assessment task – Answers and solutions: When analysing the different assessment tasks with respect to answers and solutions, first, a solution plausible to think a student would use was constructed by the researcher and written down. The solution was then looked at with mathematical glasses on and categorised according to relevant mathematical subject areas for the solution. For instance: does the solution include working with formulas, algebra, diagrams, solving equations, etc. Solutions not including any mathematical object were identified in this step and these tasks were categorised as non-mathematical reasoning (NMR) i.e. tasks solvable without any mathematical considerations. As a consequence of the added category NMR, this step is an addition to the original procedure used in previous studies. Mathematical objects refer to entities to which mathematics is applied. As mentioned above, here mathematics refers to school mathematics introduced in mathematics courses given for students at upper secondary school or mathematics assumed already to be known according to the curricula. Identification of a "real-life" event, refers to tasks where the described situation could give a clue to a known algorithm that solves the problem.

II. Analysis of the assessment task – Task variables: After identification, the next step in the procedure was to analyse the solution according to the different task variables. The first variable is the explicit formulation of the assignment. The second variable is *what* information about the mathematical objects that were given explicitly in the information, contrary to what information the students for example need to retrieve from the handbook or have to assume in order to reach a solution. The third task variable concerns *how* the information was given in the task, e.g. numerically or graphically, interwoven in the text or explicitly given afterwards. This is called representation in the procedure, step II.3. The task can also include key words, symbols, figures, diagrams or other important hints the student can use to identify the task type and which algorithm to use. These features are gathered in the fourth task variable.

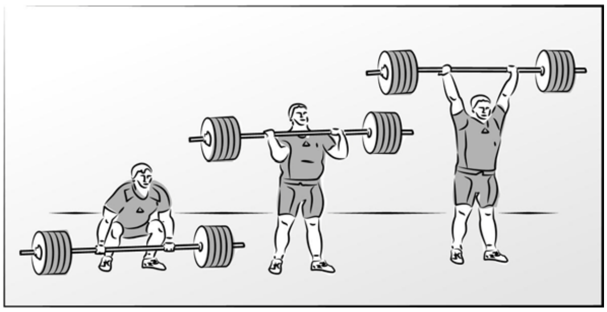
III. Analysis of the textbooks and handbook – Answers and solutions: The third step in the analysis process focuses on the textbooks and the handbook. Formulas used in the solution algorithm were looked for in the handbook and the available definitions were compared to the constructed solution of the task. The textbooks were thoroughly looked through for similar examples or exercises that were solved by a similar algorithm, as well as whether the theory text contained a description of how similar or identical problems could be solved.

IV. Argumentation for the requirement of reasoning: To motivate the concluded reasoning requirement, an argumentation based on the previous steps was made for every task. The considerations are if it is possible for a student to solve the task using a reasoning type based only on superficial mathematical properties or if it is necessary to use creative mathematical reasoning. In order to be categorised as familiar algorithmic reasoning there must have been at least three tasks considered as similar in the textbooks. If the task is similar to a formula or definition given in the handbook, it is assumed that the student can use this as guidance in order to solve the task. It is then enough with only one similar, previously encountered, example or exercise for the task to be regarded requiring text-guided algorithmic reasoning. To be categorised as requiring memorised reasoning, tasks including the same answer or solution should have been encountered at least three times in the textbooks. If none of the above reasoning types are sufficient for solving the task and there is a need to consider some intrinsic mathematical property, the task is categorised to require some kind of creative mathematical reasoning.

The examples below are chosen to represent and illustrate the different types of analysis and the categorisation of the tasks in the national physics tests[2]. All of the tasks are chosen from public tests. Normally, subtasks are treated separately because the task variables and the analysis of the textbooks can be different. The outline of all tasks in a test begins in the same way; first is the number of the task in the test given and after that, enclosed in brackets, the task's number in the National Test Bank for physics. On the next line are the maximum scores for the task given. The scores are divided into two different categories, G-scores and VG-scores. The maximum scores for each category are separated with a slash, for example 2/0 tells that a student can get a maximum of two G-scores and zero VG-scores on that particular task. In the same way, 1/1 tells that the maximum is one G-score and one VG-score. If the task consists of subtasks: a, b, etc.; the total scores for the subtasks are separated with commas.

Task no. 3 (1584)
2/0, 1/0

A weightlifter is lifting a barbell that weighs 219 kg. The barbell is lifted 2.1 m up from the floor in 5,0 s.



a) What is the average power the weightlifter develops on the barbell during the lift?

Short account for your answer:

[2] The original format of the physics tests from the National Test Bank is A4. If this thesis is read on A5 format, the size of the examples below is √2 times larger

b) What is the average power the weightlifter develops on the barbell when he holds it above the head during 3.0 s?

Short account for your answer:

Analysis of 3a

I. Analysis of the assessment task – Answers and solutions: A typical solution of an average student could be derived by the relation between power and change in energy during a specific time. Change in energy is here the same as change in potential energy for the barbell. Multiply the mass of the barbell with the acceleration of gravity and the height of the lift, and then divide by time to get the power asked for. The mathematical subject area is identified as algebra, in this case to work with formulas. The identification of the situation to lift a barbell can trigger the student to use a certain solution method and is therefore included in this analysis as an identified "real-life" situation.

II. Analysis of the assessment task – Task variables: The assignment is to calculate the average power during the lift. The mass of the barbell, the height of the lift and the time for the lift are all considered as mathematical objects. As mentioned above, an object is the entity one is doing something with. In this example, all the objects are given explicitly in the assignment in numerical form. In the presentation of the assignment there is also an illustrative figure of the lift.

III. Analysis of the textbooks and handbook – Answers and solutions: The formula for power, $P=\Delta W/\Delta t$, with explanation "ΔW = the change in energy during time Δt" is given in the handbook (p.105). So is the formula for "work during lift" $W_l = mg\cdot h$, with the explanatory text "A body with weight mg is lifted the height h. The lifting work is…" (p.104), and the formula for potential energy with the text "A body with mass m on the height h over the zero level has the potential energy $W_p= mg\cdot h$" (p.104). In the mathematics book Ma3000AB, there are plenty of examples and exercises of how to use formulas, e.g. on pages 28-30. In the physics book ErgoA, power is presented as work divided by time and work is in one instance exemplified as lifting a barbell. On page 130 there is an identical example where the power during the lift of a barbell is calculated. There is one example on page 136 where work during a lift is calculated in relation to change in potential energy. Exercise 5.05 is solved by calculating work during a lift and exercise 5.10 is also solved by a similar algorithm.

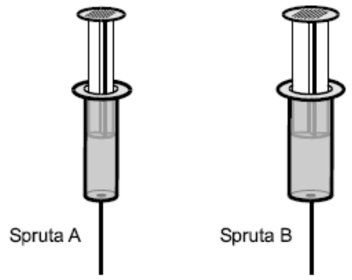
IV. Argumentation for the requirement of reasoning: The analysis of the textbooks shows that there are more than three tasks similar to the task up for categorisation with respect to the task variables, and these tasks can be solved with a similar algorithm. As mentioned in the method section, if the students have met tasks solvable with a similar algorithm at least three times, it is assumable that they remember the solution procedure. This task is then categorised as solvable using imitative reasoning, in this case FAR.

Analysis of 3b

I. Analysis of the assessment task – Answers and solutions: It is not necessary to use any mathematical argumentation in order to solve this task. The solution can be based only on physical reasoning; there is no lifting and therefore no work is done, which in turn implies that no power is developed. The task is therefore categorised as solvable with non-mathematical reasoning. This task is a typical example of an analysis resulting in the NMR categorisation.

Task no. 13 (1184)
0/2

A patient is going to get an injection. The medical staffs are reading in the instructions that they are supposed to use a syringe which gives as low pressure as possible in the body tissue. Which of the syringes A or B shall the staff choose if the same force, F, is applied and the injection needles have the same dimensions? *Argue for the answer*



I. Analysis of the assessment task – Answers and solutions: To solve this task the student can use the relation between pressure, force and area, p=F/A. Neglect the hydrostatic pressure from the injection fluid. If the force that the staff applies to the syringe is the same, it is the area of the bottom that affects the pressure; the larger the area the less the pressure. The staff should choose syringe B. The mathematical subject area is identified as algebra, to work with formulas, and also proportionality.

II. Analysis of the assessment task – Task variables: The assignment is to argue for, and to choose which syringe that gives the minimum pressure. Only the force is given as a variable, represented with a letter. Key words for the students can be *force* and *pressure*. The situation is illustrated with a figure where it appears that syringe B has a greater diameter than A.

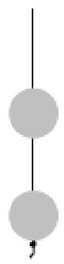
III. Analysis of the textbooks and handbook – Answers and solutions: In the handbook, the relation p=F/A is defined. Proportionalities are discussed and exemplified in Ma3000AB, but are not used for general comparisons. There is one example in Ergo A about how different areas affect the pressure and also one exercise that is solved in a similar way, using a general comparison between different areas and pressure.

IV. Argumentation for the requirement of reasoning: There is only one example and one exercise that can be considered somehow similar with regard to task variables and solution algorithm. The formula is in the handbook, but there has to be some understanding of the intrinsic properties in order to be able to use the formula in the solution. This task was therefore considered requiring some creative mathematical reasoning, in this case GCR, in order to be solved.

During the analysis process situations occurred when the analysis was not as straight forward as in previous examples. All these tasks were discussed in the reference group and below is one example of the borderline cases that arose.

Task no. 12 (1214)
1/2

In order to determine the charge on two small light silver balls, the following experiment was conducted. The balls, which were alike, weighed 26mg each. The balls were threaded on a nylon thread and were charged in a way that gave them equal charges. The upper ball levitated freely a little distance above the other ball. There were no friction between the balls and the nylon thread. The distance between the centres of the balls was measured to 2.9 cm. What was the charge on each of the balls?



I. Analysis of the assessment task – Answers and solutions: To derive at a solution, the forces acting on the upper ball must be considered. Because it is levitating freely, it is in equilibrium and according to Newton's first law the net force on the ball is then zero. The forces acting on the ball are the gravitational force, $F = mg$, (downwards) and the electrostatic force from the ball below, $F = kQ_1Q_2/r^2$, (upwards). Put these expressions equal and solve for Q_1, using that $Q_1 = Q_2$, which will give the charges asked for. The mathematical subject area is identified as algebra, to work with formulas and to solve quadratic equations.

II. Analysis of the assessment task – Task variables: The assignment is to calculate the charges on the balls. The mass on each of the balls and the distance between their centres are mathematical objects given numerically and explicitly in the assignment. The information of the charges' equal magnitude is textual and a part of the description of the situation. There is also an additional figure of the balls on the thread, illustrating the experiment.

III. Analysis of the textbooks and handbook – Answers and solutions: Coulomb's law, $F = k \cdot Q_1Q_2/r^2$ (the formula for electrostatic force), is given in the handbook (p. 108) with explanation "r=distance between the charges and … k =… ≈ $8.99 \cdot 10^9$ $Nm^2/(As)^2$". In the mathematics book Ma3000AB there are plenty of examples and exercises of how to use formulas, e.g. on pages 28-30 and of solving quadratic equations on page 269. In the physics book ErgoA, Coulomb's law is introduced and exemplified and there are at least three exercises of calculating the charge on different objects using this law. There is one example of a levitating charge (p. 227), but in this case in a homogeneous electrical field instead of due to the electrostatic force from another charged particle. There are also two exercises of similar situations as in the example. Newton´s first law is formulated in the theory text (p. 91) where it reveals that the net force has to be zero if an object for example is at rest, and this relation is used on several different occasions in ErgoA. The gravitational force is introduced on pages 92 and is then used throughout the book.

IV. Argumentation for the requirement of reasoning: Considering the mathematical reasoning, there are more than three examples or exercises in the textbooks where the same algorithm has been used, i.e. to put two expressions equal, solve for one unknown variable, including taking the square root. But there are not three or more occasions considering the physics context. To solve the task the student must first identify the force situation in order to know which expressions to equate. After having discussed this task in the reference group, it was concluded that analysing the physics context does not belong to the mathematical reasoning. Although mathematical reasoning is necessary to be able to solve the task, it is not enough, and although the mathematical reasoning can be considered as some kind of algorithmic, the task was categorised to require local creative reasoning LCR, where the minor step is to analyse the physics.

Tasks categorised as solvable with FAR, like 3a above, are hence forward called FAR-tasks and tasks solvable with GAR are called GAR-tasks. Together these kinds of tasks are referred to as IR-tasks. Tasks categorised solvable by only using physics, that is no mathematics were required, like 3b above, are hence forward called NMR-tasks. Tasks requiring GCR to be solved, like 13 above, will be called GCR-tasks and in the same way will tasks requiring LCR, like 12 above, be called LCR-tasks. Tasks requiring either LCR or GCR will be called CR-tasks.

7.2 Reflections about the method for categorising mathematical reasoning

The construction of a typical solution in the first step of the procedure is one of the reductions in the method. Identification of the mathematical subject area and the task variables depends on this typical solution and the results from the identification effect the categorisation of the required mathematical reasoning. Hence, how this typical solution is chosen can affect the findings of the analysis, i.e. the distribution of the mathematical reasoning requirements could differ from the one presented in this study. Justification for that the constructed solution is a *plausible* student solution comes from the researchers experience as a teacher, as well as considering the solution proposals and the scoring rubric provided with each test.

The third step of the procedure comprises an analysis of the textbooks in mathematics and physics. As mentioned earlier, one textbook each for mathematics and physics have been chosen to represent an average upper secondary students' available literature. There are about four different textbooks for each of the courses and which books to use in mathematics and physics is often decided locally at each school. The combination of the textbooks students in one school use could differ from the combination used by students in other schools. Although the textbooks cover essentially the same subject areas, examples and exercises could vary between the books. This influences the number of similar tasks, as the one analysed, that can be found in the textbooks, which in turn effects the categorisation of the analysed task and eventually the presented distribution of the mathematical reasoning requirements. If examples the teachers discuss during classes were included in the analysis, the number of similar tasks might be higher than when only textbooks are used as a representation of the learning history. The number of IR-tasks would then be higher and the number of CR-tasks would be lower than the findings in this study.

In the last step of the procedure, a task is argued to be a FAR-task if similar tasks have been met at least three times before. That three is an appropriate assumption is supported by a study by Boesen et al. (2010). They use three as a minimum to categorise a task as FAR and found that when students are put in front of FAR-tasks in national mathematics tests, the students try to recall appropriate algorithms to solve the test tasks. It is clear that another choice than three as the minimum number will affect the number of tasks categorised as FAR in general. It is also likely that the number of similar tasks different students need to have met to be able to remember a solution differ.

7.3 Comparing grades with kinds of tasks solved

The grades a student can receive on a test vary between *Not Pass* (IG), *Pass* (G), *Pass with distinction* (VG) and *Pass with special distinction* (MVG). For each test there are certain score levels the students need to reach to get a certain grade. To get the grade MVG, students also need to fulfil certain quality aspects besides the particular score level. To decide if it is possible for a student to get one of the higher grades without using any kind of CR, each test is first analysed separately. First the score level for each grade was compared with the maximum scores that are possible to obtain, given that the student only has solved (partly or fully) IR- and/or NMR- tasks. The available student data do not give any information about which of the qualitative aspects required for MVG the students have fulfilled, but the data sheets include students grades, thus MVG can be included in the analyses as one of the higher grades. After analysing if it is possible at all to receive the grades VG or MVG without solving any CR-tasks, students' actual results on the categorised tasks for those particular tests are summed up. The proportion of students who only got scores from IR- and NMR-tasks is then graphed with respect to the different grades.

7.4 Statistical method

Before describing the method used for the more quantitative analysis, the concept odds ratio will be explained as well as the Mantel-Haenszel procedure.

7.4.1 Odds ratio

Odds ratio can be used to measure the dependency between different nominal variables. It is commonly used in various clinical research (Haynes, Sacket, Guyatt, & Tugwell, 2006) or in biological statistics (McDonald, 2009). Because odds ratio can be used for qualitative data and the results only show influence from one variable and remain undisturbed from others, this model is used in various social science research (Ribe, 1999). Keeping the groups fixed, *odds* is defined as the probability p for an event to happen divided by the probability for the same event not to happen, $O=p/(1-p)$. *Odds ratio* is then defined as the ratio between the different odds for the event with respect to different groups (see below).

Table 1. Example of a probability matrix

Probability for Y wrt groups	Y happens	Y does not happen
Group 1	p_1	$1 - p_1$
Group 0	p_0	$1 - p_0$

Odds ratio: $\theta = \frac{\frac{p_1}{1-p_1}}{\frac{p_0}{1-p_0}} = \frac{p_1(1-p_0)}{(1-p_1)p_0}$

If the odds ratio is equal to 1, the probability for the event to happen does not depend on the factor differentiating the groups. Calculations of the odds ratio can thus tell how the probability for success in one group differs from the probability for success in another group.

Ribe (1999) describes one example where the odds ratio is used to see how the risk to be unemployed is affected by the country of birth. First the data is stratified so that other variables that also might affect unemployment are held constant. The two groups that are compared are people born in Iran and people born in Sweden.

Situation 1: woman 27 – 39 years, no upper secondary education, single.

Country of birth	Probability to be unemployed	Odds to be unemployed
Iran	0,78365	3,62214
Sweden	0,32274	0,47654

Odds ratio = 3,62214/0,47654 = 7,6

Situation 2: Man 40 – 49 years, higher education, married.

Country of birth	Probability to be unemployed	Odds to be unemployed
Iran	0,24359	0,32203
Sweden	0,04065	0,04237

Odds ratio = 0,32203/0,04237 = 7,6

The conclusion in this example is that the country of birth affects the risk to be unemployed and that the probability to be unemployed is much higher if one is born in Iran than in Sweden.

7.4.2 The Mantel-Haenszel procedure

The MH-procedure was originally developed for data analyses from retrospective studies in the clinical epidemiology area. The purpose of the MH-procedure was to test if there were any relations between the occurrence of a disease and some factors. The disease could for instance be lung cancer and one factor could be cigarette smoking (Mantel & Haenszel, 1959). A retrospective study can be performed on already collected data and does not require as big sample size as a forward study (also called prospective study) does. In a retrospective study of a disease one looks for unusually high or low frequency of a factor among the diseased persons, while in a forward study it is the occurrence of the disease among persons possessing the factor that is looked at (ibid.). The calculations involved in the MH-procedure are quite simple and this is probably a contributing factor to that the method is commonly used in various areas today e.g. epidemiology (Rothman, Greenland & Lash, 2008), biology/biological statistics (McDonald, 2009) and social/educational sciences (Fidalgo & Madeira, 2008; Guilera, Gómez-Benito & Hidalgo, 2009; Holland and Thayer, 1988; Ramstedt, 1996). One of the most common uses of the MH-procedure in educational studies seems to be for detecting existence of *differential item functioning* (DIF). DIF exists if people with the same knowledge/ability, but belonging to different groups, have different probabilities to give the right answer to an item/task. Ramstedt (1996) used a modified version of the MH-procedure to analyse if there were differences between how boys and girls succeeded on national physics tests depending on their sex, according to their personal identity number. According to Ramstedt, Holland and Thayer (1988) were the first ones to use the MH-procedure to detect DIF.

To use the MH-procedure, data should first be stratified into 2x2 contingency tables. In these tables the rows and the columns represent the two nominal variables that will be tested for dependence. The variable that is placed in the rows is the one that is tested whether it explains/affects the outcome of the variable placed in the columns. The row variable is thus sometimes called explanatory variable and the column variable is called response variable. The different contingency tables represent a third nominal variable that identifies the repeat. The two nominal variables could for example be: a disease and a factor; a plant and a habitat; group belonging and success on tasks. Examples of the repeat variable are different medical centers, different seasons, different teachers etc.

Table 2. Contingency table for repeat *i*.

Table i	Y = 1	Y = 0	Totals
X = 1	a_i	b_i	n_{i1}
X = 0	c_i	d_i	n_{i0}
Totals	m_{i1}	m_{i0}	n_i

In Table 2 X and Y represent the two nominal variables. Both variables are coded by the values 0 and 1 for the respective object included in the study. Belonging to the group of diseased persons might then be represented by X = 1 and not being diseased with X = 0. In the same way, the occurrence of a factor may be represented by Y = 1 and non-existence of the factor with Y = 0. The letters a_i, b_i, c_i and d_i denote the frequencies for respective occurrence and $n_i = a_i + b_i + c_i + d_i$. A diseased person possessing the factor will then be one of those contributing to the frequency a_i. The probability p for an event is estimated by the relative frequency $\hat{p}$. For example, the relative frequency for the event X = 1 and Y = 1 is $\hat{p} = a_i/n_i$.

