The Mathematics Enthusiast

Volume 9 | Number 3 Article 8

7-2012

University Mathematics Teachers' Views on the Required Reasoning in Calculus Exams

Ewa Bergqvist

University Mathematics Teachers' Views on the Required Reasoning in Calculus Exams

Ewa Bergqvist[1]

Abstract: Students often use imitative reasoning, i.e. copy algorithms or recall facts, when solving mathematical tasks. Research shows that this type of imitative reasoning might weaken the students' understanding of the underlying mathematical concepts. In a previous study, the author classified tasks from 16 final exams from introductory calculus courses at Swedish universities. The results showed that it was possible to pass 15 of the exams, and solve most of the tasks, using imitative reasoning. This study examines the teachers' views on the reasoning that students are expected to perform during their own and others mathematics exams. The results indicate that the exams demand mostly imitative reasoning since the teachers think that the exams otherwise would be too difficult and lead to too low passing rates.

Keywords: reasoning; creative vs. imitative; Calculus; University Calculus courses; Swedish exams

Introduction

The purpose of this study is to better understand university teachers' rationale when they create calculus exams, especially concerning the reasoning that the students are expected to perform in order to pass the exams. Earlier research indicates that students often use imitative reasoning when they solve mathematical tasks (Schoenfeld, 1991; Tall, 1996; Palm, 2002; Lithner, 2003). Imitative reasoning is a type of reasoning that is founded on copying task solutions, for example by looking at a textbook example or by remembering an algorithm or an answer. The students seem to choose[1] imitative reasoning even when the tasks require creative reasoning, i.e. during problem solving when imitative reasoning is not a successful method (the concepts of “imitative” and “creative” reasoning are thoroughly defined in Section *The Reasoning Framework*). The use of algorithms

[1] Department of Mathematics and Mathematical Statistics
Umeå University
SE–90187 Umeå
Sweden
Phone: +46 90 786 92 24 *or* +46 73 073 54 54
Fax: +46 90 786 52 22
E-mail: ewa.bergqvist@math.umu.se

saves time for the reasoner and minimizes the risk for miscalculations, since the strategy implementation only consists of carrying out trivial calculations. Thus using algorithms is in it self not a sign of lack of understanding, but several researchers have shown how students that work with algorithms seem to focus solely on remembering the steps, and some argue that this focus weakens the students' understanding of the underlying mathematics (e.g. Leinwand, 1994; McNeal, 1995) and that it might eventually limit their resources when it comes to other parts of mathematics, e.g. problem solving (Lithner, 2004). This is an important observation because it might be one of the reasons for students' general difficulties when learning mathematics.

An important question related to this situation is *why* the students so often choose imitative reasoning instead of creative reasoning. Several possible explanations are indicated by research. Hiebert (2003) states that students learn what they are given the *opportunity to learn*. He argues that the students' learning is connected to the activities and processes they are engaged in. It is therefore important to examine the different types of reasoning that the students perform during their studies. In a previous study (henceforth referred to as 'the classification study') more than 200 items from 16 calculus exams produced at 4 different universities were analysed and classified (Bergqvist, 2007). The results showed that only one of the exams required the students to perform creative reasoning in order to pass the exam.

The general question in this study is therefore: *Why are introductory level calculus exams designed the way they are, with respect to required reasoning?* This question is examined through interviews with the teachers that constructed the exams analysed in the classification study (Bergqvist, 2007). The same conceptual framework (Lithner, 2008) used to classify the items is used in this study to analyze the teachers' statements.

Background

Students that are never engaged (by the teacher) in practising creative reasoning are not given the opportunity to learn (Hiebert, 2003) creative reasoning. Students also seem to follow a "minimal effort" principle, i.e. they choose processes that are as short and easy as possible (Vinner, 1997). This principle indicates that the students would rather choose a method that rapidly provides an answer than a method that demands more complicated analytic thinking. It is therefore crucial to examine to what extent the students encounter creative reasoning in e.g. *textbooks*, *teachers' practice*, and *tests*. Firstly, this section presents research concerning these three parts of the students' environment. Secondly, the research framework (Lithner, 2008) is described, and thirdly, the results from the classification study (Bergqvist, 2007) are presented since it is the basis for the present study. The section is concluded with a presentation of the purpose and research questions of the present study, now possible to formulate based on the results from the classification study and using concepts from the framework.

Textbooks

There are several reasons to believe that the textbooks have a major influence on the students learning of mathematics. The Swedish students—both at upper secondary school and at the university—seem to spend a large part of the time they study mathematics on solving textbook exercises (e.g. Johansson & Emanuelsson, 1997). The Swedish TIMSS 2003 report also shows that Swedish teachers seem to use the textbook as main foundation for lessons to a larger extent than teachers from other countries (Swedish National Agency of Education, 2004). The same study notes that Swedish students, especially when compared to students from other countries, work independently (often with the textbook) during a large part of the lessons. Furthermore, Lithner (2004) shows that it is possible to solve about 70 % of the exercises in a common calculus (university) textbook using imitative reasoning. All these results and circumstances imply that the textbooks play a prominent role in the students' learning environment and that it focus on imitative reasoning.

Teachers' practice

The teachers' practice, especially what they do during lectures, is another factor that affects the students' learning. Teachers also often argue that relational instruction is more time-consuming than instrumental instruction (Hiebert & Carpenter, 1992; Skemp, 1978). There are however empirical studies that challenge this assumption, e.g. Pesek and Kirshner (2000). One way for the teachers to give the students the opportunity to learn creative reasoning could be to 'simulate' creativity during lectures. Bergqvist and Lithner (2005) show that creative reasoning is only simulated by the teacher to a small extent.

Vinner (1997) argues that teachers may encourage students to use analytical behaviour by letting them encounter tasks that are not solvable through pseudo-analytical behaviour. This is similar to giving the students the opportunity to learn creative reasoning by trying to solve tasks that demand creative reasoning. Vinner comments, however, that giving such tasks in regular exams will often lead to students raising the 'fairness issue,' which teachers try to avoid as much as possible. This limits the possible situations in which students can be compelled to use analytical behaviour (Vinner, 1997).

