Comparing national and teacher-made tests, explaining differences and examining impact

Jesper Boesen
ASSESSING MATHEMATICAL CREATIVITY
– Comparing national and teacher-made tests, explaining differences and examining impact

© Jesper Boesen
Doctoral thesis no. 34, 2006. Department of Mathematics and Mathematical Statistics, Umeå university
ISBN 91-7264-136-3
ISSN 1102-8300
Print: Print & Media, Umeå university, Umeå
Cover design: Anders Wallby & Jesper Boesen
To my wife Maria and our son Elliot
Preface

A few years ago, attending a conference in honour of professor Hans Wallin’s retirement, I remember presenting my research intentions about the impact of national assessments – how a national assessment system could work as an assisting curriculum, if teachers changed any of their practise as a result of impact from the national tests and so on – ideas that at the time were severely vague. I also remember being the only doctoral student in the education group at the department who didn’t use the ideas from Pólya, Schoenfeld and Lithner. I felt like a black sheep in a white herd. A few years later – now finishing my thesis and having had the time to catch up my reading – I’ve been using some of the ideas from Pólya, Schoenfeld and Lithner to compare one part of teachers’ practice, their development of written classroom assessment with the national course tests in mathematics. This comparison serves as a ground for investigating a proposed impact from the latter.

This interest is strongly connected to my former work as research assistant at the Department of Educational Measurement, Umeå university, and at PRIM-gruppen, Stockholm Institute of Education, where I worked with the National tests for the upper secondary school mathematics for some years. As I saw it, it would be wise to use the experience I’ve accumulated during my years as test developer, which meant that my interest should handle some aspect of the national tests. There is a lot to be explored in the vast field of assessment, of course, but one thing that we seemed to know too little about were the effects of testing. To be more specific: The effects the tests had on teachers practice. How did the tests work as a tool for the implementation of a new curriculum? We often heard teachers saying that the tests had such great influence, by setting an agenda for types of items to handle, topics to deal with more extensively and so on. Others said that the textbooks
had the greatest influence. I was convinced that the tests had some influence (this was founded partly on my own experience as a teacher at a college and as a colleague hearing what other teachers claimed), but I was not sure we knew how it influenced and exactly what. One trivial effect was that the test took time to carry out, both the actual test-taking time and the teacher assessment. More interesting, however, was that we didn’t know in what way, in what magnitude, or if at all the tests influenced the teaching or testing of mathematics in the Swedish schools. What made me spend nearly five years working with these questions? In my opinion the national course tests have great importance as some sort of model, showing good examples of what to deal with in the practice of teaching, giving examples to help implementing the curricula or future curricula to come. I believed, and still believe it can be a good agent for change. My general interest thus concerns how the classroom mathematics changes (or not) by the influence of the national course tests. My intention is to specify this concern in the following text and in my research papers.

Why is this a concern for a mathematics educator? Why not a scholar in general pedagogy? To answer this question a short personal declaration is in place. In my view – in a very simplified version – the tests, as a result of the curriculum Lpf94 (Swedish National Agency for Education, 2001a), represents parts of a shift in view of knowledge. One interpretation is that focus has shifted from procedural knowledge to conceptual knowledge (Hiebert, 1986). What is believed as being important mathematical competences has changed. It’s not only which content that is covered, it’s also what kind of competences that are required and the kind of quality in the knowledge that have come in focus. Some key words that characterise these competences could be, reasoning, modelling, generalising, communicating and the ability to critically examine things, and so on. These might be competences that by mathematicians and mathematics educators always have been seen to be important features, but which not earlier have been focused in the curricula. What I would like to know is if there has been a similar focus shift within the teaching community, or at least one part of it; written assessments. And if so, has teacher-made tests changed as a consequence of the national tests? This is of course a too big question for me to fully answer, but it is behind me as a driving force. To identify the use of different competences (or the intention to use) advocated
by teachers I would certainly need to use my own mathematical competences. (For instance when examining the different requirements put up by test tasks, i.e. basic skill items, reasoning, modelling, problem solving, profound understanding and so on.) This is why I see this as a concern for the mathematics education researcher.

Side projects

During the years as a PhD. student, I’ve not managed to stay totally focused on my work on the thesis. There have always been other interesting projects going on and I tend not to say no. However, although these projects cannot be seen as part of my formal education, all projects which I have attended to, have in one way or another enriched my experiences within the field. Without going into details I simply list some of the things I have been involved in:

- The Swedish National Agency for Education conducted a study on the effects of the National Tests in the lower secondary School. I was engaged as a mathematics education expert and carried out many interviews with teachers, principals and students about their views concerning impact from the National Tests. For results from that project see Naeslund and Swedish National Agency for Education (2004).

- During my work at the Department of Educational Measurements, Umeå university, I conducted a study where the assessment inter-rater reliability of the so-called aspect assessed tasks on the National Course Tests was investigated. For results see Boesen (2004).

- At the National Centre for Mathematics Education, Göteborg university, I was part of the editorial team for the translation of the book 'International perspectives on learning and teaching mathematics' (Clarke and NCM, 2004) into a revised Swedish version, available in autumn 2006, see 'Lära och undervisa matematik' (Boesen et al., 2006).

- Due to an earlier approved planning grant (2005-7599) I was on leave from my PhD. studies for three month in the beginning
of this year and worked on applications for new research grants directed to The Swedish Research Council (2006-2602 and 2006-2519). These projects will (if approved) continue to both extend and deepen the research about the impact of the national test system.

Acknowledgements

This work is not the result of one man’s isolated hard work, it’s mainly a collaborative work. This thesis wouldn’t have become what it is today if it wasn’t for my supervisors and colleagues Johan Lithner and Torulf Palm. They have both supported me throughout the years and every time I got stuck in one or another way. I have always felt that they had time to talk to me, especially this summer, when I called or e-mailed at the most odd hours. Many of the ideas presented or investigated in the following papers have sprung out of joint discussions, some are mine, some were placed at my disposal, but most are probably the result of us having a good time. Thanks Johan and Torulf! I also want to thank my third supervisor Hans Wallin for giving me inspiration and well-needed guidance in pure mathematics.

There is one person who also have made this thesis possible, Jan-Olof Lindström. If it wouldn’t have been for his support with travels in the beginning of my studies and by hiring me I wouldn’t been able to accept the offer to start this Ph. D. education. I still remember the day at Sigtuna, JO.

If it hadn’t been for all teachers who have helped us by sending in tests, and all teachers who despite a heavy workload, let me interview them, there would not have been any material to research. You all have made this thesis possible.

I would like to thank The Bank of Sweden Tercentenary Foundation (2000-1003) and The Swedish Research Council (2001-3989 and 2003-4389) for financial support. Deep gratitude is directed to the Graduate School in Mathematics Education, my fellow doctoral students and especially Gerd Brandell, for arranging courses, seminars and conferences way beyond what can reasonably be expected.

Throughout these years I have been moving around and as a consequence I have had the opportunity to get to know different departments.
I would like to thank all involved at the Centre for Mathematical Sciences, Göteborg university and Chalmers for letting me use an office there for nearly two years, it was a pleasure.

During the last two years I have been located and partly worked at the National Centre for Mathematics Education, NCM, Göteborg university. This centre is a place where I feel at home and as an honour to my colleagues I’ve colored the thesis’ front side in an ’espresso style’, you all know what that mean. Special thanks goes to Calle Flognman and Lars Mouwitz who carefully read and corrected many faulty sentences throughout previous versions of the thesis and to Bengt Johansson for always sharing valuable references and ’lending’ me time to finalise this thesis.

Finally, and with my deepest gratitude, I would like to dedicate this thesis to my family, both my large family, my late mother, father, all my sisters and my in-laws, but especially to my beloved wife Maria and our wonderful son Elliot, I love you both so very much!

/Jesper Boesen, August 2006
Disposition of the thesis

This thesis consists of two parts. Part I is a preamble – often referred to as a coat in Sweden – and Part II contains four papers. The main purposes of Part I are to summarise, synthesise and discuss the aggregated results of Part II, where the main scientific work is presented.

Part I:

Introduction In this chapter the two underlying tenets, which constitutes the points of departure for the whole thesis, are presented. Further the overall aim of the thesis is presented.

Summary of papers Each of the studies presented in papers I to IV are shortly summarised.

Theoretical background In this chapter two different frameworks are presented. One concerns different types of mathematical reasoning. This part of the chapter can also be seen as extending the description of tenet one presented in the introductory chapter. The second framework comprises descriptions of six mathematical competences, which are meant to capture most of the essential goals described in the Swedish mathematics policy documents. The text in this chapter is mainly the same as presented in the papers respectively, but are extended with collections of examples to clarify the nature of the different reasoning types and the different competences.

Methodological considerations In this chapter some of the choices that had to be made throughout the research process are presented. The different methods used and connections to the theories are shortly presented and it is argued why the latter were
used. It is also shortly discussed what the methods cannot capture and what is excluded.

Conclusions and discussion In this chapter I try to sum up and merge the results of the four studies. The chapter is ended by a discussion of what may be the consequences of these conclusions.

Part II:
List of papers


1All papers are available in the pre-print series Research Reports in Mathematics Education, but are recently submitted to international journals.
Contents

Preface i
Disposition of the thesis vii

I Preamble – The Coat 1

1 Introduction 3
  1.1 Background ............................................. 3
  1.2 The Swedish assessment system ....................... 7
  1.3 General aim ............................................ 9

2 Summary of papers 11
  2.1 Coherence of the papers ............................... 11
  2.2 Paper I .................................................. 13
  2.3 Paper II .................................................. 13
  2.4 Paper III ............................................... 14
  2.5 Paper IV ............................................... 15

3 Theoretical background 17
  3.1 Mathematical reasoning ............................... 17
  3.2 Mathematical competences ........................... 30

4 Methodological considerations 39

5 Conclusions and discussion 45
  5.1 Conclusions ............................................. 45
  5.2 Discussion ............................................. 46
  5.3 Implications .......................................... 50
II The Papers 53

6 Paper I: 55
6.1 Introduction ................................................. 57
6.2 Mathematical reasoning ................................... 58
6.3 Research questions ......................................... 64
6.4 Method ......................................................... 65
6.5 Results ......................................................... 77
6.6 Discussion .................................................... 80

7 Paper II: 85
7.1 Introduction .................................................. 87
7.2 Mathematical reasoning ................................... 89
7.3 Research questions and method .......................... 93
7.4 Results ......................................................... 103
7.5 Conclusions and Discussion ............................... 110

8 Paper III: 115
8.1 Introduction .................................................. 117
8.2 Theory .......................................................... 120
8.3 Aim .............................................................. 127
8.4 Method .......................................................... 128
8.5 Results .......................................................... 134
8.6 Discussion ...................................................... 148

9 Paper IV: 153
9.1 Introduction .................................................. 155
9.2 Educational goals and reform ............................. 157
9.3 The tested mathematics ..................................... 158
9.4 Aim .............................................................. 164
9.5 Method .......................................................... 165
9.6 Results .......................................................... 170
9.7 Conclusion and discussion .................................. 182

Bibliography 187
Part I

Preamble – The Coat
Chapter 1

Introduction

1.1 Background

This thesis is based on four separate papers, which all from different angles treat assessment and different forms of mathematical reasoning in the upper secondary school. Two underlying tenets are departed from and can be seen as parts of the background for conducting this research.