The MH-procedure includes an estimation of the *common odds ratio*, $\hat{\theta}_{MH}$, for the different contingency tables. From Table 2 follows that the odds for X = 1 and Y = 1 is estimated by a_i/b_i and the odds for X = 0 and Y = 1 is estimated by c_i/d_i. This gives that the odds ratio for contingency table i is estimated by

$$\hat{\theta}_i = \frac{{a_i}/{b_i}}{{c_i}/{d_i}} = \frac{a_i d_i}{b_i c_i}.$$

The common odds ratio calculated in the MH-procedure is defined as

$$\hat{\theta}_{MH} = \frac{\sum_j a_j d_j/n_j}{\sum_j b_j c_j/n_j} = \frac{\sum_j w_j \hat{\theta}_j}{\sum_j w_j} ,$$

where $\hat{\theta}_i$ is the odds ratio for table i and

$$w_i = \frac{b_i c_i}{n_i}$$

is the weight associated to $\hat{\theta}_i$. The summations run over all contingency tables, i.e. j = 1,…,k, where k is the number of contingency tables. Thus $\hat{\theta}_{MH}$ is a weighted average of the individual odds ratios. The assumed null hypothesis, H_0, is that there is no dependence between the variables X and Y, i.e. $\theta_{MH} = 1$.

The most important step in the procedure is the calculation of a MH test statistic, which tells whether $\hat{\theta}_{MH}$ differs sufficiently from 1 so that H_0 can be rejected. The most commonly used test statistic, χ^2_{MH}, is approximately chi-square distributed, and is compared to a chi-square distribution with one degree of freedom (Mantel & Haenszel, 1959; Ramstedt, 1996; Mannocci 2009; McDonald, 2009). The definition of χ^2_{MH} is

$$\chi^2{}_{MH} = \frac{(|\sum_i \mathrm{a}_i - \sum_i E(\mathrm{a}_i)| - ½)^2}{\sum_i Var(\mathrm{a}_i)} ,$$

where $E(\mathrm{a}_i) = {}^{n_{i1}m_{i1}}/_{n_i}$ is the expected value for a_i under H_0 and

$$Var(\mathrm{a}_i) = \frac{n_{i1}n_{i0}m_{i1}m_{i0}}{n_i^2(n_i - 1)}$$

is the variance for a_i (Mantel & Haenszel, 1959).

Instead of χ^2_{MH}, a test statistic, Z_{MH}, which is approximately normal distributed, can be used (McCullagh & Nelder, 1989). The advantage of using Z_{MH} is that the direction of a possible dependence is detected. Therefore this test statistic is used in this thesis. The definition of Z_{MH} is

$$Z_{MH} = \frac{\sum_i\{\mathrm{a}_i - E(\mathrm{a}_i)\} - ½}{\sigma_{\sum_i\{\mathrm{a}_i - E(\mathrm{a}_i)\}}} ,$$

where $E(\mathrm{a}_i)$ is as above and

$$\sigma_{\sum_i\{\mathrm{a}_i - E(\mathrm{a}_i)\}} = \sqrt{var(\mathrm{a}_i - E(\mathrm{a}_i))} = \sqrt{\frac{\sum_i\{n_{i1}n_{i0}m_{i1}m_{i0}\}}{n_i^2(n_i - 1)}}$$

is the standard deviation of $\mathrm{a}_i - E(\mathrm{a}_i)$ (McCullagh & Nelder, 1989). The value ½ that is subtracted in the numerator for each of the statistics is a continuity correction value (Mantel & Haenszel, 1959; McCullagh & Nelder, 1989).

7.4.3 Method

To answer the third research question, if there is a dependence between students' success on IR-tasks and how they succeed on GCR-tasks, the MH-procedure was first used on one randomly chosen pair of IR/GCR-tasks from one of the eight physics tests. For each of the chosen tasks, students' scores were collected from the data sheet, as well as the ID number of respective student's teacher. The two different categories of tasks, IR and GCR, are the two nominal variables tested for dependence. To control for a possible influence from students different teachers, teacher is chosen as the variable that identifies the repeat. One influence may be on the correcting and scoring of the tasks, since the correcting involves some interpretation of the scoring rubrics and could thus result in some differences in how to score a specific answer. It is assumed that the individual teacher is consistent in the correcting of his/her students' solution to respective task. Another influence from the teachers is that different teachers show different examples on the blackboard, which influence what become familiar solutions to students. There can also be a difference in what kind of mathematical reasoning different teachers give their students the possibility to practice on. Some teachers may be more focused on working with creative mathematical reasoning than others.

To see if there was any consistency in the result from the MH-procedure, more tasks were chosen from the eight tests. The number of GCR tasks varied between two and six for the different tests, so it was decided to choose two GCR tasks from each test. The GCR tasks were randomly selected where it was possible. It was further decided to test how the success on these two chosen GCR-tasks depended on success on two of the simpler IR-tasks and two of the harder IR-tasks on the same test. The difference between a simpler and a harder IR-task turned out to mainly depend on how many steps that were needed in the solution algorithm. For a simpler IR-task, the solution consisted mostly of one step; and for a harder IR-task, there were often three or more steps to remember. The number of IR-tasks varied between six and nine in the different tests. Each GCR-task was then tested for dependence against all four of the IR-tasks.

MATLAB was used to arrange the contingency tables needed in the MH-procedure[3]. The rows in a contingency table represent the students who have succeeded (1) and not succeeded (0) on the particular IR-task. The columns represent in the same way the students who have succeeded (1) and not succeeded (0) on the GCR-task. Success on a task is defined as to have solved the task completely, i.e. to have attained the maximum score. For each entry, MATLAB calculated the number of students that fulfilled that particular combination e.g. a_i is the number of students who have succeeded on both the IR-task and the GCR-task. The row and column totals were summarized, as well as the total number of students for teacher i.

Table 3. Contingency table for IR and CR with respect to teacher *i*.

Teacher i	GCR (1)	GCR (0)	Totals
IR (1)	a_i	b_i	n_{i1}
IR (0)	c_i	d_i	n_{i0}
Totals	m_{i1}	m_{i0}	n_i

After this, Z_{MH} the approximately normal distributed test statistic, was calculated for every pair i.e. 64 test statistics were calculated. The obtained value is compared to critical values for a two-tailed test and 5 % significance level, to decide if H_0 can be rejected or not. For a table to be included in the calculation of the test statistic, each of the calculated expected values has to be 5 or more. Because this is a pilot study, no correction for the multiple comparisons was done (cf. McDonald, 2009).

8 Summary of the studies

8.1 Paper I

The study in Paper I addresses the first of the research questions, "What is the character of the mathematical reasoning that is required of students in upper secondary school to solve the tasks in physics tests from the Swedish national test bank?". To answer this question, 209 tasks from ten

[3] See Appendix for the MATLAB code.

different physics tests, randomly selected from the National Test Bank in Physics, were analysed. The analysis comprised of a thoroughly qualitative examination of the tasks as well as of the textbooks in mathematics and physics. The textbooks were chosen to represent the students' learning history. Certain task variables were identified and the test tasks were compared to the tasks met in the textbooks. The method used is the one called "categorisation of mathematical reasoning requirements" and described in the method section. The tasks were categorised according to the different kind of mathematical reasoning (i.e. FAR, GAR, LCR or GCR) required to reach a solution; or if the task were solvable by only using knowledge from physics (i.e. NMR). The analysis was made by one researcher and often the process was straight forward, but occasionally some border-line cases arose. Thus various examples were continuously discussed in a reference group consisting of one mathematician and one mathematics education researcher. Examples of the different kinds of analyses are presented in Section 7.1.

The main results from the analysis of the mathematical reasoning requirements show that three-fourth of the physics tasks required mathematical reasoning. It is also shown that a third of the total number of the tasks require some kind of creative mathematical reasoning to reach a solution, which corresponds to 46 % of the tasks requiring mathematical reasoning. Comparing requirements in tests for Physics A and for Physics B, there are more tasks in Physics B that can be solved by only using NMR. However, when comparing tests from different years the findings reveal a notable variation in the distribution of the different reasoning categories. No consistency could be seen among the tests with respect to this analysis. These results answer the first of the research questions, to what extent and of what kind mathematical reasoning is required when solving physics tests.

8.2 Paper II

The second and the third research questions were studied in Paper II. The categorisation of tasks regarding their mathematical reasoning requirement from Paper I was used as data, as well as students' scores on the different tasks together with their grades on the different tests. To examine the second research question, "Is it possible for a student to get one of the higher grades without using creative mathematical reasoning and if it is, how common is it?", each test was analysed separately. The requirements for the various grades were compared to the total score able to receive in each category (i.e. IR, CR or NMR). For those tests for which it was possible to get a higher grade than G without using CR, the proportion of students who had not solved any CR tasks completely were graphed with respect to their grades on the different tests, the method is described more thoroughly in Section 7.3.

The result shows that in three of eight tests it is possible to receive one of the higher grades without solving any tasks requiring CR. Nevertheless, when graphing the proportions, it turns out this is not to frequently occurring. Only in one of the eight tests a larger number of the students get a higher grade than G without solving CR tasks. This test, however, differs from the other tests in the way that the number of NMR tasks is larger and could be a reason for the larger number of students with higher grades. This answers the second of the research questions, whether it is possible and to what extent higher grades are received without solving tasks requiring creative mathematical reasoning. The analysis of the tests also revealed that mathematical reasoning was needed to attain a G on six of the eight physics tests. Another finding in the analysis of the scores with respect to reasoning category is that the average score on CR tasks is the same on both Physics A and Physics B tests.

To answer the last of the research questions, “Does a student’s success on tasks categorised as requiring creative mathematical reasoning depend on the student’s success on tasks requiring imitative reasoning?”, the Mantel-Haenszel procedure was used. In the procedure different normal distributed test statistics, Z_{MH}, are calculated to decide if there is any relation between how a student succeeded on an IR-task and on a GCR-task. The values for Z_{MH} are compared to the critical values for a two-tailed test and 5 % significance level to see if there is any dependency between success on GCR-tasks and success on IR-tasks. The MH-procedure is described in Section 7.4.2 and the method in Section 7.4.3. The study’s first goal is to decide if the MH procedure is an appropriate method to use to examine this kind of relation. If the procedure seems to give results to trust on, the second goal will be to analyse these results in order to answer the third research question.

The results from the study show that the MH-procedure can be used for the desired examination, but that the MH-procedure is sensitive to the number of students each teacher is responsible for, with respect to the correcting and grading of the tests. The calculated test statistics from the MH-procedure indicates that there is a positive dependence between success in IR-tasks and GCR-tasks.

9 Discussion of the results

The results from both papers confirm that the ability to reason mathematically is important when solving tasks in physics tests from the National Test Bank. In Paper I is shown that 75 % of the tasks require mathematical reasoning, and almost half of these 75 % require creative mathematical reasoning. Showing that a majority of the tasks in physics tests require mathematical reasoning is one way to support the assumption that mathematics is important when learning physics. Further support for the assumption is given by the result in Paper II. The analysis of the score levels for the grades revealed that it was impossible to pass six of the eight tests without reasoning mathematically. This is a main result backing up the conclusion above. Also worth commenting on is the result that students can get one of the higher grades without using any kind of creative mathematical reasoning. Counting scores for the tests with respect to reasoning category shows that higher grades are possible to attain in three out of eight tests. However, comparing this result with student data tells that this occurs rarely. Thus the importance of being able to reason mathematically, in particular the ability to use creative mathematical reasoning, to pass and to do well on physics tests is strengthened further. As the small review in Section 5.2 about physics reasoning shows, previous research has studied sense making in physics without involving or focusing on the mathematical aspects. This thesis is a contribution by considering one part of the role of mathematics students are confronted when learning physics in upper secondary school. The findings should be interesting both for education researchers in mathematics and in physics, as well as for teachers in both subjects.

The importance of mathematics is explicated in the syllabuses for both physics and mathematics, as well as in the curriculum for upper secondary school. In this thesis is one of the aspects of mathematics studied, that is mathematical reasoning, and this aspect is studied in relation to national tests in physics. As mentioned in Section 3.3, the purpose of national tests is to be an assessment support to teachers, and also a guiding of how to interpret the syllabuses/curricula. The tests may thus be viewed as an extension of the syllabuses/curricula and in this respect the conclusion from the results is consistent with what to expect from the tests. The necessity of being able to reason mathematically is, with respect to the results discussed in previous paragraph, one of the things communicated by the tests to the teachers. According to the well-known saying "*What you test is what you get*", tests stress what is focused on. Thus the necessity of mathematical reasoning is also communicated to the students. The results in this thesis clarify, at least implicitly, to both teachers and students that it is unlikely to attain a higher grade than *Pass* without having some understanding of intrinsic mathematical properties. From the discussion above about procedural and conceptual knowledge we know that some intrinsic understanding may follow from working with exercises involving standard procedures. At the same time it is clear that one cannot fully understand the underlying concepts if the focus only is on the procedures. Therefore, viewing mathematics only as the formulas in the handbook is not fruitful, neither for students striving to succeed on physics tests, nor for teachers in their aim to help the students to reach their goals, something most teachers are aware of.

As shown in Paper I, there were on average more CR-tasks in the Physics B tests than in the tests for Physics A. Scrutinising this result for each test shows that this result varies over the years and that the variation between tests for the same course sometimes is bigger than between the different courses. Further analysis in Paper II reveals that the average proportions of the scores for the different reasoning categories are the same for tests in Physics A and Physics B. Assuming that the average result is general and drawing on the results from both papers, it seems that solving a CR-task in a Physics A test scores higher than solving a CR-task in a Physics B test. One conclusion could be that being able to perform more demanding mathematics is more valued in Physics A than in Physics B. Comparing this to the text in the syllabuses, which states that there is a higher demand on the mathematical processing in Physics B, one could ask if not giving as many points for the creative mathematical processing in Physics B as in Physics A is a consequence of the test developers interpretation of the syllabuses. Although there were no intentions to evaluate the tests from the National Test Bank in Physics, some results with this character came up that seems important to comment on. The difference in the scores for CR-tasks in Physics A and Physics B might as well be a consequence of the test developers' use of another framework for categorising the tasks according to the goals in the syllabuses. This thesis' description of the mathematical reasoning requirements in physics tests can provide a complement for national test developers to decide whether the tests assess what is intended according to the curriculum.

According to the framework, a task is categorised to require creative mathematical reasoning if some novelty is necessary to come up with a solution. The method used equates this condition with if there are too few exercises and examples in the textbooks corresponding to the task up for analysis with respect to the task variables. By changing one or more of the task variables for a task categorised as IR, for example changing the context, a task can be made requiring CR instead of IR. This awareness among the national test developers likely influences the amount of the different reasoning categories. A comparison between national tests and teacher made tests in mathematics shows a significant difference in the mathematical reasoning requirements in the different mathematics tests (Palm et al., 2011). However, the analysis in this thesis does not tell if there exist physics tasks that always require CR due to their physics content, i.e. a kind of intrinsic property. It would be interesting to modify the conceptual framework that is used in this study and use it to analyse physics reasoning. The results from the study presented in this thesis and the result from a study on physics reasoning could then be compared and maybe contribute to a deeper understanding of the reasoning requirements students are put in front of.

The small study about the statistical method shows that the Mantel-Haenszel procedure can be used for analysing the relation between how students succeed on different kind of tasks with respect to the reasoning requirements. The test statistic from the MH-procedure confirms that students who can solve tasks requiring imitative reasoning more likely can solve tasks requiring creative reasoning. Although the marking of the test is supposed to be objective, it is known that different teachers interpret the solution manual and the criteria for a test differently. The possibility to control for different teachers with the different contingency tables is a major reason why the MH-procedure was chosen. The constraint on the expected value in each of the entries in the contingency tables limited the number of teachers contributing to the calculation of the test statistic. In some cases the constraint was not fulfilled for any of the teachers and the test statistic could not be calculated. There were also occasions when students' results from very few teachers contributed to the calculation of the statistic; this may influence the generality of the result. To decide if the MH-procedure is an appropriate method with respect to separating the contingency tables according to teachers, further studies are required. McCullagh & Nelder (1989) discuss how a logistic linear model could be considered when for example analysing the effect of a new treatment of a disease, conducted on different medical centres that differs in some respects that are hard to measure. By doing the logistic parameterisation it is possible to investigate if the treatment has the same direction (better or poorer) at all medical centres. If it does not have the same direction, one has to be very careful in the conclusions that are drawn from analyses in which results from different centres are contributing to the calculation of the result. Different directions can indicate that the MH-procedure is not appropriate for the analysis of dependence between a treatment and getting well. This linear logistic parameterisation could be one method to use in order to decide if the MH-procedure is appropriate for analysing dependence between the success on tasks requiring different kinds of reasoning.

The analysis of NMR-tasks was often straightforward as exemplified in Paper I, but on a few occasions borderline cases occurred, shown in Section 7.1 and in Paper I. The same can be applied to the other categories as well. All borderline cases were discussed in the reference group, as well as several of the other tasks. The thorough description in Paper I of the analysis process for six of the tasks is included to ensure a high reliability and validity for the method used.

As mentioned in Section 7.2, textbooks are used to represent the learning history. It is assumed that all students have read and worked themselves through all examples and exercises in the textbooks. If they have not done that, this could influence the number of IR-tasks on a test for particular students. It is not excluded that a solution a student present to a task is a creative solution for that particular student. The categorisation is a minimum requirement under the assumption above. If more tasks can be considered as IR for particular students this could also influence the actual number of tests on which it is possible to get a higher score than G without using any kind of CR. If there were more IR-scores, there could be more students receiving a VG or a MVG by only solving IR- or NMR-tasks. It would be interesting in a future study to let students solve some of the analysed tasks and see what kind of mathematical reasoning they actually use (cf. Boesen et al. 2010).

This thesis fits best among the *mathematics in physics* research, also mentioned in the introduction. Going back to the discussion about how the tests from the National Test bank could be viewed as an extension of the national curriculum, one could assume that students' results on the tests are a measure of their knowledge in physics, at least in school physics. The scores and grade on a test should help teachers to decide which goals a student has attained and the level of the achievements of the goals. In this respect, students' grades on the physics tests could be viewed as a measure of their attained knowledge in physics. The result that 75 % of the tasks in the physics tests require mathematical reasoning and that it is impossible to pass six out of eight tests without mathematical reasoning, show that students' ability to reason mathematically is an integral part of their knowledge in physics. This in turn likely influences how students study and prepare themselves for tests in physics. It is well known that a focus on IR can explain some of the learning difficulties that students have in mathematics (see for example Lithner, 2008). The results above show that a focus on IR when learning physics in upper secondary school will make it hard for the students to do well on the physics tests. A reasonable conclusion is that focusing on IR can give students learning difficulties in physics, as it does in mathematics. To be able to say more about the relation between mathematical reasoning and learning in physics further studies are needed; for example a comparison between how students have succeeded on tests in mathematics and on the tests analysed in this thesis.

As accounted for in Section 2.2, much of previous research has focused on students' use of mathematics while learning and solving problems in physics. By instead focusing on the mathematical requirements in physics tests, the results presented in this thesis hopefully are a complement and a contribution to the research about how students' knowledge in mathematics influence their success and learning in physics.

Bibliography

Ball, D. & Bass, H. (2003). Making mathematics reasonable in school. In J. Kilpatrick, G. Martin & D. Schifter (Eds.), *A Research Companion to Principles and Standards for School Mathematics,* (pp.27-44). Reston VA: National Council of Teachers of Mathematics.

Baroody, A. J., Feil, Y., Johnson, A. R. (2007). An Alternative Reconceptualization of Procedural and Conceptual Knowledge. *Journal for Research in Mathematics Education,* 38(2), 115-131. Stable URL: http://www.jstor.org/stable/30034952

Basson, I. (2002). Physics and mathematics as interrelated fields of thought development using acceleration as an example. *International Journal of Mathematical Education in Science and Technology*, 33(5), 679–690.
DOI: 10.1080/00207390210146023

Bergqvist, T., Lithner, J. & Sumpter, L. (2008). Upper Secondary Students' Task Reasoning, *International Journal of Mathematical Education in Science and Technology,* 39 (1), 1-12.
DOI: 10.1080/00207390701464675

Bing, T. (2008). *An epistemic framing analysis of upper level physics students' use of mathematics.* Dissertation. University of Maryland. Retrieved 2010-02-10 from
http://www.physics.umd.edu/perg/dissertations/Bing/

Björk, L-E. & Brolin, H. (2001). *Matematik 3000, kurs A och B.* Stockholm, Sweden: Natur och Kultur.

Björk, L-E. & Brolin, H. (2006). *Matematik 3000, kurs C och D.* Stockholm, Sweden: Natur och Kultur.

Bliss, J. (2008). Commonsense reasoning about the physical world, *Studies in Science Education*, 44(2), (pp. 123-155).
DOI: 10.1080/03057260802264149

Blum, W. & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects — State, trends and issues in mathematics instruction. *Educational Studies in Mathematics*, 22(1), (pp. 37-68).
DOI: 10.1007/BF00302716

Boesen, J., Lithner, J. & Palm, T. (2010). The relation between types of assessment tasks and the mathematical reasoning student use. *Educational studies in mathematics*, 75 (1), (pp.89-105).
DOI: 10.1007/s10649-010-9242-9

Cui, L. (2006). *Assessing college students' retention and transfer from calculus to physics* (Doctoral dissertation). Kansas State University. Retrieved 2012-06-27 from
http://krex.k-state.edu/dspace/handle/2097/183

Department of Applied Educational Science. (2011). Some earlier given test in Physics. Retrieved 2011-10-06 from http://pb-fy.edmeas.napb.se/information/tidigare_prov.asp diSessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction,* 10 (2-3), 105-225. DOI:10.1080/07370008.1985.9649008

Doorman, L. M. & Gravemeijer K. P. E. (2009). Emergent modeling: discrete graphs to support the understanding of change and velocity. *ZDM – The International Journal on Mathematics Education,* 41, 199-211.
DOI 10.1007/s11858-008-0130-z

Duval, R. (2002). Proof understanding in mathematics: What ways for students? In F.L. Lin (Ed.), *Proceedings of the international conference on mathematics-"Understanding proving and proving to understand"* (pp. 61-77). Taipei, Taiwan: National Science Council and National Taiwan Normal University.