Tests

Several studies show that assessment in general influence the way students study (Kane, Crooks, & Cohen, 1999). Palm, Boesen, and Lithner (2005) examine teacher-made tests and Swedish national tests for upper secondary school. The focus of the study was to classify the test tasks according to what kind of reasoning that is required of the students in order to solve the tasks. Palm et al. (2005) showed that the national tests require the students to use creative reasoning to a much higher extent (around 50 %) than the teacher made tests (between 7 % and 24 % depending on study programme and course) in order to get a passing grade. The results from these studies indicate that upper secondary school students are not required to perform creative reasoning to any crucial extent, since the national test does not determine the students' grades to any higher extent than the teacher-made tests. This result is in alignment with other studies indicating that teacher-made tests mostly seem to assess some kind of low-level thinking. One example is an analysis of 8800 teacher-made test questions, showing that 80 % of the tasks were at what is called the "knowledge-level" (Fleming & Chambers, 1983). Senk, Beckmann, and Thompson (1997) classify a task as *skill* if the "solution requires applying a well-known algorithm" (p. 193) and the task does not require translation between different representations. This definition of skill has many obvious similarities with Lithner's definition of algorithmic reasoning (see the following section). Senk et al. (1997) report that the emphasis on skill varied significantly across the analysed tests—from 4 % to 80 % with a mean of 36 %. The authors also classified items as requiring reasoning if they required "justification, explanation, or proof." Their analysis showed that, in average, 5 % of the test items demanded reasoning (varying from 0 % to 15 %). Senk et al. (1997) also report that most of the analysed test items tested low-level thinking, meaning that they either tested the students' previous knowledge, or tested new knowledge possible to answer in one or two steps.

This presentation of research on textbooks, teachers' practice, and tests, suggests that students use imitative reasoning since that is what they are given the opportunity to learn.

The Reasoning Framework

Lithner (2008) has developed a conceptual research framework that specifically considers and defines different types of mathematical reasoning. In the framework, *reasoning* is defined as "the line of thought adopted to produce assertions and reach conclusions in task solving" (p. 3). Reasoning is therefore *any* way of thinking that concerns task solving—it does not have to be based on formal deductive logic—and can denote even as simple procedures as recalling facts. Lithner (2008, p. 257) points out that with this definition the line of thought is called reasoning "even if it is incorrect as long as there is some kinds of sensible (to the reasoner) reasons backing it." The framework defines two basic reasoning types; *imitative reasoning* and *creative mathematically founded reason-*

ing. The classification study (Bergqvist, 2007) uses two types of imitative reasoning, *algorithmic reasoning* and *memorized reasoning*, and two types of creative mathematically founded reasoning, *local* and *global creative mathematically founded reasoning* (Figure 1). These four types of reasoning are defined shortly below and are presented more thoroughly in the classification study and in the framework itself.

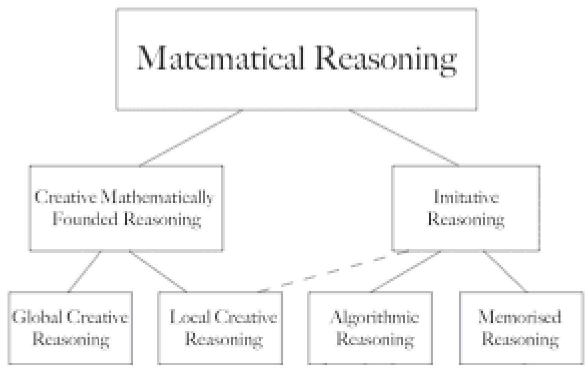

Figure 1: Overview of reasoning types in the framework used in this study

Imitative reasoning

Imitative reasoning (IR) is a type of reasoning based on copying task solutions. For example, a student can solve a task using imitative reasoning by remembering a fact or by using an example in the textbook as a guide. The two main types of imitative reasoning are *memorized* and *algorithmic* reasoning.

The reasoning in a task solution is denoted ***memorized reasoning*** (MR) if it is founded on recalling a complete answer by memory and it is implemented only by writing down the answer.

The reasoning in a task solution is denoted ***algorithmic reasoning*** (AR) if it is founded on remembering a set of rules that will guarantee that a definite solution can be reached, and consists of carrying out trivial (to the reasoner) calculations or actions by following the set of rules.

Creative mathematically founded reasoning

Lithner (2008) uses the concept of creative mathematically founded reasoning (from now on abbreviated 'creative reasoning' or CR) to characterize the creation of new and reasonably well-founded task solutions. To be classified as CR the reasoning has to be new (to the reasoner) and, in order to be mathematically founded, to be supported by arguments anchored in intrinsic mathematical properties of the components involved and that these arguments motivate why the conclusions are true or plausible. If a task is almost completely solvable with IR and requires CR only in a very local modification of e.g. the used algorithm, the task is said to require *local creative reasoning* (LCR). If a task has no solution that is globally based on IR and therefore demands CR all the way, it is said to require *global creative reasoning* (GCR).

The classification study

The purpose of the classification study was to examine the reasoning that Swedish university students in mathematics have to perform in order to solve exam tasks and pass exams. The study aimed at answering two research questions:

- *In what ways can students solve exam tasks using imitative and creative reasoning?*
- *To what extent is it possible for the students to solve exam tasks using imitative reasoning based on surface properties of the tasks?*

The results were obtained through analysis and classification of more than 200 items from 16 calculus exams produced at 4 different Swedish universities. These exams each completely determined the students' grades at the courses that they were given. The grades were *failed*, *passed*, and *passed with distinction*. The students at all courses had a maximum time available of approximately 6 hours. This setting varies between universities, courses, and teachers, but is quite common in Sweden.

A task is only possible to solve using imitative reasoning (IR) if the students have had a chance to learn and familiarise themselves with the assignment given in the task and its solution. Whether a task demands creative reasoning (CR) of the students or not is therefore directly connected to what type of tasks the students have practiced solving before the exam. Thus a task was classified based on what the students have had a chance to practice and what is familiar to the students. The concept of familiarity is connected to imitative reasoning and is presented more in detail in Bergqvist (2007).

The analysis of the classification showed that more than 65 % of the tasks were solvable by imitative reasoning (IR), and all exams except one were possible to pass without using creative reasoning (CR). The qualitative analysis shows that all task solutions demanding local creative reasoning (LCR) consist of a familiar global algorithm that has to be adjusted in some small way. The fact that the global algorithms in these tasks are familiar gives the students a reasonable chance of solving at least parts of the given tasks, even without using CR. It is in fact possible for the students to completely (IR tasks), or to a large extent (LCR tasks), solve 91 % of the tasks without using CR since only 9 % of the tasks demand global creative reasoning (GCR). Using IR and LCR, i.e. basically memorized solutions or solutions based on a familiar global algorithm, it is possible to get the grade passed with distinction on all exams except one.

Purpose and questions

The purpose of this study is to better understand university teachers' rationale when they create calculus exams, especially concerning the reasoning that the students are expected to perform in order to pass the exams. The aim is to answer the general question (GQ): *Why are introductory level calculus exams designed the way they are, with respect to required reasoning?* The study therefore examines why the exams analysed in the classification study (Bergqvist, 2007) were constructed with the particular proportion of types of tasks that they have but also how the teachers who constructed the exams view exam construction in general and the results of the classification study. In order to examine the possible answers to the general question, this study aims at answering the following research questions.