Findings from earlier research have shown that one crucial reason behind many students’ difficulties in mathematics is their strive for superficial solution strategies when solving tasks and that they consequently conduct ways of reasoning that can be seen in a mathematical perspective as both non-intrinsic and non-creative. Several studies have shown that students in a variety of different situations use what throughout this thesis is called imitative reasoning (Bergqvist et al., 2003; Boesen et al., 2005; Lithner, 2000b, 2003b; Lithner and Långström, 2006). This way of reasoning, or way of solving tasks, is mainly characterised by a search for familiar examples, solutions or other forms of ‘masters’ from where either the whole or parts of a solution can be copied or reproduced. This way of reasoning is, on one hand, often successful since students and teachers seem to reproduce mutual expectations. Teachers let their students work with tasks, which they expect them to manage and students use solution methods that mainly are imitated from their previous work, which they in turn know will work in most cases (cf. Brousseau (1997)). And since they as it
seem very seldom meet other types of tasks than the familiar ones, they never get a real chance to develop their problem solving skills. What this kind of imitative reasoning cannot do is to serve as a foundation or means to solve unfamiliar tasks, i.e. problems where ready-made solutions aren’t available. The term ‘problem’ is used with many different meanings, and can in principle mean everything from a dressed up exercise to front-line research. In this thesis the following meaning will be used. Whether a task constitutes a problem or not depends both on the solver and the task. What might be a problem for one student might not be so for another. This way of specifying a problem originates from Schoenfeld (1985) who formulated this idea in the following way:

"Being a ‘problem’ is not a property inherent in a mathematical task. Rather, it is a particular relationship between the individual and the task that makes the task a problem for that person. The word problem is used here in this relative sense, as a task that is difficult for the individual who is trying to solve it. Moreover, that difficulty should be an intellectual impasse rather than a computational one. [...] If one has ready access to a solution schema for a mathematical task, that task is an exercise and not a problem.”

(p. 74)

When attempting to solve a problem, the use of imitative reasoning often, if not by sheer luck, leads to a dead end. What is needed is the production of something new, a type of reasoning that extends from what can be imitated. It is this superficial strive for solutions to recognise that may explain why so many students have difficulties when solving problems. They are simply not used to, and probably have had very limited opportunities to learn and to construct their own ways of reasoning. This phenomenon, i.e. the strive for superficial solutions, represents the first of the two tenets. In this thesis creative mathematically founded reasoning is used to denote the kind of reasoning required in order to be able to solve problems. How this kind of reasoning is defined, characterised and related to other reasoning types will be thoroughly described in Chapter 3.1.

Because of this first tenet, it’s important to continue the investigation of the learning environment to find explanations for this heavy
focus on superficial solution strategies, which seem to hamper the development of more sustainable mathematical abilities. One part of this environment, which is also believed to influence learning a lot, is the kind of assessments students meet during their education. The second tenet is thus represented by the idea that assessment can influence practice.

It is often stated that tests and assessments influences school practise and have, or at least should have, a central role in reforming education (Barnes et al., 2000; Bass, 1993; Gipps, 1994; Kulm, 1993; Leder, 1992; Niss, 1993). The rhetoric behind can simply be formulated: If teachers are to change their teaching their assessment practises must change as well. Conversely it’s assumed that change in assessment practises might lead to changed teaching (Kulm, 1993; Bass, 1993). The old ‘folklore’

"What you test is what you get"

is a well-used representative for this idea. Centrally administered tests can influence teacher practice, such as subject coverage, use of time and treatment of subject matters. It can also influence assessment, the higher the stake the higher the impact (Smith, 1991). Barnes et al. (2000) claim that curriculum reform often tend to result in only new paperwork if not accompanied by well aligned test systems. Further, they claim that national assessments in Australia clearly worked as a catalyst for curriculum reform. By a study of local working-plans, teacher-made tests etc., partly by interviews and inquiries, they found distinct similarities between the national assessments and teacher practise.

Black and Wiliam (1998) do not specifically treat influence from centrally administered tests, but how changes in classroom assessment practices can lead to better student learning. They claim by their extensive review of the literature concerning formative assessment, that the answers are uniform. They show e.g. that traditional teacher’s assessments and test traditions are closely connected to reproducing facts and procedures, tasks are seldom analysed and discussed with colleagues, for similar results see also Palm et al. (2005); Cooney et al. (2001); Björklund (2004). Black and Wiliam argue that an improvement of this practise and assessments aligned with the aims of education, result in improved student learning, see also de Lange and Romberg (2005).
These results point to the great potential that lies in the use of well-constructed assessments.

Returning to centrally administered tests there are those who express more uncertainty concerning impact than e.g. Barnes et al. (2000). Cimbricz (2002) claims that

"although a relationship between the state-mandated testing and teachers’ beliefs and practice does exist, testing does not appear to be an exclusive or a primary lever of change". (Discussion)

By an extensive review of the literature Mehrens (2002) seeks answers to the question Consequences of Assessment: What is the Evidence? His results are summarised in a dozen points concerning potential impact, out of which two are shown below:

"3. It would profit us to have more research. 4. The evidence we do have is inadequate with respect to drawing any cause/effect conclusions about the consequences.” (p. 173)

In Sweden this research is essentially lacking. There are some official reports from the Swedish National Agency for Education (2004; 2005b; 2005a). In short, it’s reported that teachers say that they are influenced, younger teachers to a higher degree than elder, teachers say that the national tests influence their planning and course coverage to relatively high extent, ways of working to less extent. Teachers also say that they especially treat topics thought to be covered in the national tests and work with similar tasks.

Despite these studies, both internationally and specifically in the Swedish settings, studies concerning the relation between national tests and teacher’s test-tradition essentially are missing. Most studies cover what teachers say about potential impact, little about what actually happens with their practice. Together with what beliefs-research say about weak connections between what is said and what is done (Handal, 2003; Gregoire, 2003), the evidence for real impact doesn’t seem that strong. As this very short overview above points to, a lot of the literature are about rhetoric, but irrespective of a partly polarised view (Glaser and Silver, 1994), there’s a large unanimity about the need for further research (Broadfoot, 2002; Mehrens, 2002).
1.2 The Swedish assessment system

National and professionally developed national mathematics tests have been a part of the Swedish education system since the 1940’s (Kilpatrick and Johansson, 1994). Today’s assessment system has been in use for about eleven years now, and its introduction can be seen as one part of the general educational reform introduced by a new curriculum in 1994 (Swedish National Agency for Education, 2001a). With this introduction of a new curriculum Sweden left a system with relatively detailed descriptions of what should be covered, in which order and how it should be taught. The current system is characterised by what, at least in Sweden, is called ‘participatory goal fulfillment’. What this means in reality is that we have a curriculum that isn’t very detailed and has to be locally interpreted. The policy documents serve with the guiding principles. The realisation of these ideas is up to the teachers of the field to decide. This local interpretation is of course dependent upon what the individual teacher sees as important mathematical abilities.

In short the new system could be described as involving a new knowledge view, a constructivist view on learning, and the mathematics syllabuses describe processes and different competences to a much higher degree than earlier documents (Wyndhamn, 1997). Lundgren (1999) summarises these ideas:

"The goals of curricula and syllabus are to less extent expressed in terms of subject matters and more in terms of concepts, connections and in terms of knowledge as tools for learning"\(^1\) (p. 39).

Goals in modern curricula are often of two different types. One type of goals concerns subject matters or subject content and the other abilities or competences. The introduction of the process- and competence-goals can be seen to represent the most significant educational reform introduced within mathematics education. However this introduction seems to have had limited consequences for the learning environment (Palm et al., 2005; Lithner, 2000b, 2004). Here the term learning environment includes the part organised by school, represented by teaching, textbooks, study plans, tasks students work with and teacher-made tests.

\(^1\)Authors translation.
Simply put, if curricula states ‘algebra’ teachers will provide students with algebra, but despite statements about problem solving few traces of that can be found in the learning environment. The competence-goals summarises well what research in mathematics education and other educational actors consider being the most urgent reform. The goals are expressed slightly different in different contexts, perhaps most elaborated are the process-goals in the NCTM Principles and Standards (NCTM, 2000) or the competences in (Niss and Jensen, 2002). Central in this reform is that it’s not enough to formulate goals in terms of subject content, which has been done traditionally both in catchall terms (arithmetic, algebra, geometry, statistics etc.) and in detailed terms (tables of multiplication, solving of specific equations, area calculation, calculations of means, etc.). It is necessary to also formulate goals in terms of abilities or competences that students should learn and which are general for different ages and subject contents.

This change in curriculum, and view of knowledge led to the need for a new assessment system, with slightly different purposes (Lindström, 1994; Pettersson and Kjellström, 1995). There are a number of official aims of the national test material. In a government decision (Ministry of Education, 1999) it is stated that the different test materials aim at:

i) acting as support for analysis and assessment of pupils strong and weak sides as basic data for decisions concerning necessary actions;

ii) acting as support when assessing whether or not the pupils have reached the postulated goals in the syllabuses;

iii) acting as support for the principle of equal assessment and fair grading.  

In addition to these aims there are other aims that have been expressed by the Swedish National Agency for Education in different publications/materials e.g.

"All tests shall be seen as a tangible proposal of the curriculum’s view of knowledge and the syllabus’ subject view." (Swedish

\footnote{Authors translation.}
The national course tests are in this perspective seen as potential exemplary models by pointing at which qualities are due for different grades and also exemplary by themselves as tests, by showing how different qualities of knowledge can be assessed. Indirectly the national course tests can be seen as instruments for in-service training for teachers, regarding the intentions of the curriculum and the view of knowledge they mediate. This idea, about assessments as a vehicle for change, is included in a framework developed in order to facilitate research about reform movements, e.g. 'The Standards' NCTM (2000). In this framework (Weiss et al., 2001) reform messages are assumed to influence the educational system through three main channels 'Curriculum', 'Teacher development' and 'Assessment and accountability'. These channels are assumed to be indirect, intertwined, complex and interactive, both within themselves as well as with different layers in the educational system. They are also assumed to operate slowly, decades rather than years.

