Constructing Exams in Mathematics – a Norwegian Experience

Gunnar Gjone, University of Oslo, Norway

Introduction - An overview of traditions

In this paper we will focus on written national exams in mathematics. Norway has written exams at the end of the basic school (grade 10), and there are also written exams in mathematics at the end of every year of upper secondary school (grades 11, 12 and 13)[1]. There are also different versions of the exam for different programs in the academic stream in upper secondary. There are also local exams for the vocational streams in upper secondary. However, we will not go into these here. The system of exams in mathematics in the Norwegian school system is quite extensive. However, little research has been carried out concerning exams in mathematics. What is presented here is mainly based on personal experiences.

Organisation

The written exams are very important elements in the Norwegian educational system. Even if their importance has diminished over the years, they have a large influence on the teaching during the school year. The question: *To which extent does the final exam influence your teaching* was answered in the following way in 1990:

To a large extent	37%
To some extent	49%
To a small extent	9%
Not sure	1%

This question was given to teachers in the last year of basic school. It was at that time relating the exam at the end of year 9. It is likely that in upper secondary school the influence would be even greater.

The purpose of the exams

The written exams have traditionally a high status in the educational system and have several purposes. For the school authorities the exams give feedback on how students perform. The results of the students might also be used to get information on the different schools. Exams also play a role for implementing a curriculum – this was especially the case in the reform of 1994, when graphic calculators were introduced in upper secondary school. It was then communicated that the exams required that students had access to graphic calculators during the exam.

The written exams also have purposes for the teachers. From the exam questions the teacher can see what is valued in the curriculum. Also the teacher has the possibility to see how his grades are comparing to the exam results. Moreover, in the grading instructions provided for the graders, the teachers might learn which elements are taken into accounts by the graders, e.g. that the reasoning leading up to an answer is important for the students to communicate.

[1] With the new curriculum going into effect in the fall of 2006, there will be no written national exam at the end of the first year of upper secondary. The first will be after the second year in upper secondary.

Perhaps the most visible purpose of the exams is for ranking or selecting students for further progress in the educational system. The students will also learn which elements are considered important in the curriculum, and therefore should be in focus of the students' learning. The traditional role of the exams as ranking the students was considered a necessary evil by many teachers, especially at the end of compulsory education.

Earlier a student at the end of lower secondary, and at the end of upper secondary might expect to have written exams in the main subjects. For the basic school this would be Norwegian, Mathematics and English. This is no longer the case. Now the student can expect to have a written exam in only one of the main subjects. This is organised such that the exam is distributed over the main regions. It is only published 2 weeks before the actual exam date which region will have which subject for the written exam. So teachers and students will have to prepare for all alternatives. This is even the case for some of the main subjects in upper secondary. If the student does not have the exam in mathematics, his grade will be the school based grade assigned by the teacher.

A typical exam

Written national exams for a certain grade take place on the same day all over the country, usually in the end of May. A typical written exam lasts for 5 hours, and has from 4 to 6 numbered tasks. A task consists of several questions, usually 4 to 6. The questions might all relate to the same situation or they might be unrelated. Lately one of the tasks has consisted of 5 or more unrelated questions so as to cover more of the different areas of the course. Some of the problems are in a context, others could be more abstract mathematical problems. All questions have the form of constructed response. The first task is usually comparatively easy, and one might expect to find harder problems towards the end of the set. The exam is constructed to cover a reasonably large part of the course.

Types of exam questions
In a study by Kay Piene, (Piene,1961) exam questions are categorized:

A General single questions
1. Definition- and (short) reproduction questions
2. Descriptive or referencing questions
3. Connections and modelling
B Special single questions
1. Reasoning
2. Calculations

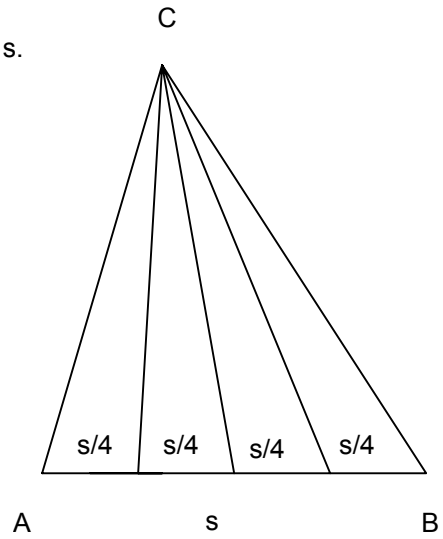
Even if this study is quite old, I will argue that it is still valid. However, we will find that there has been a shift in focus. There have been fewer problems on reasoning in later years. The only additional problem category consists of problems that are more experimental. The student might be expected to draw or experiment in a situation. The following problem – especially the second question – I consider to be an example of this.

Problem 4 (Spring 2002)

a) On the figure Δ*ABC* is an arbitrary triangle with side *AB* = s.
Δ*ABC* is divided into four smaller triangles.
Explain why the four triangles have the same area.

b) Divide an arbitrary triangle in four triangles that all have the same area, in at least two other different ways.