Research question 1: *What factors affect university mathematics teachers when they construct exams?* (RQ1)

Research question 2: *In what ways, and to what extent, is the aspect of required reasoning considered by university mathematics teachers when they construct exams?* (RQ2)

Research question 3: *What are university mathematics teachers' views on the existing reasoning requirements in exams?* (RQ3)

Method

Choice of method

The research questions concern the teachers' rationale for constructing exam tasks so the answers are sought through interviews with teachers. The interview situation is especially appropriate since it concerns difficult concepts and it provides the interviewees with possibilities to ask questions and the interviewer with possibilities to clarify concepts in case of any misunderstandings. The interview outline is semi-structured, with pre-defined questions, but allowing deviations with an aim to explore statements by the teachers that are especially relevant, e.g. as described in Kvale (1996).

Collection of data

The interviewees

The eight teachers whose exams were classified during the previous study (Bergqvist 2007) were selected to participate in this interview study, and six of them accepted. The two remaining teachers declined due to a heavy work-load. The six teachers are in some sense selected randomly, since the exams originally were chosen by randomly selecting three Swedish universities, and adding the author's own, and then studying all introductory calculus exams produced at these universities during one academic year (2003/2004). The group is too small to completely represent all Swedish university teachers in mathematics, but the teachers work at four universities of different size, location, and age. Five of the six teachers in the study are senior lecturers and have Ph. D. degrees in mathematics. The sixth teacher is an instructor which implies an educational level equivalent to a master in mathematics. The teachers' ages vary between 30 and 60 years. Five of the teachers were male, one was female.

The Interview

To facilitate communication of previous results and of concepts, the interviews use a type of stimulated recall (Calderhead, 1981). From the beginning of each interview the interviewer and the interviewee has an exam constructed by the interviewee at hand. The interviewee is invited to exemplify any statements with tasks from the exam if appropriate. Later on during the interview the classification of each task in the exam is presented to the interviewee, and any discussion of concepts and classification can be given concrete form through examples from the teacher's own exam. The method of using existing data from the teachers' practices (in this case their exams) is inspired by Speer (2005), who argues that some previously reported inconsistencies between teachers' professed and attributed beliefs might be a consequence of a lack of shared understanding.

Each interview starts with a brief introduction concerning the study and is then divided into four parts, each described below. The interview outline was piloted on two teachers, not included in the final version of the study, and the outline was adjusted according to the resulting experiences. Each interview question is based on at least one research question (see each part of the interview). Each interview took approximately one hour.

Part One. The teacher is asked about his/her views on exam construction, how he/she constructs exams and exam tasks, and about choices, difficulties, goals and practical circumstances. Two examples of practical circumstances were given to all teachers: “e.g. the *time* or the *students*”. All interview questions are connected to the first research question (RQ1: *What factors affect university mathematics teachers when they construct exams?*) and some, that specifically consider the exam tasks, are also connected to the second research question (RQ2: *In what ways, and to what extent, is the aspect of required reasoning considered by university mathematics teachers when they construct exams?*). The questions are posed without any specific vocabulary or concepts being introduced, so that the teacher him-/herself can make the judgement on what is important in this context. The teacher is in the beginning of this part explicitly invited to use his/her own exam to exemplify any statements.

Part Two. Three concepts regarding the requirement of reasoning are introduced: *memorized reasoning*, *algorithmic reasoning* and *creative reasoning*. This is done by letting the teacher read a few pages of text that described these different types of mathematical reasoning, including relevant examples from the teacher’s own exam. The difference between local and global creative reasoning is mentioned but not elaborated on. The purpose of introducing some specific concepts is to create a basis for common understanding of a vocabulary that enables communication between the interviewer and the interviewees. The teacher is then asked if he/she takes the required reasoning of the tasks into consideration during exam construction, a question that is directly related to the second research question (RQ2).

Part Three. The teacher is first informed of the classification results for one of his/her own exams (Bergqvist, 2007) and is then asked about his/her opinion of these results, e.g. why the exam had the particular distribution and if he/she is pleased with the distribution. The interview questions are all directly connected to the second and/or third research questions (RQ2 and RQ3: *What are university mathematics teachers' views on the existing reasoning requirements in exams?*).

Part Four. The teacher is informed of the general results from the classification study (see Section *The classification study*). The interview questions all concern the teachers' views on the existing reasoning requirements (RQ3); the teachers are e.g. asked about his/her opinions of these results, what he/she thinks are possible reasons for them, and whether the teacher thought that the aspect of reasoning is important or not.

Method of Analysis

Main analysis

The main analysis includes the transcriptions of all interviews and are performed in several steps according to the following description, similar to suggestions by e.g. Glesne (1999). The transcribed interviews are divided into sections based on 1) the subject treated in each section and 2) which research question that is most closely related to that subject. Each section is headlined using a description of the subject (1). For example, if the teacher talks about *when* she constructs the exams this will be seen as connected to RQ1 (*What factors affect university mathematics teachers when they construct exams?*) and therefore headlined with "When the exam is constructed (RQ1)". For each particular RQ there will be several different headlines. All sections (from all the teachers) are then grouped according to their headlines and the teachers' opinions in the grouped sections summarised. The summaries are then used to answer each of the three RQs and the general question. If some section treats a subject not related to the RQs and therefore cannot be given a headline that is connected to any of the RQs, that section will be treated separately after the rest of the analysis was done. There will probably be very few such sections since the interview is semi-structured and each interview question is connected to at least one RQ.

Additional method of analysis for RQ1

The first research question (RQ1) is "*What factors affect university mathematics teachers when they construct exams?*" The answers to this question may concern anything from practical conditions to opinions and intentions and there are probably many factors. To enable a better overview of the factors that affect the teachers the factors are grouped. Each group will consist of one or several connected factors where each connection is motivated by direct argumentation and/or by several explicit teacher statements. Direct argumentation means here for example that the factor *student group* is connected to the exam's *degree of difficulty*, with the argument that the students' previous knowledge will affect how difficult the exam is to them. Motivation via the teachers' statements could instead be that these two factors are connected since several teachers said that when they teach a low-performing student group they, more or less subconsciously, make the exam less difficult. The factor that is most strongly connected to the others in the group and/or most relevant to the purpose of this study is chosen to represent the group and is called the *main factor*.

Analysis

Answers to RQ1

The first research question is "What factors affect university mathematics teachers when they construct exams?" (RQ1). A lot of different factors were mentioned by the teachers: time, students' proficiency and previous knowledge, course content, tradition, degree of difficulty, task type, etc. The analysis resulted in four groups of factors, and accordingly four main factors.

There are a few weak connections ignored in this structure. An example of a weak connection is that two teachers think that the degree of difficulty of the exams is linked to the factor tradition via the examiners at their department. This connection is ignored since the examiners are part of a system present at only one department and the connection is not mentioned by any of the other teachers.