It is with these more implicit goals, presented above and within the assessment channel in mind this thesis is conducted.

1.3 General aim

Each of the papers in part II has its own aims and research questions. What they all have in common is that they all, directly or indirectly, continue the survey concerning tenet one presented in the introduction, i.e. they all investigates the learning environment for reasons for the high emphasis on imitative reasoning. By this survey, by comparing teacher-made tests with the national course tests, impact from the latter on teachers’ development of written classroom assessment can be investigated. The overall aims of this thesis are thus to:

1. Extend the knowledge about the learning environment by seeking answers to the question – What kind of mathematical abilities are required to solve written assessments?
2. Extend the knowledge about the assessment systems’ potential role as a reformer by seeking answers to the question – In what ways do national course tests influence teachers’ development of written assessments?

On top of these questions, as within most research, ways of how to reach answers to these questions have to be developed. Thus, a secondary and indirect aim is to suggest ways of investigating how different requirements of tasks can be established and how impact from the national assessment system can be approached.
Chapter 2

Summary of papers

In this chapter each research study will be shortly summarised, their combining ideas as well as how these papers are connected to each other will be presented.

2.1 Coherence of the papers

As presented in the previous chapter, students’ use of superficial reasoning seems to be a main reason for learning difficulties in mathematics (cf. tenet one). It is therefore important to continue the investigation for the reasons for this use and the components that may affect students’ mathematical reasoning development. Assessments have been claimed to be a component that significantly may influence students’ learning (cf. tenet two).

Therefore, the purpose of the study in Paper 1 was to investigate the kind of mathematical reasoning that is required to successfully solve tasks in the written tests students encounter in their learning environment. This study showed that a majority of the tasks in teacher-made assessment could be solved successfully by using only imitative reasoning. The national tests however, required creative mathematically founded reasoning to a much higher extent.

The question about what kind of reasoning the students really use, regardless of what we theoretically claimed to be required on these tests, still remains. This question is investigated in Paper 2. Here is also the relation between the theoretically established reasoning requirements,
i.e. the kind of reasoning the students have to use in order to successfully solve included tasks, and the reasoning actually used by students studied. The results showed that the students to large extent did apply the same reasoning as were required, which means that the framework and analysis procedure can be valuable tools when developing tests. It also strengthens many of the results throughout this thesis. A consequence of this concordance is that as in the case with national tests with high demands regarding reasoning also resulted in a higher use of such reasoning, i.e. creative mathematically founded reasoning. Paper 2 can thus be seen to have validated the used framework and the analysis procedure for establishing these requirements.

Paper 3 investigates the reasons for why the teacher-made tests emphasises low-quality reasoning found in paper I. In short the study showed that the high degree of tasks solvable by imitative reasoning in teacher-made tests seems explainable by amalgamating the following factors: (i) Limited awareness of differences in reasoning requirements, (ii) low expectations of students abilities and (iii) the desire to get students passing the tests, which was believed easier when excluding creative reasoning from the tests. Information about these reasons is decisive for the possibilities of changing this emphasis. Results from this study can also be used heuristically to explain some of the results found in paper 4, concerning those teachers that did not seem to be influenced by the national tests.

There are many suggestions in the literature that high-stake tests affect practice in the classroom. Therefore, the national tests may influence teachers in their development of classroom tests. Findings from paper I suggests that this proposed impact seem to have had a limited effect, at least regarding the kind of reasoning required to solve included tasks. What about other competences described in the policy documents? Paper 4 investigates if the Swedish national tests have had such an impact on teacher-made classroom assessment. Results showed that impact in terms of similar distribution of tested competences is very limited. The study however showed the existence of impact from the national tests on teachers test development and how this impact may operate.
2.2 Paper I

Title: The requirements of mathematical reasoning in upper secondary level assessments
Authors: Torulf Palm, Jesper Boesen, Johan Lithner
See page 55 and forward.

In this study a stratified and random selection of teacher-made tests that students encounter during their studies in mathematics were collected. In addition eight national tests were included in the sample, as representatives for the national tests. These two groups of tests were compared regarding the kind of reasoning that were required to solve included tasks, i.e. the type of reasoning the students have to use in order to successfully solve included tasks. The different types of reasoning are thoroughly described in Chapter 3.1. The results show that only a small proportion of the tasks in the teacher-made tests required the students to produce new reasoning and to consider the intrinsic mathematical properties involved in the tasks. In contrast, the results also show that the national tests include a large proportion of tasks to which memorisation of facts and procedures are not sufficient.

2.3 Paper II

Title: The relation between test task requirements and the reasoning used by students – An analysis of an authentic national test situation
Authors: Jesper Boesen, Johan Lithner, Torulf Palm
See page 85 and forward.

In the study reported in paper II eight students, two per course A-D in the upper secondary mathematics courses were video-taped in a think-aloud study. This was an authentic national high-stake assessment for the students and they were actually solving a real National Test with its consequences. Findings of this study showed that the theoretically established reasoning requirements worked fully satisfactory. It was also found that students’ very seldom used reasoning ‘more creative’ than what was required, and thus if a tests is to assess e.g. conceptual understanding, tasks requiring only imitative reasoning is not enough, i.e.
tasks sharing familiar features with textbook tasks should be avoided. Students used the same kind of reasoning in 74% of the cases. Large differences between courses were found, where the A-course stood out as the high demander, i.e. that course held both higher demands and the students who solved it also used high-quality reasoning to a much higher extent when compared to the other courses and the other students. The A course in the Swedish system is mainly a repetition of the lower secondary school, whereas courses B-D builds upon mainly new subject matter. This led us to the conclusion that a very content-dense curriculum may hamper the successful use and development of creative mathematically founded reasoning.

2.4 Paper III

Title: Why emphasis imitative reasoning? – Teacher-made tests

Author: Jesper Boesen

See page 115 and forward.

Paper I found large discrepancies between reasoning requirements in Teacher Made Tests (TMTs) and National Course Tests (NCTs). These discrepancies consisted of a narrow focus on imitative reasoning in the TMTs and a relatively high focus on creative reasoning in the NCTs (Palm et al., 2005). Paper III aims at finding explanations to the reasons for this discrepancy. A stratified sample of teachers from the first study were interviewed about: 1) their awareness about creative mathematically founded reasoning, 2) their general beliefs about creative mathematically founded reasoning and finally, 3) their specific beliefs and concerns about creative mathematically founded reasoning in their own written tests. Findings of this study show that the high degree of tasks solvable by imitative reasoning in TMTs seems explainable by an amalgam of the following factors: (i) Limited awareness of differences in reasoning requirements, (ii) low expectations of students abilities and (iii) the desire to get students passing the tests, which was believed easier when excluding creative mathematically founded reasoning from the tests.
2.5 Paper IV

Title: National course tests’ impact on teachers’ written classroom assessment

Author: Jesper Boesen

See page 153 and forward.

It is often claimed that high-stake tests influence teacher practise. An earlier study (paper I) revealed large discrepancies between teacher made tests (TMTs) and national course tests (NCTs) regarding one central aspect of the tested mathematics – the kind of mathematical reasoning required to solve included tasks (Palm et al., 2005). This implies that the supposed impact from the NCTs perhaps isn’t that strong after all. What about other aspects of the tested mathematics, other abilities?

In this last study the same sample of TMTs as was used in paper I was used and they were categorised by the use of another framework capturing most of the central competences, as described in the policy documents.

The results show that the NCTs and TMTs in general are focusing on very different aspects of mathematics, where the latter focus on algorithmic skills and the former to much greater extent require competences such as modelling, problem solving, communication and so on. The real impact of the NCTs on how teachers’ develop their own tests seem in this perspective to be fairly modest. However, the relatively few teachers with NCT-similar tests confirmed impact from the NCTs, and two ways of how this impact may operate have been exemplified; consciously or indirectly trying to imitate the NCTs. Thus, the conclusive reason for some teachers development of tests similar to the national tests seem to be their conscious decision to do so.
Chapter 3

Theoretical background

3.1 A framework for analysing mathematical reasoning

The framework for analysing mathematical reasoning is a summary of Lithner (2005) which is a theoretical structuring of the outcomes of a series of empirical studies aiming at analysing characteristics of the relation between reasoning types and learning difficulties in mathematics (Bergqvist et al., 2003; Lithner, 2000b, 2003b, 2004). The framework defines different types of mathematical reasoning found in the empirical studies. These comprise rich problem solving (in terms of Creative mathematically founded reasoning) and a family of reasoning types characterised by a strive for a recall of algorithms or facts (in terms of Imitative reasoning). The reasoning types will be defined below (see Figure 3.1 for an overview). Before defining the different reasoning types a number of terms that are used in the definitions will be treated. Reasoning throughout this thesis is the line of thought, the way of thinking, adopted to produce assertions and reach conclusions. It is not necessarily based on formal deductive logic, and may even be incorrect as long as there are some kind of sensible (to the reasoner) arguments that guide the thinking. Argumentation is the substantiation, the part of the reasoning that aims at convincing oneself, or someone else, that the reasoning is appropriate. In particular in a task-solving situation, which is called a problematic situation if it is not clear how to proceed, two types of argumentation are central: (1) The strategy
choice, where ‘choice’ is seen in a wide sense (choose, recall, construct, discover, guess, etc.), can be supported by predictive argumentation: Will the strategy solve the difficulty? (2) The strategy implementation can be supported by verificative argumentation: did the strategy solve the difficulty?