Ekbom, L., Lillieborg, S., Larsson, S., Ölme, A. & Jönsson, U. (2004). *Tabeller och formler för NV- och TE-programmen, 5 ed.* Stockholm: Liber.

Engström, S. (2011). *Att vördsamt värdesätta eller tryggt trotsa. Gymnasiefysiken, undervisnings-traditioner och fysiklärares olika strategier för energiundervisning* (Doctoral dissertation) Mälardalen University. Retrieved 2011-10-07 from
http://mdh.diva-portal.org/smash/record.jsf?searchId=1&pid=diva2:310

Fidalgo, A. M. & Madeira, J. M. (2008). Generalized Mantel-Haenszel Methods for Differential Item Functioning Detection. *Educational and Psychological Measurement*, 68(6), 940-958.
DOI: 10.1177/0013164408315265

Forbus, K. D. (1981). Qualitative reasoning about physical processes. *Proceedings of the 7th IJCAI,* (pp. 326-330). Vancover, BC.

Forbus, K. D. (2004). Qualitative Reasoning. In A. B. Tucker (ed.) *Computer science handbook 2nd ed* (62-1 – 69-19). Boca Raton, Fla. : Chapman & Hall/CRC

Franken, Robert E. (2002). *Human motivation 5 ed*. Belmont, California: Wadsworth/Thompson Learning.

Gray, E. M. & Tall, D. O. (1994). Duality, Ambiguity and Flexibility: A Proceptual View of Simple Arithmetic. *The Journal for Research in Mathematics Education,* 26(2), 115-141.
Stable URL: http://www.jstor.org/stable/749505

Guilera, G., Gómez-Benito, J., Hidalgo, M. D. (2009). Scientific production on the Mantel-Haenszel procedure as a way of detecting DIF. *Psicothema* 21(3) 493-498.
ISSN 0214 – 9915

Hanna, G. (2000). Proof and its classroom role: A survey. In M.J. Saraiva et al (Eds.), *Proceedings of Conference en el IX Encontro de Investigacao en Educacao Matematica* (pp. 75-104). Funado.

Hanna, G. & Jahnke, H. N. (2002). Another Approach to Proof: Arguments from Physics. *ZDM – The International Journal on Mathematics Education*, 34 (1), 1–8
DOI: 10-1007/BF02655687

Harel, G. (2006). Mathematics Education Research, Its Nature, and Its Purpose: A Discussion of Lester's Paper. *ZDM – The International Journal on Mathematics Education,* 38 (1), 58-62.
DOI: 10.1007/BF02655906

Haylock, D. (1997). Recognising mathematical creativity in schoolchildren. *ZDM – The International Journal on Mathematics Education*, 29(3), 68-74.
DOI: 10.1007/S11858-997-0002-Y

Haynes, R., Sacket, D. L., Guyatt, G. H., & Tugwell, P. (2006). *Clinical Epidemiology: How to Do Clinical Practice Research 3rd edition*. Philadelphia: Lippincott Williams & Wilkins.

Holland, P. W. & Thayer, D. T. (1988). *Differential item performance and the Mantel Haenszel Procedure*. In H. Wainer & H. I. Braun (Eds) Test validity. Hillsdale, NJ: Lawrence Erlbaum Associates.

Howson, A. G. (ed.) (1988). *Mathematics as a service subject* ICMI study series, 99-0639469-2 based on the work of a symposium held in Undine (Italy), from 6 to 10 April, 1987. Cambridge: Cambridge University Press.

Johansson, H. (2013a). *Mathematical Reasoning Requirements in Swedish National Physics Tests.* Pre-print. Gothenburg.

Johansson, H. (2013b). *Mathematical Reasoning when solving physics tests. Relations between grades and kinds of tasks solved and seeking dependence between different kinds of reasoning.* Pre-print. Gothenburg.

Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education*, 16(2/3), 57-66.
ISSN: 10774610. Accession Number: 17971216

Leng, M. (2010). Preaxiomatic Mathematical Reasoning: An Algebraic Approach. In G.Hanna et al (Eds.), *Explanation and Proof in Mathematics, Philosophical and Educational Perspectives*. (pp. 47-57). New York, USA: Springer Science + Business Media, LLC 2010

Lesh, R., & Zawojewski, J. (2007). Problem Solving and modeling. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 763-804). Charlotte, NC: Information Age Publishing.

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics,* 67(3), 255-276.
DOI: 10.1007/s10649-00-9104-2

Lpf94. (2006). *Curriculum for the non-compulsory school system Lpf 94*. Stockholm: Fritzes.

Mannocci, A. (2009). The Mantel-Haenszel procedure. 50 years of the statistical method for confunders control. *Italian Journal of Public Health*, 6(4), 338-340.
DOI: 10.2427/5765

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute,* 22(4), 719-748.
DOI:10.1093/jnci/22.4.719

Marongelle, K. A. (2004).How Students Use physics to Reason About Calculus Tasks. *School Science and Mathematics,* 104(6), 258-272.
DOI: 10.1111/j.1949-8594.2004.tb17997.x

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models – 2nd edition.* Cambridge: University Press.

McDonald, J. H. (2009). *Handbook of biological statistics, 2nd edition*. University of Delaware, Baltimore: Sparky House Publishing

Michelsen, C. (1998). Expanding Context and Domain: A cross-Curricular Activity in Mathematics and Physics. *ZDM – The International Journal on Mathematics Education*, 30(4), 100-106.
DOI: 10.1007/BF02653149

Michelsen, C. (2005). Expanding the domain – variables and functions in an interdisciplinary context between mathematics and physics. In Beckmann, A., Michelsen, C., & Sriraman, B (Eds.). *Proceedings of the 1st International Symposium of Mathematics and its Connections to the Arts and Sciences* (pp.201-214)*.* The University of Education, Schwäbisch Gmünd, Germany.

Ministry of Education and Research (2001). Departementsserien (Ds) Ds 2001:48. *Samverkande styrning: om läroplanerna som styrinstrument*. Retrieved 2011- 10- 28 from
http://www.regeringen.se/content/1/c4/26/49/601d5449.pdf

National Council of Teachers of Mathematics (NCTM) (2000). *Principles and standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Nguyen, N-L. & Meltzer, D. (2003). Initial understanding of vector concepts among students in introductory physics courses. *American Journal of Physics,* 71(6), 630-638
DOI: 10.1119/1.1571831

Niss, M. (1993?). Mathematics in Society. In Biehler, R., Scholz, W., Sträβer, R. & Winkelmann, B. (eds.) *Didactics if Mathematical as a Scientific Discipline* (pp. 367 – 378). Dordrecht: Kluwer Academic Publisher.

Niu, W & Sternberg, R. J. (2006). The philosophical roots of Western and Eastern conceptions of creativity. *Journal of Theoretical and Philosophical Psychology*, Vol 26(1-2), 18-38. DOI: 10.1037/h0091265

Palm, T., Boesen, J., & Lithner, J. (2011). Mathematical Reasoning Requirements in Upper Secondary Level Assessments. *Mathematical Thinking and Learning, 13*(3), 221-246. DOI: 10.1080/10986065.2011.564994

Pólya, G. (1954). Mathematics and plausible reasoning (Vols. I and II). Princteton, NJ: Princeton University Press.

Pålsgård, J., Kvist, G. & Nilson, K. (2005a). *Ergo Fysik A*. Stockholm: Liber.

Pålsgård, J., Kvist, G. & Nilson, K. (2005b). *Ergo Fysik B*. Stockholm: Liber.

Ramstedt, K. (1996). *Elektriska flickor och mekaniska pojkar. Om gruppskillnader på prov – en metodutveckling och en studie av skillnader mellan flickor och pojkar på centrala prov i fysik.* (Doctoral dissertation), Umeå University. Umeå: UmU Tryckeri.

Redish, E. F. & Gupta, A. (2009). Making Meaning with Math in Physics: A Semantic Analysis. Contributed paper presented at GIREP 2009, Leicester, UK, August 20, 2009. http://www.physics.umd.edu/perg/talks/redish/GIREP2009/GIREP%20Semantics.pdf

Ribe, M. (1999). Oddskvoter berättar. *Välfärdsbullentinen*, 4, 14-15. Retrieved 2012-11-08 from http://www.scb.se/Grupp/valfard/_dokument/A05ST9904_10.pdf%E2%80%8E

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando: Academic Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for Research on Mathematics Teaching and Learning* (pp. 334-370). New York: MacMillan.

Sfard, A. (1991). On the dual nature of mathematical conceptions: reflections on process and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22(1), 1-36. DOI: 10.1007/BF00302715

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education,* 36(5), 404-411. Stable URL: http://www.jstor.org/stable/30034943

Star, J. R. (2007). Foregrounding procedural knowledge. *Journal for Research in Mathematics Education,* 38(2), 132-135. Stable URL: http://www.jstor.org/stable/30034953

Silver, E. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *ZDM – The International Journal on Mathematics Education*, 29(3), 75-80. DOI: 10.1007/s11858-997-0003-x

Swedish National Agency for Education. (2000a). Kursplan Fysik A Retrieved 2010-10-12 from http://www.Swedish National Agency for Education.se/forskola_och_skola/gymnasieutbildning/gymnasieskola_fore_ht_2011/2.3034/sok_amnen_och_kurser/13845/func/kursplan/id/3053/titleId/FY1201%2520-%2520Fysik%2520A

Swedish National Agency for Education. (2000b). Kursplan Fysik B Retrieved 2010-10-12 from http://www.Swedish National Agency for Education.se/forskola_och_skola/gymnasieutbildning/gymnasieskola_fore_ht_2011/2.3034/sok_amnen_och_kurser/13845/func/kursplan/id/3054/titleId/FY1202%2520-%2520Fysik%2520B

Swedish National Agency for Education. (2000c). Ämnet Fysik, gymnasieskolan Retrieved 2010-10-12 from http://www.skolverket.se/forskola-och-skola/gymnasieutbildning/gymnasieskola-fore-ht-2011/kursplaner/sok-amnen-och-kurser/13845/func/kursplan/id/3054/titleId/FY1202%2520-%2520Fysik%2520B?_xurl_=http%3A%2F%2Fsvcm.skolverket.se%2Fsb%2Fd%2F2503%2Fa%2F13845%2Ffunc%2Famnesplan%2Fid%2FFY%2FtitleId%2FFysik

Swedish National Agency for Education. (2001). *Programme Manual (GY 2000:19). Programme goal and structures, core subjects, subject index for upper secondary school*. Stockholm: Fritzes.

Swedish National Agency for Education. (2003). *Lusten att lära – Med fokus på matematik: Nationella kvalitetsgranskningar 2001-2002.* Stockholm, Sweden: Swedish National Agency for Education

Swedish National Agency for Education. (2005). *Skolverkets Provbank. Hur används den och vad tycker användarna?* Retrieved 2012-11-06 from http://www.skolverket.se/prov-och-bedomning/rapporter-och-film

Swedish National Agency for Education. (2009). *TIMSS Advanced 2008. Svenska gymnasieelevers kunskaper i avancerad matematik och fysik i ett internationellt perspektiv*. Stockholm: Fritzes

Swedish National Agency for Education. (2011) Tabell 3B, Retrieved 2011-10-05 from http://www.Swedish National Agency for Education.se/statistik_och_analys/2.1862/2.4391/2.4392

Swedish Schools Inspectorate. (2010). *Fysik utan dragningskraft. En kvalitetsgranskning om lusten att lära fysik i grundskolan.* Retrieved 2011-11-23 from from http://www.skolinspektionen.se/Documents/Kvalitetsgranskning/fysik-i-grundskolan/slutrapport-undervisningen-fysik.pdf

Sriraman, B. (2009). The Characteristics of Mathematical Creativity. *ZDM-The International Journal on Mathematics Education*, 41(1&2), 13-27.
DOI: 10.1007/s11858-008-0162-4

Tasar, M. F. (2010). What part of the concept of acceleration is difficult to understand: the mathematics, the physics, or both? *ZDM-The International Journal on Mathematics Education,* 42(5), 469–482.
DOI: 10.1007/s11858-010-0262-9

Tuminaro, J. (2002). How Student Use Mathematics in Physics: A Brief Survey of the Literature. *University of Maryland Physics Education Research Group*. Retrieved 2010-02-10 from http://www.physics.umd.edu/perg/math/UsingMath.pdf

Winter, H. (1978). Geometrie vom Hebelgesetz aus – ein Beitrag zur Integration von Physik- und Mathematikunterricht der Sekundarstufe I. *Der Mathematikunterricht* 24 (5), 88-125.

Wittman, M. C. (2002). The object coordination class applied to wave pulses: Analysing student reasoning in wave physics. *International Journal of Science Education*, 24 (1), 97-118. DOI:10.1080/09500690110066944

Wyndham, J., Riesbeck, E., Schoult, J (2000). *Problemlösning som metafor och praktik*. Linköping: Institutionen för tillämpad lärarkunskap, Linköpings univ.

Yackel, E., & Hanna, G. (2003). Reasoning and proof. In J. Kilpatrick, G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 227–236). Reston, VA: National Council of Teachers of Mathematics.

Paper 1

Mathematical Reasoning Requirements in Swedish National Physics Tests

Helena Johansson

Abstract

The ability to reason mathematically is widely accepted to be of great importance when learning physics. In this study, a sample of physics tests was analysed with respect to the following question: "What is the character of the mathematical reasoning that is required of students in upper secondary school to solve the tasks in physics tests from the Swedish national test bank?". In the analysis it was determined whether mathematical reasoning was required at all and if it was, that reasoning was characterised. The framework used distinguishes between imitative and creative mathematical reasoning. The analysis process consisted of structured comparisons between a representative student solution and students' learning history. The learning history was delimited to the textbooks and the content was compared to various key characteristics in the solution. Of the 209 analysed tasks, three-fourths required mathematical reasoning to be solved. Creative mathematical reasoning was required for one third of the tasks and two-fifths were solvable using imitative reasoning. The result confirms that mathematical reasoning is an important component when taking tests from the National Test Bank since a majority of the tasks requires mathematical reasoning.

Introduction

Mathematics and physics are historically closely intertwined and many mathematical concepts have been developed when needed in the description of the laws of nature. How this relation becomes apparent in a school context and how it might affect students' learning is discussed from different point of view in educational research. Some of the discussions depart from the learning of mathematics and how the relation to physics could influence this learning, below called *physics in mathematics.* Other discussions take a starting point in the learning of physics and discuss various aspects of the relation to mathematics, referred to as *mathematics in physics*.

Physics in Mathematics

Blum and Niss (1991) discuss the great value of keeping a close contact between mathematics and physics in school, since physics can provide good examples for validating mathematical models. They continue to discuss how separation between the subjects can lead to unnatural distances between the mathematical models and the real situation intended to model. In their paper, Doorman and Gravemeijer (2008) discuss the advantage of learning mathematical concepts through mathematical model building and how examples from physics are beneficial to symbolize the concepts. Hanna (2000) and Hanna and Jahnke (2002) propose that it is advantageous to use arguments from physics in mathematical proofs to make them more explanatory. They refer to Polya (1954) and Winter (1978) and continue discussing the benefits of integrating physics in mathematics education while learning and dealing with mathematical proofs. The importance of using physics to facilitate students' learning of various mathematical concepts is also discussed by Marongelle (2004). Using events from physics can help students to understand different mathematical representations.

Mathematics in Physics

Tasar (2010) discusses how a closer relation between the school subjects mathematics and physics can contribute to the understanding of physics concepts. A closer relation might also prevent the assumption that students already understand the mathematical concepts needed in physics (ibid.). A closer relation, noticed by Basson (2002), might also decrease the amount of time physics teachers spend on redoing the mathematics students need in physics. The "redoing" is likely a consequence of e.g. that *"physics teachers claim that their students do not have the pre-requisite calculus knowledge to help them master physics"* (Cui, 2006, p.2). Michelsen (2005) discusses how interdisciplinary modelling activities can help students to understand how to use mathematics in physics and to see the links between the two subjects. Redish and Gupta (2009) emphasize the need to understand how mathematics is used in physics and also the cognitive components of expertise, in order to teach mathematics for physics more effectively to students. Basson (2002) mentions how learning problems in physics not only depends on the complexity of the subject, but also on improper mathematical knowledge. Bing (2008) discusses the importance of learning the language of mathematics when studying physics. Nguyen and Meltzer (2003) analyse students' knowledge of vectors and conclude that there is a gap between students' intuitive knowledge and how to apply their knowledge in a formal way, which can be an obstacle when learning physics. Tuminaro (2002) has classified a lot of research on how students are using mathematics in physics according to the different approaches in the research papers.

In a study by Mulhall and Gunstone (2012) two major types of physics teachers are distinguished and grouped, the *Conceptual* and the *Traditional.* Mulhall and Gunstone conclude that a typical teacher in the conceptual group presumes that students can solve numerical problems in physics without a deeper understanding of the underlying physics. A typical opinion among teachers in the traditional group is that physics is mathematical and that a student develops an understanding of the physics by e.g. working with numerical problems. Doorman and Gravemeijer (2008) notice, (with reference to Clement 1985 and Dall'Alba et. al. 1993), that most of the effort both in physics and mathematics is on the manipulations of formulas, instead of focusing on why they work.

Mathematics in the syllabuses

The upper secondary school in Sweden is governed by the state through the curriculum and the syllabuses. In the curriculum are laid down the fundamental values that are to permeate the school's activities and also the goals and guidelines that are to be applied. The syllabuses, on the other hand, detail the aims and objectives of each specific course. They also indicate what knowledge and skills students must have acquired on completion of the various courses. The usefulness of mathematics is expressed in the syllabuses for mathematic –a core subject– as e.g. "The school in its teaching of mathematics should aim to ensure that pupils: develop confidence in their own ability to … use mathematics in different situations, …, develop their ability with the help of mathematics to solve … problems of importance in their chosen study orientation" (Swedish national Agency for Education, 2001, p.112).

According to the syllabus in physics, the teaching should aim to ensure that the students e.g. *develop their ability to quantitatively and qualitatively describe, analyse and interpret the phenomena and processes of physics in everyday reality, nature, society and vocational life; develop their ability with the help of modern technical aids to compile and analyse data, as well as simulate the phenomena and processes of physics* (Swedish National Agency for Education, 2000c). Explicitly, mathematics is important when making quantitative descriptions, and implicitly when analysing data, although the analysing part is mentioned in relation to technical aids. In the syllabuses for the two courses Physics A and Physics B, mathematics is mentioned more explicitly. In Physics A, the students should *be able to make simple calculations using physics models*[1] (Swedish National Agency for Education, 2000a). In Physics B there is more than one aim that includes mathematics. The student should *be able to handle physical problems mathematically*. They should also *be able to make calculations in nuclear physics using the concepts of atomic masses and binding energy* [1](Swedish National Agency for Education, 2000b). Physics B has Physics A as a prerequisite and the students should attain a deeper understanding for certain physics concepts when studying Physics B. It is also explicated that there are higher demands on the mathematical processing in Physics B (Swedish National Agency for Education, 2000c).

The literature review shows that there is a lot of educational research on the relation between the school subjects mathematics and physics. However, no studies on what kind of mathematical reasoning (see Theoretical framework for the definition) is required when learning physics were found. It seems natural and important to get a picture of these requirements. As a first step to this end, in this paper is studied what is the extent of and what kind of mathematical reasoning is required when taking a physics test from the Swedish National test bank. With this departure this study is suited among *mathematics in physics* research.

Theoretical framework

The definition of mathematical reasoning and the framework that is used for the analyses in this paper are developed by Lithner (2008). The framework has been developed through empirical studies on how students are engaging in various kinds of mathematical activities. The initial purpose of these studies was to analyse students' rote thinking and how this may affect their learning difficulties in mathematics. As a result, *reasoning* as "the line of thought adopted to produce assertions and reach conclusions in task solving" was defined (ibid. p. 257). *Mathematical reasoning* is used as an extension of a strict mathematical proof to justify a solution and is seen as a product of separate reasoning sequences. Each sequence includes a choice that defines the next sequence and the reason is the justification for the choice that is made. The mathematical foundation of the reasoning can either be *superficial* or *intrinsic*. The accepted mathematical properties of an object are of different relevancy in different contexts e.g. what type of problem one is trying to solve. This leads to a distinction between *surface* properties and *intrinsic* properties, where the former ones have little relevance in the actual context and leads to superficial reasoning, and the latter ones are central and have to be regarded (ibid. pp. 260 – 261). Depending on whether the reasoning is superficial or intrinsic, the framework distinguishes between *imitative* and *creative mathematical founded reasoning*.