Main factor 1: Time

The first group consists of only one factor and that is ***time***. This factor is mentioned to all teachers when the question is asked (as an example) and all teachers confirms explicitly that time plays a prominent role in the exam construction process. The teachers do not specifically connect time, or the lack of time, to any other of the factors mentioned in the study. Two of the teachers do not feel that they lack time considering the way the exams look today. Both of them, however, have the opinion that the type of examination most common today, the written exam, is a product of a general lack of time. Several of the alternative examination types, e.g. oral exams, demand more time, which the teachers simply do not have. The four other teachers view the situation in different ways. The first states that time is a factor in exam construction but he does not specify how. The second states that consequence of the lack of time is that he does not spend as much time constructing exams as he thinks that he should. The third teacher believes that the access of time admittedly affects the process, but that constructing exams takes whatever time the teacher has at hand, and that sometimes the exam will not get any better no matter how much time he or she spend working on it. The fourth teacher does at first not seem to think that time is a factor, but later in the interview, during the discussion on alternative forms of examination, he states that he would choose to administer an oral exam instead of a written if he did not lack the time.

Main factor 2: Degree of difficulty

The second group consists of four factors, three tightly connected and one more loosely related. The first three factors are the *student group*, the *degree of difficulty*, and a *high passing rate*. The last factor is *whether the exam is a make-up exam or a regular exam.* The connections between factors in this group are the following. A group of high performing students will have a higher passing rate than a group of low performing students. If the degree of difficulty of the exam is relatively low, the passing rate will be relatively high. If the students lack certain previous knowledge, a task might be more difficult for them than for other students. The first three factors are therefore directly connected. The last factor is connected to the student group, since most high performing students do not take a make-up exam and the student group taking a make-up exam is therefore in average probably lower performing than a group taking a regular exam. The factor chosen to represent the whole group of factors is ***degree of difficulty*** since it is a property not only of exams but also of tasks and therefore is most closely connected to the focus of the study. Also, all teachers mention the degree of difficulty as an affecting factor. The first three factors will be more extensively discussed below since the teachers did so during the interview.

The teachers all agree that the composition of the *student group* is a factor that affects the exams. A couple of teachers state explicitly that there is a direct link between the students' proficiency and the exam's degree of difficulty.

> “Of course, if you have very poor students at a course then... (...) I don't do it deliberately, it's not as if I try to construct an easier exam, but I still believe that... you're affected by it.” (Teacher D)

The students' previous knowledge is a property of the student group and is mentioned by five of the teachers. Two of them mention that when students lack certain proficiencies, it limits the degree of difficulty possible to set for the exam. One adds that it would not help to change the form of the examination, since a lot of other circumstances, e.g. the extent of the contents that is included and the need for a high passing rate, also limit what is possible to test. The three other teachers discuss the issue of students' previous knowledge in connection to the distribution of tasks of different reasoning types in their own and other teachers' exams. They all express the view that the students' previous knowledge, or rather their lack of such, is directly connected to the amount of creative reasoning (CR) that is reasonable to test at an exam (or even to work with during teaching). One of these teachers suspect that it would be possible to test a higher amount of CR at a course in Linear Algebra because it is not based on the students' presumed previous knowledge to the same extent as the Calculus courses. The teachers' views on this connection are further described in the section related to the second research question (RQ2).

All teachers mention in different contexts that the *degree of difficulty* of the tasks is something that they consider to some extent during exam construction. Four of the teachers say that it is often, if not always, difficult to produce (or choose) tasks with a reasonable degree of difficulty. Some specific examples mentioned by the teachers are: the first time a teacher is teaching a course; if the teacher tries to make the tasks interesting and/or fun; when a course is graded with more than two grade levels[2]; and when the course is more advanced. One of these four teachers adds that even though the curriculum is intended to point out what the students are supposed to master at the end of the course, it is very difficult to use the curriculum in order to set a reasonable difficulty level.

Two teachers point to the need for *a high passing rate* as a factor influencing exam construction.

Two teachers remark that they usually do not spend as much time constructing a make-up exam as they do with a regular exam. One of them saves the "really interesting" tasks for the ordinary exams since most of the capable students do not take the make-up exam (they already passed the regular exam). The other teacher simply states—without explaining how—that this factor affects him during exam construction.

Main factor 3: Task type

Group three consists of four factors with the factor ***task type*** as main factor. Directly connected to task type are the factors *textbook*, *previous exams*, and *tradition*. The task type, i.e. whether a task is a standard task or not, is according to several of the teachers determined by what type of tasks the students have encountered before. Previous encounters are connected to the textbook and to some extent also to previous exams. (The subject matter presented in the textbook is not focused here, only the *type* of tasks in the textbook.) One teacher points to the tradition within the engineering students courses, where mostly tasks solvable by well-known algorithms (task type) are expected. Another teacher mentions previous exams as a part of the tradition in mathematics. The factor tradition is therefore connected to both task type and previous exams. The teachers' wish to vary the contents of the exams within its traditional form is directly connected to previous exams and variation is therefore also a factor included in this group. ***Task type*** was chosen to be the main factor in this group since it is, according to the analysis of RQ2, directly connected to required reasoning in tasks.

All the teachers consider *task type,* whether a task is a standard task or if it somehow requires more from the students, when constructing exams. There is a separate terminology for almost every teacher in the study, but when the teachers are asked to specify what they mean it turns out that it is possible to connect their definitions to the definitions in the framework via the concept of familiarity (see Section framework). The connection is more thoroughly analysed in relation to RQ3 and is presented in the sub-section *Common understanding of concepts.*

The *textbook* is mentioned by three of the teachers in the study and one of them suggests a connection between the type of tasks in the textbook and in the exam. He is of the opinion that the textbook gives a frame for what the students have been working with during the course and therefore indirectly affects the exam. Three teachers feel that it is difficult to produce an interesting, and preferably enjoyable, exam that is not a copy of *previous exams*, which at the same time does not differ too much from the older exams. Two other teachers state that previous exams from the course affect the design and contents of the exam and, according to one of them, also the difficulty level.

Four teachers mention *tradition* as a factor. One of the teachers explicitly states that the tradition within teaching mathematics is very strong, and both he and another teacher mention that the typical written exam, containing 6--8 tasks and being presented to the students at the end of a course, is a product of this tradition. The third teacher mentions another type of tradition, the engineering students' strong traditions that are inherited from one age group to the next. He states that it strongly affects the teaching and the examination of these students. Concerning this culture among the engineering students (studying to be Masters of engineering), this teacher says that:

> "[These students] are in some way prepared to learn a lot of algorithms. It is a part of the tradition. (...) [They] are supposed to squeeze in massive amount of credits into a small number of years, there isn't much time for reflection (...) [they] adapt, that's understandable. At the

> same time as you try to get at it, you feel that the whole system is working against you then." (Teacher C)

The fourth teacher declares that the existence of an examiner at his department naturally influences the form and degree of difficulty of the exams. This can be regarded as a type of formalised tradition affecting the exams.