(translated by the author)



Assessment of exams

As soon as the exam has been completed the school will send the papers to the graders. National exams are graded by two independent graders, and it is usually organised in such a way that each of the graders is grading her papers with two different other persons. All the graders in a region will meet on a certain date to discuss and finalize grades for the students. They will start with some initial discussion of problems they have encountered during assessing the exam.

We are for the time being using a 7 point scale (0 – 6-best), 0 and 1 is failing. This system is under continuous revision. Grading is usually performed by the graders giving "points" for the solutions of the problems. Cut-off points between the grades are usually determined by tradition. It also used to be the case that the number of points for problems was stated in the exam papers, this is no longer the case, and the graders have to determine the points themselves. We have also seen a movement towards an assessment focusing on the exam paper in its entirety (more than the separate questions).

There is also a strong view that the students should be presented with the criteria used by the graders. To some extent this is written in the beginning of the exam paper. The following is an example of such statements:

> The grade is determined after an evaluation of your solution in its entirety.
> This means that the grader assess to which degree that you:
> - show basic knowledge
> - show understanding by applying knowledge in new situations
> - carry through logical reasoning
> - can use tools
> - see connections and are inventive
> - can evaluate if an answer is reasonable
> - explain both the process and your solution
> - write up your solution clearly and carefully
>
> (translated by the author)

A wrong answer on one question should not have consequences for solutions of further questions in the same task. In general one is very careful to have solutions which will build on

solutions of earlier parts of the task. It is stated in the guidelines to the graders that a wrong answer should not have consequences for grading solutions to subsequent parts of the task. However, this is only to a certain degree. If the wrong answer makes the subsequent questions "substantially" easier, this should not be the case, i.e. there should be consequences.

Another way is to use the formulation "show that". In that case, the students can use the answer, even if they have not derived it.

Work in the exam commission

Membership in an exam commission is confidential. In a given year the work in the commissions start early in the fall, they are usually given information about the outcome of the corresponding exam given in the spring, as well as collected experiences from a few select graders. A typical commission consists of 4 – 6 people, and there is a wish for gender as well as geographical balance among the members.

The first task is the construction of the problems. After an initial discussion the groups are producing problems. Now follows a period of revisions and discussions. The groups can meet at regular intervals, or more irregular. It is up to the groups to decide. A time limit is also given. The production process is also extensive and the deadline for completing a problems set is usually late in the fall.

The individuals members will come up with original problems or problems inspired by other problems. Some years back it was a duty for all the graders to suggest future exam problems. Since this lead to a large number of often uninteresting problems being submitted, this practise was abandoned. Textbooks and problem collections are checked that they do not contain problems close to the problems suggested by the commission members.

In general problems should be in a context. There are a few "guidelines" for the commissions – some formal and many informal – for construction of the problems: gender balance, problems should not consider physical characteristics of people, e.g. weight, balance with respect to minorities, problematic social issues should also not be the subject of exam problem, e.g. smoking, drinking, accidents etc. No open piloting is carried out, but some of the committee members might try them out in their own classes – if their classes will not have the examination later.

Some principles for exam construction that has been active at some point
The first question (or even the first problem) of the set should be mastered by "everyone". Consider the following first problem from the exam in first year upper secondary in 2001:

Problem 1 (spring 2001)
Siv and Are are going to get married. They will invite 30 guests for dinner. From a restaurant they have received the following suggestions for dinner:

First course		**Main Course**		**Dessert**	
Shrimp salad	40 NOK	Reindeer	180 NOK	Berries with cream	50 NOK
Mushroom soup	50 NOK	Steak	130 NOK	Ice-cream	40 NOK
Smoked salmon	30 NOK	Leg of lamb	150 NOK	Caramel pudding	35 NOK

Choose one menu, and make a clear presentation of the costs of the price of the dinner, if:

- All the guests will have the same first course, main course and dessert
- Drinks are 50 NOK per person
- If the amount is more than 6000 NOK the restaurant will deduct 15%

(translated by the author)

The rationale of such a problem is that the students should be able to start on the set. How are challenges introduced in an exam?

Challenges

In my opinion it is correct to state that in recent years the makers of exams have been more concerned with constructing problems for all, than problems for the few. We find few tasks that we might label as challenges in exams in later years. The tradition in Norway was that a set should contain at least one very difficult problem or challenge. To solve this problem was sometimes seen as a requirement to get the best grade. However, in recent years it has been acknowledged that the best grade also is an interval and the requirement for having the solution of a very hard problem has been more or less dropped.

Traditionally we also found increasing difficulty of the problems throughout the set. The more recent sets, however, show a variation concerning difficulty over the tasks. An argument for this would be that if challenging problems came towards the end, many students would also be pressed for time – since the natural approach for many students at an exam is to start at the beginning. A challenging problem should not be the first problem, but it could be the 3^{rd}, 4^{th} or 5^{th}. There is no advice given to the groups that they should avoid challenging problems for the exam.