[1] Authors translation

One example, described in Bergqvist, Lithner and Sumpter (2008), of when only surface properties are considered, is a student who tries to solve a max-min problem: “Find the largest and the smallest values of the function $y = 7 + 3x - x^2$ on the interval [-1, 5]”. This task can be solved with a straightforward solution procedure: One first uses that the function is differentiable on the whole interval to find all possible extreme points in the interval, (i.e. solve $f'(x) = 0$). If there are extreme points, the value at this point is calculated and compared with the values at the endpoints. In the situation described, the student does not remember the whole procedure, but reacts on the words largest and smallest and starts differentiating the function and solves $f'(x) = 0$. This calculation only gives one value and the answer demands two. Instead of considering intrinsic mathematical properties, the student seeks a method that will provide two values and instead solves the second degree equation $7 + 3x - x^2 = 0$. Two points are now obtained and the function values at these points are accepted as the solution by the student. Although the student gives these values as an answer, it is with some hesitation because the method used did not involve any differentiation, something remembered by the student to be related to a max-min problem.

Creative mathematically founded reasoning

Creativity is an expression often used in different contexts and without an unequivocal definition, (for a discussion see Lithner (2008, pp. 267-268)). Creativity in the framework that is used in this paper adopts the perspective of Haylock (1997) and Silver (1997), where creativity is seen as a thinking process that is novel, flexible and fluent.

Creative mathematical founded reasoning[4] *(CR)* fulfils all of the following criteria. (Lithner, 2008, p.266)

i. Novelty. A new reasoning sequence is created or a forgotten one is recreated.
ii. Plausibility. There are arguments supporting the strategy choice and/or strategy implementation motivating why the conclusions are true or plausible.
iii. Mathematical foundation. The arguments are anchored in intrinsic mathematical properties of the components involved in the reasoning.

Imitative reasoning

Imitative reasoning is categorised in *memorised reasoning* and *algorithmic reasoning*. The arguments that motivate the chosen solution method (i.e. the reasoning) could be anchored in surface mathematical properties.

Memorised reasoning (MR) fulfils the following conditions (Lithner, 2008, p. 258)

i. The strategy choice is founded on recalling a complete answer.
ii. The strategy implementation consists only of writing it down.

If some kind of calculations is required to solve the task, there is often no use in remembering an answer. Instead it is more suitable to recall an algorithm. The term algorithm is here used in a wide sense and refers to all the procedures and rules that are needed to reach the conclusion to a specific type of tasks, not only the calculations.

[4] Henceforward called *creative mathematical reasoning*.

Algorithmic reasoning (AR) fulfils the following conditions (ibid. p.259)

i. The strategy choice is to recall a solution algorithm, which, if it is followed step by step, will give the right answer without any demands of novelty.
ii. The remaining parts of the strategy implementation are trivial for the reasoner and just a careless mistake can obstruct the reaching of an answer.

Depending on the argumentation for the choice of the used algorithm, algorithmic reasoning is subdivided into three different categories: *Familiar algorithmic reasoning, Delimiting algorithmic reasoning* and *Guided algorithmic reasoning*. In this study, only the categories familiar algorithmic reasoning and guided algorithmic reasoning are used.

Familiar AR (FAR) fulfils (ibid. p. 262)

i. The reason for the strategy choice is that the task is identified as belonging to a familiar class of tasks which are known to be solved by a specific algorithm.
ii. The algorithm is implemented.

Text- guided AR (GAR) fulfils (ibid. p.263)

i. The strategy choice concerns identifying surface similarities between the task and an example, definition, theorem, rule or some other situation in a text source.
ii. The algorithm is implemented without verificative argumentation.

Local and global creative mathematical reasoning

Lithner (2008) introduces a refinement of the category *creative mathematical reasoning* (CR) into *local CR* (LCR) and *global CR* (GCR) that captures some significant differences between tasks categorised as LCR and GCR. This differentiation has been more elaborated by other scholars that have used the framework, e.g. Boesen, Lithner and Palm (2010) and Palm, Boesen and Lithner (2011). The difference between LCR and GCR is that in LCR, the reasoning is mainly MR or AR but contains a minor step that requires CR. If instead there is a need for CR in several steps, it is called GCR, even when some parts contain AR and/or MR. Important to stress is that as soon as CR is involved there has to be some understanding of the intrinsic mathematical properties in the task.

Non-mathematical reasoning

In the application of the framework for the analyses described in this thesis, an additional category is defined. This category consists of those tasks that can be solved by only using physics knowledge; and this category is called *non-mathematical reasoning* (NMR). Physics knowledge is here referred to as relations and facts that are discussed in the physics courses and not in the courses for mathematics, according to the syllabuses and textbooks, e.g. that angle of incidence equals angle of reflection. In the same way, a solution that requires mathematical reasoning refers to mathematics taught in courses at upper secondary school or assumed already to be known by the students according to the curricula.

Research Question

As mentioned in the introduction, several studies about the relations between the school subjects mathematics and physics have been found. However, there seem to be few studies specifically treating mathematical reasoning in physics. Neither regarding what kind of mathematical reasoning are required nor how different kinds might affect the learning of physics. By analysing the mathematical reasoning requirements of students when taking physics tests, this study attempts to contribute to parts of this domain. The motivations for this choice are that a requirement of mathematics when learning physics is explicated in the syllabuses and that students focus on learning what they meet in tests. With the definitions in the theoretical framework the research question is the following:

- What is the character of the mathematical reasoning that is required of students in upper secondary school to solve the tasks in physics tests from the Swedish national test bank?

Physics tests from the National Test Bank

About 12% of all students in the upper secondary school in Sweden attend the Natural Science Programme or the Technology Programme (Swedish National Agency for Education, 2011). In both programmes, the course Physics A is compulsory whereas the advanced course Physics B is elective. Physics B can be required when applying to the university or other higher educations and this has to be considered by the students when making their course decisions. The aim of the physics courses is that the students should attain various goals specified in the syllabuses. A substantial part of the physics education consists of performing laboratory work and solving problems (Mulhall & Gunstone, 2012). Written tests are commonly used as an assessment of the students' achievements. A student's grade on a course depends on how well the student has achieved the goals for the course (Swedish National Agency for Education, 2000a, 2000b). The descriptions in the syllabuses of the goals and the different grade levels are quite brief and the intention is that the syllabuses and curricula should be processed, interpreted and refined locally at each school. In order to accomplish equivalent assessment in physics, the Swedish National Agency of Education provides assessment supports, including the National Test Bank in Physics. In this respect the developed physics tests can be considered as a governmental concretisation of the syllabuses for physics. The character and the design of the tasks in tests stress what is covered in the taught curriculum. The tests also influence the teachers' interpretation of the syllabuses, which by extension stress what students focus on (Ministry of Education and Research, 2001; Swedish National Agency for Education, 2003). There might be a difference between the national tests and teacher made tests; this is not investigated in this study.

The material in the Test Bank is classified as secret and could be accessed via the Internet by teachers who have received a password that lasts until the summer vacation the current academic year. The material consists of single tasks to choose from or complete tests that comprise the goals for Physics A or Physics B. The test material in the National Test Bank for Physics is developed by the Department of Applied Educational Science at Umeå University, who has had this commission since shortly after the new national curriculum was implemented in 1994. In total there are so far 847 tasks to choose from and 16 complete tests for each of the courses Physics A and Physics B, all classified as secret. The first tests are from 1998 and the latest is from spring 2011. In addition, there are five tests for each course open for students to practice on. In this way the students get an idea of what the tests look like and what is required when taking a test. (Department of Applied Educational Science, 2011). As opposed to national tests in e.g. mathematics, the teachers are not obligated to use the test from the National Test Bank in e.g. physics. However, it is common that these tests are used as a final exam in the end of the physics courses (Swedish National Agency for Education, 2005).

Since the start there has been a change in the design of the tasks. In the beginning there was more or less only one correct solution to each task. This has evolved to a higher degree of open tasks which could be solved using different approaches. The last ten years the final task has been an "aspect-task" that is assessed according to the achieved level in different assessment groups. The task should be easy to start with, but it should also include a challenge to more proficient students. The first three years, i.e. 1998- 2000, there was an experimental part included in the tests (this part is not included in the analysis in this study).

Methods

In this study, the following sample of the tests in the National Test Bank was analysed. Tests for the course Physics A; December 1998, May 2002 December 2004, May 2005 and December 2008; and for Physics B; May 2002, May 2003, May 2005, February 2006 and April 2010. The tests were randomly selected with respect to two constraints. First, that there should be five tests for each course Physics A and Physics B. Second, that there had to be two or three tests from each one of the courses that were not classified as secret; this in order to have the possibility to discuss examples in the analysis.

Categorisation of Mathematical Reasoning requirements

The objects of study were the mathematical reasoning required from students to solve the tasks in the physics tests from the National Test Bank. No students with their actual solutions were included in the study. Required reasoning refers to what kind of reasoning that is sufficient to solve a task, and the chosen framework gives the possibility to determine this. In order for a task to be categorised as requiring Algorithmic reasoning (AR) or Memorised reasoning (MR), the student should be able to recognize the type of task. This could be assumed if identical or similar tasks have been encountered on several occasions in the textbooks. Whether the solution to a task requires creative mathematical reasoning or is just a routine procedure depends on whether the solution requires some novelty. This in turn depends on the education history of the solver (cf. Björkqvist, 2001; English & Sriraman, 2010; Lesh & Zawojewski, 2007; Schoenfeld, 1985; Wyndhamn et. al., 2000.)

When analysing the tasks in the tests and categorising the required reasoning, it is necessary, (as in all research), to reduce the complexity. Choosing the framework that is used for the analysis is one of the first reductions that are made. Another reduction is due to the fact that no students are present in this study and, therefore, there is no actual learning history to consider. According to studies on what the education in mathematics looks like, it seems that most of the learning activities consist of students working with their textbooks (Swedish National Agency for Education, 2003). In an evaluation of the physics education in secondary school, one of the results is that the teaching is guided by the textbooks. The evaluation is done on 35 schools with a lower score than average on the national physics test for grade nine, so no claims are made that it represents the general procedure of the physics education (Swedish Schools Inspectorate, 2010). Also, The Ministry of Education and Research (2001) discusses the fact that textbooks and assessment are seen as two of the most important control factors of the education. In a qualitative study of a physics class, Engström (2011) shows that the textbook still occupies a main part of the guidance of the education and the report of TIMMS advanced 2008, shows that teachers mostly use the textbook in physics to choose and solve problems from (Swedish National Agency for Education, 2009). In a study by Boesen et. al. (2010) real students´ actual type of mathematical reasoning on categorised tasks on tests in mathematics was compared with the categorisation. The result showed that when the tasks were similar to the tasks in the textbooks, students were trying to recall algorithms in order to reach a solution. If the tasks did not have the same properties as the ones in the textbooks, mostly creative mathematical reasoning was used. With reference to the results above, the students' prior learning in physics and mathematics is in this study equated with the content in the textbooks.

In this study textbooks in both mathematics and physics were considered. When writing the tests, the students are allowed to use a handbook in physics designed for the physics courses in upper secondary school. The access to formulas and definitions in this handbook also has to be taken into account when analysing the tasks. The textbooks and the handbook were chosen among the books commonly used in the physics courses in upper secondary school. The books used for categorisation of the tasks in Physics A tests where "Ergo Fysik A" (Pålsgård et. al., 2005a) and "Matematik 3000 Kurs A och B" (Björk & Brolin, 2001). For tests in Physics B "Ergo Fysik B" (Pålsgård et. al., 2005b) and "Matematik 3000 Kurs C och D" (Björk & Brolin, 2006) where used. The books from Physics A were also considered because Physics A is a prerequisite to be able to take Physics B. The handbook chosen was ”Tabeller och formler för NV- och TE- programmen" (Ekbom et. al., 2004). Even if not all students in the Swedish upper secondary school are using the books above, it is a reasonable assumption of the learning situation for a conceivable student.

The procedure for analysing the tasks is given by the chosen framework and an analysis sheet was used to structure the procedure. The steps comprised in the procedure are outlined below and are used earlier in e.g. Palm et.al. (2011).

I. Analysis of the assessment task- Answers and solutions
    d) Identification of the answers (for MR) or algorithms (for AR)
    e) Identification of the mathematical subject area
    f) Identification of the real life event

II. Analysis of the assessment task- Task variables
    1. Assignment
    2. Explicit information about the situation
    3. Representation
    4. Other key features

III. Analysis of the textbooks and handbook – Answers and solutions
    a) In exercises and examples
    b) In the theory text

IV. Argumentation for the requirement of reasoning

Analysis of the assessment task- Answers and solutions: When analysing the different assessment tasks with respect to answers and solutions, first, a solution plausible to think a student would use was constructed by the researcher and written down. This is another reduction of the complexity, due to the fact that students were not the object of study. Justification for the fact that the constructed solution is a *plausible* student solution comes from the researcher's experience as a teacher, as well as from considering the solution proposals and the scoring rubric provided with the test. The solution was then looked at with mathematical glasses on and categorised according to relevant mathematical subject areas for the solution, e.g. does the solution include working with formulas, algebra, diagrams, solving equations, etc. Solutions not including any mathematical object were also identified in this step and these tasks were categorised as non-mathematical reasoning (NMR) i.e. tasks solvable without any mathematical considerations. As a consequence of the added category NMR, this step is an addition to the original procedure used in previous studies. Mathematical objects refer to entities to which mathematics is applied. As mentioned above, in the present study mathematics refers to school mathematics introduced in mathematics courses given for students at upper secondary school or mathematics assumed already to be known according to the curriculums. Identification of a "real-life" event refers to tasks where the described situation could give a clue to a known algorithm that solves the problem (see the analysed example below)

Analysis of the assessment task- Task variables: After identification, the next step in the procedure was to analyse the solution according to different task variables. The first variable is the explicit formulation of the assignment. The second variable is *what* information about the mathematical objects that were given explicitly in the information, contrary to what information the students for example need to receive from the handbook or have to assume in order to reach a solution. The third task variable concerns *how* the information was given in the task, e.g. numerically or graphically, interwoven in the text or explicitly given afterwards. This is what above (II.3) is referred to as representation. The task can also include key words, symbols, figures, diagrams or other important hints the student can use to identify the task type and which algorithm to use. These features are gathered in the fourth task variable.

Analysis of the textbooks and handbook – Answers and solutions: The third step in the analysis process focuses on the textbooks and the handbook. Formulas used in the solution algorithm were looked for in the handbook and the available definitions were compared to the constructed solution of the task. The textbooks were thoroughly looked through for similar examples or exercises that were solved by a similar algorithm. The theory text was also regarded to see whether it contained a description of how similar or identical problems could be solved.

Argumentation for the requirement of reasoning: Finally an argumentation based on the previous steps was made by the researcher for every task, to motivate the concluded reasoning requirement, i.e. is it possible for a student to solve the task using a reasoning type based only on superficial mathematical properties or is it necessary to use creative mathematical reasoning? In order to be categorised as familiar algorithmic reasoning there must have been at least three tasks considered as similar in the textbooks. It can then be assumed likely that the students remember the algorithm, which might not be the case if there are fewer occasions. That three is an appropriate assumption is supported in the study by Boesen et.al. (2010). If the task is similar to a formula or definition given in the handbook, it is assumed that the student can use this as guidance in order to solve the task. It is then enough with only one similar, previously encountered, example or exercise for the task to be regarded requiring text-guided algorithmic reasoning. To be categorised as requiring memorised reasoning, tasks including the same answer or solution should have been encountered at least three times in the textbooks. It is then assumable that the student can copy the answer. If none of the above reasoning types are sufficient for solving the task and there is a need to consider some intrinsic mathematical property, the task is categorised to require some kind of creative mathematical reasoning.

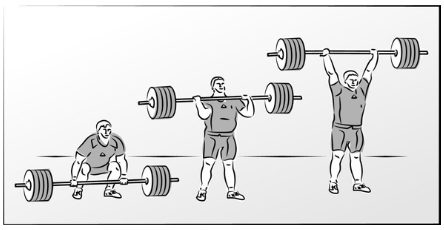
During the analysing process, both tasks where the categorisation was straightforward and tasks where the categorisation could be considered as border-line cases occurred. Examples of different types of analyses are given in the "Data and Analysis" section. Typical examples of the different kinds of categorisation were continuously discussed in a reference group consisting of a mathematician and a mathematics education researcher.

Data and Analysis

The examples below are chosen to represent and illustrate the different types of analysis and the categorisation made of the tasks in the national physics tests. The idea is that the required reasoning is represented by the constructed solutions. All of the tasks are chosen from public tests. Normally, subtasks are treated separately because the task variables and the analysis of the textbooks can be different. The first three tasks are examples of typical categorisations, where no hesitation concerning the categorisations' occurred.

Task 1 ("The Weightlifter (a)")

"A weightlifter is lifting a barbell that weighs 219 kg. The barbell is lifted 2.1 m up from the floor in 5,0 s.



a) What is the average power the weightlifter develops on the barbell during the lift?"

Analysis of the assessment task- Answers and solutions: A typical solution of an average student could be derived by the relation between power and change in energy during a specific time. Change in energy is here the same as change in potential energy for the barbell. Multiply the mass of the barbell with the acceleration of gravity and the height of the lift, and then divide by time to get the power asked for. The mathematical subject area is identified as algebra, in this case to work with formulas. The identification of the situation to lift a barbell can trigger the student to use a certain solution method and is therefore included in this analysis as an identified "real-life" situation.

Analysis of the assessment task- Task variables: The assignment is to calculate the average power during the lift. The mass of the barbell, the height of the lift and the time for the lift are all considered as mathematical objects. As mentioned above, an object is the entity one is doing something with. In this example, all the objects are given explicitly in the assignment in numerical form. In the presentation of the assignment there is also an illustrative figure of the lift.

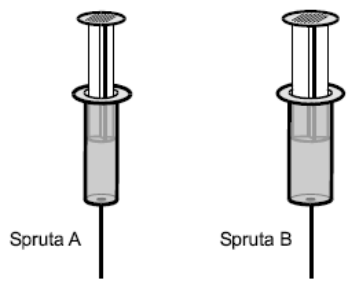
Analysis of the textbooks and handbook – Answers and solutions: The formula for power, $P=\Delta W/\Delta t$, with explanation "ΔW = the change in energy during time Δt" is given in the handbook (p.105). So is the formula for "work during lift" $W_l = mg \cdot h$, with the explanatory text "A body with weight mg is lifted the height h. The lifting work is…" (p.104), and the formula for potential energy with the text "A body with mass m on the height h over the zero level has the potential energy W_p= mg·h" (p.104). In the mathematics book Ma3000AB, there are plenty of examples and exercises of how to use formulas, e.g. on pages 28-30. In the physics book ErgoA, power is presented as work divided by time and work is in one instance exemplified as lifting a barbell. On page 130 there is an identical example where the power during the lift of a barbell is calculated. There is one example on page 136 where work during a lift is calculated in relation to change in potential energy. Exercise 5.05 is solved by calculating work during a lift and exercise 5.10 is also solved by a similar algorithm.

Argumentation for the requirement of reasoning: The analysis of the textbooks shows that there are more than three tasks similar to the task up for categorisation with respect to the task variables, and these tasks can be solved with a similar algorithm. As mentioned in the method section, if the students have met tasks solvable with a similar algorithm at least three times, it is assumable that they remember the solution procedure. This task is then categorised as solvable using imitative reasoning, in this case FAR.

Task 2 ("The Weightlifter (b)")
"A weightlifter is lifting a barbell that weights 219 kg. The barbell is lifted 2.1m up from the floor in 5.0 s.
b) What is the average power the weightlifter develops on the barbell when he holds it above the head during 3.0 s?"

Analysis of the assessment task- Answers and solutions: It is not necessary to use any mathematical argumentation in order to solve this task. The solution can be based only on physical reasoning; there is no lifting and therefore no work is done, which in turn implies that no power is developed. The task is therefore categorised as solvable with non-mathematical reasoning. This task is a typical example of an analysis resulting in the NMR categorisation.

Task 3 ("The Syringes")
"A patient is going to get an injection. The medical staffs are reading in the instructions that they are supposed to use a syringe which gives as low pressure as possible in the body tissue. Which of the syringes A or B shall the staff choose if the same force, F, is applied and the injection needles have the same dimensions?" *Argue for the answer*



Analysis of the assessment task- Answers and solutions: To solve this task the student can use the relation between pressure, force and area, p=F/A. Neglect the hydrostatic pressure from the injection fluid. If the force that the staff applies to the syringe is the same, it is the area of the bottom that affects the pressure; the larger the area the less the pressure. The staff should choose syringe B.

The mathematical subject area is identified as algebra, to work with formulas, and also proportionality.

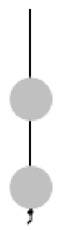
Analysis of the assessment task- Task variables: The assignment is to argue for, and to choose which syringe that gives the minimum pressure. Only the force is given as a variable, represented with a letter. Key words for the students can be *force* and *pressure*. The situation is illustrated with a figure where it appears that syringe B has a greater diameter than A.

Analysis of the textbooks and handbook – Answers and solutions: In the handbook, the relation p=F/A is defined. Proportionalities are discussed and exemplified in Ma3000AB, but are not used for general comparisons. There is one example in Ergo A about how different areas affect the pressure and also one exercise that is solved in a similar way, using a general comparison between different areas and pressure.