Main factor 4: Task content domain

The last group consists of three factors: the *task content domain*, the *course content*, and the *textbook*. The connection between these three factors is the following: the content domains of the tasks in an exam are directly connected to the course contents, which usually are presented in the textbook (in this case it is the subject matter present in the textbook that is focused, not the type of tasks included). The main factor is chosen to be ***task content domain*** since it too represents a property of the exam tasks.

Three teachers say that the *content domains* (e.g. differentiation, related rates, or limits) of the exam tasks should be a representative selection of the course contents. Three of the teachers point out that the *course itself* naturally affects the exam's structure and contents, e.g. via the syllabus. The *textbook* is mentioned by three of the teachers in the study and two of them mention aspects connected to the content of the course. One states that he uses the textbook as a starting point for exam construction. The other says that even though the textbook influences the content of the exams, the tasks in the textbook are often too easy and uncomplicated in comparison to the exam tasks. All teachers mention the *course content* in at least one of the themes: the exam construction process, the practical considerations, or the task properties, although then as the factor task content domain.

According to the analysis there are more than ten different factors, each affecting at least two of the teachers in the study, and connected either to the exam construction process, to the task properties, or to practical considerations. The analysis further shows that there are four main factors that each represents a group of factors: the *time*, the *task type*, the *task content domain*, and the task's or exam's *degree of difficulty*.

Answers to RQ2

Here the answer to research question 2, *In what ways, and to what extent, is the aspect of required reasoning considered by university mathematics teachers when they construct exams?*, is presented.

All teachers state during the first part of the interview (before the concepts of imitative and creative reasoning has been introduced to them) that they consider task type, i.e. whether a task is a standard task or not, during exam construction (see the concept of *task type* introduced in the Section Answers to RQ1). Their descriptions of standard tasks differ, but can in every case be connected to the aspect of required reasoning and the connections are in all cases but one explicitly confirmed by the teachers themselves (see section Answers to RQ3, last sub-section). These connections imply that the teachers do consider the aspect of required reasoning when they construct exams.

Five teachers agree that the distribution of tasks of different reasoning types varies, and should vary, with the level of the course. They all think that the proportion of tasks demanding creative reasoning (CR) should be larger in courses on higher level mathematics, but have different ways of motivating their opinion. One teacher says for example that on the lowest university level, the students are supposed to learn calculation skills, and that if they want to “learn mathematics” they should keep on studying. Some teachers also guess that the distribution would be different on courses on the same level but with different mathematical contents. Three of the teachers think that there are a smaller percentage of algorithmic reasoning (AR) tasks in a linear algebra exam than in these calculus exams, and one teacher points to Geometry as a subject where the part CR would be particularly high. All four teachers believe that it is the lower demand on the students' previous knowledge in these courses that makes it possible for the teacher to work with more CR during teaching and assessing.

All the teachers also say, more or less explicitly, that tasks where the students themselves are supposed to construct a part of the solution, a common example is a short proof, usually are more difficult for the students than the tasks they solve using well-known algorithms. Translated into the terminology of the reasoning framework this means that the teachers view tasks that demand creative reasoning (CR) as more difficult than tasks solvable with algorithmic reasoning (AR).

Answers to RQ3

Research question 3 reads: *What are university mathematics teachers' views on the existing reasoning requirements in exams?* In order to draw reasonable conclusions of the teachers' statements, it is important to determine if their views, as they are expressed in the interviews, actually concern the same properties of the tasks that the theoretical framework defines. The first sub-section below therefore discusses if there is a common understanding between interviewer and interviewees of the concepts involved. That discussion is concluded in the fourth sub-section after more relevant information is presented in sub-section two and three. It would also be difficult to have a fruitful discussion with the teachers concerning the existing reasoning requirements in their exams if they saw the classification results as unreasonable. The second sub-section therefore takes a closer look at the teachers' views of the classification of the tasks in their own exam. The third sub-section focuses on the teachers' views on the general results from the classification study. As mentioned, the fourth sub-section concludes the discussion on the common understanding.

Common understanding of concepts

Five of the teachers explicitly say that they think that the aspect of required reasoning is important and relevant. All six teachers use an informal and personal terminology (presented before the formal concepts are introduced by the interviewer) to describe task properties, and for every teacher this terminology is possible to link to the definitions of algorithmic and creative reasoning (AR and CR). The connection between the teachers' personal vocabulary and the formal terminology is in most cases based on the notion of familiarity of answers and algorithms connected to imitative reasoning. Furthermore, all teachers but one (Teacher B) did explicitly say that the presented definitions (in part 2 of the interview) mainly coincided with what they intended to describe with their own terminology. The specific case of teacher B is presented in the end of this sub-section.

AR tasks

The teachers' personal terminology describing AR tasks are "standard tasks", "routine tasks", "typical tasks", "template tasks" and "tasks without a twist." During the interviews the teachers explained what they meant by these terms, and their explanations were very similar to the definitions of AR tasks. The similarities were particularly obvious when the teachers' descriptions were compared to the classification tool and the concept of familiar tasks (Bergqvist, 2007; Lithner, 2008). One teacher for example said that: "Most of the tasks are typical routine tasks, standard tasks, that you should be able to solve if you've solved most of the tasks in the textbook." (Teacher E) When this teacher later in the interview was introduced to the definitions of the reasoning types, he pointed out that AR tasks were the same tasks as the ones he called routine tasks. Another example is one teacher that describes the concept of "template tasks" as "tasks that are more or less a copy of tasks that the students have solved before" (Teacher C). The teachers generally defined different types of tasks very concretely and in direct connection to what the students have done during the course.

CR tasks

The teachers' descriptions of non-routine tasks were possible to connect to CR tasks in a similar way. Teacher E describes as follows:

“You can always complicate one [task], of course, it's very easy, you just have to change a formulation, and the task is all of a sudden significantly more difficult. Add an arbitrariness, or calculate a task backwards, or... almost whatever.”

Typical examples of the teachers' personal terminology for CR tasks are “tasks with a twist,” “tasks that contain something tricky or unexpected,” and “types of tasks that the students have not seen before.” This is in direct alignment with the definition of CR tasks.

MR tasks

Memorized reasoning (MR) is not defined or commented upon by the teachers to the same extent as algorithmic and creative reasoning. All of the exams discussed with the teachers during the interviews contain MR tasks according to the classification, so the type is used by the teachers. Only a couple of the teachers, however, mention or describe this type of tasks.

The first teacher points out that even if a student has memorized something, e.g. a proof of a theorem, the student's presentation at the exam might reveal information on what the student really understands. This opinion is not shared by all of the other teachers. During other parts of the interview (not in connection to the definitions) one teacher said that the memorized reasoning (MR) tasks are practically uninteresting, since they only measure the students’ capacity of memorising information. Another teacher thinks that memorising definitions and theorems definitely has its place in mathematics, the same way that memorising vocabulary is a part of learning a foreign language. In spite of this opinion he does not think that MR tasks measure anything more than what the students have memorized.