3.1.1 Creative Mathematically Founded Reasoning

When contrasting creative reasoning in mathematics to imitative reasoning there are two types of considerations to make that will be briefly discussed below: What makes it creative and what makes it mathematical? Creativity. According to Haylock (1997) there are at least two major ways in which the term is used: i) thinking that is divergent and overcomes fixation and ii) the thinking behind a product that is perceived as grandiose by a large group of people. Silver (1997) argues that “although creativity is being associated with the notion of ‘genius’ or exceptional ability, it can be productive for mathematics educators to view creativity instead as an orientation or disposition toward mathematical activity that can be fostered broadly in the general school population”. Thus, a notion of creativeness limited to ii), the thinking of geniuses or the creation of great ideas with large impact on our society, is not suitable for the purposes of this paper. Instead, central are the creative aspects of ordinary students’ everyday task solving thinking, the reasoning that goes beyond just following strict algorithmic paths or recalling ideas provided by others. Regarding i) Haylock (1997) sees
two types of fixation. Content universe fixation concerns the range of elements seen as appropriate for applications to a given problem: useful knowledge is not seen as useful. Algorithmic fixation is shown in the repeated use of an initially successful algorithm that becomes an inappropriate fixation. According to Silver (1997), a new research based view of creativity suggests that it is related to deep, flexible knowledge in content domains and associated with long periods of work and reflection rather than rapid and exceptional insights. The framework of this paper amalgamates Haylock’s and Silver’s views and sees fluency, flexibility and novelty as key qualities of creativity. However, in the analyses of the reasoning in the empirical studies that resulted in the framework (Lithner, 2005) a need arose to characterise what distinguishes creative mathematical reasoning from general creative reasoning. This resulted in complementing the description of creativity by adding the aspects of plausibility and mathematical foundation. Plausibility. One could claim that the argumentation in creative mathematical reasoning should be logically strict as in proof, but this is inappropriate in the school context. In school tasks, one of the goals is also to achieve a high degree of certainty, but one crucial distinction from professional tasks is that within the didactic contract (Brousseau, 1997) of school it is allowed to guess, to take chances, and use ideas and reasoning that are not completely firmly founded. Even in exams it is often accepted to have only, for example 50% of the answers correct, while it is absurd if mathematicians or the engineers are correct in only 50% of their conclusions. This implies that is is allowed, and perhaps even encouraged, within school task solving to use forms of mathematical reasoning with considerably reduced requirements on logical rigour. Pólya (1954, p. iv) stresses the important role of reasoning that is less strict than proof: “In strict reasoning the principal thing is to distinguish a proof from a guess, a valid demonstration from an invalid attempt. In plausible reasoning the principal thing is to distinguish a guess from a guess, a more reasonable guess from a less reasonable guess”. Mathematical foundation. In this framework, well-founded arguments are anchored in intrinsic properties of components involved in the reasoning. Before specifying this foundation it is necessary to briefly discuss the components one is reasoning about, which consists of objects, transformations, and concepts. The object is the fundamental entity, the ‘thing’ that one is doing something with or the result of doing something, e.g. numbers,
variables, functions, graphs, diagrams, matrices, etc. A *transformation* is what is being done to an object (or several), and the outcome is another object (or several). Counting apples is a transformation applied to real-world objects and the outcome is a number. To calculate a determinant is a transformation on a matrix. A well-defined sequence of transformations, e.g. finding a local maximum of a third-degree polynomial will be called a *procedure*. A concept is a central mathematical idea built on a related set of objects, transformations, and their properties, for example, the concept of function or the concept of infinity. A property of a component is *mathematical* if it is accepted by the mathematical society as correct. Since a property of a component may be more or less relevant in a particular context and problematic situation, it is necessary to distinguish between *intrinsic* properties that are central and *surface* properties that have no or little relevance in a specific situation. In deciding which of the fractions $\frac{99}{120}$ and $\frac{3}{2}$ that is the largest, the size of the numbers (99, 120, 3 and 2) is a surface property that is insufficient to consider in this particular task while the quotient captures the intrinsic property.

With the preparations made above it is now possible to define *Creative mathematically founded reasoning (CR)* as fulfilling the following conditions: I. Novelty. A new (to the reasoner) sequence of solution reasoning is created, or a forgotten sequence is a recreated. To imitate an answer or a solution procedure is not included in CR. II. Flexibility. It fluently admits different approaches and adoptions to the situation. It does not suffer from fixation that hinders the progress. III. Plausibility. There are arguments supporting the strategy choice and/or strategy implementation, motivating why the conclusions are true or plausible. Guesses, vague intuitions and affective reasons are not considered. IV. Mathematical foundation. The argumentation is founded on intrinsic mathematical properties of the components involved in the reasoning.

### 3.1.2 Imitative Reasoning

Different versions of imitative reasoning are types of reasoning that are more frequently used by students than CR is. In these types of reasoning students copy or follow a model or an example without any attempts at originality. Learning difficulties are partly related to a reduction of complexity that appears as a procedural focus on facts
and algorithms and a lack of relational understanding (Lithner, 2005). Hiebert (2003) finds massive amount of converging data showing that students know some basic elementary skills but there is not much depth and understanding. Leron and Hazzan (1997) argue that analyses of task solving behaviour should not only consider attempts to understand the task, and successes and failures in such attempts. They emphasise additional non-cognitive means of trying to cope: attempts to guess and to find familiar surface clues for action, and the need to meet the expectations of the teacher or researcher. The definitions below aims at characterising superficial reasoning that may be based on such attempts to cope, where the two main types found in the empirical studies mentioned above are defined as Memorised and Algorithmic reasoning.

**Memorised reasoning (MR)** fulfils the following conditions: i) The strategy choice is founded on recalling by memory and answer. ii) The strategy implementation consists only of writing it down. One can describe any part of the answer without having considered the preceding parts. An example is to recall every step of a proof.

An *algorithm* is a set of rules that will if followed solve a particular task type. The most common algorithms consist of procedures. **Algorithmic reasoning (AR)** fulfils the following conditions: i) The strategy choice is founded on recalling by memory, not the whole answer in detail as in MR, but an algorithm that will guarantee that a correct solution can be reached. ii) After this algorithm is given or recalled the reasoning parts that remain in the strategy implementation are trivial for the reasoner and only a careless mistake can hinder that an answer to the task is reached.

Fundamental in AR is how to identify a suitable algorithm. If this can be done, the rest is straightforward. AR based on surface property considerations is common, often dominating, and Bergqvist et al. (2003); Lithner (2000b, 2003b, 2004) have distinguished three (partly overlapping) families of common reasoning:

- **Familiar AR/MR** (*FAR*). This reasoning consists of strategy choice attempts to identify a task as being of a familiar type with a corresponding known solution algorithm or complete answer. The simplest example is a version of the Key word strategy where the word ‘more’ in a text is connected to the addition algorithm and ‘less’ to subtraction (Hegarty et al., 1995). Another example can be found in (Lithner,
2000), which describes how students make a holistic but superficial inter-
pretation of the task text and reach a clear but faulty image that it
is of a particular familiar type.

Delimiting AR (DAR). The algorithm is chosen from a set of algo-
rithms that are available to the reasoner, and the set is delimited by the
reasoner through the included algorithms’ surface property relations to
the task. For example, if the task contains a second-degree polynomial
the reasoner can choose to solve the corresponding equation even if the
task asks for the maximum of the polynomial (Bergqvist et al., 2003).
Here the reasoner do not have to see the task as a familiar one.

Guided AR (GAR). An individual’s reasoning can be guided by a
source external to the task. The two main types empirically found are:
(i) Person-guided AR, when someone (e.g. a teacher or a peer) pilots
a student’s solution. (ii) Text-guided AR, where the strategy choice is
found on identifying similar surface properties in an example, defini-
tion, theorem, rule, or some other situation in a text source connected
to the task.

Using the the definitions of AR and MR we can now also define
two different categories of CR; Local and global CR. Reasoning that
is mainly based on MR or AR but contains minor, local elements of
CR will be called local CR (LCR) while reasoning that contains large
elements of CR is called global CR (GCR). The latter may still contain
large elements of MR or AR. One difference between LCR and MR/AR
is that the latter may be possible to carry out without considering any
intrinsic mathematical properties. CR cannot be done arbitrarily, not
even locally, and it may in LCR be necessary to understand large parts
of the task in order to make the required local decisions.

3.1.3 Examples

The following examples are borrowed and originates from Lithner (2006).

Examples of CR

For the sake of comparability, several examples of reasoning types in
this section are solutions to the same task. Some examples are authen-
tic and some constructed, but in the latter cases they are similar to
authentic examples from the studies mentioned earlier. A constructed
example of CR is Anne’s work with the following task:

**Task 1.** Find the largest and smallest values of the function \( y = 7 + 3x - x^2 \) on the interval \([-1, 5]\).

Anne draws the graph on her calculator, see Figure 3.2, she wants to know what it looks like.

![Graph of the function](image)

**Figure 3.2: Anne’s drawing.**

“I recognise this function, I have seen that an \( x^2 \)-function (meaning second degree polynomial) can look like a valley if it is \( +x^2 \) and a hill if it is \( -x^2 \). I can see that the minimum is at the endpoint, at \( x = 5 \).”

She sees that the maximum seems to be at \( x = 1.5 \), but recalls that one can not determine such a value from a graph. Anne calculates several function values for \( x \) close to 1.5, but seems unable to use them. She is silent for 2 minutes.

“We have just learnt about derivatives... It says what the slope is. And, the maximum is the only place where the slope is zero. I can do that, I think... I think that the derivative of an \( x^2 \)-function is an \( x \)-function (meaning first degree polynomial).”

Anne easily calculates \( y' = 3 - 2x \), finds its zero at \( x = 1.5 \) and evaluates \( y(1.5) = 9.25 \). She looks at the graph and points at \((1.5, 9.25)\).
"Yes... this is what I said. It fits with the graph."

i) Novelty: Anne has maximised second degree functions in earlier courses by completing the square, but has forgotten the method. She has just encountered derivatives in her present mathematics course, but not yet seen the maximisation algorithm that she constructs through her key strategy choice: The maximum is at the top where the derivative is zero, and this can be calculated. She is not just following an algorithmic procedure provided by someone else.

ii) Flexibility: Anne does not only have the necessary resources but also master proper heuristics and control, and have developed the belief (Schoenfeld, 1985) that this kind of reasoning can be useful. She thus is able to take the initiative to analyse the situation and adapt to its conditions, a type of initiative that seems to be uncommon among students that focus on algorithmic approaches (Lithner, 2000b, 2003b).

iii) Plausibility: She has mathematically based arguments (see i)) supporting the plausibility of the strategy choice and the conclusion. They are not mainly based on similarities to a given algorithm or on familiarity, as in imitative reasoning.

iv) Mathematical foundation: Anne has a well-developed conceptual understanding of functions and is basing her CR on an intrinsic property, the relation between derivative, slope and maximum, which makes her able to construct a mathematically well-founded solution.