Argumentation for the requirement of reasoning: There is only one example and one exercise that can be considered somehow similar with regard to task variables and solution algorithm. The formula is in the handbook, but there has to be some understanding of the intrinsic properties in order to be able to use the formula in the solution. This task was therefore considered requiring some creative mathematical reasoning, in this case GCR, in order to be solved.

As mentioned in the "Method" section, during the analysis process situations occurred when the analysis was not as straight forward as in previous examples. All these tasks were discussed in the reference group and below are some examples of the borderline cases that arose.

Task 4 ("Charges on a thread")
"In order to determine the charge on two small light silver balls, the following experiment was conducted. The balls, which were alike, weighed 26mg each. The balls were threaded on a nylon thread and were charged in a way that gave them equal charges. The upper ball levitated freely a little distance above the other ball. There were no friction between the balls and the nylon thread. The distance between the centres of the balls was measured to 2.9 cm. What was the charge on each of the balls?"



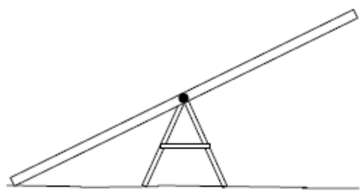
Analysis of the assessment task- Answers and solutions: To derive at a solution, the forces acting on the upper ball must be considered. Because it is levitating freely, it is in equilibrium and according to Newton's first law the net force on the ball is then zero. The forces acting on the ball are the gravitational force, $F = mg$, (downwards) and the electrostatic force from the ball below, $F = kQ_1Q_2/r^2$, (upwards). Put these expressions equal and solve for Q_1, using that $Q_1 = Q_2$, which will give the charges asked for. The mathematical subject area is identified as algebra, to work with formulas and to solve quadratic equations.

Analysis of the assessment task- Task variables: The assignment is to calculate the charges on the balls. The mass on each of the balls and the distance between their centres are mathematical objects given numerically and explicitly in the assignment. The information of the charges' equal magnitude is textual and a part of the description of the situation. There is also an additional figure of the balls on the thread, illustrating the experiment.

Analysis of the textbooks and handbook – Answers and solutions: Coulomb's law, $F = k \cdot Q_1Q_2/r^2$ (the formula for electrostatic force), is given in the handbook (p. 108) with explanation "r=distance between the charges and … k =… ≈ $8.99 \cdot 10^9$ $Nm^2/(As)^2$". In the mathematics book Ma3000AB there are plenty of examples and exercises of how to use formulas, e.g. on pages 28-30 and of solving quadratic equations on page 269. In the physics book ErgoA, Coulomb's law is introduced and exemplified and there are at least three exercises of calculating the charge on different objects using this law. There is one example of a levitating charge (p. 227), but in this case in a homogeneous electrical field instead of due to the electrostatic force from another charged particle. There are also two exercises of similar situations as in the example. Newton´s first law is formulated in the theory text (p. 91) where it reveals that the net force has to be zero if an object for example is at rest, and this relation is used on several different occasions in ErgoA. The gravitational force is introduced on pages 92 and is then used throughout the book.

Argumentation for the requirement of reasoning: Considering the mathematical reasoning, there are more than three examples or exercises in the textbooks where the same algorithm has been used, i.e. to put two expressions equal, solve for one unknown variable, including taking the square root. But there are not three or more occasions considering the physics context. To solve the task the student must first identify the force situation in order to know which expressions to equate. After having discussed this task in the reference group, it was concluded that analysing the physics context does not belong to the mathematical reasoning. Although mathematical reasoning is necessary to be able to solve the task, it is not enough, and although the mathematical reasoning can be considered as some kind of algorithmic, the task was categorised to require local creative reasoning LCR, where the minor step is to analyse the physics.

Task 5 ("The seesaw")
"How can Lars, 70 kg, and his son Anton, 28 kg, place themselves on a 3.5 m long seesaw so that it stays in equilibrium?"



Analysis of the assessment task- Answers and solutions: This task can be solved using the equilibrium of torque (moment of force), $M = Fr$, i.e. the torque with respect to Anton must have the same magnitude as the torque with respect to Lars. The forces that act on the seesaw is of the same magnitudes as the gravitational forces, $F= mg$, on Lars and Anton respectively. Assuming that Anton is placed 1.60 m from the rotation axis, one gets the equation $F_Lr = F_A \cdot 1.60$, which will give the answer together with the assumption of Anton's position. As in the examples above, the mathematical subject area was identified as algebra, more specifically to work with formulas and equations. To swing a seesaw is a real-life situation often used as examples in mechanics and therefore included in the analysis.

Analysis of the assessment task- Task variables: The assignment is to show where on the seesaw Anton and Lars can sit when it is in equilibrium. Mathematical objects that are given numerically in the assignment are the masses of Anton and Lars. In addition, the total length of the seesaw is given and there is a picture of a seesaw without any people on it.

Analysis of the textbooks and handbook – Answers and solutions: The formula for Torque is given in the handbook (p. 100), M=Fr with explanatory text "r is the perpendicular distance from the rotation axis to the line of action of the force. At equilibrium $\sum Fr = \sum M = 0$". There is also a figure in the handbook of M around a rotation axis with F and r marked. In Ma3000AB there are plenty of examples and exercises on how to use formulas (e.g. on pages 28-30) and of how to solve equations. In ErgoA, the relation for torque is formulated with words in the theory text. When introducing torque, the theory also refers to swinging a seesaw, both in text and with images (p. 105). There are two examples that use the formula for torque, as defined in the handbook. One of the examples is similar to this task except that one does not have to assume any distance. There are some exercises using a similar algorithm, but for calculating masses (via force) from given distances instead of distances from given masses.

Argumentation for the requirement of reasoning: The algorithmic procedure to solve a task involving a seesaw has been met both in the theory text and in examples. There are plenty of exercises of how to handle expressions and solving equations with one unknown variable. The difference in this case is that none of the distances are given in the assignment. There are therefore two unknown variables in the expression and one of the distances has to be assumed, by using the information about the total length of the seesaw. After discussion about this task, it is categorised as requiring local creative mathematical reasoning (LCR). The minor step in this case is to realise that one has to make an assumption of one of the distances in order to be able to solve the task, and this is regarded as demanding some intrinsic mathematical understanding.

Task 6 ("Walking in water")

"You are walking out into the water at a beach with a stony bottom. In the beginning, it hurts very much under your feet when you are walking on the stones, but when the water gets deeper it starts to feel less. When the water reaches you up to the chest, the stones do not feel as painful anymore. Explain this."

Svar: ____

____

____

____

____

Analysis of the assessment task- Answers and solutions: To solve this task the students are supposed to refer to Archimedes' principle. The more of the body being under the water, the bigger is the buoyant force from the water. Assuming the body is in equilibrium at each step, the bigger the buoyant force is, the smaller is the normal force from the stones, and consequently the pressure from the stones. Therefore, it hurts less when the water level reaches higher on the body. This relation can be argued for, using the formulas for Archimedes' principle, for pressure and the equilibrium of forces. The mathematical area could then be considered involving formulas and proportionality. Following the solution proposal and the scoring rubric provided with the test, there is no need to use any mathematical relations/formulas to argue for the answer.

Analysis of the assessment task- Task variables: The assignment is to explain why it does not hurt as much when you are in deeper water. No mathematical objects are given explicitly in the task. The situation refers to a real-life event, walking in the water. Bathing is a common situation referred to when discussing Archimedes' principle. The depth of the water is also indicated in the assignment as important.

Analysis of the textbooks and handbook – Answers and solutions: In the handbook, Archimedes' principle is formulated with the words "The buoyant force on an object is equal to the weight of the displaced fluid" (p. 98). On the same page is the formula for pressure, p=F/A, given. In Ma3000AB there are plenty of examples and exercises on how to use formulas, e.g. on pages 28-30, and exercises on proportionality on pages 73 and 75, but not used for general comparison. In ErgoA Archimedes' principle is formulated with words and as an expression (p. 171) and there is one example that relates volume to the buoyant force.

Argumentation for the requirement of reasoning: When following the scoring rubric of what is demanded of a student to solve this task, there is no need to refer to the formulas and use them to argue for the given explanation. The student needs to mention Archimedes' principle and that the buoyant force increases when the volume of the body in the water increases, but he/she does not need to explain why or show how the volume increase implies the force increase. They also have to mention something about how this increased buoyant force decreases the normal force, but according to the scoring rubric there is no need to use the relation for pressure to show why this decreased normal force makes it hurt less. The space given to write the answer on also indicates that a few lines are enough as an answer. After discussing this task and the minimum solution that is required of a student, it is decided that the reasoning is mainly physical and that mathematical reasoning is not necessary to solve this task. It is then categorised as solvable with non-mathematical reasoning, NMR.

Results and Conclusion

The main result shows that 76 % of the 209 analysed tasks required mathematical reasoning to be solved. A majority of these were solvable using imitative reasoning (IR) and the remaining ones required creative mathematical reasoning (CR) (Table 1).

Table 4. Categorisation result, overview

	Number of tasks	Non-mathematical reasoning (NMR) %	Creative mathematical reasoning (CR) %	Imitative reasoning (IR) %
Physics A	103	21	33	46
Physics B	106	26	38	36
Total	209	24	35	41

The result also shows some differences in the categorisation with respect to the various courses Physics A and Physics B. There were slightly more non-mathematical reasoning (NMR) tasks in the Physics B tests than in the tests for Physics A and the same goes for the CR-tasks. A more distinct difference was notable among the IR-tasks, but here with the greater lot in Physics A tests (Table 4). A majority of the IR-tasks (78%) were solvable with familiar algorithmic reasoning (FAR) and the rest with guided algorithmic reasoning (GAR). The CR-tasks were separated into local creative mathematical reasoning (LCR) and global creative mathematical reasoning (GCR). In general, Physics B tests consisted of more GCR-tasks than Physics A tests while the amount of LCR-tasks was almost the same (Table 5).

Table 5. Categorisation result, detailed.

	Number of tasks	Nbr of NMR	NMR %	Nbr of FAR	Nbr of GAR	IR %	Nbr of LCR	Nbr of GCR	CR %	GCR %	IR+LCR %
Physics A Dec98	20	3	15	6	6	60	4	1	25	5	80
Physics A May02	20	4	20	3	3	30	5	5	50	25	55
Physics A Dec04	19	4	21	7	1	42	2	5	37	26	53
Physics A May05	19	5	26	6	2	42	4	2	32	11	63
Physics A Dec08	25	6	24	12	1	52	4	2	24	8	68
Total Physics A	103	22	21	34	13	46	19	15	33	15	64
Physics B May02	18	2	11	7	0	39	5	4	50	22	67
Physics B May03	19	5	26	8	1	47	3	2	26	11	63
Physics B May05	23	7	30	4	3	30	5	4	39	17	52
Physics B Feb06	23	10	43	8	0	35	2	3	22	13	43
Physics B April10	23	4	17	5	2	30	4	8	52	35	48
Total Physics B	106	28	26	32	6	36	19	21	38	20	54
Total	209	50	24	66	19	41	38	36	35	17	59

When comparing tests from different years, the analysis showed a notable variation in the proportions of the different mathematical reasoning types. There is no consistency among the tests with respect to this analysis (Table 5). From the analysis it was possible to subdivide the tasks, initially categorised as LCR, into five subcategories depending on the nature of the minor creative step that was required (some are exemplified in the Data and Analysis section).

1. If it is obvious that a familiar formula/relation is going to be used, but for reasoning about a situation with respect to the variables instead of calculating some values.
2. If a familiar diagram is used in a slightly different way from what is earlier met in examples/exercises.
3. If there is guidance in the handbook, but some intrinsic mathematical understanding is needed to be able to use this.
4. If some values that have to be used in a familiar algorithm/formula depend on earlier calculations or have to be assumed.

5. If a solution requires that “two” simpler algorithms need to be combined in a new way. In these cases it is often not the mathematical aspect that is the difficulty; more often it is the physical understanding/knowledge that is the local step. Task 4 in the Data and Analysis section is a typical example of this subcategory.

Table 6. LCR subcategories.

Five subcategories of LCR	1. If it is obvious that a familiar formula/relation is going to be used, but for reasoning about a situation with respect to the variables instead of calculating some values **Number**	2. If a familiar diagram is used in a slightly different way from what is earlier met in examples/exercises **Number**	3. If there is guidance in the handbook, but some intrinsic mathematical understanding is needed to be able to use this **Number**	4. If some values that have to be used in a familiar algorithm/formula depend on earlier calculations or have to be assumed **Number**	5. If a solution requires that “two” simpler algorithms need to be combined in a new way. In these cases it is often not the mathematical aspect that is the difficulty; more often it is the physical understanding/knowledge that is the local step **Number**
Physics A	6	2	1	3	7
Physics B	3	3	3	6	4
Total	9	5	4	9	11

Discussion and implications

The fact that a majority of the tasks requires mathematical reasoning shows that the ability to reason mathematically is important when taking physics tests from the test bank. To be able to solve a major part of the tasks in a test, only imitative reasoning is not enough but creative mathematical reasoning is required. Different kinds of tasks require different kinds of reasoning. Tasks requesting analysing answers more often require CR, while tasks solvable with a calculation more often are categorised as IR-tasks. Choosing tasks with different properties will then influence the reasoning requirements. A comparison between national tests and teacher made tests in mathematics shows a significant difference in the mathematical reasoning requirements (Palm et. al., 2011). The subcategorisation of the LCR-tasks indicates that the mathematical reasoning requirements in the physics tests could depend on other properties of the physics tasks. To say more about how different task properties influence the reasoning requirements, further studies are needed. One purpose of the national tests is to help teachers interpret the syllabuses for the physics courses. Thus, a national test in physics could be assumed to cover most of the different domains in physics for upper secondary school in Sweden. The data from the analysis could then be used to investigate whether there are certain mathematical subject areas that are required to be able to solve physics tasks in the national tests. And further, to analyse if different subject areas requires different kinds of mathematical reasoning. Is there for example more tasks requiring CR if the subject area is calculus than if the area is algebra? As mentioned in the introduction, certain mathematical subjects could be troublesome for students when learning physics. In order to say more about what kind of mathematical reasoning requirements students are put in front of as well as what kind of mathematical reasoning they have the opportunity to practice when learning physics in school, more comprehensive studies of students’ learning environment have to be made.

The total scores on the analysed tests vary between 38 to 48 points. The points are divided in “pass” points and “passed with distinction” points. To pass one of the tests 12 points are required and the kind of the points does not matter. To get one of the higher grades, “Pass with distinction” or “Pass with special distinction”, the limit varies between 24 to 26 points for the different tests. Some of these points have to be “passed with distinction” points where the lower limit varies between five and seven points for the grade “Pass with distinction” and between 12 and 13 for the grade “Pass with special distinction”. The result then implies that it is possible to pass a test only using IR or IR together with NMR. In this analysis no concern to the different kinds of points has been taken when categorising the tasks. However, from the result that a third of the tasks require CR, it is reasonable to assume that to get a higher grade some kind of CR will often be required. The consistency in this conclusion is examined further in another study (Johansson, 2013b). This paper also includes an analysis of a possible dependence between how students succeed on tasks requiring CR if they have solved tasks requiring IR. To be able to say more about the relation between mathematical reasoning and learning in physics, a comparison between how students succeeded on tests in mathematics and on the test analysed in this study would be an interesting future study.

This study does not have the intention to be an evaluation of the physics test from the National Test Bank. During the analysis, however, some observations with this character came up that seem important to comment on. The result shows that the analysed tests are aligned with the syllabuses and curriculums, in which the importance of mathematics is explicated (see Introduction subsection Mathematics in the syllabuses). This result might be an expected result, as national tests provide assessment support to teachers and indicate how to interpret the syllabuses/curriculums. According to the fact that tests stress what is focused on, the necessity of mathematical reasoning is thus one aspect communicated to teachers and students. The possibility for students to solve two-fifths of the tasks in the tests with IR might influence them to focus on various standard procedures and on surface properties/similarities in tasks met in the physics courses, rather than learning the underlying mathematics/physics. In this respect the result increases the knowledge of what kind of mathematical reasoning that is communicated through the national tests in physics and how this might influence the focus of the students’ learning.

It is clear that the analysing procedure cannot capture everything in a learning situation. One reduction is to equate the learning history with the textbooks. The actual number of IR-tasks could therefore be larger for some students than shown in the result; if, for example, the students have met additional tasks in the classroom. At the same time, categorising a task to be solvable with IR does not exclude a student from using a creative solution. The categorisation IR represents a minimum requirement under the assumption that a student has read the textbooks and gone through all of the examples and exercises.

The analysis of NMR-tasks was often straightforward as exemplified in ”The Weightlifter (b)”- task (Data and Analysis section), but on a few occasions borderline cases occurred as shown in the “Walking in water”-task (Data and Analysis section). The same holds for the other categories as well. All borderline cases were discussed in the reference group and also several of the other tasks. The thorough description of the analysis process for six of the tasks is included to ensure a high reliability and validity for the method used. A limitation in the analysing procedure is revealed in the sub-categorisation of the LCR-tasks, while the researcher’s experience as a teacher tells that several of these tasks have more in common with IR-tasks than with CR-tasks. Most tasks of types 2, 3 and 5 (Table 6) are of this kind. Considering the IR-tasks together with the LCR-tasks belonging to subcategories 2, 3 and 5; then approximately two-thirds of the analysed tasks requiring mathematical reasoning can be solved using these kinds of reasoning. At the same time, the omitting of experience could be considered a strength, as it prevents possible preconceptions to influence the result.

As noticed in Table 5, a comparison between tests from different years showed a notable variation in the proportions of the different mathematical reasoning types. The fluctuation might be a consequence of the test developers' use of another framework for categorising the tasks according to the goals in the syllabuses. On average there were more CR-tasks in the Physics B tests than in the tests for Physics A. Scrutinising this result for each test shows that this result varies over the years. This variation/fluctuation needs to be analysed further before making any general conclusions. If, however, the average result holds, then CR can be argued to be more important in Physics B. The syllabuses explicate that there are higher demands on the mathematical processing in Physics B, but say nothing of the kind of processing. Thus this mathematical processing might as well be IR. The same can be applied to a comparison of NMR-tasks for the various courses over different years. On average, there are more NMR-tasks in Physics B tests than in Physics A tests. If this result is general it could be considered consistent with the syllabuses, which state that students should develop a deeper understanding for some of the physical concepts in Physics B compared to Physics A. Describing the mathematical reasoning requirements in the physics tests can provide an alternative framework for national test developers to decide whether the tests assess what is intended according to the curriculum.

References

Basson, I. (2002). Physics and mathematics as interrelated fields of thought development using acceleration as an example. *International Journal of Mathematical Education in Science and Technology*, 33(5), 679-690.
DOI: 10.1080/00207390210146023

Bergqvist, T., Lithner, J. & Sumpter, L. (2008). Upper Secondary Students' Task Reasoning, *International Journal of Mathematical Education in Science and Technology,* 39 (1), 1-12.
DOI: 10.1080/00207390701464675

Bing, T. (2008). *An epistemic framing analysis of upper level physics students' use of mathematics* (Doctoral dissertation). University of Maryland. Retrieved 2010-02-10 from http://www.physics.umd.edu/perg/dissertations/Bing/

Björk, L-E. & Brolin, H. (2001). *Matematik 3000, kurs A och B.* Stockholm, Sweden: Natur och Kultur.

Björk, L-E. & Brolin, H. (2006). *Matematik 3000, kurs C och D.* Stockholm, Sweden: Natur och Kultur.

Björkqvist, O. (2001). Matematisk problemlösning. In B. Grevholm (Ed.), *Matematikdidaktik: ett nordiskt perspektiv* (pp. 115-132). Lund, Sweden: Studentlitteratur.

Blum, W. & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects — State, trends and issues in mathematics instruction. *Educational Studies in Mathematics*, 22(1), 37-68.
DOI: 10.1007/BF00302716

Boesen, J., Lithner, J. & Palm, T. (2010). The relation between types of assessment tasks and the mathematical reasoning student use. *Educational studies in mathematics*, 75 (1), 89-105.
DOI: 10.1007/s10649-010-9242-9

Clement, J. (1985). Misconceptions in graphing. In L. Streefland (Ed.), *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education*, Utrecht: Utrecht University (pp. 369–375).

Cui, L. (2006). *Assessing college students' retention and transfer from calculus to physics* (Doctoral dissertation). Kansas State University. Retrieved 2012-06-27 from http://krex.k-state.edu/dspace/handle/2097/183

Dall'Alba, G., Walsh, E., Bowden, J., Martin, E., Masters, G., Ramsden, P. & Stephanou, A. (1993). Textbook treatments and students' understanding of acceleration. *Journal of Research in Science Teaching,* 30(7), 621-635.
DOI: 10.1002/tea.3660300703

Department of Applied Educational Science. (2011). Some earlier given test in Physics*.* Retrieved 2011-10-06 from http://pb-fy.edmeas.napb.se/information/tidigare_prov.asp

Doorman, L. M. & Gravemeijer K. P. E. (2009). Emergent modeling: discrete graphs to support the understanding of change and velocity. *ZDM – The International Journal on Mathematics Education,* 41, 199-211.
DOI 10.1007/s11858-008-0130-z

Ekbom, L., Lillieborg, S., Larsson, S., Ölme, A. & Jönsson, U. (2004). *Tabeller och formler för NV- och TE- programmen, 5 ed.* Stockholm: Liber.