Teacher B

The only teacher that did not say that his own view of task properties coincided with the presented definitions explained that he does not specifically think in those terms. Compared to the other teachers, he is also the one most hesitant to use the definitions after they were presented. It is, however, apparent from his description of tasks and how he constructs tasks that he in many cases does not want the students to use AR. He says about task construction: “That is also something that I try to think about, that you [the student] won't be able to just take a formula from the textbook and just insert...” (Teacher B) He says that he wants the students to learn the line of reasoning behind the solutions, and that he often constructs tasks that are similar to the textbook tasks but includes small changes. According to him these changes do not “really” make the tasks more difficult, but he also acknowledges that they are harder to solve for the students. His strive for the students to perform local creative reasoning corresponds with the result from the classification study where he is the only teacher who constructed an exam that is not possible to pass using only imitative reasoning.

Teacher B also notes that a certain type of tasks might first belong to one category and later on to another. He takes the introduction of a task in keeping with “What does it mean that a function f is differentiable in a point?” as an example. He describes how he introduced such a task in his exams one year, and then let it be included in each exam for a few years to follow. The task was difficult for the students the first year and their results were poor, but after a few years the students seemed to know how to answer the task. He suggested that the student at first needed some type of creative reasoning (CR) to solve the task, but that in later exams it was solved by memorized or algorithmic reasoning (MR or AR). This is in alignment with the classification of AR and MR tasks: when the students have encountered a task several times, it becomes familiar and therefore CR is no longer necessary in order to solve it. Teacher B is discussed again in the following sub-section and the question of common understanding of concepts is therefore concluded in a summary in the end of the section Answers to RQ3.

The teachers' views on the classification

All teachers in the study accept the classification of the tasks in their own exam, two of them straight away, and four after some discussion. The discussions concerned for example why a particular related rates task was classified as local creative reasoning and not algorithmic reasoning. In this case the reason was that the specific real world situation was new to the students and they had to perform creative reasoning when they created a mathematical model. In all cases but one the teacher agreed with the classification after the discussion, and in that single case the task was in fact misclassified. This implies that the teachers have embraced the definitions and it supports the reliability of their opinions of the definitions and the classifications of their exam tasks.

All teachers expressed that they were satisfied with the distribution of tasks of different reasoning types in their own exam. An opinion presented by several teachers in different contexts, and by two teachers in this specific context, is that it is not reasonable, or maybe not even fair to the students, to believe that the students will be able to solve tasks requiring creative reasoning during the exam of this course. One teacher says for example that it is not possible for the students to be creative if they do not master the basics first. Another teacher states, based on the opinion that the students' previous knowledge has deteriorated, that:

”We can't demand from the students that they, in addition to learning to handle the basic concepts, also manage to put them in context, at least not as soon as at the exam.” (Teacher E)

This is in line with the fact that most of the teachers believe that there could, and maybe should, be a larger percentage of creative reasoning demanded by students taking higher level courses and also on courses that demand less previous knowledge by the students.

Four of the teachers said that tasks solvable with memorized and algorithmic reasoning (MR and AR) are important and should not be regarded as less valuable. Two of those teachers were of the opinion that approximately half of the points on the exams should be MR or AR. A third teacher said that even though it is not explicitly discussed among the teachers, it is clear to him that they try to measure calculation skills in the introductory calculus exams. The teacher later confirms that his concept *calculation skills* is directly connected to AR. He also believed that the percentage of AR in this course is increasing over the years, and that this increase is a sign of the teachers adapting to the deterioration of the students' previous knowledge.

The teachers' views on the results from the previous study

The five teachers whose exams were possible to pass using only IR were not surprised by the results from the classification study and also viewed the results as reasonable. One of them said:

”If I were to guess, then 16 of 16 exams would be like that, where you can get more than 50 % of the points with those two categories [MR and AR] So it is definitely not a surprise that it is such a large percentage.” (Teacher A)

When the teachers were asked what they think is the reason for the resulting distribution of tasks of different types, their answers varied to some extent. One teacher suggested that what the students primarily need is to learn mathematical tools that will help them with the upcoming courses. Another stressed that it is important that the students are able to implement algorithms. Some of the teachers meant that over the last ten years the percentage of algorithmic reasoning in the exams in this type of course has increased, and that this is due to the fact that for each year the students have poorer and poorer previous knowledge. Four of the teachers agreed that the main reason for the distribution is that a larger proportion of CR tasks would make the exams too difficult, which in turn would result in too few students getting passing grades.

Teacher B again

The teacher whose exam was not possible to pass without creative reasoning (CR), Teacher B, was the one most surprised by the results of the classification of the teachers' own exams. He was a bit puzzled since his opinion was that it should be possible to pass the exams using only IR, but he still did not think that his own exam was too difficult. Based on a quote presented earlier (in Section *Teachers' views on the aspect of reasoning*) he however said: "That is also something that I try to think about, that you [the student] won't be able to just take a formula from the textbook and just insert..." (Teacher B) This could indicate that his conception of imitative reasoning differs some-what from the definition used by the researcher; that imitative reasoning tasks are very familiar to the students and that the students have solved more or less identical tasks several times. It is possi-ble that what he consider to be imitative reasoning (e.g. in his own exam) would be defined as local creative reasoning (LCR) in a formal classification (which is what his exam requires according to the classification study). The difference between these two concepts is indeed small and probably the most difficult to grasp in such short time as during the interview.

Summary of common understanding of concepts

There is a common understanding between interviewer and all interviewees of the basic concepts *imitative* (IR) and *creative reasoning* (CR), even if the finer details concerning the difference be-tween algorithmic (AR) and local creative reasoning (LCR) might be lost in such a short interview. The presentation of the concepts in the second part of the interview was indeed focused on memo-rized, algorithmic, and creative reasoning, without going into detail considering local and global creative reasoning, something that might be the cause of the difference in views. For five of the teachers the difference between AR and LCR does not affect the classification of their exam: the students can pass it using only IR (memorized and algorithmic reasoning). For one teacher (Teacher B) the results might be more confusing since his exam requires the students to solve IR tasks *and* some LCR tasks in order to pass. If the difference between these concepts is small, it is not surprising that he is more hesitant than the other teachers to accept the suggested terminology. Teacher B's statements concerning the difference between CR and IR in general however suggests that these two basic concepts are clear to him. On the other hand, Teacher B also noted that a CR task could change into an IR task when it became familiar to the student, which implies that he has grasped the basic idea of the difference between imitative and creative reasoning.

