Changes in grades on the Norwegian lower secondary school mathematics exam

Aina Fossum[1] and Bjørn Smestad[2]

[1,2] Oslo Metropolitan University, Norway; Aina.Fossum@oslomet.no; Bjorn.Smestad@oslomet.no

Exams based on Norway's 2006 math curriculum were given every spring from 2009 to 2019, and the distribution of grades varied widely. Based on document analyses of the mathematics exams, we identify some traits of the exams' form that may contribute to these variations in grades. The answer formats and weighting of the two parts of the exam seem to be factors that should be taken into account in an in-depth analysis of the exams.

Keywords: Mathematics exam, assessment, language.

Background

The Norwegian Curriculum for Knowledge Promotion was implemented from 2006, and the first corresponding lower secondary school mathematics written exam was given in 2009.[3] Since then, there have been changes in the distribution of grades, with a remarkable decline in the lowest grades and an increase in the highest grades in the last four years (Figure 1).

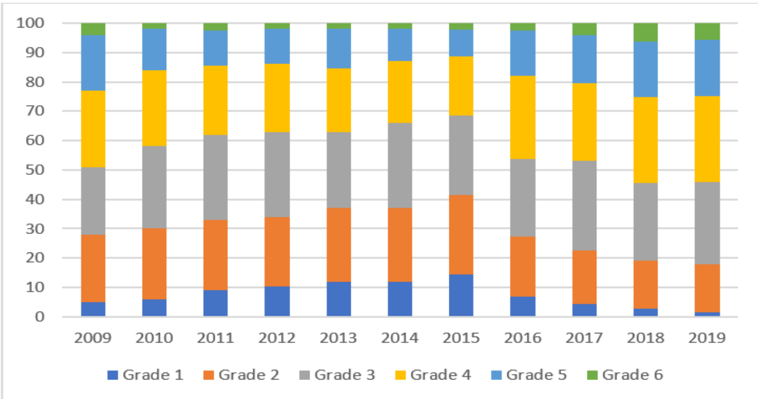

Figure 1. Distribution of grades for examinations in Year 10[4]

The proportion of examinees who were awarded the three lowest grades (1, 2, or 3) increased from about half to over 60 percent during the period from 2009 to 2015, while there has been an increase in the three highest grades (4, 5 or 6) since 2016. The proportion who achieves these grades is now approaching 60 percent, while the lowest grade has hardly been used in the last two years (Figure 1).

2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
3.4	3.2	3.1	3.1	3.1	3.0	2.9	3.3	3.4	3.6	3.6

Table 1. Average grade for the mathematics exam in Year 10[5]

[3] Although the exam in 2008 was based on this curriculum, it had the same format as exams based on the previous curriculum. It is therefore not included in our analysis. Due to the Covid-19 pandemic, the last ordinary exam under this curriculum was administered in 2019. All exams are available at https://matematikk.net/side/Eksamensoppgaver. English version of the mathematics curriculum: https://www.udir.no/kl06/MAT1-04?lplang=http://data.udir.no/kl06/eng

[4] Results for the period 2015-2019 were retrieved from www.skoleporten.udir.no, while the results for the period 2009-2014 were taken from graphs presented in articles from the Directorate for Education and Training.

[5] Source: https://www.ssb.no/statbank/table/07498/

As Table 1 shows, the average grade has changed throughout the curriculum period. An increase of one tenth means that, on average, every tenth student has moved up one grade, so both the decrease from 2009 to 2015 and the later increase must be characterized as noteworthy changes.

Exams can have more roles than those formally enshrined in legislation. They certify competence and rank students, and they are used by different groups for a variety of purposes (Newton, 2007). Exam results are published at the school level and influence the standing of the school in the community. They are also used in aggregate form as the basis for political decisions about mathematics teaching (Nortvedt & Buchholtz, 2018). In addition, we know that most teachers use previous exams in their teaching (Andresen et al., 2017). The tests represent the ceiling, in terms of what is expected in most classrooms (Burkhardt & Schoenfeld, 2018), so exams in mathematics may be of great importance to everyday teaching.