To illustrate the meaning of part iii) of the definition of CR, consider the following example. It is characterised as CR since it is founded on mathematical properties of the components and since the more often a pattern is repeated, the higher the probability that the pattern is generally true (Pólya, 1954). Ted claims:

"The formula $f(n) = n^2 + n + 41$ gives a prime number for all $n$ I have tried, from $n = 0$ to $n = 30$. Therefore $f(n)$ is probably a prime for all whole numbers $n$."

24
Ted’s assertion is wrong since $f(40)$ is not a prime. The quality of the reasoning must not be determined only by its correctness, but also by the context: even the most skilled mathematicians have proposed well-founded conjectures that proved to be false. It can be seen as a fairly good performance for a grade nine student to produce Ted’s reasoning, but very poor if it came from a graduate mathematics student.

Examples of MR

To memorise an algorithmic solution of Task 1 would mean memorising, not the algorithm but every word and symbol so that the solution could (for example) be written by starting from the bottom. This is not realistic in calculational tasks like Task 1, but is in other tasks like asking for facts (”How many $cm^3$ is a litre?”), definitions (”What is a polynomial?”), and proofs:

At an undergraduate exam the following task was given: ”State and prove the Fundamental Theorem of Calculus.” The students had been given in advance a list of 12 proofs they were supposed to learn. The two-page proof can be found in any calculus textbook. The answer handed in by Frank (and almost all of the 50% of the 150 students who got correct answers) looked the same: An identical copy of the textbook proof. In addition, most of the faulty answer attempts contained large parts of the same proof, but with some parts missing or in reversed order which resulted in broken logical-deductive chains. It seemed that the students to a large extent tried to memorise the proof without managing to understand it.

In a post-test Frank was asked to explain six equalities included in the proof. Most of them are elementary in relation to the complexity of the proof:

\[ F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \to 0} \frac{1}{h} \left( \int_{x}^{x+h} f(t) \, dt - \int_{x}^{x} f(t) \, dt \right) = \lim_{h \to 0} \frac{1}{h} \int_{x}^{x+h} f(t) \, dt = \lim_{h \to 0} \frac{1}{h} hf(c) = \lim_{c \to x} f(c) = f(x) \]

Frank was only able to explain two of them, which is a strong indication that he did not understand the long and technically difficult proof, but had managed to memorise it. Most of the students with correct exam answers that participated achieved similar results as Frank in the post-test (Lithner, 2005).
In order to be classified as MR every detail of the answer should be recalled, which is not very economic except in a few task types. Most common school tasks asks for calculational transformations, and there it is more appropriate to recall not the answer but a solution algorithm as presented in the next example.

**Examples of FAR**

As an authentic example of Familiar AR consider Bill who has done many exercises of the same type as Task 1, finding absolute maxima and minima for polynomials \( f(x) \) of degree at most 3, so Task 1 and the corresponding algorithm is familiar to him.

"It is one of those... Maximisation tasks should be done by... first do the derivative then set it equal to zero. Then look at the sign of the derivative at... [he hesitates] No... it has an interval, then you just have to check the endpoints."

Bill has memorised, without understanding, the solution procedure as an algorithm that will solve this particular task type:

1. Differentiate \( f(x) \). This in turn is an algorithm that can be memorised.
2. Solve \( f'(x) = 0 \) by the algorithms for first and second degree equations.
3. Evaluate \( f(x) \) at the endpoints and where \( f'(x) = 0 \) if that \( x \)-value belongs to the interval of definition.
4. The maximum is the largest value and the minimum the smallest.

Bill follows this algorithm, and without difficulties solves the task. Bill’s reasoning is not based on mathematical properties of functions and maxima. The argument that makes him think that the solution is correct is that he sees the task as familiar, he has solved many tasks using the algorithm above. He makes no attempts to verify the answer. The algorithm does not work if the task is slightly different, e.g. finding local maxima of a third degree polynomial, and since Bill does not understand the rational behind the algorithm he can not modify it but has to know another suitable algorithm or he is in a dead end.
Example of DAR

Often Delimiting AR is carried out only in one step, either because the reasoner only knows one algorithm within the delimited set or because the first attempt yields an acceptable conclusion.

Sally’s delimited algorithmic reasoning when working with Task 1 is authentic (Bergqvist et al., 2003). She starts in a suitable way and differentiates $y$. She finds the zero of the derivative ($x = 1.5$) and evaluates $y(1.5)$ to obtain $y = 9.25$. Sally hesitates:

“I think I should have got two values, and I don’t know what I did wrong.”

She does not know how to proceed and abandons the method without reflection on potential reasons why she only got one value (while the task asks for two).

Sally instead draws the graph on the calculator and tries to apply the built-in minimum function. The user is supposed to move the cursor along the curve to give the left bound, right bound and a rough guess of the minimum’s location. Sally wants to find a local minimum where she could mark the guess, but there is none since $y = 7 + 3x - x^2$ only has one critical point, a maximum. She abandons the method without reflection. Sally then uses the calculator’s table-function:

<table>
<thead>
<tr>
<th>X</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>3</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>3</td>
<td>-3</td>
</tr>
</tbody>
</table>

Table 3.1: Sally’s table

Sally finds $-3$ as the smallest value and and $9$ as the largest value, but is troubled since she found a larger value ($y = 9.25$) earlier. She does not understand that the table method can miss the largest value as in this case (at $x = 1.5$). She once again abandons the method without reflection.

Sally now tries to solve the task by setting the function equal to zero ($7 + 3x - x^2 = 0$). She uses the familiar algorithm to solve the second degree equation and gets two values, $x_1 \approx 4.54$ and $x_2 \approx -1.54$. She is somewhat uncertain, but since she obtained two values she believes that this is the solution.
Sally makes five strategy choices: To find the critical points, the three maximisation algorithms, and the equation.

a. All choices concern trying familiar algorithms. She has used the attempted methods in many textbook exercises.

b. The arguments behind her choices are based only on surface connections between the task and known algorithms. 'Maximise' is related to the first two methods. Functions are related to tables. The second degree polynomial is (to her) related to the algorithm for solving the corresponding equation.

c. There is no analysis, evaluation or other consideration whether the algorithms are suitable and generate any useful knowledge or not.

d. When she is doubtful she quickly searches for another algorithm.

Her strategy is to apply algorithms, but she knows several hundred and delimits the set of algorithms to try by choosing only algorithms that have some connection to the information in the task. Since she is unable (or reluctant) to found these connections on intrinsic properties the algorithms are chosen on surface considerations only. The task and the corresponding algorithm is not familiar and well known as it is to Bill above. There is nothing wrong in itself with testing different algorithms, but Sally’s problem is that her weak conceptual and procedural understandings are too incomplete to help her make proper choices. They become random choices from a delimited set of known algorithms.

**Examples of GAR**

**Person-guided AR** In the following real example (Lithner, 2002), the 9th grade student Moa does not know how to calculate 15 % of 90 and asks the teacher for help. The teacher starts to help Moa by writing in her book, without saying anything to start with (the complete dialogue is included in the transcript):

\[
90 \cdot 0.15 \quad (*)
\]
Teacher: “What is 5 \cdot 0?” Moa: “0.” [The teacher writes 0 under (*)]
Teacher: “What is 5 \cdot 9?” Moa: “45.” [The teacher writes 45 under (*)]
Teacher: “What is 1 \cdot 0?” Moa: “0.” [The teacher writes 0 under (*)]
Teacher: “What is 1 \cdot 9?” Moa: “9.” [The teacher writes 9 under (*)]
Then the addition that the teacher wrote under (*) shall be carried out:
\[
\begin{array}{c}
450 \\
+ 90 \\
\end{array}
\]
\[
\begin{array}{c}
\text{(**)} \\
13.50 \\
\end{array}
\]
Teacher: “What is 5+0?” Moa: “5.” [The teacher writes 5 under (**)]
Teacher: “What is 4+9?” Moa: “13.” [The teacher writes 13 under (**)]
Teacher: Where shall the decimal point be placed?” Moa: [Silence]

The teacher marks the decimal point at the correct place and leaves.

Moa does not participate in any other activities than adding and multiplying one-digit numbers (which is primary school mathematics and irrelevant for Moa’s difficulties). No intrinsic properties of the mathematics that is problematic to her, percentage and multiplication of decimal numbers, are considered. The teacher does not try to find out what Moa’s difficulties really are, does not discuss the fundamental principles behind the algorithm, does not help Moa to by herself reflect over these principles, does not help Moa to consider the strategy choices, etc. It seems that Moa has to try to memorise the algorithm and the connection to this task type.

Text-guided AR In this realistic example (see Lithner, 2000b, 2003b; Lithner and Långström, 2006) for similar data), Task 1 is one of the exercises in John’s textbook.

"I always start by looking at the examples. It should be one that is the same... similar as the exercise... Yes, it could be this one... it is a second degree function, and it is a max task.”

"EXAMPLE 6.
Find the maxima and minima of \( f(x) = x^2 + 4x + 1 \) if \( x \in [-3, 0] \).
SOLUTION: \( f'(x) = 2x + 4 \). Solve the equation \( f'(x) = 0 \).
\[ 2x + 4 = 0 \Rightarrow x = -2. \] Since \( f'(x) \) exists for all \( x \) and
\[ x = -2 \] is the only zero of \( f'(x) \) we just have to evaluate
the function there and at the endpoints of the interval.
\[ f(-3) = -2, \quad f(-2) = -3, \quad \text{and} \quad f(0) = 1. \]
Thus we have
\[ \text{Maximum } f(0) = 1 \quad \text{and minimum } f(-2) = -3. \]

John says: "Yes, I will try this, to see what I get."

John copies the procedure given in Example 6 in every detail and
reaches a solution: Maximum \( f(1.5) = 9.25 \) and minimum \( f(5) = -3. \)
John compares his answer to the one in the textbook’s solution section.
They are the same and he turns to the next exercise.

To carry out this solution it is necessary to: i) Know that 'maxima
and minima' is equivalent to 'largest and smallest value' (but not to
know what these expressions mean mathematically) in order to identify
the task and the example as similar. ii) Know that the two representa-
tions of intervals are equivalent. iii) Know the differentiation algorithm
for a second-degree polynomial. iv) Evaluate a second-degree polyno-
mial. v) Solve a linear equation. The rest can be copied.
The key feature is that none of the solution procedure steps consider
the intrinsic properties, the underlying meaning, of the function, the
derivative, and extreme values, and the reasons behind choosing these
particular steps do not have to be known. In fact, no mathematical
properties central to 'applications of the derivative' (the topic of the
textbook section) are used, or need to be understood. Only knowledge
of elementary facts and basic algebra are required.