English, L., & Sriraman, B. (2010). Problem solving for the 21st century. In L. English & B. Sriraman (Eds.), *Theories of mathematics education* (pp. 263-290). Berlin, Heidelberg: Springer.

Engström, S. (2011). *Att vördsamt värdesätta eller tryggt trotsa. Gymnasiefysiken, undervisnings-traditioner och fysiklärares olika strategier för energiundervisning* (Doctoral dissertation). Mälardalen University. Retrieved 2011-10-07 from
http://mdh.diva-portal.org/smash/record.jsf?searchId=1&pid=diva2:310

Hanna, G. (2000). Proof and its classroom role: A survey. In M.J. Saraiva et al (Eds.), *Proceedings of conference en el IX Encontro de Investigaçao en Educaçao Matematica* (pp. 75–104), Funado.

Hanna, G. & Jahnke, H. N. (2002). Another Approach to Proof: Arguments from Physics. *ZDM-The International Journal on Mathematics Education*, 34 (1), 1-8.
DOI: 10.1007/BF02655687

Haylock, D. (1997). Recognising mathematical creativity in schoolchildren. *ZDM-The International Journal on Mathematics Education*, 29(3), 68-74.
DOI: 10.1007/s11858-997-0002-y

Johansson, H. (2013b). *Mathematical Reasoning when solving physics tests. Relations between grades and kinds of tasks solved and seeking dependence between different kinds of reasoning.* Pre-print. Gothenburg.

Lesh, R., & Zawojewski, J. (2007). Problem Solving and modeling. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 763-804). Charlotte, NC: Information Age Publishing.

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics,* 67(3), 255-276.
DOI: 10.1007/s10649-007-9104-2

Marongelle, K. A. (2004).How Students Use physics to Reason About Calculus Tasks. *School Science and Mathematics,* 104(6), 258-272.
DOI: 10.1111/j.1949-8594.2004.tb17997.x

Michelsen, C. (2005). Expanding the domain – variables and functions in an interdisciplinary context between mathematics and physics. In A. Beckmann, C. Michelsen & B. Sriraman (Eds.). *Proceedings of the 1st International Symposium of Mathematics and its Connections to the Arts and Sciences* (pp.201-214). The University of Education, Schwäbisch Gmünd, Germany.

Ministry of Education and Research (2001). Departementsserien (Ds) Ds 2001:48. *Samverkande styrning: om läroplanerna som styrinstrument*. Retrieved 2011- 10- 28 from http://www.regeringen.se/content/1/c4/26/49/601d5449.pdf

Mulhall. P. & Gunstone. R. (2012). Views About Learning Physics Held by Physics Teachers with Differing Approaches to Teaching Physics. *Journal of Science Teacher Education*, 23 (5), 429-449.
DOI: 10.1007/s10972-012-9291-2

Nguyen, N-L. & Meltzer, D. (2003). Initial understanding of vector concepts among students in introductory physics courses. *American Journal of Physics,* 71(6), 630-638.
DOI: 10.1119/1.1571831

Palm, T., Boesen, J., & Lithner, J. (2011). Mathematical Reasoning Requirements in Upper Secondary Level Assessments. *Mathematical Thinking and Learning, 13*(3), 221-246.
DOI: 10.1080/10986065.2011.564994

Pólya, G. (1954). Mathematics and plausible reasoning (Vols. I and II). Princteton, NJ: Princeton University Press.

Pålsgård, J., Kvist, G. & Nilson, K. (2005a). *Ergo Fysik A*. Stockholm: Liber.

Pålsgård, J., Kvist, G. & Nilson, K. (2005b). *Ergo Fysik B*. Stockholm: Liber.

Redish, E. F. & Gupta, A. (2009). *Making Meaning with Math in Physics: A Semantic Analysis.* Contributed paper presented at GIREP 2009, Leicester, UK, August 20, 2009. Retrieved 2012-06-25 from
http://www.physics.umd.edu/perg/talks/redish/GIREP2009/GIREP%20Semantics.pdf

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando: Academic Press.

Silver, E. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *ZDM-The International Journal on Mathematics Education*, 29(3), 75-80.
DOI: 10.1007/s11858-997-0003-x

Swedish National Agency for Education. (2000a). *Kursplan Fysik A*. Retrieved 2010-10-12 from http://www.Swedish National Agency for Education.se/forskola_och_skola/gymnasieutbildning/gymnasieskola_fore_ht_2011/2.3034/sok_amnen_och_kurser/13845/func/kursplan/id/3053/titleId/FY1201%2520-%2520Fysik%2520A

Swedish National Agency for Education. (2000b). *Kursplan Fysik B*. Retrieved 2010-10-12 from http://www.Swedish National Agency for education.se/forskola_och_skola/gymnasieutbildning/gymnasieskola_fore_ht_2011/2.3034/sok_amnen_och_kurser/13845/func/kursplan/id/3054/titleId/FY1202%2520-%2520Fysik%2520B

Swedish National Agency for Education. (2000c). *Ämnet Fysik, gymnasieskolan*. Retrieved 2010-10-12 from http://www.skolverket.se/forskola-och-skola/gymnasieutbildning/gymnasieskola-fore-ht-2011/kursplaner/sok-amnen-och-kurser/13845/func/kursplan/id/3054/titleId/FY1202%2520-%2520Fysik%2520B?_xurl_=http%3A%2F%2Fsvcm.skolverket.se%2Fsb%2Fd%2F2503%2Fa%2F13845%2Ffunc%2Famnesplan%2Fid%2FFY%2FtitleId%2FFysik

Swedish National Agency for Education. (2001). *Programme Manual (GY 2000:19). Programme goal and structures, core subjects, subject index for upper secondary school*. Stockholm: Fritzes.

Swedish National Agency for Education. (2003). *Lusten att lära – Med fokus på matematik: Nationella kvalitetsgranskningar 2001–2002.* Stockholm: Fritzes.

Swedish National Agency for Education. (2005). *Skolverkets Provbank. Hur används den och vad tycker användarna?* Retrieved 2012-10-21 from http://www.skolverket.se/prov-och-bedomning/rapporter-och-film

Swedish National Agency for Education (2009). *TIMSS Advanced 2008. Svenska gymnasieelevers kunskaper i avancerad matematik och fysik i ett internationellt perspektiv.* Stockholm: Fritzes.

Swedish National Agency for Education. (2011). Tabell 3B. Retrieved 2011-10-05 from http://www.Swedish National Agency for Education.se/statistik_och_analys/2.1862/2.4391/2.4392

Swedish Schools Inspectorate. (2010). *Fysik utan dragningskraft. En kvalitetsgranskning om lusten att lära fysik i grundskolan.* Retrieved 2011-11-23 from http://www.skolinspektionen.se/Documents/Kvalitetsgranskning/fysik-i-grundskolan/slutrapport-undervisningen-fysik.pdf

Tasar, M. F. (2010). What part of the concept of acceleration is difficult to understand: the mathematics, the physics, or both? *ZDM-The International Journal on Mathematics Education,* 42, 469-482.
DOI: 10.1007/s11858-010-0262-9

Tuminaro, J. (2002). *How Student Use Mathematics in Physics: A Brief Survey of the Literature.* University of Maryland Physics Education Research Group. Retrieved 2010-02-10 from http://www.physics.umd.edu/perg/math/UsingMath.pdf

Winter, H. (1978). Geometrie vom Hebelgesetz aus – ein Beitrag zur Integration von Physik- und Mathematikunterricht der Sekundarstufe I. *Der Mathematikunterricht* 24 (5), 88-125.

Wyndham, J., Riesbeck, E. & Schoult, J. (2000). *Problemlösning som metafor och praktik*. Linköping: Institutionen för tillämpad lärarkunskap, Linköpings univ.

Paper II

Mathematical Reasoning when solving physics tests

Relations between grades and kinds of tasks solved and seeking dependence between different kinds of reasoning

Helena Johansson

Abstract

In this two-folded study, the first part examines the relation between students' grades and what kind of tasks, with respect to mathematical reasoning requirements, they have solved in national physics tests. The second part works as a pilot-study to try out if the Mantel-Haenszel procedure is an appropriate statistical method to answer questions about if there is a dependence between students' success on different physics tasks that requires two different kinds of mathematical reasoning. The analysis shows that in three out of eight national physics tests, it is possible to get a higher grade than "Pass" without solving tasks that require students to consider intrinsic mathematical properties and produce new reasoning. The result reveals though, that it is uncommon that a student gets a higher grade without solving tasks that require the student to come up with not already familiar solutions. The pilot-study shows that the Mantel-Haenszel procedure is sensitive to the number of students each teacher accounts for. If there are not too few students, the procedure can be used. The result from the pilot-study indicates that there is a dependence between success on tasks requiring different kinds of reasoning. It is more likely that a student manages to solve a task that requires the produce of new reasoning if the student has solved a task that is familiar from before.

Introduction

Relations between learning physics and knowing mathematics

Many scholars discuss the importance to understand how mathematics is used in physics and how students' mathematical knowledge affects their learning of physics (e.g. Basson, 2002; Bing, 2008; Nguyen & Meltzer, 2003; Redish & Gupta, 2009). diSessa (1993) notices that students who have studied physics and can solve a quantitative task in physics still can give an inconsistent qualitative analysis of the same task. A q*uantitative* task refers to when the task is posed in explicitly quantitative terms and the solution can be attained through calculations using appropriate physics laws. A *qualitative* task refers to when the solution requires an analysis of the posed physical situation i.e. what will occur and/or why. According to Swedish National Agency for Education (2009) a common activity in physics classes is students using physics laws and formulas to solve routine tasks. The most common homework is to read in the textbook and/or to solve various tasks posed in the book, and sometimes to memorize formulas and procedures (ibid.). Redish (2003) state that practice, in the meaning that students just solve various tasks, is necessary but not enough to get a deeper understanding of the underlying physics concepts. Students must learn both *how* to use the knowledge and *when* to use it. The same conclusion also holds for learning mathematics, shown by e.g. Schoenfeld (1985) in his study of how students become good problem solvers in mathematics; as well as by Lesh and Zawojewski (2007), who discuss how working with mathematical modelling develop students' understanding and learning in mathematics. Rote learning can be a main factor behind learning difficulties in mathematics (Lithner, 2008; Schoenfeld, 1992).

During studies of how students are engaging in different mathematical activities, Lithner (2008) has gradually developed a framework for characterising students mathematically reasoning. The framework distinguishes between *creative mathematical founded reasoning* (CR) and *imitative reasoning* (IR). The former one refers to a reasoning that is anchored in intrinsic mathematical properties and that includes some novelty to the reasoner. If instead the anchoring is in surface properties and the reasoning consists of remembering an answer or following a process step by step, it is IR. There has been a discussion in the mathematical educational research society whether procedural knowledge should be considered only as superficial and rote learned or viewed from a wider perspective (Baroody, Feil & Johnson, 2007; Star, 2007). There is an agreement that procedural knowledge is important, but not enough, when students learn mathematics (Baroody et al., 2007; Gray & Tall, 1994; Sfard, 1991; Star, 2007). However, there is also an argumentation about if *deep procedural knowledge* could exist without involvement of conceptual knowledge (Baroody, Feil & Johnson 2007; Star, 2005, 2007). In the description of the framework used for characterising required mathematical reasoning, Lithner (2008) discusses different aspects of procedures and concepts. Although the definitions of the reasoning categories do not include references to procedural or conceptual knowledge, one could assume some relations between CR and conceptual knowledge on one hand and IR and procedural knowledge on the other hand.

The studies in this paper are based on the assumption that students' ability to reason mathematically affects how they succeed to solve tasks in physics.

Physics in the Swedish school

There are mainly two different physics courses in the Swedish upper secondary school. Physics A that is compulsory for all natural science and technology students and Physics B that is an optional continuation. The final grade a student is awarded after completion of a physics course depends on the achieved level of proficiency (Swedish National Agency for Education, 2000a, 2000b). The grades vary between *Not Pass* (IG), *Pass* (G), *Pass with distinction* (VG) and *Pass with special distinction* (MVG). The descriptions in the syllabuses of the different grade levels are quite vague and the intention is that the syllabuses should be processed and interpreted locally at the schools. To accomplish equivalent assessment in physics, assessment supports are provided by the Swedish National Agency of Education. One of these supports is the National Test Bank in Physics to which teachers can get access after they have registered and received a password. Once logged in, teachers can download course tests for both Physics A and Physics B. After a test is used, the teachers are intended to report back students' results on the test to the Test Bank. A majority of all registered teachers use the provided tests as a final exam in the end of the physics courses (Swedish National Agency for Education, 2005). The character and the design of the tasks in national tests stress what is covered in the taught curriculum. The tests guide teachers' interpretation of the syllabuses and by extension influence what students focus on (Ministry of Education and Research, 2001; Swedish National Agency for Education, 2003).

Conceptual framework

The definition of mathematical reasoning and of the different reasoning categories used to analyse and categorise physics tasks in the author's previous paper/study are developed by Lithner (2008). This framework is also used in this paper since the categorised physics tasks serve as a basis for the conducted analyses. *Reasoning* is defined as ”the line of thought adopted to produce assertions and reach conclusions in task solving” (Lithner, 2008 p. 257). *Mathematical reasoning* refers to a product of separate reasoning sequences where the justification for the choice of the next sequence is mathematically founded. The mathematical foundation of the reasoning can either be *superficial* or *intrinsic.* Superficial and intrinsic refer to the relevance of the mathematical argument that is used. An object's mathematical properties are of different relevance in different contexts e.g. what type of problem one is trying to solve. This leads to a distinction between *surface* properties and *intrinsic* properties. The former ones have little relevance in the actual context and give rise to superficial reasoning, and the latter ones are central and the ones to be regarded (ibid. pp. 260-261).

Creative mathematically founded reasoning

Creativity is an expression often used in different contexts and without an unequivocal definition (for a discussion see Lithner (2008, pp. 267-268)). For the definitions of the different kinds of reasoning the perspective of Haylock (1997) and Silver (1997) is adopted. This implies that creativity is seen as a thinking process that is novel, flexible and fluent (Lithner, 2008).

Creative mathematical founded reasoning[5] *(CR)* fulfils all of the following criteria. (Lithner, 2008, p.266)

i. Novelty. A new (to the reasoner) reasoning sequence is created or a forgotten one is recreated.
ii. Plausibility. There are arguments supporting the strategy choice and/or strategy implementation motivating why the conclusions are true or plausible.
iii. Mathematical foundation. The arguments are anchored in intrinsic mathematical properties of the components involved in the reasoning.

Imitative reasoning

Imitative reasoning (IR) is divided into *memorised reasoning* and *algorithmic reasoning*. The arguments that motivate the chosen solution method (i.e. the reasoning) could be anchored in surface mathematical properties.

Memorised reasoning (MR) fulfils the following conditions (Lithner, 2008, p. 258)

i. The strategy choice is founded on recalling a complete answer.
ii. The strategy implementation consists only of writing it down.

If some kind of calculations is required to solve the task, there is often no use in remembering an answer. Instead it is more suitable to recall an algorithm. The term algorithm is here used in a wide

[5] Henceforward called *creative mathematical reasoning*.

sense and refers to all the procedures and rules that are needed to reach the conclusion to a specific type of tasks, not only the calculations.

Algorithmic reasoning (AR) fulfils the following conditions (ibid. p.259)

i. The strategy choice is to recall a solution algorithm, which if it is followed step by step will give the right answer without any demands of novelty.
ii. The remaining parts of the strategy implementation are trivial for the reasoner and just a careless mistake can obstruct the reaching of an answer.

Local and global creative mathematical reasoning

Lithner (2008) introduces a refinement of CR into *local CR* (LCR) and *global CR* (GCR) that captures some significant differences between tasks categorised as LCR and GCR. This differentiation has been more elaborated by other scholars that have used the framework e.g. Boesen, Lithner and Palm (2010) and Palm, Boesen and Lithner (2011). The difference between LCR and GCR is that in LCR, the reasoning is mainly MR or AR but contains a minor step that requires CR. If instead there is a need for CR in several steps, it is called GCR, even when some parts contains AR and/or MR. Important to stress is that as soon as CR is involved there has to be some understanding of the intrinsic mathematical properties in the task.

Non-mathematical reasoning

Because it was physics tasks that were analysed an additional category were defined, *non-mathematical reasoning* (NMR), i.e. tasks that can be solved by only using physics knowledge. Physics knowledge refers to relations and facts that are discussed in the physics courses and not in the courses for mathematics according to the syllabuses and textbooks. One example is that angle of incidence equals angle of reflection. In the same way, a solution that requires mathematical reasoning refers to mathematics taught in courses at upper secondary school or assumed already to be known by the students according to the curriculums.

Research questions

As an approach to the assumption that students' ability to reason mathematically affects how they succeed to solve tasks in physics, this study analyses the mathematical reasoning requirements students are put in front of when solving tasks in physics tests. The author therefore used Lithner's (2008) distinction of mathematical reasoning, IR and CR, in a previous study (Johansson, 2013a) that addressed the question "what are the mathematical reasoning requirements to solve physics tasks in physics tests from the Swedish national test bank?" The result showed that students must use some kind of mathematical reasoning to solve three-fourth of the tasks in a test and that one-third of the tasks required creative mathematical reasoning. One approach to examine the assumption further is to combine the previous result with data representing the practice. In this paper students' scores on the previously categorised tasks are, together with their grades on the tests, used as data. Each student's grade is compared to what kind of tasks respective student has received scores for. Moreover, students' scores are used as a measure of their success on respective tasks. A comparison of how students succeed on different kinds of tasks (i.e. tasks requiring IR or GCR) will hopefully answer if there is a dependence between IR and GCR on physics tasks, and if so, of what kind. This paper consists of two parts that both use the definitions in the conceptual framework. The purpose of the parts are though some different. The first part addresses the following question:

- Is it possible for a student to get one of the higher grades without using creative mathematical reasoning, and if so, how common is it?

The purpose of the second part is mainly to examine if the used statistical test method is appropriate to use to answer the following question:

- Does a student's success on tasks categorised as requiring global creative mathematical reasoning depend on the student's success on tasks requiring imitative reasoning?

In this respect the purpose of the second part is to work as a pilot study. If the test method works out well it could be used for further analyses of relations between success on IR- and CR/GCR-tasks on mathematics tests that have been categorised according to Lithner's (2008) conceptual framework.

Because the characters of the two questions are of different kinds, two different methods have been used and the input data differ in some respect. In the Method section below, there will first be a description of the different data that have been used and after that follows the outline of the respective method.

Methods

Input data

In a study by Johansson (2013a), tasks in physics tests from the National Test Bank have been categorised according to the conceptual framework above. The results from that study are, in addition with student data, used as data in this paper. The student data are used by permission from Department of Applied Educational Science at Umeå University, the department in charge of the National Test Bank in Physics. Student data come as excel sheets, one sheet for each test. The sheets contain information about students' scores on each task separated in G- and VG-scores; students' total score on the tests; a calculated grade on the tests; the program each student attend as well as an ID-number of their respective teacher. No names of the students are present in the sheets, instead each student has got an ID-number. The IDs for both the teachers and the students are unidentifiable for anyone outside the Department of Applied Educational Science at Umeå University, so data could be considered anonymous. The number of student data for each test varies from 996 to 3666.

Examples

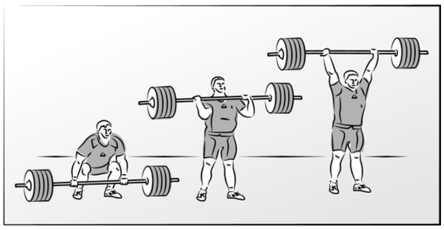
Below are three examples of tasks from one of the physics tests[6]. The first, 3a, were categorised as solvable with imitative reasoning. Similar tasks are hence forward called IR-tasks. The second example, 3b, illustrates a task categorised solvable by only using physics, that is no mathematics were required. Tasks like this are called NMR-tasks. The last example is of a GCR-task, i.e. tasks requiring global creative mathematical reasoning to be solved. In the same way will tasks requiring LCR be called LCR-tasks and tasks requiring either LCR or GCR will be called CR-tasks. The outline of all tasks in a test begins in the same way; first is the number of the task in the test given and after

[6] The original size of the physics tests is A4, i.e. the size of the examples below is originally √2 times larger if this paper is read on A5 format.

that, enclosed in brackets, the task's number in the National Test Bank. On the next line are the maximum scores for the task given. As mentioned above, the scores are divided into two different categories, G-scores and VG-scores. The maximum scores for each category are separated with a slash, for example 2/0 tells that a student can get a maximum of two G-scores and zero VG-scores on that particular task. In the same way, 1/1 tells that the maximum is one G-score and one VG-score. If the task consists of subtasks: a, b, etc.; the total scores for the subtasks are separated with commas.

Task no. 3 (1584)
2/0, 1/0

A weightlifter is lifting a barbell that weighs 219 kg. The barbell is lifted 2.1 m up from the floor in 5,0 s.



c) What is the average power the weightlifter develops on the barbell during the lift?

Short account for your answer:

d) What is the average power the weightlifter develops on the barbell when he holds it above the head during 3.0 s?