Discussion

The main purpose of this study is to find answers to the general question: *Why are the exams designed the way they are, with respect to required reasoning?* In order to answer this question, three research questions were formulated and the first four sub-sections below focus on answering each of these three research questions and the general question. How the teachers connect task difficulty to required reasoning, and the consequences thereof, is discussed in a separate sub-section after that. The last sub-section focuses on the teachers' views of the students' competences.

What factors affect university mathematics teachers when they construct exams? (RQ1)

The analysis shows that there are four main factors that affect the teachers, and they each represents a group of factors: the *time*, the *task type*, the *task content domain*, and the task's or exam's *degree of difficulty*. Time and student group was suggested as possible factors to all the teachers and all of them agreed that time was an important factor. They did not specifically connect this particular factor to any of the other ones. The three other main factors mentioned above represent the properties of tasks that the teachers say they consider and all these were connected to several other factors according to the grouping described in the analysis.

In what ways, and to what extent, is the aspect of required reasoning considered by the teachers when they construct exams? (RQ2)

According to the teachers' statements, they consider the aspect of required reasoning during exam construction. They verbalise the aspect through descriptions of properties and types of tasks that they choose to include, or not include, in their exams. An important result is that these teachers see an obvious connection between the type of reasoning required of a task and its degree of difficulty. They strongly feel that creative reasoning is more difficult than imitative reasoning for the students, even though some teachers comment that AR tasks can be difficult as well. This result points to a strong link between two of the main factors in the analysis of RQ1: *task type* and *degree of difficul-ty*.

What are the teachers' views on required reasoning in exams? (RQ3)

The analysis shows that there is a common understanding between the interviewer and the interviewees concerning the basic concepts imitative and creative reasoning. The teachers generally accept the definitions and also the classifications presented to them, and they view the aspect of required reasoning as important. They also feel that the main result from the previous study, that it is possible to pass most of the exams using only imitative reasoning, is reasonable. Some of the teachers meant that over the last ten years the percentage of algorithmic reasoning in the exams in this type of course has increased, and that this is due to the fact that for each year the students have poorer and poorer previous knowledge. A larger percentage of tasks demanding creative reasoning would therefore result in too difficult exams and too low passing rates. Most of the teachers seemed to feel, however, that the aspect of required reasoning is both relevant and important, and that it could and should be demanded to a higher extent in e.g. courses in more advanced mathematics.

Why are the exams designed the way they are? (GQ)

The analyses of the three research questions show that the teachers are in general quite content with the situation as it is. During exam construction, they primarily focus on the tasks' content domain and degree of difficulty. Since the teachers generally regard a task demanding creative reasoning as more difficult than a task solvable with imitative reasoning, they use the tasks' types to regulate the difficulty. The lack of time, both during exam construction and during teaching, in connection to the students' poor previous knowledge makes it unreasonable to the teachers to demand that the students master creative reasoning in order to pass the exams. The teachers' referral to the students' poorer previous knowledge over the years seems to be based on their personal experiences, but is also supported by several local and national studies (Högskoleverket, 1999; Boo & Bylund, 2003; Petterson, 2003; Brandell, 2004). Four teachers suggested (Answers to RQ2) that a lower demand on the students' previous knowledge in some courses (e.g. linear algebra) makes it possible for the teacher to work with more CR during teaching and assessing during these courses. According to the teachers it can therefore be more problematic to require the students to perform CR in the courses focused on in this study (introductory calculus courses) than in introductory courses in general.

The described situation supplies a possible answer to the general question of why the exams are designed the way they are, with respect to required reasoning: *The exams demand mostly imitative reasoning since the teachers believe that they otherwise would, under the current circumstances, be too difficult and too difficult exams would lead to too low passing rates*.

Task difficulty and required reasoning

The teachers use the task type (the required reasoning) as a tool for regulating the tasks' degree of difficulty, and they believe that tasks demanding creative reasoning (CR) are more difficult than tasks solvable with imitative reasoning (IR). A relevant and important question is therefore if CR tasks really *are* more difficult than IR tasks. To answer this question it is necessary to more specifically define the term *difficulty*. A task's difficulty is not a constant property since a task that is difficult to solve for a student may be easy to solve for a teacher. Task difficulty therefore has to be determined in relation to the solver. There are many task properties that may affect a task's degree of difficulty, e.g. the task's complexity and its wording. It is therefore not easy to describe task traits that completely determine a task's degree of difficulty. In the following discussion, an exam task's degree of difficulty is therefore determined by the success rate for the students participating in the course. Although the success rate is a blunt and partly unstable instrument, it provides a useful definition in this context.

In a study on the relation between required reasoning in test tasks and the reasoning actually used by students, Boesen, Lithner, and Palm (2010) noted that the success rate in the students' solution attempts was higher for tasks classified as having high relatedness to textbook tasks (cf. IR tasks). Combined with the teachers' statements this makes it reasonable to believe that CR tasks generally are more difficult for the university students than IR tasks. One of the teachers in the present study had the following comment on whether CR is more difficult or not:

”It's probably not more difficult for everybody. It depends on how good you are perhaps. For the weaker students it is definitely more difficult, for the good students it might be easier. Because, I mean... they don't have to learn that much, they can do it anyway. Because if you solve that one [points at a definite integral task] well, then you have to learn the method of partial fractions and so on, you know, and that isn't something you figure out if you haven't seen it [before].” (Teacher B)

A CR task can be uncomplicated and simply put, and the solution might be based on resources that are well established with the solver. So the obvious question follows: do CR tasks *have to* be more difficult than IR tasks? Are the CR tasks more difficult due to some inherent property of creative reasoning, or due to circumstances connected to the situations in which the students meet the different types of reasoning?

The CR tasks that the students encounter in tests and textbooks are often difficult for them to solve. The two task properties—demanding CR and being difficult—often appear simultaneously. If CR tasks do not *have* to be difficult, then one reason for this coincident could be that it is easier to construct difficult CR tasks than difficult IR tasks. A related circumstance is that most of the uncomplicated and easy tasks in the beginning of exams and textbook chapters are solvable with IR. This could result in students never practising creativity in connection to fairly simple task settings. As a consequence, the students would not develop the competences necessary to perform CR and would not consider CR as a first choice option during task solving. Another reason that CR is more difficult than IR could simply be that the students do not practice CR. If easy tasks are often solvable with IR, and difficult tasks often demands CR, than this could also be a reason that the teachers use required reasoning to regulate the task difficulty. The teachers, and the students, might simply be used to CR tasks being difficult.

A way of changing this situation could be to let the students become more familiar with encountering unfamiliar tasks. Since a particular task is solvable with imitative reasoning if it is familiar to the students, they cannot practise using CR simply by becoming familiar with more types of tasks. The situation of *trying to solve an unfamiliar task using CR* could however become familiar in itself. Vinner (1997) argues similarly that teachers may encourage students to use analytical behaviour by letting them encounter tasks that are not solvable through pseudo-analytical solution processes. Hiebert (2003) points to the concept *opportunity to learn* and argues that what the students learn is connected to what activities and processes they engage in. Engaging the students in solving unfamiliar tasks using CR can be seen as giving the students the opportunity to learn creative reasoning.