Changes in results may be due to an increase in students' competence, in the exams' content, or in the formats of the exams. It is therefore of interest to investigate the changes in exams and grades more closely.

Theory and previous research

The curriculum used in Norway during the period addressed in this paper was competency-based. Among other things, competency frameworks should demonstrate that learning mathematics comprises more than just acquiring an array of facts, and that mathematics involves more than merely carrying out well-rehearsed procedures (Kilpatrick, 2014). To assess the total competency required by the exam sets, a comprehensive analysis based on a framework for mathematical competence is required. Turner et al. (2015) give examples of how this can be done based on the PISA Mathematics Frameworks, and they describe theoretical and practical issues relating to these analyses.

The formulation of tasks can also affect their difficulty. Andresen et al. (2017) include a summary of international research on language traits that make math tasks difficult to understand and show that these traits interact in complex ways. Such traits include many words, many low-frequency words, passive voice, and long noun clauses. Illustrations can sometimes be helpful but might also make the text more challenging. In their article investigating changes in high-stakes mathematics examinations, Morgan and Sfard (2016) developed an analytic framework with lexico-grammatical aspects and visual mediators as a central part. These are much in line with the literature review presented by Andresen et al. (2017).

Sangwin and Jones (2017) refer to research on answer formats that report impact both on students' results and on possible discrepancies between what the examiners intend to assess and what is actually assessed. They investigated the impact of answer formats for reversible mathematical tasks (e.g., verifying a solution, versus solving an equation) and found that students performed better on multiple-choice items than on constructed-response items, and that this effect was larger than the effect of guessing. How the mathematics exam is perceived by the students is also dependent on the knowledge they have acquired from their teaching and learning materials, such as textbooks. Lithner (2008) developed a framework to determine whether creative mathematically founded reasoning is necessary, or whether imitative reasoning is sufficient for solving a task. To be able to use imitative reasoning, the task must be familiar to the student to some extent. Non-identical tasks regarded by a mathematically well-versed person as having ”the same mathematical content” may be seen by the student as anything but equivalent (Morgan & Sfard, 2016, p. 89).

The distribution of tasks over the main mathematical domains in the curriculum may contribute to the total difficulty of the exams. In the PISA tests, there are four content categories with an even score distribution, and items in each category have a range of difficulty and mathematical demand (OECD, 2019). Exams in Norway shall give the student the opportunity to show competence in as much of the subject as possible (Forskrift til opplæringslova, 2006, §3–22), but there are no guidelines for the distribution of tasks over the mathematical domains like those found in PISA tests.

The full Norwegian examination sets are published on the day of the examination. While the PISA tests are composed of clusters of trend items in mathematics that have been kept confidential (OECD, 2019), a similar approach is not possible for the Norwegian national mathematics exams.

Despite this, many aspects of the exams can be compared from year to year, and the significance of the exam makes it interesting to investigate these changes more closely. To paint a full picture of these changes, a document analysis based on one or more of the frameworks mentioned above is needed (Lithner, 2008; Morgan & Sfard, 2016; Turner et al., 2015). Such analyses are extremely time-consuming, so, in a preliminary phase, we found it useful to carry out some simple document analyses of selected traits in *the form* of the exams that could contribute to explaining the changes in grades. We do not intend to draw conclusions about the extent to which each of the traits contributes to the changes in grades. Rather, we use the variations in grades to look at and discuss factors that should be taken into account in such analyses.

Our research question is: Which traits in the form of the exams are possible explanations for the changes in grades on the mathematics exam given in Norwegian lower secondary schools?