### 3.2 Mathematical competences

There are different ways of comparing what kind of mathematics that
are assessed by tests. Comparison can be done by either subject mat-
ters or by requirements in terms of cognitive behaviours e.g. by required
competences, both included and implicitly described in the policy docu-
ments (Swedish National Agency for Education, 2001a,b). In this thesis
the tested mathematics are compared by either reasoning requirements
(Chapter 3.1) as in Paper I-III or by competences as in Paper IV.
The following competences description origin from the development
and construction of the Swedish NCTs. There are many different
ways to organise mathematical knowledge as a ground for describing
goals e.g. by taxonomies for cognitive categorisation, where the work of Bloom (1956) and Biggs and Collis (1982) perhaps are the most well known or as in the work of Kruteskii (1976) another significant attempt, with a different approach to provide a theory of mathematical abilities. For more recent attempts to systematise mathematical knowledge by different processes or competences, see e.g. NCTM (2000); Niss and Jensen (2002); Gjone (1993) or Kilpatrick et al. (2001), yet another way is done by Ernest (2004) who distinguishes between relevance and utility in his description of teaching and learning goals. The following description of mathematical competences is a compromise between different needs and considerations, where it was prioritised that the competences should be fairly easy to communicate and to understand by task constructors, test developers and by teachers. Their main purpose, however, are to facilitate test development and secure alignment between the NCTs and the Swedish policy documents (Palm et al., 2004, p. 2). All competences have their counterpart in described intentions within the policy documents (Palm et al., 2004, p. 3-8). For a similar approach, but with the NCTM Standards in mind, see e.g. Thompson (1997). Below the competences used in this thesis (Paper IV) are described and exemplified, for further examples and more elaborate descriptions see Palm et al. (2004). The competences are to some degree overlapping and interconnected they are also slightly different in character, but each competence has its own core identity.

3.2.1 Problem solving competence

This competence is required when a student meets a problem, i.e. a task in which she or he doesn’t know how to proceed and no complete known solution procedure can be used. The student has to construct something new (something non-routine), apply their knowledge in a new situation. Whether a task requires this competence or not depends both on the solver and the task, what might be a problem for one student might not be so for another. *Types of tasks that might require this competence:* 1. Unusual tasks, 2. Tasks where given information is different than what the students are used to, and 3. Complex tasks.

Example 1: (NCT Course B 1998 autumn, task 10) At ice-hockey matches at Globen in Stockholm, anyone who wants
to can buy a match programme for 25 SEK. At the end of the game prizes are raffled and the match programmes are the raffle tickets. At a match between Djurgården and Brynäs, three cruises to Helsinki were raffled. Calculate the probability that you win one of these cruises if you buy a match programme. You have to make up the information you need to be able to carry out your calculations.

Comment: In this task students don’t have available the information normally supplied in similar tasks. They will have to make appropriate assumptions, which mean that they cannot proceed to place given information in preexisting formulas (as in the normal case) before these assumptions have been made.

3.2.2 Algorithm competence

To be familiar with and be able to use relevant algorithms within the specific course comprises the core essence in this competence. This means that the student should be familiar with – and in a task-solving situation be able to use – procedures in one or several steps where all involved steps and the overall sequence are well known to the student. Every step in the procedure can often be described as a sequence of more elementary steps. Types of tasks that might require this competence: Routine tasks in which the student can use well known theorems associated with the task type or other typical tasks, e.g. solving equations, simplifying expressions, seeking maxima’s and minima’s of familiar functions etc.

Example 1: (NCT Course E 2002 spring, task 10) Determine the general solution to the differential equation \(3y'' + 6y' - 24y = 0\)

Comment: In order to solve this task the student can connect it to and use a well known algorithm.

Example 2: (NCT Course C 2000 spring, task 6) Determine by the aid of the derivative possible maximum, minimum or terrace points to the curve \(y = 2x^3 - 3x^2\)

Comment: In this task the student can execute a well exercised procedure.
Example 3: (NCT Course D 2000 spring, task 5) The triangle $ABC$ is given. Calculate the area of the triangle.

![Triangle ABC](image)

**Figure 3.3: Triangle $ABC$.**

*Comment:* The student can relate the task to and use well exercised standard theorems. Even if the student doesn’t know in the beginning which theorem to use, the choice is of routine character since the overall procedure is familiar.

### 3.2.3 Concept competence

This means proficiency with a concept’s or notion’s definition, including the ability to define and use the *meaning* of the concept. In order to get a clear picture of a students’ concept competence it’s generally necessary to use several tasks with different angles of approach. A single student solution can however, to various degrees, indicate the students’ concept competence. *Types of tasks that might require this competence:* 1. Tasks in which the student are asked to explain, or interpret a central notion or a relation or parts of it, 2. Tasks in which the student is given information that should explicitly be used to reach assertions or conclusions, 3. Unusual tasks, 4. Open tasks, 5. Tasks that requires understanding for different representations of the same mathematical concept.

Example 1: (NCT Course A 1995 autumn, task 7) You’re about to build an aquarium in glass in roughly 160 litres. Propose suitable measures.

*Comment:* In this task the posed question is reversed in contrast to a traditional task where the sides usually are given and the volume
is asked for. This means that a student have to consider the intrinsic properties of the volume concept (and formula relations) and the solution can therefore give an indication of a students’ understanding of the concept of volume.

Example 2: (NCT Course E 1998 spring, task 8) A solution to an equation usually is a number satisfying some condition. e.g. the number three is a solution to the equation $x^3 - 27 = 0$. What is meant by a solution to a differential equation?

Comment: A correct solution of this task, considering the intrinsic relationship between a differential equation and its solution, may give an indication of the student’s knowledge of the concept of differential equations.

Example 3: (NCT Course A 1995 spring, task 8) (a) Give an example of a number somewhere between $5 \cdot 10^{-3}$ and $5 \cdot 10^{-2}$. (b) Give an example of a number, in fraction, bigger than $3/4$ but smaller than 1.

Comment: These two sub-task are somewhat unusual in the way the questions are formulated and a student really has to consider fundamental properties of different representations of numbers, thus a correct solution may give an indication of the student's number sense.

3.2.4 Modelling competence

This competence comprises, to create (or to formulate) and use a mathematical model from a non–mathematical situation, interpret the results of it and to evaluate the mathematical model by identifying its limitations and prerequisites (Blomhøj, 2004; Palm et al., 2004). It should be noted that the modeling competence in this text does not imply that a task have to be a problematic situation for the solver, which often is the case in similar frameworks (cf. Blomhøj (2004) where this is the case). Types of tasks that might require this competence: 1. Tasks that test the whole modelling process, 2. Tasks that test some parts of the modelling process.
Example 1: (NCT Course C 2002 spring, task 13) Year 1960 the number of grey seals where estimated to 20 000 in the East Sea. Due to high levels of toxic substances the number started strongly to decline. The decline was exponential and year 1980 there were only 2000 grey seals left. (a) Which was the average yearly decline in percents between the years 1960 and 1980? (b) Exponential declining models, as in this case, have a limitation in the long run. Which is this limitation?

Comment: In this task the student have to go through the whole modelling process, and evaluate its limitations (sub-task b). Sub-task a is of familiar type and could be solved routinely.

Example 2: (NCT Course C 1996 autumn, task 4) In 1996 a family bought a cottage by a norrland’s river. The land on which the cottage is placed they had to rent. In the lease it’s stated that the yearly rent year 1991 was 1420 SEK and that this rent should follow index for January. Which is the yearly rent in 1996?

Table 3.2: Index, example 2.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Index (January)</td>
<td>218.9</td>
<td>230.2</td>
<td>241.0</td>
<td>245.1</td>
<td>251.3</td>
<td>255.6</td>
</tr>
</tbody>
</table>

Comment: The task tests parts of the modelling procedure.

3.2.5 Reasoning competence

With reasoning we here have in mind an argumentation done on general logical and specific theoretical grounds, including deductive reasoning where assertions are done based on specific assumptions and rules and where the strictest version constitutes a mathematical proof. The competence also includes inductive reasoning, where general statements can be reached based on observations of patterns and regularities. This means that the competence could involve an element of investigative
activity in finding patterns, formulating, improving and examine different hypotheses. It also involves critical examinations of proofs and other mathematical statements. Important is however, that the reasoning may be carried out as a routine activity, with mainly known arguments and memorised proofs as well as a problem solving activity where no known arguments is accessible for the student. *Types of tasks that might require this competence:* 1. Tasks where the students are asked to formulate and investigate hypotheses, analyse and reach conclusions, 2. Tasks where the student should strengthen or prove something, 3. Evaluative tasks, 4. Tasks in which the student has to generalise, 5. Tasks where the student has to connect old findings to new, 6. Tasks where the student should explain something.

Example 1: (NCT Course A 2005 spring, Task 10) Martin and Johanna are going to buy a new car. Johanna likes a car which costs 194 000 kr. Martin says that the value of this type of car falls by about 17% per year. They wonder how much the car would be worth in 3 years and each of them calculates in their own way. In figure 3.4, you can see their respective solution attempts. Which of them has interpreted the problem correctly? How might Martin and Johanna have reasoned in their calculations?

![Figure 3.4: Johanna’s and Martins’ solutions.](image)

**Comment:** In solving this task the student has to interpret both solutions and argue for her or his choice.
Example 2: (NCT Course B 2005 spring, Task 16) Figure 3.5 shows the letter M, placed on a horizontal surface. The two vertical “supporting legs” are equally long. Show that \( v = 2x \).

![Figure 3.5: The letter M.](image)

Comment: In this task the student has to prove a statement.

### 3.2.6 Communication competence

The communication competence treats the ability to communicate mathematical ideas and thoughts in written as well as in oral form. This means to be able to receive and understand information with a mathematical content and to produce and convey such information. It also includes to understand mathematical terminology and notions and to be able to use them accordingly in communication with others. *Types of tasks that might require this competence:* 1. Tasks that explicitly asks for descriptions or explanations of concepts, notions, rules or methods, 2. Tasks that stipulate special requirements regarding mathematical language.

Example 1: (NCT Course C 1998 spring, task 5) Explain, by giving an example, the notion *decline* in a statistical survey.

Comment: A correct solution of this task gives an indication of the student’s knowledge of the concept of missing values. This task could however be seen as requiring both the concepts and communication competences.

Example 2: (NCT Course C 1997 spring, task 11) A friend of yours, reading the same math-course, approaches you and
asks "I don’t get this derivative-stuff". Help your friend by explaining the concept of derivative. Explain in as great detail as possible, also explain in as many different ways you can. *Your assignment is not to deduce or describe the rules of derivatives.*

*Comment:* In this task the student has to explain a concept and is thereby given the opportunity to show her/his ability to communicate mathematics. Tasks of this type coincides with the concept competence.