The authorities closely follow how the distribution of the grades develops over the years. There is a certain tendency to compare the grades from one year to the next, and even if we no longer have a distribution of grades following certain percentages it is my clear impression that the grades should not deviate substantially from one year to the next. This is handled in the following way: After having graded the first half of their exam papers, the graders report to the Directorate for Education and Training. The directorate make up statistics based on the grades submitted, makes a distribution of grades and then might give advice to the graders. This might be like the following: If you are on the borderline between two grades, you should not “go up”.

Another important element is that hard problems/challenges are not marked in the test. At some university exams in mathematics in the 1960s some problems had the information “can be skipped” meaning that if you wanted a good mark you should definitely try, if to pass the exam was your main concern you should skip this. This way of marking problems has not been used in school mathematics courses in recent years.

Some of the organisational attempts that have been made to renew the exams might allow for more challenging problems. At the basic school there has been given 3 parts exams. In one part the students get some information as to which topic will be covered in the task. They can work and try to formulate questions, which they expect may appear on the exam. Hence they will be better prepared and hence there could be more challenging questions. There are also

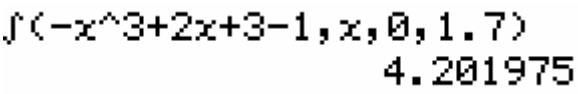
developed booklets that the students can use in exams. These booklets contain basic formulas and definition. This type of material, however, has not led to more challenging problems.

Challenging problems and the use of technology
With the curriculum revision in 1994 graphic calculators were introduced as "almost compulsory". The message given was that the exam would presuppose that the students had a graphic calculator available at the whole or parts of the exam. The principle introduced was that the calculator could be used in the solutions of all the problems. Hence this marked a thorough revision of exam problems. In upper secondary school this has been the natural situation up to now.

When students have the possibility to use technology during the exam, the situation changes concerning what will be a challenging problem. The use of technology allows the students to use a variety of methods. Consider the following problem in the exam for the first year in upper secondary school:

> **Problem 4** (spring 2006)
> Let the function f be given by: $f(x) = -x^3 + 2x + 3$
> a) Use the calculator, and sketch the graph of f. Use x-values in the interval [-2, 3].
> b) Find the coordinates of the maximal and minimal points of f.
> c) For which value of x does the graph intersect the line $y = 1$?
> d) Find the area bounded by the y-axis, the graph of f and the line $y = 1$.
> (translated by the author)

This is a fairly standard problem from a first year upper secondary exam. The students at this level did not really know integration. They knew the concept, but not methods of integration. A majority of the students I graded used the integration function available on the graphing calculator. Their answer looked like the following:



This problem would have been challenging without the calculator. If we allowed using approximations to find the solution of the cubic $-x^3 + 2x + 3 = 1$ students would be able to find an approximate solution, but they would have to put some work into the process. With technology, the problem is almost trivial.

Several students were trying to solve the problem by approaching the area with rectangles. This was clearly seen as a challenge. The majority of the students however, solved the problem by finding the integral on their calculator as indicated above. This problem shows that the status of the problem as originally thought of as being a challenge has a changed status with the use of technology. This also points to another problem for the commission: The graphic calculator is a strong mathematical tool, and it seems that the commissions might have difficulties constructing challenging problems. The question they might ask themselves is if there exists a simple solution with any use of the technology.

Trends

It has been a quite visible trend in Norwegian that the questions should "approach" a

realistic or authentic situation. As an example of this trend all information given in a problem should be presented in the beginning of the problem formulation and not stepwise just before each question. The context should "approach" a realistic situation.

More experimental mathematics has also appeared in the exam problems. This I see as a consequence of the use of technology. With a graphic calculator the students have the possibility to perform a fairly large amount of calculations, looking for patterns and connections. However, there have not been challenging problems of this type in my view

The case of CAS[2]

The discussion in Norway these days concerns the use of CAS for the written examinations. At this time there seems to be a solution of a two-part exam: one part without any tools or written material, and one part where such resources are allowed. It would be restrictions on communications, either with other persons or using the internet. The question we might ask if the new situation will lead to more use of challenging problems.

One might argue that with CAS students will not have to use much time on basic skill problems, hence there would be more time to work on challenging problems. But there might be a danger that we will just get a new type of skill problems.

In my opinion the introduction of graphic calculators did not lead to more challenging problems. In the Norwegian curriculum *regression* was introduced when students could use graphic calculators, but this has not lead to more challenging mathematics – only a different kind of skill problems.

Conclusion

As mentioned I had difficulty finding problems representing challenges in the Norwegian exams. A trend has been that the exam problems are testing skills. To the question why the situation is as it is, I do not think there is a simple answer. But a possible explanation might be that the Norwegian educational system is not very competitive, and many feel that the right place for challenging mathematics would be outside of schools.

It might be necessary in our system to develop new forms of exams to be able to have challenging problems. We could introduce exams for groups or exams with more preparation time to mention a few.

Literature

Gjone, G (1993) Types of Problems and how Students in Norway Solve them. In Niss, M. (ed.) *Cases of Assessment in Mathematics Education. An ICMI Study.*Dordrecht: Kluwer Academic Publishers.

Piene, K. (1961) *Eksamenskarakterer og forhåndskarakterer.* Oslo: J.W. Cappelens forlag. (in Norwegian)

[2] CAS – Computer Algebra System