Method

Both mathematical and non-mathematical aspects may contribute to an exam’s degree of difficulty. If these aspects vary, they may be possible explanations for variations in grades. Language traits and illustrations can shed light on how easy the task is to understand, while variations in the main areas covered may tell us something about how predictable the exam is. The answer formats provide information about the degree to which mainly procedures and specific knowledge are tested. The mathematics exam given in Norwegian compulsory schools consists of two parts. Part 1 must be completed in two hours using no tools other than pen and paper, and Part 2 must be handed in after another three hours. In Part 2, students are free to use whatever tools they wish, except communication. Adopting the framework developed by Lithner (2008), an analysis of the exams from 2017–2019 showed that creative mathematically founded reasoning was necessary to solve more tasks in Part 2 than in Part 1 of these exams (Bjørnset et al., 2020), so the relative weighting of Parts 1 and 2 could therefore be of interest. These aspects are investigated through a document analysis of all 11 ordinary Year 10 exam sets under the current curriculum.[6]

In the exams, tasks are numbered and sub-tasks are labeled a, b, c, and so on. Although multiple answers are required in some sub-tasks, we still count them as a single answer. For each main task, the maximum score that can be awarded is stated on the exam paper. The pre-grading report shows

[6] The exams given in fall 2013 and in spring 2020 and 2021 have not been included, since they were only distributed to external candidates.

how the points are distributed among the sub-tasks. We had access to these reports for the period 2014–2019; for the period 2009–2013, we distributed the points for the sub-tasks based on experience from recent years. Other researchers might distribute the points differently, but this would only affect the analyses to a small extent. Our analyses are based on the percentage of the total score achievable for the exam.

Since different language traits interact in complex ways, looking at only a few traits will not give a complete picture of how language affects the exam's degree of difficulty. For the purposes of this article, we have chosen some aspects as examples. The amount of text in the tasks contributes to their difficulty, especially for weak readers; therefore, the proportion of sub-tasks that has little text is of interest. We have identified sub-tasks with no more than 10 words, including the introduction to the task, if there is one. Of these sub-tasks, some come late in a main task and after sub-tasks or introductions containing more than 10 words. We call the sub-tasks that have no more than 10 words and which do not come after tasks/introductions with more than 10 words ”real short tasks.” Tables are included when analyzing illustrations. We distinguish between illustrations crucial to solving the task and those that are not.

Each sub-task is compared to competency objectives in the curriculum to determine to which main area it belongs. If a sub-task covers competency objectives from several main areas, the points are distributed among them. The tasks are also divided into different answer formats: multiple-choice, answer only, open, drawing, spreadsheet, and digital graphing tool. True/false tasks are categorized as ”answer only,” as is filling in certain cells in a pre-arranged table. We categorize lines of symmetry, finding the vanishing point, drawing a graph without digital tools, and coloring parts of a figure under ”drawing.”

Results

There have been major changes in the answer formats for the exams which have affected what is required regarding showing calculations, mathematical reasoning, and communication. Figure 2 shows that the proportion of multiple-choice tasks in the exams has increased markedly in the last three years, at the expense of tasks where the students have to justify their answers (open tasks). The changes in answer formats and grading can be exemplified by how linear equations are tested in Part 1. Typically, there have been two equations that must be solved by hand, with the most difficult one counting twice as much as the simplest one. This was changed in 2017, so the two equations were given equal scores, and then changed again in 2018, when the simplest one was presented as a multiple-choice task. The same pattern is seen for testing students' ability to manipulate algebraic expressions.