Example 3: (NCT Course C 2002 spring, task 6)
(a) Explain, by using a graph, why the derivative of a constant function is zero.
(b) Explain, by the definition of derivatives, why the derivative of a constant function is zero.

*Comment:* In these tasks a student have to perform explicit arguments for given statements. Tasks of this type also coincides with the concept competence.
Chapter 4
Methodological considerations

In this section I will describe and make some small reflections upon the use of the different research approaches used within this thesis. The methods used in the different papers differ from each other, which is a direct consequence of the nature of the posed research questions in the papers. I will describe the choice of both the reasoning framework and the development of the task variables in the analysing procedure used in the first two papers. In contrast to these approaches semi-structured interviews were used in the last two papers, and some comments on the approaches of those papers will be made as well.

The quotation below illustrates the complexity of many educational phenomena.

“The astonishing complexity of mathematical learning: An individual student’s mathematical learning often takes place in immensely complex ways, along numerous strongly winding and frequently interrupted paths, across many different sorts of terrain. Many elements, albeit not necessarily their composition, are shared by large classes of students, whereas others are peculiar to the individual.” (Niss, 1999, p. 13)

In order to study any educational phenomenon, a reduction of complexity is needed. Reduction in this thesis is done within many different layers. The reduction is a composite of choices of theory, phrasing research questions, choosing and sampling data-material, choosing method, how
to structure data for analysis, choosing how to write the introduction and so on. It all boils down to what is researchable. All the way deliberate compromises between fidelity and manageability have to be made. Research is to great extent about finding harmony between the different research components.

The uses of the research methods are by no means isolated from the considerations taken in various steps throughout the whole research process. I therefore start by a short justification of the theoretical framework used in paper I-III. The framework presented in chapter 3.1 is chosen for this thesis since it

- is firmly anchored in empirical data in the sense that it captures key characteristics of reasoning which in empirical studies have been found to be used by students.

- does not consist of vague abstract definitions and sets of examples, but is built on well-defined concepts and can be used as a concrete tool for determining characteristics of empirical data.

- is uniform in the sense that characterisations of different reasoning types are comparable within the framework. This enables connections between different learning and achievement situations to be identified, in order to explain origins of reasoning types and their relations to learning difficulties.

- can be used to communicate key phenomena between researchers, students, teachers, textbook authors, curricula developers etc., and provide a basis for analysing, understanding and constructing learning environment components like tests, textbooks and teaching.

Two different applications of this framework are used in this thesis. In paper I the framework was used to categorise the kind of reasoning that was required to solve included tasks in both teacher-made and national course tests. In paper II it was used to both establish reasoning requirements on a test and to determine the reasoning used by the students who took the tests. In paper III it was used to communicate key phenomena with the interviewed teachers.

Returning to paper I, it should be recognised that the reasoning framework is an aid in describing and understanding different forms of
reasoning. Not necessarily a way of establishing the kind of reasoning that is required in order to solve tasks, which was the purpose in the study. In order to determine whether or not a task can be solved by imitative reasoning one must take into account what may reasonably be known to the particular task solver at hand. In the study reported in paper I, this was particularly intricate since no real students were at hand, just imaginary average students. Besides the reduction of complexity that follows by the choice of theoretical framework above, the choice to choose only the textbooks as representatives for students' learning experiences was the first reduction of complexity. In this study each test was paired with the textbook used by the same teacher who sent in the test. This mean that each test had its own learning history reference. It was also chosen that three similar tasks, examples or theory sections in the textbook could be enough for a student to use some form of imitative reasoning to solve a test task. This is also a reduction of complexity. All these reductions, which in some sense reduce fidelity, were conducted to increase manageability.

A great part of the scientific work in this study dealt with the development of the analysis procedure for deciding whether or not a specific task reasonably could be solved by some form of imitative reasoning or not, see page 68. This was seen as a significant part of the approach. In order to illustrate this significance I’d like to tell a short story:

The phone rings. A man answers, and from the telephone receiver it’s heard: *Is this 031-16 38 46?* Whereby the man replies: *Yes, yes, yes, yes, yes, no, no, yes!*

What I’m trying to illustrate is that it’s very easy to, at hand, say that this or that task should be familiar to a student, by our old experiences as teachers. What we, the authors of the article, did by developing and using the analysis procedure is that the decisions about familiarity were done after having compared each and every task variable. This procedure was extremely time consuming, we spent more than a year analysing these tests and their respective used textbooks. This can be seen as a drawback and implies that if this part of the approach is to be used by e.g. teachers in the field, there is need for a less time consuming approach to be developed. However, scientific rigor is high. Further considerations can be found in the paper, see page 55 and forward. The method used in this paper can be summarised as a classifications
study, with qualitative analyses of the key-characteristics of both the
task variables and the type of reasoning that theoretically were required
to solve the test tasks.

In Paper II the same framework was used, but this time students’
reasoning were analysed as well. This was a think-aloud study (cf. Lith-
ner (2003b)), where eight students taking the national course test were
video taped and instructed to speak out all their thinking. The session
was ended by an interview, which aimed at both giving feedback to
the student as well as following up parts hard to interpret (stimulated
recall). In order to simulate a high-stake situation, i.e. not reducing
complexity all to much, but avoiding an ethical dilemma; a student’s
bad result on a national test due to nervousness in this extreme sit-
uation, students were given the opportunity to re-do a similar test at
a later time. The analyses of the students’ reasoning were facilitated
by structuring data, both the students written solutions and the video-
material, according to the reasoning structure described in Chapter 3.1.

The approaches in Papers III and IV are different from the previous
two. Paper III is mainly an interview study whereas Paper IV consists of
both a classification of tests, however with the aid of another framework,
and interviews with teachers. Important however in both these studies
are the explicit link to the respective teachers own practice, i.e. their
own tests.

In Paper III teachers were interviewed about: 1) their awareness
about creative reasoning, 2) their general beliefs about creative rea-
soning and finally, 3) their specific beliefs and concerns about creative
reasoning in their own written tests. In order to approach the first
theme, a set of tasks, some possible to solve with imitative reasoning
and some requiring creative mathematically founded reasoning, were
sent to the teacher with the instruction to identify the most significant
differences between them. The idea was that if they did not see different
kind of reasoning requirements as an important feature, then this might
to some extent have explained why the did not include other require-
ments than imitative reasoning in their tests. In the last part of the
interviews the teachers own tests, the ones they submitted to us in the
first study (Paper I) was shown (stimulated recall). The purposes were
mainly to confirm that we spoke about the same phenomena and to
avoid a situation where the teachers confessed something that weren’t
possible to validate.
In the last paper, no. IV, a different framework for categorising the tests was used, a competence framework (Palm et al., 2004). The main purpose was to capture other parts of the goals described in the policy documents, which the former framework did not cover. This second framework, was chosen since it

- captures most of the main features described in the Swedish policy documents, and thereby makes explicit the relation between the goals and the competences,

- is similar to other frameworks that encompass goals of mathematics education (cf. NCTM (2000); Niss and Jensen (2002)), which enables comparison with other studies.

This study approached impact from the national course tests, specifically impact on teachers’ test development. Within the frame of my Ph. D. studies there was neither room for nor financial support for large scale before and after scenarios. The idea used in this study was that by comparing the national tests with the teacher-made tests (a sample which I already had) by the kind of competences tested, 'non'-impact could at least be determined. That is, in the case with very different test profiles it should be possible to conclude that, regardless of how teachers may have been influenced by the national tests, this potential influence hadn’t resulted in any change of their own test development. This idea was followed and in addition two groups of teachers with tests either very similar to the national tests or very different from them were interviewed about impact and the reasons for similarity. This means that this study, in some sense, excluded a potentially large group in the middle, who might have been influenced. However, this is the result of all sampling, and an unavoidable part in reducing complexity.

Another aspect of the methodology in this study concerns the categorisation. The competence framework is neither developed for, nor are the nature of tasks such that only one competence always can be seen as applicable. It is more often so that a specific task can be seen as requiring more than one competence (cf. Romberg et al. (1990)). In this study, however, only one competence was categorised for each task in order to reduce complexity and facilitate the comparison. Due to this choice an order of priority had to be constructed (cf. page 167). It is hard, if not impossible, to establish a clear and precise one to one
relation between a task and what kind of competence is required to solve it. The categorisation done in paper IV represents one of several possible interpretations, the priority order could have been constructed differently. At task-level these categorisations may therefore sometimes be debatable. But at an aggregated test-level, since consequently handled, they can reveal important information on differences in testing cultures and on what kinds of competences are tested.

In both paper III and IV where the interviews generated large parts of the data, I have included relatively many quotations in the respective papers. The intention was to make it possible for the reader to a) make her or his own interpretations of the data and b) thereby see if she or he agrees upon my interpretations of the same data. The main problem, however, still remains, the reader only have access to a restricted set of the data, whereas I had all data available.

**Summary:** In this thesis two main approaches have been used and complemented each other; categorisation by qualitative analyses of tasks and students’ reasoning and interviews with teachers. On the one hand I have used different forms of theoretical structuring, where either students’ used reasoning or the requirements of tasks have been categorised and characterised. On the other hand when seeking explanations to results found by the categorisation the interviews have added information which could not have been attained otherwise.

The methodology is only as good as its usefulness in generating appropriate data and guiding useful analyses relative to the research questions (Simon, 2004, p. 159)

The use of these methods together can, since they generated appropriate data that were possible to analyse according to the posed questions, be seen as appropriate.
Chapter 5

Conclusions and discussion

In this chapter I will shortly retell the main conclusions from the studies reported in paper I-IV, synthesise some of the results and finally discuss them, both in relation to earlier research as well as what their practical implications may, or perhaps should be.

5.1 Conclusions

- Teacher-made tests mainly focus on tasks that can be solved without having to consider intrinsic mathematical properties.

- National tests focus to much higher degree on tasks which cannot be solved without having to consider intrinsic mathematical properties.

- In the case where a test requires creative mathematically founded reasoning, as in the national tests, students also use non-imitative solution strategies to a much higher extent.

- A very content-dense curriculum may hamper the development of creative mathematically founded reasoning and conceptual understanding.

- The focus on tasks possible to solve by imitative reasoning in teacher-made tests seems explainable by teachers limited awareness about creative mathematically founded reasoning and their wish to get as many students as possible to reach the grade level
Passed as possible, which was believed difficult to do when requiring high quality reasoning from weaker students.