Short account for your answer:

Task no. 10 (1588)
1/1

A lens of a digital camera has a focal length of 8.0 mm. The lens' distance to the film plane may vary between 8.0 mm and 10.0 mm.

Which is the closest distance you can take a photo of an object and still get a sharp image?

Part one

In order to answer the first research question, eight different physics tests from the Swedish National Test Bank in Physics were used. Together the tests comprised 169 categorised tasks. Students' grades on the tests and their scores on the categorised tasks were also used. The availability of student data decided which tests that were used in the analysis. As mentioned above, grades vary between IG, G, VG and MVG. For each test there are certain score levels the students need to attain to get a certain grade. To get the grade MVG, students need to fulfil certain quality aspects besides the particular score level. To decide if it is possible for a student to get one of the higher grades without using any kind of CR, each test is first analysed separately. This first analysis consists in comparing the score level for each grade with the maximum scores that are possible to obtain, given that the student only has solved (partly or fully) IR- and/or NMR- tasks. The available student data do not give any information about which of the qualitative aspects required for MVG the students have fulfilled, but the data sheets include students grades, thus MVG can be included in the analyses as one of the higher grades. After analysing if it is possible at all to receive the grades VG or MVG without solving any CR-tasks, the second analysis started. Now students' actual results on the categorised tasks for those particular tests are summed up. The proportion of students who only got scores from IR- and NMR-tasks is then graphed with respect to the different grades.

Part two

To answer the second research question, the *Mantel-Haenszel (MH) procedure* has been chosen as statistical test method. As stated above, the main purpose of the second part in this paper is to try out and decide if the MH-procedure is appropriate to use for answering this kind of question. Hence this part is a pilot study. Before describing data used in this second part and outlining the method, a description of the MH-procedure will be given.

The Mantel-Haenszel procedure

The MH-procedure was originally developed for data analyses from retrospective studies in the clinical epidemiology area. The purpose of the MH-procedure was to test if there were any relations between the occurrence of a disease and some factors. The disease could for instance be lung cancer and one factor could be cigarette smoking (Mantel & Haenszel, 1959). A retrospective study can be performed on already collected data and does not require as big sample size as a forward study (also called prospective study) does. In a retrospective study of a disease one looks for unusually high or low frequency of a factor among the diseased persons, while in a forward study it is the occurrence of the disease among persons possessing the factor that is looked at (ibid.). The calculations involved in the MH-procedure are quite simple and this is probably a contributing factor to that the method is commonly used in various areas today e.g. epidemiology (Rothman, Greenland & Lash, 2008), biology/biological statistics (McDonald, 2009) and social/educational sciences (Fidalgo & Madeira, 2008; Guilera, Gómez-Benito & Hidalgo, 2009; Holland and Thayer, 1988; Ramstedt, 1996). One of the most common uses of the MH-procedure in educational studies seems to be for detecting existence of *differential item functioning* (DIF). DIF exists if people with the same knowledge/ability, but belonging to different groups, have different probabilities to give the right answer to an item/task. Ramstedt (1996) used a modified version of the MH-procedure to analyse if there were differences between how boys and girls succeeded on national physics tests depending on their sex, according to their personal identity number. According to Ramstedt, Holland and Thayer (1988) were the first ones to use the MH-procedure to detect DIF.

To use the MH-procedure, data should first be stratified into 2x2 contingency tables. In these tables the rows and the columns represent the two nominal variables that will be tested for dependence. The variable that is placed in the rows is the one that is tested whether it explains/affects the outcome of the variable placed in the columns. The different contingency tables represent a third nominal variable that identifies the repeat. The two nominal variables could for example be: a disease and a factor; a plant and habitats; group belonging and success on tasks. Examples of the repeat variable are different medical centers, different seasons, different teachers etc.

Table 7: Contingency table for repeat *i*.

Table i	Y = 1	Y = 0	Totals
X = 1	a_i	b_i	n_{i1}
X = 0	c_i	d_i	n_{i0}
Totals	m_{i1}	m_{i0}	n_i

In Table 7, X and Y represent the two nominal variables. Both variables are coded by the values 0 and 1 for the respective object included in the study. Belonging to the group of diseased persons might then be represented by X = 1 and not being diseased with X = 0. In the same way, the occurrence of a factor may be represented by Y = 1 and non-existence of the factor with Y = 0. The letters a_i, b_i, c_i and d_i denote the frequencies for respective occurrence and $n_i = a_i + b_i + c_i + d_i$. A diseased person possessing the factor will then be one of those contributing to the frequency a_i. The probability p for an event is estimated by the relative frequency $\hat{p}$. For example, the relative frequency for the event X = 1 and Y = 1 is $\hat{p} = a_i/n_i$.

The MH-procedure includes an estimation of the *common odds ratio*, $\hat{\theta}_{MH}$, for the different contingency tables. *Odds*, O, is defined as the probability p for an event divided by the probability for the same event not to occur i.e. O=p/(1-p). *Odds ratio,* θ, is defined as the ratio between the different odds for the event with respect to the different row variables, that is, keep X fixed in the table. From above follows that the odds for X = 1 and Y = 1 is estimated by a_i/b_i and the odds for X = 0 and Y = 1 is estimated by c_i/d_i. This gives that the odds ratio for contingency table i is estimated by

$$\hat{\theta}_i = \frac{a_i/b_i}{c_i/d_i} = \frac{a_i d_i}{b_i c_i}.$$

The common odds ratio calculated in the MH-procedure is defined as

$$\hat{\theta}_{MH} = \frac{\sum_j a_j d_j / n_j}{\sum_j b_j c_j / n_j} = \frac{\sum_j w_j \hat{\theta}_j}{\sum_j w_j},$$

where $\hat{\theta}_i$ is the odds ratio for table i and

$$w_i = \frac{b_i c_i}{n_i}$$

is the weight associated to $\hat{\theta}_i$. The summations run over all contingency tables, i.e. j = 1,…,k, where k is the number of contingency tables. Thus $\hat{\theta}_{MH}$ is a weighted average of the individual odds ratios. The assumed null hypothesis, H_0, is that there is no dependence between the variables X and Y, i.e. $\theta_{MH} = 1$.

The most important step in the procedure is the calculation of a MH test statistic, which tells if $\hat{\theta}_{MH}$ differs sufficiently from 1 so that H_0 can be rejected. The most commonly used test statistic, χ^2_{MH}, is approximately chi-square distributed, and is compared to a chi-square distribution with one degree of freedom (Mantel & Haenszel, 1959; Ramstedt, 1996; Mannocci 2009; McDonald, 2009). The definition of χ^2_{MH} is

$$\chi^2_{MH} = \frac{(|\sum_i a_i - \sum_i E(a_i)| - ½)^2}{\sum_i Var(a_i)},$$

where $E(a_i) = {n_{i1} m_{i1}}/{n_i}$ is the expected value for a_i under H_0 and

$$Var(a_i) = \frac{n_{i1} n_{i0} m_{i1} m_{i0}}{n_i^2 (n_i - 1)}$$

is the variance for a_i (Mantel & Haenszel, 1959).

Instead of χ^2_{MH}, a test statistic, Z_{MH}, that is approximately normal distributed can be used (McCullagh & Nelder, 1989). The advantage of using Z_{MH} is that the direction of a possible dependence is detected. Therefore this test statistic is used in the study in this paper. The definition of Z_{MH} is

$$Z_{MH} = \frac{\sum_i \{a_i - E(a_i)\} - ½}{\sigma_{\sum_i \{a_i - E(a_i)\}}},$$

where $E(a_i)$ is as above and

$$\sigma_{\sum_i \{a_i - E(a_i)\}} = \sqrt{var(a_i - E(a_i))} = \sqrt{\frac{\sum_i \{n_{i1} n_{i0} m_{i1} m_{i0}\}}{n_i^2 (n_i - 1)}}$$

is the standard deviation of $a_i - E(a_i)$ (McCullagh & Nelder, 1989). The value ½ that is subtracted in the numerator for each of the statistics is a continuity correction value (Mantel & Haenszel, 1959; McCullagh & Nelder, 1989).

Method

To answer the second research question, if there is a dependence between students' success on IR-tasks and how they succeed on GCR-tasks, the MH-procedure was first used on one randomly chosen pair of IR/GCR-tasks from one of the eight physics tests. For each of the chosen tasks, students' scores were collected from the data sheet, as well as the ID number of respective student's teacher. The two different categories of tasks, IR and GCR, are the two nominal variables tested for dependence. To control for a possible influence from students different teachers, teacher is chosen as the variable that identifies the repeat. One influence may be on the scoring of the tasks, since the scoring involves some interpretation of the scoring rubrics and could thus result in some differences in how to score a specific answer. It is assumed that the individual teacher is consistent in the scoring of his/her students' solution to respective task. Another influence from the teachers is that different teachers show different examples on the blackboard, which influence what become familiar solutions to students. There can also be a difference in what kind of mathematical reasoning different teachers give their students the possibility to practice on. Some teachers may be more focused on working with creative mathematical reasoning than others.

To see if there was any consistency in the result from the MH-procedure, more tasks were chosen from the eight tests. The number of GCR tasks varied between two and six for the different tests, so it was decided to choose two GCR tasks from each test. The GCR tasks were randomly selected as far as possible. It was further decided to test how the success on these two chosen GCR-tasks depended on success on two of the simpler IR-tasks and two of the harder IR-tasks on the same test. The difference between a simpler and a harder IR-task turned out to depend mainly on how many steps that were needed in the solution algorithm. For a simpler IR-task, the solution consisted mostly of one step; and for a harder IR-task, there were often three or more steps to remember. The number of IR-tasks varied between six and nine in the different tests. Each GCR-task was then tested for dependence against all four of the IR-tasks.

MATLAB was used to arrange the contingency tables needed in the MH-procedure. The rows in a contingency table represent the students who have succeeded (1) and not succeeded (0) on the particular IR-task. The columns represent in the same way the students who have succeeded (1) and not succeeded (0) on the GCR-task. Success on a task is defined as to have solved the task completely, i.e. to have attained the maximum score. For each entry, MATLAB calculated the number of students that fulfilled that particular combination e.g. a_i is the number of students who have succeeded on both the IR-task and the GCR-task. The row and column totals were summed up, as well as the total number of students for teacher i.

Table 8: Contingency table for IR and CR with respect to teacher *i*.

Teacher i	GCR (1)	GCR (0)	Totals
IR (1)	a_i	b_i	n_{i1}
IR (0)	c_i	d_i	n_{i0}
Totals	m_{i1}	m_{i0}	n_i

After this, Z_{MH}, the approximately normal distributed test statistic, was calculated for every pair, i.e. 64 test statistics were calculated. The obtained value was compared to critical values for a two-tailed test and 5 % significance level, to decide whether H_0 can be rejected or not. For a table to be included in the calculation of the test statistic, each of the calculated expected values has to be 5 or more. Since this is a pilot study, no correction for the multiple comparisons was done (cf. McDonald, 2009).

Analyses and results

In this section, the results from the analyses are presented. The outline follows the outline in the previous section and is divided in two parts, one for each of the research questions.

Part oneTable 9 shows how the scores, possible to obtain on each of the eight tests that were analysed, are distributed among the reasoning categories IR and NMR. The table also includes the levels for the grades G, VG and MVG. The notation for the scores follow the convention described above, i.e. G/VG.

Table 9: Analysis of the distribution of G and VG scores among IR- and NMR-tasks.

Test	Max score	Min required score for G	Min required score for VG	Min required score for MVG	Max scores for IR-tasks	Max scores for NMR-tasks	Max score possible without CR-tasks
Physics A May 02	26/17	12	25 (with at least 6 VG scores)	25 (with at least 12 VG scores)	12/0	3/3	18 (with 3 VG)
Physics A Dec 04	23/17	12	24 (with at least 5 VG scores)	24 (with at least 12 VG scores)	14/3	3/3	23 (with 6 VG)
Physics A May 05	22/16	12	24 (with at least 6 VG scores)	24 (with at least 12 VG scores)	12/3	8/4	27 (with 7 VG)
Physics B May 02	23/25	12	27 (with at least 7 VG scores)	27 (with at least 13 VG scores)	11/4	2/0	17 (with 4 VG)
Physics B May 03	23/20	12	25 (with at least 6 VG scores)	25 (with at least 13 VG scores)	12/8	5/1	26 (with 9 VG)
Physics B May 05	22/22	12	25 (with at least 6 VG scores)	25 (with at least 12 VG scores)	8/5	7/2	22 (with 7 VG)
Physics B Feb 06	22/21	12	25 (with at least 7 VG scores)	25 (with at least 13 VG scores)	11/7	9/9	36 (with 16 VG)
Physics B April 10	24/20	12	25 (with at least 6 VG scores)	25 (with at least 12 VG scores)	9/4	4/1	18 (with 5 VG)

In three of the eight tests, highlighted above, it is possible to get the grade VG by solving tasks not requiring any CR. In one of the tests, Physics B from February 2006, it is with respect to score level possible to obtain the grade MVG by solving only IR- and NMR-tasks. The analysis does not reveal anything about if the requirements of the qualitative aspects for MVG are possible to fulfil by solving only these kinds of tasks. The proportion of students who only had solved IR- and/or NMR-tasks was graphed with respect to their grades on the tests (see the figures below). It turned out that it is not frequently occurring that a student gets a higher grade than G by only solving these kinds of tasks. In the test for Physics A from 2005, only 0.17 % of the students got a higher grade (Figure 2); and in the Physics B test from 2003 none of the students got higher grades than G (Figure 3). The Physics B test from 2006 seems to be an exception though, since 25% of the students taking this test got a VG and 17% got a MVG. The analysis of how the scores are distributed among the reasoning categories for the different tests shows that the Physics B test from 2006 contains a lot more scores in the NMR category than any of the other tests (see Table 9). The total scores possible to obtain by only solving NMR-tasks are 18; nine of these are VG-scores, which is more than enough to fulfill the requirement for a VG (minimum 7 VG).

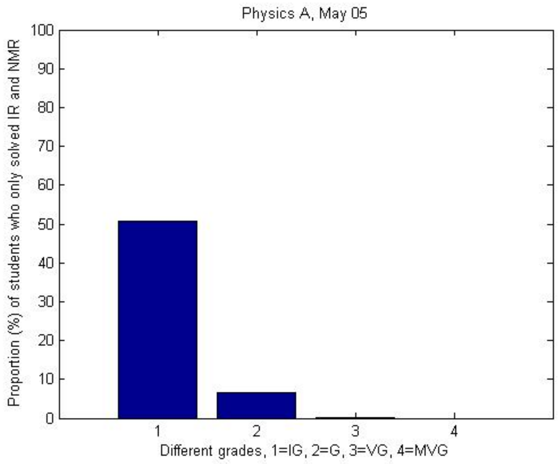


Figure 2: Proportion of students who only solved IR- and NMR-tasks with respect to the different grades for the Physics A test in 2005.

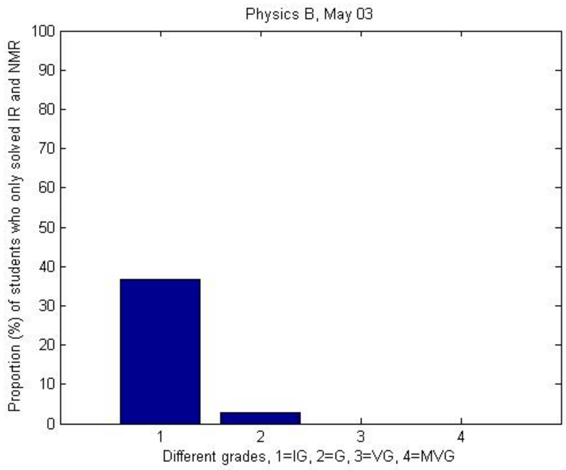


Figure 3: Proportion of students who only solved IR- and NMR-tasks with respect to the different grades for the Physics B test in 2003.

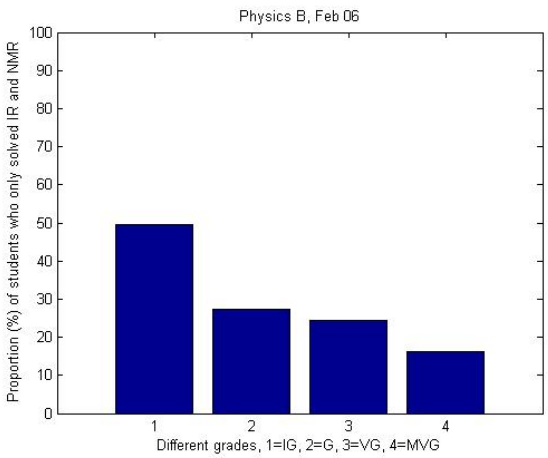


Figure 4: Proportion of students who only solved IR- and NMR-tasks with respect to the different grades for the Physics B test in 2006.

Below is a table of the proportion of the scores among the different reasoning categories. The result shows that about one-fifth of the scores could be obtained by solving tasks not requiring mathematical reasoning. It is also revealed that the majority of the scores are among the CR-tasks, but the difference from the proportion of scores among IR-tasks is not very big.

Table 10: Proportions of scores in the reasoning categories with respect to the physics courses.

Test	% max IR scores	% max NMR scores	% max CR scores
Physics A May 02	28	14	58
Physics A Dec 04	43	15	42
Physics A May 05	39	32	29
Physics A	37	20	43
Physics B May 02	39	5	56
Physics B May 03	47	14	39
Physics B May 05	30	20	50
Physics B Feb 06	42	42	16
Physics B April 10	30	11	59
Physics B	38	18	44

Part two

The Mantel-Haenzsel procedure resulted in 64 calculations of the MH-test statistic Z_{MH}. The critical values for a two-sided Z test with respect to 5% significance level are ±1.96, i.e. if the calculated value is greater than 1.96 or lesser than -1.96, H_0 can be rejected.

Table 11: Value of Z_{MH} for each MH-test in respective Physics test

	SimpleIR-1 -> GCR-1	SimpleIR-1 -> GCR-2	SimpleIR-2 -> GCR-1	SimpleIR-2 -> GCR-2	HardIR-1 -> GCR-1	HardIR-1 -> GCR-2	HardIR-2 -> GCR-1	HardIR-2 -> GCR-2
PhysicsA May02	5.8337	7.2634	5.6848	6.0494	9.3400	9.7978	8.2621	13.8206
PhysicsA Dec04	8.1920	5.7853	9.1352	12.2009	9.4944	10.1191	6.6884	11.3628
PhysicsA May05	8.8723	8.7645	4.9397	3.5521	15.7600	17.0689	11.9232	10.1675
PhysicsB May02	6.4192	4.4152	6.9744	5.8487	12.6614	10.7029	14.7344	8.5469
PhysicsB May03			8.5434	4.0272	10.5086	8.3243	9.5079	5.2267
PhysicsB May05	13.3128	6.7442	11.2680	10.0132	14.3362	20.9337	7.6338	12.1267
PhysicsB Feb06	8.6249	10.0267	6.3660	9.9268	7.9677	10.0952	8.3594	11.8029
PhysicsB April10							2.8897	

If the condition regarding expected value is not fulfilled for any of the contingency tables in the MH-procedure, the calculation does not give any value for Z_{MH}. This is represented by empty entries in Table 11 above. It can be seen in the table that the expected value was less than 5 for all teachers in most of the calculations for the test Physics B from April 10. The result shows that all of the 43 calculated Z_{MH} are greater than 1.96, which indicates that if a student succeeds on the IR-task it is more likely that he/she will succeed on the GCR-task. It is further seen that the Z_{MH} values are greater for most of the pairs Hard IR/GCR-tasks than for the Simple IR/GCR-tasks. This indicates a stronger dependence between success on a Hard IR-task and success on a GCR-task, than the dependence between success on a Simple IR-task and success on a GCR-task.

To be able to say more about the calculations of the Z_{MH}, and eventually the appropriateness of the MH-procedure, it was examined how many different teachers, (contingency tables), that fulfilled the condition regarding expected value. It is seen in Table 12 that some of the calculations of the Z_{MH} are based on results from very few teachers, e.g. in Physics B from May 03. Few contingency tables in the calculations of Z_{MH} can increase the risk of Type II errors, i.e. that H_0 is not rejected although it is false.

Table 12: Number of groups/teachers (i) that fulfill $E \geq 5$ (under H_0) for each MH-test in respective physics test.

	SimpleIR-1 -> CR-1	SimpleIR-1 -> CR-2	SimpleIR-2 -> CR-1	SimpleIR-2 -> CR-2	HardIR-1 -> CR-1	HardIR-1 -> CR-2	HardIR-2 -> CR-1	HardIR-2 -> CR-2
PhysicsA May02	15	17	4	6	18	20	19	21
PhysicsA Dec04	6	7	15	16	21	22	12	10
PhysicsA May05	17	14	7	5	20	16	28	22
PhysicsB May02	8	3	13	5	12	5	11	3
PhysicsB May03	0	0	7	2	4	1	7	2
PhysicsB May05	20	8	32	14	24	6	35	14
PhysicsB Feb06	14	12	11	12	11	10	14	13
PhysicsB April10	0	0	0	0	0	0	2	0

The expected value in each entry in a contingency table has to be 5 or greater; therefore number of students for each teacher has to be at least 20. Number of different teachers/tables and the total number of students were calculated for each of the tests.

Table 13: Number of teachers and students for each physics test.