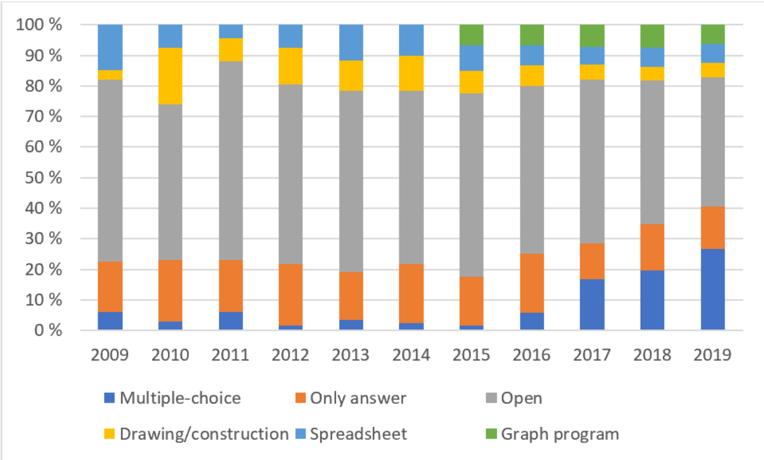

Figure 2. Proportion of different answer formats in exams, 2006–2019

Until 2016, the proportion of the total time allocated to Part 1 (two-fifths), was reflected in the scoring of the two parts, with 40 percent of the points being awarded for Part 1. From 2017, the weighting shifted, with more weight being given to Part 1; in the 2019 exam, 48 percent of the points were allocated to this part. The limits for awarding individual grades remained stable during the same period.

Figure 3 shows that the proportion of ”real short” sub-tasks was stable, at about 20 percent, from 2010 to 2017, but has decreased in recent years. In 2019, students were presented only six ”real short tasks.” That represents 9 percent of the points, which is a pronounced decrease from 2017.

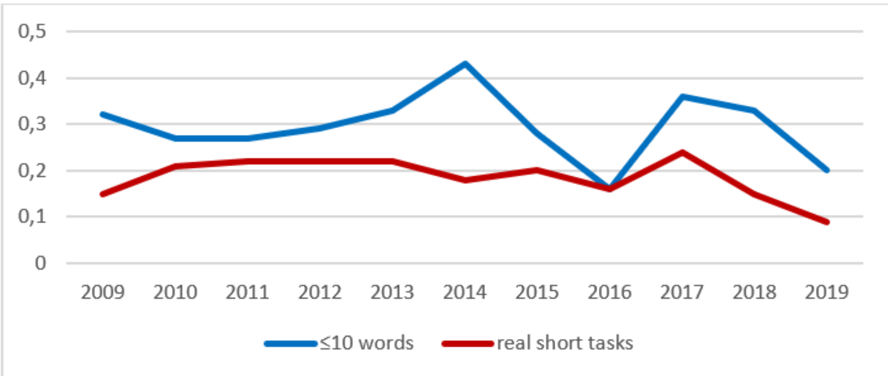

Figure 3. Proportion of sub-tasks with few words and ”real short tasks” in exams, 2009–2019

Figure 4 shows the development of the use of illustrations over the period. The numbers have varied, with a decrease in the number of illustrations from 2009 to 2015, and then a jump in 2016. The number of illustrations that are not crucial for solving the problem has also increased in recent years.

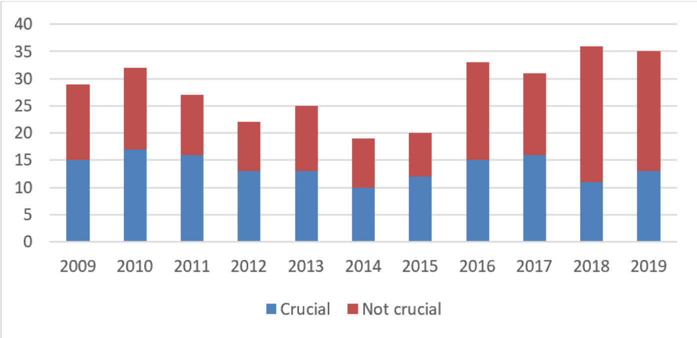

Figure 4. Distribution of illustrations deemed crucial and not crucial for solving the task

The curriculum is divided into five main areas, and Figure 5 shows that these are tested to varying degrees in the exam. There has been less variation in recent years.