The teachers' views of the students' competences

That the teachers use the task type (the required reasoning) as a tool for regulating the tasks' degree of difficulty is also relevant in light of how mathematics is structured in several international frameworks for school mathematics. One example is NCTM (2000), often called the Standards, that describes mathematics using two dimensions: *content standards* and *process standards*. According to the Standards, mathematics should be seen as having these two dimensions, and task type is part of the process dimension. When designing tests and exams, it is possible to consider a third dimension: the tasks' degree of difficulty. That the teachers use the tasks' required reasoning as a tool to regulate the tasks' degree of difficulty, reduces this three dimensional setting to an more simple setting with only two dimensions, content and difficulty/process, at least when reasoning is the process considered.

The two task properties that the teachers mainly consider, task content domain and degree of difficulty, concern relatively 'concrete' aspects of the tasks. The task content domain is directly visible in the formulation of a task; it concerns e.g. integration or drawing graphs. The degree of difficulty is obvious from the students' results on the exam and can partly be predicted by experienced teachers. Also, the task type is determined by the extent to which the students have previously encountered very similar tasks. Since the teachers use task type to regulate degree of difficulty, this makes the degree of difficulty even more concrete for the teachers. A possibility is that the teachers primarily focus on content and difficulty *because* they are, or can be seen as, more concrete properties of the tasks than other aspects of the process dimension.

References

Bergqvist, E. (2007). Types of reasoning required in university exams in mathematics. *Journal of Mathematical Behavior* 26(4) 348-370.

Bergqvist, T. & Lithner, J. (2005). Simulating creative reasoning in mathematics teaching. *Department of Mathematics and Mathematical statistics, Umeå University, Research Reports in Mathematics Education*, no. 2.

Boesen, J., Lithner, J., & Palm T. (2010). The relation between types of assessment tasks and the mathematical reasoning students use. *Educational Studies in Mathematics*. Published online: 27 April 2010.

Boo, P-A. & Bylund, P. (2003). Studenternas förkunskaper. *Nämnaren* 30(3): 46–51. (In Swedish)

Brandell, L. (2004). *Matematikkunskaperna 2004 hos nybörjarna på civilingenjörsprogrammen vid KTH.* Stockholm.

Calderhead, J. (1981). Stimulated recall: A method for research on teaching. *British Journal of Educational Psychology* 51 211–217.

Fleming, M. & Chambers, B. A. (1983). Teacher-made tests: Windows on the classroom. In Hathaway, W. (ed), *Testing in the schools: New directions for testing and measurement*. Jossey-Bass, San Francisco.

Glesne, C. (1999). *Becoming qualitative researchers: an introduction*. New York, Longman.

Hiebert, J. (2003). What research says about the NCTM standards. In Kilpatrick, J., Martin, G., & Schifter, D. (eds) *A Research Companion to Principles and Standards for School Mathematics,* pp 5-26. Reston Va., National Council of Teachers of Mathematics.

Hiebert, J. & Carpenter, T. (1992). Learning and teaching with understanding. In Grouws, D., (ed) *Handbook for Research on Mathematics Teaching and Learning*, pages 65–97. New York: Macmillan.

Högskoleverket (1999). *Räcker kunskaperna i matematik?* Stockholm, Högskoleverket. (In Swedish)

Johansson, B. & Emanuelsson, J. (1997). *Utvärdering i naturkunskap och matematik: lärare i grundskolan berättar: slutrapport från UNO-LÄR-projektet* (In Swedish). The Swedish National Agency for Education: Liber distribution, Stockholm.

Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Eductional Measurement: Issues and Practice*, 18(2) 5–17.

Kvale, S. (1996). *An Introduction to Qualitative Research Interviewing.* SAGE Publications, Inc. Thousand Oaks, California.

Leinwand, S. (1994). It's time to abandon computational algorithms. *Education Week*, 9(36).

Lithner, J. (2003). Students' mathematical reasoning in university textbook exercises. *Educational studies in mathematics* 52 29–55.

Lithner, J. (2004). Mathematical reasoning in calculus textbook exercises. *Journal of Mathematical Behaviour* 23 405–427.

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics* 67(3).

McNeal, B. (1995). Learning not to think in a textbook-based mathematics class. *Journal of Mathematical Behavior* 14 18–32.

NCTM (2000). *Principles and Standards for School Mathematics*. Reston, Va., National Council of Teachers of Mathematics.

Palm, T. (2002). *The Realism of Mathematical School Tasks—Features and Consequences*. Department for Mathematics and Mathematical statistics. Ph.D. Thesis. Umeå University.

Palm, T., Boesen, J., & Lithner, J. (2005). The requirements of mathematical reasoning in upper secondary level assessments. *Department of Mathematics and Mathematical statistics, Umeå University, Research Reports in Mathematics Education*, no. 5.

Pesek, D. & Kirshner, D. (2000). Interference of instrumental instruction in subsequent relational learning. *Journal for Research in Mathematics Education*, 31(5) 524–540.

Petterson, R. (2003). *Resultat av Diagnostiska prov i Matematik för nyantagna teknologer vid civilingenjörslinjerna Chalmers 1973-2005*. Göteborg: Chalmers tekniska högskola. (In Swedish)

Schoenfeld, A. H. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce of formal and informal mathematics. *Informal Reasoning and Education*. J. a. P. Voss, D. & Segal, J. Hillsdale, NJ., Lawrence Erlbaum Associates, 311–344.

Senk, S. L., Beckmann, C. E., & Thompson, D. R. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education* 28(2) 187–215.

Skemp, R. (1978). Relational understanding and instrumental understanding. *Arithmetic Teacher*, 26(3) 9–15.

Speer, N. M. (2005). Issues of methods and theory in the study of mathematics teachers' professed and attributed beliefs. *Educational Studies in Mathematics* 58 361–391.

Swedish National Agency of Education (2004). TIMSS 2003: Swedish pupils knowledge in mathematics and science in school year 8 in a national and international perspective (in swedish). *Swedish National Agency of Education reports*, no. 255.

Tall, D. (1996). Functions and calculus. In Bishop, A., Clements, K., Keitel, C., Kilpatrick, J., & Laborde, C. (eds.) *International Handbook of Mathematics Education*. Dordrecht, Kluwer: 289–325.

Vinner, S. (1997). The pseudo-conceptual and the pseudo-analytical thought processes in mathematics. *Educational Studies in Mathematics* 34 97–129.

Notes

1 In this context the word choose does not necessarily mean that the students make a conscious and well considered selection between methods, but just as well that they have a subconscious preference for certain types of procedures.

2 Then usually *failed*, *3*, *4*, and *5* instead of *failed*, *passed*, and *passed with distinction*.