	#Different Teachers	Tot #Students
PhysicsA May02	44	1900
PhysicsA Dec04	61	2198
PhysicsA May05	56	2612
PhysicsB May02	79	3666
PhysicsB May03	74	2734
PhysicsB May05	101	3454
PhysicsB Feb06	41	1458
PhysicsB April10	34	996

Table 13 reveals that the number of students in Physics B from April 2010 is quite few compared to the other tests. In Table 11 it is seen that only one Z_{MH} is obtained for this particular test, few students on too many teachers is one indication that the MH-procedure might not generate any reliable result, or any result at all.

Discussion and implications

The analyses from part one show that it is possible to receive a higher grade than G by using only IR and NMR on three out of eight tests. Comparing this result with student data reveals that not using any CR and still receive a higher grade only occurs on one of the eight tests. This particular test is, compared to the other tests, slightly different with respect to how the scores are distributed among the reasoning categories. Further analysis of the test shows that tasks regarded as MVG-tasks are all in the NMR category, i.e. tasks where it is possible to show the qualitative aspects required for MVG. This may be an explanation to the higher frequency of students receiving VG and MVG by using only IR and NMR, compared to the other tests. The analysis of the tests also shows that it is impossible to pass the test without solving any tasks requiring mathematical reasoning in six of the eight tests. These results strengthen the outcome from the previous study that creative mathematical reasoning is important when learning physics in upper secondary (Johansson, 2013a).

The results from part two shows that the MH-procedure can be used to determine if there is a dependence between students' success on tasks requiring different kinds of mathematical reasoning. If the MH-procedure shall give reliable results, one has to control that there are not too few students for each of the teachers included. Too few students can result in that the condition regarding the expected value is not fulfilled and thus result in too few contingency tables included in the calculation of the MH test statistic, which eventually may result in incorrect conclusions. Instead of calculating the total number of teachers and students, as was done in this pilot study, calculations of the number of students for each of the teachers is suggested to be done if the MH-procedure is going to be used in further studies. Correction for the multiple testing is also required if results in further studies shall be reliable. Although correction for multiplicity has not been done in this study, the result from the MH-procedure is a significant indication on a positive dependence between success in IR-tasks and GCR-tasks. According to the above discussion about similarities between IR – procedural knowledge and CR – conceptual knowledge, the acquired result can indicate a dependence between procedural and conceptual knowledge.

Viewing the physics tests from the National Test bank as an extension of the national curriculum, one can assume that students' results on the tests are a measure of their knowledge in physics. The scores and grade on a test should help teachers decide which achievements a student has attained and the level of the achievements. In this respect, students' grades on the physics tests could be viewed as a measure of their attained knowledge in physics. The result that mathematical reasoning is required to pass six of the eight tests support the assumption that students' ability to reason mathematically affects their success in solving physics tasks. The result shows that students' ability to reason mathematically is an integral part of their knowledge in physics. This in turn likely influences how students study and prepare themselves for tests in physics. Further support for the assumption is that students not solving tasks requiring CR will not likely attain a higher grade than G. It is well known that a focus on IR can explain some of the learning difficulties that students have in mathematics (see for example Lithner, 2008). The results above show that a focus on IR when learning physics in upper secondary school will make it hard for the students to do well on the physics tests. A reasonable conclusion is that focusing on IR can give students learning difficulties in physics, as it does in mathematics.

References

Baroody, A. J., Feil, Y. & Johnson, A. R. (2007). An Alternative Reconceptualization of Procedural and Conceptual Knowledge. *Journal for Research in Mathematics Education,* 38(2), 115-131. Stable URL: http://www.jstor.org/stable/30034952

Basson, I. (2002). Physics and mathematics as interrelated fields of thought development using acceleration as an example. *International Journal of Mathematical Education in Science and Technology*, 33(5), 679-690.
DOI: 10.1080/00207390210146023

Bing, T. (2008). *An epistemic framing analysis of upper level physics students' use of mathematics* (Doctoral dissertation). University of Maryland. Retrieved 2010-02-10 from http://www.physics.umd.edu/perg/dissertations/Bing/

Boesen, J., Lithner, J. & Palm, T. (2010). The relation between types of assessment tasks and the mathematical reasoning student use. *Educational studies in mathematics*, 75 (1), 89-105.
DOI: 10.1007/s10649-010-9242-9

diSessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction,* 10 (2-3), 105-225.
DOI:10.1080/07370008.1985.9649008

Doorman, L. M. & Gravemeijer K. P. E. (2009). Emergent modeling: discrete graphs to support the understanding of change and velocity. *ZDM – The International Journal on Mathematics Education,* 41, 199-211.
DOI 10.1007/s11858-008-0130-z

Fidalgo, A. M. & Madeira, J. M. (2008). Generalized Mantel-Haenszel Methods for Differential Item Functioning Detection. *Educational and Psychological Measurement*, 68(6), 940-958.
DOI: 10.1177/0013164408315265

Gray, E. M. & Tall, D. O. (1994). Duality, Ambiguity and Flexibility: A Proceptual View of Simple Arithmetic. *The Journal for Research in Mathematics Education,* 26(2), 115-141.
Stable URL: http://www.jstor.org/stable/749505

Guilera, G., Gómez-Benito, J. & Hidalgo, M. D. (2009). Scientific production on the Mantel-Haenszel procedure as a way of detecting DIF. *Psicothema* 21(3), 493-498.
ISSN 0214 - 9915

Holland, P. W. & Thayer, D. T. (1988). Differential item performance and the Mantel Haenszel Procedure. In H. Wainer & H. I. Braun (Eds), *Test validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Johansson, H. (2013a). *Mathematical Reasoning Requirements in Swedish National Physics Tests.* Pre-print. Gothenburg.

Lesh, R., & Zawojewski, J. (2007). Problem Solving and modeling. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 763-804). Charlotte, NC: Information Age Publishing.

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics,* 67(3), 255-276.
DOI: 10.1007/s10649-007-9104-2

Mannocci, A. (2009). The Mantel-Haenszel procedure. 50 years of the statistical method for confunders control. *Italian Journal of Public Health*, 6(4), 338-340.
DOI: 10.2427/5765

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute,* 22(4), 719-748.
DOI:10.1093/jnci/22.4.719

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models – 2nd edition.* Cambridge: University Press.

McDonald, J. H. (2009). *Handbook of biological statistics, 2nd edition*. University of Delaware, Baltimore: Sparky House Publishing

Ministry of Education and Research (2001). Departementsserien (Ds) Ds 2001:48. *Samverkande styrning: om läroplanerna som styrinstrument*. Retrieved 2011- 10- 28 from http://www.regeringen.se/content/1/c4/26/49/601d5449.pdf

Nguyen, N-L. & Meltzer, D. (2003). Initial understanding of vector concepts among students in introductory physics courses. *American Journal of Physics,* 71(6), 630-638.
DOI: 10.1119/1.1571831

Palm, T., Boesen, J., & Lithner, J. (2011). Mathematical Reasoning Requirements in Upper Secondary Level Assessments. *Mathematical Thinking and Learning, 13*(3), 221-246.
DOI: 10.1080/10986065.2011.564994

Ramstedt, K. (1996). *Elektriska flickor och mekaniska pojkar. Om gruppskillnader på prov – en metodutveckling och en studie av skillnader mellan flickor och pojkar på centrala prov i fysik.* (Doctoral dissertation). Umeå: UmU Tryckeri.

Redish, E. F. & Gupta, A. (2009). *Making Meaning with Math in Physics: A Semantic Analysis.* Contributed paper presented at GIREP 2009, Leicester, UK, August 20, 2009. Retrieved 2012-06-25 from
http://www.physics.umd.edu/perg/talks/redish/GIREP2009/GIREP%20Semantics.pdf

Redish, .E. F. (2003). *Teaching Physics.* USA: John Wiley & Sons, Inc.

Rothman, K. J., Greenland, S. & Lash, T. L. (2008). *Modern epidemiology 3rd edition*. Philadelphia: Lippincott Williams & Wilkins.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando: Academic Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for Research on Mathematics Teaching and Learning* (pp. 334-370). New York: MacMillan.

Sfard, A. (1991). On the dual nature of mathematical conceptions: reflections on process and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22(1), 1-36. DOI: 10.1007/BF00302715

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education,* 36(5), 404-411.
Stable URL: http://www.jstor.org/stable/30034943

Star, J. R. (2007). Foregrounding procedural knowledge. *Journal for Research in Mathematics Education,* 38(2), 132-135.
Stable URL: http://www.jstor.org/stable/30034953

Swedish National Agency for Education. (2000a). *Kursplan Fysik A*. Retrieved 2010-10-12 from http://www.Swedish National Agency for Education.se/forskola_och_skola/gymnasieutbildning/gymnasieskola_fore_ht_2011/2.3034/sok_amnen_och_kurser/13845/func/kursplan/id/3053/titleId/FY1201%2520-%2520Fysik%2520A

Swedish National Agency for Education. (2000b). *Kursplan Fysik B*. Retrieved 2010-10-12 from http://www.Swedish National Agency for Education.se/forskola_och_skola/gymnasieutbildning/gymnasieskola_fore_ht_2011/2.3034/sok_amnen_och_kurser/13845/func/kursplan/id/3054/titleId/FY1202%2520-%2520Fysik%2520B

Swedish National Agency for Education. (2003). *Lusten att lära – Med fokus på matematik: Nationella kvalitetsgranskningar 2001–2002.* Stockholm: Fritzes.

Swedish National Agency for Education. (2005). *Skolverkets Provbank. Hur används den och vad tycker användarna?* Retrieved 2012-10-21 from http://www.skolverket.se/prov-och-bedomning/rapporter-och-film

Swedish National Agency for Education (2009). *TIMSS Advanced 2008. Svenska gymnasieelevers kunskaper i avancerad matematik och fysik i ett internationellt perspektiv.* Stockholm: Fritzes.

Appendix

MATLAB code for calculating the normal distributed Mantel-Haenszel tests statistic, Z_{MH}.

```
close all
clear all

numberTeachers=zeros(8,1);
numberStudents=zeros(8,1);
antalGrupper=zeros(8);
Z_MH=zeros(8);
Z_MHsimpleIR=zeros(8,1);
Z_MHhardIR=zeros(8,1);

alfaMH=zeros(8);
chi2_MH=zeros(8);
p_chi2_MH=zeros(8);
p_Z_MH=zeros(8);

for m=1:8
    if m==1
        [siffror,text]=xlsread('ResultatFyAvt02_konvTeach2tal.xls');
        %FyAvt02: "SimpleIR" 1a(1190a)=(:,19) 2/0, 2b(1322b)=(:,22) 2/0;
        %"HardIR" 6(1237)=(:,26) 2/0, 11(1304)=(:,32) 3/0;
         %GCR 13(1184)=(:,35) 0/2, 16a(1218a)=(:,40+41) 1/1.

        %One match for each IR
        IR=[19,19, 22,22,26,26,32,32];
        GCR=[35,41,35,41,35,41,35,41];
        pIR=[2,2,2,2,2,2,3,3]; %maximum score on IR
        pGCR=[2,1,2,1,2,1,2,1];  %maximum score on CR

    elseif m==2
        [siffror,text]=xlsread('ResultatFyAht04_konvTeach2tal.xls');
        %FyAht04: "SimpleIR" 1(1407)=(:,19) 1/0,  3(1577)=(:,21) 2/0;
        %"HardIR" 8b(1404b)=(:,31) 2/0, 9(1035)=(:,32) 1/0;
        %GCR 7a(1458a)=(:,28) 2/0, 11(1574)=(:,36) 0/2.

        %One match for each IR
        IR=[19,19, 21,21,31,31,32,32];
        GCR=[28,36,28,36,28,36,28,36];
        pIR=[1,1,2,2,2,2,1,1]; %maximum score on IR
        pGCR=[2,2,2,2,2,2,2,2];  %maximum score on CR

    elseif m==3
        [siffror,text]=xlsread('ResultatFyAvt05_konvTeach2tal.xls');

        %FyAvt05: "SimpleIR" 3a(1584a)=(:,22) 2/0, 8a(1519a)=(:,29) 1/0;
        %"HardIR" 11(1590)=(:,34+35) 1/2, 12a(1497a)=(:,36) 0/1;
        %GCR 10(1588)=(:,32+33) 1/1, 13(1118)=(:,38) 0/2/¤.

        %One match for each IR
        IR=[22,22,29,29,35,35,36,36];
        GCR=[33,38,33,38,33,38,33,38];
        pIR=[2,2,1,1,2,2,1,1]; %maximum score on IR
        pGCR=[1,2,1,2,1,2,1,2]; %maximum score on CR

    elseif m==4
        [siffror,text]=xlsread('ResultatFyBvt02_konvTeach2tal.xls');
        %FyBvt02: "SimpleIR" 2(1231)=(:,20) 2/0, 4(836)=(:,22) 2/0;
        %"HardIR" 5(1324)=(:,23) 2/0, 9b(1325b)=(:,31) 0/2;
        %GCR 14(1238)=(:,38+39) 1/3, 15(1227)=(:,40+41) 3/4/¤.

        %One match for each IR
        IR=[20,20,22,22,23,23,31,31];
        GCR=[39,41,39,41,39,41,39,41];
```

```
        pIR=[2,2,2,2,2,2,2,2]; %maximum score on IR
        pGCR=[3,4,3,4,3,4,3,4]; %maximum score on CR

elseif m==5
        [siffror,text]=xlsread('ResultatFyBvt03_konvTeach2tal.xls');
        %FyBvt03: "SimpleIR" 2(1339)=(:,20) 2/0, 5(1370)=(:,23) 2/0;
        %"HardIR" 13(1043)=(:,36+37) 1/2, 14b(1396b)=(:,39) 0/2;
        %GCR 15(1373)=(:,40) 0/3¤, 16(1345)=(:,41+42) 2/6/¤.

        %One match for each IR
        IR=[20,20,23,23,37,37,39,39];
        GCR=[40,42,40,42,40,42,40,42];
        pIR=[2,2,2,2,2,2,2,2]; %maximum score on IR
        pGCR=[3,6,3,6,3,6,3,6]; %maximum score on CR

elseif m==6
        [siffror,text]=xlsread('ResultatFyBvt05_konvTeach2tal.xls');
        %FyBvt05: "SimpleIR" 6(262)=(:,25) 2/0, 9a(1529a)=(:,29) 2/0;
        %"HardIR" 11(1580)=(:,33) 0/3, 12c(1581c)=(:,36) 0/2;
        %GCR 7(1440)=(:,26) 0/2 14(1537)=(:,39+40) 1/2.

        %One match for each IR
        IR=[25,25,29,29,33,33,36,36];
        GCR=[26,40,26,40,26,40,26,40];
        pIR=[2,2,2,2,3,3,2,2]; %maximum score on IR
        pGCR=[2,2,2,2,2,2,2,2]; %maximum score on CR

elseif m==7
        [siffror,text]=xlsread('ResultatFyBvt06_konvTeach2tal.xls');
        %FyBvt06: "SimpleIR" 8a(840a)=(:,26) 2/0, 10b(1678b)=(:,30) 0/2;
        %"HArdIR" 13(1470)=(:,36) 0/3, 14b(1474b)=(:,38) 0/2;
        %GCR 12a(1560a)=(:,33) 0/1 12b(1560b)=(:,34) 0/2.

        %One match for each IR
        IR=[26,26,30,30,36,36,38,38];
        GCR=[33,34,33,34,33,34,33,34];
        pIR=[2,2,2,2,3,3,2,2]; %maximum score on IR
        pGCR=[1,2,1,2,1,2,1,2]; %maximum score on CR

else
        [siffror,text]=xlsread('ResultatFyBvt10_konvTeach2tal.xls');
        %FyBvt10: "SimpleIR" 3(1642)=(:,21) 2/0, 5(1721)=(:,24) 2/0;
        %"HardIR" 7a(1715a)=(:,26) 1/0, 13(1676)=(:,39+40) 1/3/¤;
        %GCR 9c(1718c)=(:,32) 0/2 10(1724)=(:,33) 0/3.

        %One match for each IR
        IR=[26,26,30,30,36,36,38,38];
        GCR=[33,34,33,34,33,34,33,34];
        pIR=[2,2,2,2,3,3,2,2]; %maximum score on IR
        pGCR=[1,2,1,2,1,2,1,2]; %maximum score on CR

end
for p=1:8
        %Mantel-Haenszel estimate of common odds ratio, alfa_MH, as well as
        %chi2 distributed test statistic and two-sided normal distributed test statistic, Z "=sqrt(chi2)"
        %to decide if alfa_MH is stistical significant different from 1.

        t=siffror(1,3); %TeacherID, data sorted so all students with the same teacher is after each other.
        n_11=0; %calculate 2x2 entries and the total sum
        n_10=0;
        n_01=0;
        n_00=0;
        n=0;

        alfa_numerator=0;
        alfa_denominator=0;
```

```
%notation according to McCullagh och Nelder
U=0;
var_U=0;

%how many different teachers in the calculation,
%i.e. how many contingency tables have entries >5.
count=0;

%number of different teachers for each test
g=0;
%number of students for each test
students=0;

for i=1:length(siffror)
        %sum up for each teacher
        if siffror(i,3)==t
              students=students+1;
              %success on both IR och GCR
              if siffror(i,IR(p))==pIR(p) && siffror(i,GCR(p))==pGCR(p)
                      n_11=n_11+1;
              %success on IR but not on GCR
              elseif siffror(i,IR(p))==pIR(p) && siffror(i,GCR(p))==0
                      n_10=n_10+1;
              %no success on IR but success on GCR
              elseif siffror(i,IR(p))==0 && siffror(i,GCR(p))==pGCR(p)
                      n_01=n_01+1;
              %no success on IR nor GCR
              elseif siffror(i,IR(p))==0 && siffror(i,GCR(p))==0
                      n_00=n_00+1;
              end
        else
              g=g+1;

              %2x2 contingency table
              obs=[n_11 n_10;n_01 n_00];
              n=sum(sum(obs));

              n_rad=sum(obs,2);
              m_kolumn=sum(obs,1);

              %Expected values under the assumption of H_0
              E=zeros(2);
              for j=1:2
                      for k=1:2
                              E(j,k)=sum(obs(j,:))*sum(obs(:,k))/n;
                      end
              end
              %Condition E > 5
              if E(:,:)>=5
                      p;
                      count=count+1;

                      alfa_numerator=alfa_numerator+(n_11*n_00/n);
                      alfa_denominator=alfa_denominator+(n_10*n_01/n);

                      U=U+(n_11-E(1,1));
                      var_U=var_U + prod(n_rad)*prod(m_kolumn)/(n^3-n^2);
                      t=siffror(i,3);
                      n_11=0;
                      n_10=0;
                      n_01=0;
                      n_00=0;
                      n=0;

                      students=students+1;
```

```
                    %include data from the first student for the next teacher
                    %success on both IR och GCR
                    if siffror(i,IR(p))==pIR(p) && siffror(i,GCR(p))==pGCR(p)
                        n_11=n_11+1;
                    %success on IR but not on GCR
                    elseif siffror(i,IR(p))==pIR(p) && siffror(i,GCR(p))==0
                        n_10=n_10+1;
                    %no success on IR but success on GCR
                    elseif siffror(i,IR(p))==0 && siffror(i,GCR(p))==pGCR(p)
                        n_01=n_01+1;
                    %no success on IR nor GCR
                    elseif siffror(i,IR(p))==0 && siffror(i,GCR(p))==0
                        n_00=n_00+1;
                    end
                else
                    t=siffror(i,3);
                    students=students+1;
                    %include data from the first student for the next teacher
                    %success on both IR och GCR
                    if siffror(i,IR(p))==pIR(p) && siffror(i,GCR(p))==pGCR(p)
                        n_11=n_11+1;
                    %success on IR but not on GCR
                    elseif siffror(i,IR(p))==pIR(p) && siffror(i,GCR(p))==0
                        n_10=n_10+1;
                    %no success on IR but success on GCR
                    elseif siffror(i,IR(p))==0 && siffror(i,GCR(p))==pGCR(p)
                        n_01=n_01+1;
                    %no success on IR nor GCR
                    elseif siffror(i,IR(p))==0 && siffror(i,GCR(p))==0
                        n_00=n_00+1;
                    end
                end
            end
        end
        %How many teachers fulfil the condition E > = 5 for each mathing in each test
         antalGrupper(m,p)=count;

        %total number of teacher for each test
        numberTeachers(m)=g;

        %totalt number of students for each test
        numberStudents(m)=students;

        %weighted odds-ratio
        alfaMH(m,p)=alfa_numerator/alfa_denominator;

        %MantelHaenszel Chi2-statistic with ½ = Yates correction for continuity
        chi2_MH(m,p)=(abs(U)-1/2)^2/var_U;

        %MantelHaenszel Z-statistic with ½ = Yates correction for continuity
        sigma_U=sqrt(var_U);
        Z_MH(m,p)=(U-1/2)/sigma_U;
    end
    % 95% limits for chi2-distribution with 1 degree of freedom, 2x2-tables
    limit_95chi2=chi2inv(0.95,1);
    95% limits for normal distribution with mean 0 and standard deviation 1.
    limit_95_Z=norminv([0.025, 0.975],0,1);
end
```