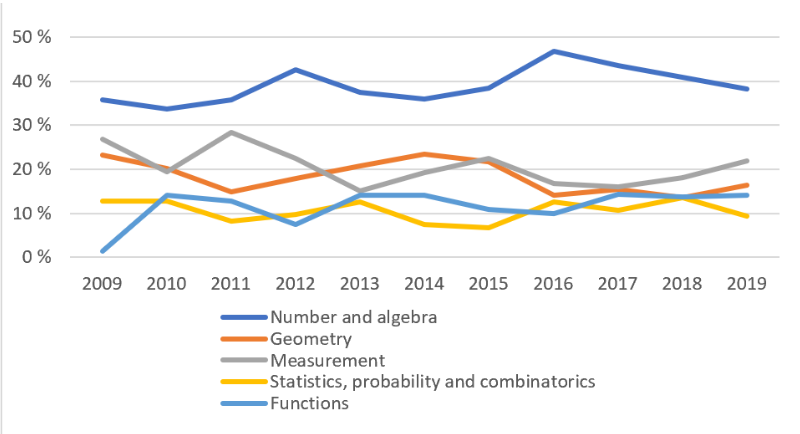

Figure 5. Proportion of tasks testing different areas, 2009–2019

Analysis and discussion

Below we discuss some of the changes presented in Figure 1 and Table 1. Many of the aspects addressed in our analyses of the mathematics exams vary considerably, which may therefore contribute to hypotheses about the reasons for the changes in grades during the period. In our materials, we find two distinct changes in the exam that could explain the decrease in the average grade from 2009 to 2010. Tasks that required drawing and construction were at their lowest in 2009 and highest in 2010 (see Figure 2). From Figure 5, we note that, while the main area ”Functions” was hardly tested in 2009, tasks in this area accounted for 14 percent of the exam in 2010. These changes might have resulted in more tasks being unfamiliar to the students, since teachers may have adjusted their teaching based on previous exams (Andresen et al., 2017; Burkhardt & Schoenfeldt, 2018).

The variations in the main areas tested from year to year was greater until 2016 than it has been in recent years. This might help to explain the further decrease in the average grade beginning in 2010 and the increase in grades since 2016. Less variation in the main areas tested may result in a more predictable mathematics exam and might lead to a narrowing of the curriculum (Gurskey, 1994), and to the exam being perceived as easier. Based on the average score on each item in the exams for the period 2017–2019, there seems to be a range of difficulty within each of the main areas in the curriculum (Andresen et al., 2017; Bjørnset et al., 2020). A more thorough analysis of the content and complexity of the tasks in each mathematical domain will be necessary to determine to what extent the main areas tested affect the overall difficulty of the exam.

The lowest average grades are associated with the 2014 and 2015 exams (see Table 1). In both of these years, we find a minimum number of illustrations, and in 2014 there was a larger proportion of tasks that used fewer words than in any other year. These are both features that could point in the direction of *less* demanding exams. Considering the other traits, we find that, in the 2015 exam, there are fewer multiple-choice tasks and tasks that only require an answer with no justification. This is also the first exam with the mandatory use of a digital graphing tool (see Figure 2). The years 2014 and 2015 were also special for having less focus on statistics, probability, and combinatorics, and placing more emphasis on geometry than the rest of the period.

In the last three years, the average grades have increased, and the lowest grade was hardly used in 2019. In the same period, we can observe a change of traditional tasks from open to multiple-choice ones, and a change in scores, with simple and more demanding tasks given the same score. Some of these items are of the reversible type, and the answer format might influence the results (Sangwin & Jones, 2017). Those who are awarded the lowest grades for the exams generally earn their points in Part 1, with a large proportion of their scores coming from multiple-choice tasks and those that require an answer with no justification (Bjørnset et al., 2020). As more tasks are converted into a multiple-choice format and assigned heavier weights, compared to those that require calculations or some other form of justification of the answers (as exemplified with equations and algebraic expressions), this will benefit these students at the exam. The decrease in the number of tasks that are short in words points in the opposite direction. The same applies to the increase in illustrations that are not crucial for solving the tasks, since this, arguably, increases the competency required of students, since they have to give more consideration to whether or not the provided illustrations are necessary.