- Impact of the national course tests on teachers’ development of tests, in terms of tested competences, seem to be fairly modest.

- The impact of the national course tests has however been proved to exist and partly shown how to operate, on some teachers test development. I.e. a causal relationship has been shown (cf. request from Mehrens (2002)).

- The frameworks and analysis procedures used throughout this thesis have shown to be useful as research tools and can be used to further evaluate and develop tests.

5.2 Discussion

What can these results say together? How do they relate to the two tenets presented in the introduction and to the overall aims of the thesis? In this section I also turn to the issue of ‘So what?’. What may be the implications of the findings of this thesis? How can the results be used?

5.2.1 Relating to tenet one

Tenet one can be seen as an explanatory idea for many students’ difficulties in their studies of mathematics. In short, a large part of their difficulties may be explained by a too narrow focus on solutions strategies that in a mathematical perspective can be seen as superficial. A focus that do not consider any of the intrinsic mathematical properties that explains, justifies and makes a solution, or the use of a strategy understandable.

To begin with, the studies together have extended the knowledge about a specific part of the learning environment, the written tests that students encounter in their mathematical studies. The first study showed that a majority of the tasks in teacher-made assessment could be solved successfully by using only imitative reasoning. The national tests on the other hand required creative mathematically founded reasoning
to a much higher extent. However, most students encounter about two to five teacher-made classroom tests during a course, compared to only one national test. This means that most test items they meet only require imitative reasoning. This implies a) students are probably not well prepared for those tasks on the national tests that are not possible to solve with imitative reasoning. b) It may explain why so many students have difficulties with the national tests. As a consequence this part of the learning environment (i.e. teacher-made tests) does not seem to promote the development of creative mathematically founded reasoning, and it may be one further influential piece in explaining many students’ learning problems.

This focus, in teachers test tradition, on tasks requiring mostly algorithmic skills and known procedures was also confirmed in the fourth study, where it was shown that the majority of the teacher-made tests included very few tasks testing other competences, such as communication, modeling, concept, problem solving and so on. Furthermore it was found, when cross-checking results from study I and IV, that many of the tasks in the teacher-made tests categorised as requiring e.g. the modeling competence, were of imitative nature, which even further confirms the algorithmic and procedural focus. The same phenomenon was not found in the national tests. The teacher-made tests seem neither to signal that non-imitative reasoning nor other competences than algorithmic ones are important to learn.

These results are in concordance with earlier research that concerned other parts of the learning environment, e.g. textbooks, classroom instruction and various laboratory settings (Lithner, 2000b, 2003b; Lithner and Långström, 2006), where this focus have been shown to constitute major parts. These results are also in line with results found by e.g. Black and Wiliam (1998) and Cooney et al. (2001) concerning teachers’ test development and the focus on facts and procedures (cf. the introduction).

One aspect of the findings of this thesis, perhaps quite remarkable, is that some of this focus on tasks solvable by imitative reasoning can be explained by teachers’ conscious and deliberate choices to exclude tasks that requires creative mathematically founded reasoning. This type of tasks was excluded since they were believed to be too difficult for a large group of students to cope with. Whether or not this deliberate choice may explain this kind of focus in other parts of the learning
environment is worth continued research.

Results within this thesis – together with earlier research at my department as well in relation to what many other researchers say (cf. introductions papers I and II) – all point at the same direction; essentially all parts of the learning environment seem to have a too narrow focus, a too far-reaching reduction of the complexity of mathematics towards a diet consisting solely of facts and algorithms. One important and fundamental conclusion from contemporary research is that students can learn what they are given opportunities to learn. This is reflected in the well-researched idea opportunity to learn (Hiebert, 2003). The narrow focus on tasks that are possible to solve without the need to consider intrinsic mathematical properties found in teacher-made tests means that students aren’t given real chances of showing their problem solving abilities as well as many of the other competences described in the Swedish policy documents, this may have severe consequences for how students are graded. The tests also signal to the students what they are to focus on in their studies (Crooks, 1988), and one thing that one probably can be quite sure about is that students do not over-emphasise things that are not included in tests (cf. Wilson (1993)). This indicates that – if teachers’ tests are reflecting their instruction as were claimed by the interviewed teachers in study III – students aren’t given real chances of developing many of the competences that are formulated as goals in current curriculum.

All together and in the light of tenet one, results from this thesis seem to further strengthen this explanatory model. This by both widening its scope (assessment tradition) and adding some new elements such as some teachers’ deliberate choice to exclude requirements beyond recalling facts and using familiar strategies. By the study in paper II another very significant element has been added, the authentic moderately high-stake national assessment situation. This is perhaps one of the most valuable results within this tenet since many of the earlier studies, which have built up this explanatory model, have been about different laboratory situations.

5.2.2 Relating to tenet two

The national course tests seem to differ greatly from many other parts of the learning environment. This has both been shown in Paper I,
II and IV, and been indirectly confirmed by teachers in the interviews in Papers III and IV. The national course tests assess a wider range of mathematical competences, and can be seen as a counterweight to many other parts of the learning environment.

This leads to the second tenet, i.e. the idea about assessments as agents of change. In what ways can the national tests’ focus on other types of goals be seen as reflected in teachers’ development of tests? This is already indirectly answered for in the previous section, the national course tests seem to have had a very limited effect on the development of teacher-made tests. However, the studies III and IV together have contributed to the knowledge about how impact from the national tests may operate and if results from paper III are used heuristically it may also contribute to the explanations of why so many teachers aren’t influenced, and help explaining why impact from the national tests are so restricted.

If the aggregated results from this thesis are used to evaluate, or at least start a discussion about the national test system according the ‘fourth aim’ (see page 8), concerning the national tests as exemplary models, it must be concluded that it cannot be seen as fully reached. Although the national tests still can be regarded as exemplary models, they do not seem to be convincing agents of change (cf. Cimbricz (2002)). As I see it there is a difference between influence and impact. One might be influenced, but that doesn’t necessarily mean it results in any effect, due to a wide range of potential restricting factors (cf. e.g. Paper III). As presented in the introduction, the national course tests in Sweden seem to have influenced many teachers. This is fairly well documented, once again see introduction. It however seems like this influence mostly concerns what is said and is not reflected in practice, at least not in one part of the practice, the development of tests. These results only concern the tested mathematics. Impact in terms of how teachers assess their students’ knowledge, how they score and grade etc. are not covered in this thesis. In a few of the interviews it was indicated that teachers had changed their ways of assessing students, and this is another part of teachers’ practice that deserves further research.

Two important aspects of the findings in this thesis relates to the second tenet and sums up the section. Firstly, impact in terms of tested competences was shown to be fairly limited. Secondly, impact may be seen as even more desirable in the light of these findings. As it appeared
the national tests were, at least in terms of the tested competences, well aligned with the policy documents, where on the other side most of the teacher-made tests only tested narrow parts of the curriculum. This means that the national tests still can be seen as important in the aspect that they constitute good models of how students abilities, in relation to current policy documents, can be assessed. Since the national tests in this regard, when compared to other parts of the learning environment, can be seen as quite unique, it’s important to continue researching not only how they may influence various parts of teachers’ practice, but rather why they don’t.

5.2.3 Relating to the development of methods

One final conclusion concerns the used methods rather than explicit results of the studies. The frameworks and methods used throughout this thesis have shown to be useful as research tools and can be used as instruments to further evaluate and develop tests. The analysis procedure developed, described and used in Paper II, was very time consuming (discussed in Chapter 4). This should perhaps be put in a perspective. If this procedure is to be used by teachers, it might not be that time consuming after all. Firstly teachers would only have to categorise one test at a time, not nearly 70 as we did, they are reasonably also much more familiar with the textbooks than we were. This implies that this procedure also could be a significant tool for deciding whether or not a task may be solved by imitative reasoning or not, and could thus also be a useful tool when developing and constructing tests.

5.3 Implications

The big questions however still remains. How to change focus towards more non-imitative solution strategies – a focus which instead considers the intrinsic properties of mathematics? Could the assessment system be changed in order to increase impact and to facilitate this shift?

In Paper IV I shortly discuss one aspect that relates to the latter question. What teachers see are only the ready-made tests, the final products. Nothing about how to reach that goal. One of the reasons for the national test developers’ success in implementing these kinds of
tasks in the national tests may partly be due to their use of frameworks of the kind presented in this thesis, this aside from the fact that they, at least in comparison with teachers, have substantial resources both in terms of available time and personnel.

Study IV showed a few examples of the possibility to construct tests that assess most of the competences without explicit knowledge about the competence framework – perhaps much of what frameworks like this and alike captures are part of teachers’ tacit knowledge? However, it should probably be much easier to construct better aligned tests with the aid of conceptual guides like the competence framework or the reasoning framework.

The results of this thesis should not be seen as a critique towards neither the national assessment system nor practising teachers. I would say that neither is the assessment system properly designed to meet the fourth and indirect aim of being an agent of change, nor are teachers doing an unqualified work. The old question about where the ‘trade-off’ between algorithmic skills and conceptual understanding (in this thesis represented by creative mathematically founded reasoning) should be placed still seems to be at play in many classrooms. The introduction of the new types of curricula-goals seems to only have been added on-top of earlier demands. Which together with many hard-to-interpret cognitive concepts that don’t naturally have the same meaning to different persons, probably have made it very hard for teachers to cover the whole curriculum. Results from Paper II strengthens that a too content-dense curriculum may be the case. This was shown by a higher proportion of both requirements and successful solutions in the A-course when compared to the courses B-D, where the latter courses comprised many more new concepts in content terms. What seem to be excluded from instruction in those cases are just the ‘novel’ goals, those who are described in qualitative terms rather than just as subject content. Jeremy Kilpatrick said nearly two decades ago:

Why is it that so many intelligent, well-trained, well-intentioned teachers put such a premium on developing students’ skill in the routings of arithmetic and algebra despite decades of advice to the contrary from so-called experts? What is it the teachers know that the others do not? (Kilpatrick, 1988, p. 274).
Irrespective of what it is teachers know, believe or decide upon, they need support if they are to carry out the very difficult task of developing tests that at the same time can assess students’ different mathematical competences, and serve as a ground for the development of more sustainable mathematical abilities. Hopefully the findings of this thesis may constitute a ground for designing such support.
Bibliography


Kuhs, T. and Ball, D. (1986). *Approaches to teaching mathematics: Mapping the domains of knowledge, skills, and dispositions*. Michigan State University, Center on Teacher Education, East Lansing, MI.