Conclusion

Our analysis of different aspects of these exams paints a diverse picture. Some aspects are likely to impact some groups of students more than others. The analysis of language and illustrations points toward a more difficult exam for weak Norwegian readers in recent years, while the other traits analyzed here mostly points towards a decreased level of difficulty for all students. Thus, it is possible that the exam has become less equitable by being more difficult for the group of weak Norwegian readers, while other aspects camouflage this. It is not possible to draw explicit conclusions about the extent to which each of the traits we have included here contributes to the changes in grades in the Norwegian lower secondary school mathematics exam; however, we argue that these traits concerning the exam items' form are relevant.

Nortvedt and Buchholtz (2018) refer to research finding that policy-makers tend to focus on overall results, rather than on utilizing the detailed information that assessments offer. If the changes in grades in recent years are not the result of increased mathematical competence among students but are due to changes in the form of the exams, the results may mislead policy-makers. The form of the exam in Norway is in a revision process at the time of writing and our findings might be considered relevant in this process.

More research is needed to explain the changes in grades on the Norwegian lower secondary school mathematics exam. A comparative analysis of mathematical competency required by the exams, and of the mathematical content and distribution of difficulty between and within the mathematical content domains represented in the exams, is needed. These will provide valuable information concerning whether or not there have been changes in the degree of difficulty. In addition, there is a need for more research on the variations in the different aspects we have pointed out and how they work together. The effects of answer formats and the weighting of tasks and parts of the exams should also be included in an extensive analysis of the Norwegian lower secondary school mathematics exam.

References

Andresen, S., Fossum, A., Rogstad, J. & Smestad, B. (2017). *På prøve. Evaluering av matematikkeksamen på 10. trinn våren 2017 [To the test. Evaluation of the tenth-grade mathematics examination in the spring of 2017].* (Fafo-rapport 36/2017). Fafo.

Bjørnset, M., Fossum, A., Rogstad, J. & Smestad, B. (2020). *På like vilkår? Evaluering av matematikkeksamen på 10. trinn våren 2017–2019 [Evaluation of the examination in mathematics for tenth-grade, 2017–2019].* (Fafo-rapport 01/2020). Fafo.

Burkhardt, H. & Schoenfeld, A. (2018). Assessment in the service of learning: challenges and opportunities or Plus ça Change, Plus c'est la même Chose. *ZDM Mathematics Education. 50,* 571–585.

Forskrift til opplæringslova. (2020). *Forskrift til opplæringslova* (FOR-2006-06-23-724) [Regulation for the Education Act]. Lovdata. https://lovdata.no/dokument/SF/forskrift/2006-06-23-724

Guskey, T. R. (1994). What you assess may not be what you get. *Educational Leadership, 51*(6), 51–54.

Kilpatrick, J. (2014). Competency frameworks in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 85–87). Springer.

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics, 67*(3), 255–276.

Morgan, C. & Sfard, A. (2016). Investigating changes in high-stakes mathematics examinations: a discursive approach. *Research in Mathematics Education, 18*(2), 92–119.

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education, 14*(2), 149–170.

Nortvedt, G. A. & Buchholtz, N. (2018). Assessment in mathematics education: responding to issues regarding methodology, policy, and equity. *ZDM Mathematics Education. 50,* 555–570.

OECD (2019), PISA 2018 Mathematics framework. In *PISA 2018 Assessment and Analytical Framework* (pp. 73–95). OECD Publishing

Sangwin, C. J. & Jones, I. (2017). Asymmetry in student achievement on multiple-choice and constructed-response items in reversible mathematics processes. *Educational Studies in Mathematics 94,* 205–222.

Turner, R., Blum, W. & Niss, M. (2015). Using competencies to explain mathematical item demand: A work in progress. In K. Stacey & R. Turner (Eds.), *Assessing Mathematical Literacy* (pp. 85–115). Springer International Publishing.