Teacher's Mathematical Task Selection

KRISTINE MOLDESTAD

SUPERVISOR
Cengiz Alacaci

University of Agder, 2022
Faculty of Teacher's Education
Department of Mathematics

Preface

I had one fear in the beginning of my work with this thesis, and that was that I would not get any teachers to meet me for interviews. Luckily, I got into contact with three lovely teachers that would help me out. I am so grateful that you would help me out. This thesis would not be possible without you. Thank you so much!

Another person that I need to show my gratefulness to is my supervisor, Cengiz Alacaci. Thank you for all the help and advise you have given me throughout this process of writing a master's thesis. It has been a pleasure to work with you.

Working on this thesis have given me clearer knowledge on a topic I until now have found to be quite abstract. I feel my work have been relevant for my future career as a mathematics teacher and I am looking forward to use what I have learnt in my own practices.

Kristiansand, May 2022

Kristine Moldestad

Abstract

The purpose of this thesis is to study teachers' task selection to answer the research question "*What considerations does mathematics teachers have for a "good" mathematical task*?" The framework of this thesis has included Documentational Approach to Didactics, Constructivism and Personal Constructs Psychology, in addition to some theory on what a good task is. For this the theory on cognitive demands in tasks, and the zone of proximal development have been used. To gather and analyse data, I have used Repertory Grid Technique. To follow this technique, I started with making sets of tasks. Three voluntary teachers then compared the tasks and offered constructs. These constructs were used as a base for a Likert-scale the teachers were to fill out. Then I used a qualitative content analysis on the ratings the teachers gave to look for perceptions the teachers had for "good" tasks. In the discussion I compare the teachers' perceptions. What I found was that different teachers have various considerations for what a "good" task is. Some of their considerations are similar, but some of them are not. Evaluating tasks appears to be a complex process for teachers where many features need to be considered.

Table of content

List of tables

1 Introduction

Most of what students learn of mathematics is limited by what they do in the classroom. *Trends in Mathematics and Science Study* (TIMSS), a survey of international extend, found in 2003 that 80% of the time in a mathematics classroom was spent doing tasks (Alacaci, 2022, p. 2).

One can describe a mathematical task as a problem, or several problems, that all are meant to focus on a particular mathematical idea (Nyman, 2016, p. 1510). A task includes a question that is meant to be answered, and it may incorporate a process to help the students with their work (Alacaci, 2022).

Since the definition of a task includes the concept of problems, the definition of a problem is needed. However, the definition of a problem varies among mathematics educators. Some distinguish between *routine* and *non-routine* problems, whereas others only refer to a problem as non-routine problems and call routine problems for "exercises" (DeGraaf, 2015, p.10). In this thesis the word "problem" will refer to both routine and non-routine problems.

A routine problem is a problem the student can solve in a recipe like way, where the student's own reflective thinking does not need to be activated in any large degree. This type of problem is usually solved fast. For a problem to be non-routine, it needs to be a problem where the student solving it is actively seeking a solution. This solution does not follow any algorithmic approach the student has prior knowledge about, so the student must investigate him/herself in the problem to be able to find a solution (DeGraaf, 2015).

Since the student's investigation and prior knowledge is a main factor in making a problem non-routine, separating the routine and non-routine problems is an individual affaire (DeGraaf, 2015, p.10-11). A problem can be a non-routine problem for some, but merely a routine problem for another. When a teacher selects a task for a class, with the intent of the task being non-routine, he/she should evaluate the problems in the tasks relative to the class in which it should be used to affirm if it is a non-routine task.

The Norwegian curriculum was recently renewed and started taking over the old curricula in 2020. A new feature of this curriculum is a section called *core elements*. The core elements describe the features the students must learn to excel in the subject. They exist to give students a feeling of continuity in the education (Utdanningsdirektoratet, 2019). All education in class should be linked to the core elements.

In the new Norwegian curricula for 1T, the first core element is *exploration and problem solving*. The word *explore* has also made its presence quite frequently in the learning goals of the course (Utdanningsdirektoratet 2020), as opposed to not being mentioned at all in the previous Norwegian curricular for 1T (Utdanningsdirektoratet 2006). The focus of the course has notably shifted to a more explorative approach of learning and inquiry-based learning is encouraged.

Exploration in mathematics 1T is about finding patterns and connections and participate in discussions to get a common understanding of mathematics. When exploring it is the strategies used to solve a task that is in focus, not the solution itself. Problem solving is about students developing their own methods to solving unknown types of problems (Utdanningsdirektoratet, 2020). Considering this new explorative way of teaching mathematics, tasks with non-routine problems are much encouraged to include in class since these are tasks that get students to investigate and explore mathematical ideas.

Since so much time of classroom education is spent doing tasks, and most of the mathematics a student learn is from what's being done in class, it should be obvious that the tasks students work with should be well suited to give the students the best possibility of developing their mathematical knowledge. The goal of this study is to investigate how teachers evaluate and compare tasks and get a glance into how teachers think when evaluating if they think a task is "good." The research question for this master's thesis is "*What considerations does mathematics teachers have for a "good" mathematical task*?"

I think task selection is an important topic to study because tasks are such a big part of mathematics education. Knowledge on how to best select suitable tasks for your students is helpful to get the most out of the mathematics education. By studying teachers' task selection, I hope, as a future teacher, to gain knowledge on how to evaluate potential tasks I want to use in the classroom to give my students the best base possible to learn mathematics.

2 Theoretical framework

2.1 Documentational Approach to Didactics

Today, resources are widely available, much thanks to the internet. One internet search can give a teacher a seemingly endless number of resources to use in class. But how does the teacher verify if the resources they find are well suited for their class? The Documentational Approach to Didactics (DAD) is an approach that studies this and looks at how teachers interact and work with resources. This approach has been the framework of this thesis.

The Documentational Approach to Didactics is an approach that focuses on the way mathematics teachers interact with the curriculum they are bound by and other resources they come across that can be use in class. The resources used for education, either by the teacher or by the students, are *mathematics curriculum resources*. These resources include both the traditional texts and the interactive digital texts. When a teacher is selecting, modifying, and creating resources for class we refer to her/his work as *teacher documentation work*. The result of this work is *teacher documentation* (Trouche, Gueudet & Pepin, 2020, p.237).

However, just looking into how the teacher chooses resources is not enough to include in the Documentational Approach to Didactics. In addition to the resource itself, the teacher also needs to have a *scheme of usage* for the resource. That is, how does the teacher plan to use the resources. The intent of a resource may vary, so a scheme of usage will be different between teachers, even if they use the same resource. The combined process of developing a resource and a scheme of usage on this resource is called *documentational genesis* and results in what is called a document (Trouche. et al., 2020, p. 239).

There are two processes to consider when looking at interaction between a teacher and a resource. The first process is the *instrumentation* process. Instrumentation describes how the affordances of a resource will have an impact on how the teacher uses this resource. The limits of what it is possible to do with a resource will naturally limit what a teacher can do with that resource. However, different minds interact with resources differently. Some teachers may see possibilities in a task that others do not see, and this can result in the teachers presenting and using the task differently from one another. This process of teacher-resource interaction is called *instrumentalization*. (Trouche et al., 2020, p. 239). Since teacher and resource have an impact on each other, a resource that works for one teacher in his/her teaching, might not work for another.

2.2 A good mathematical task (Task characteristics)

2.2.1 Cognitive demands in tasks

According to Margaret Schwan Smith and Mary Kay Stein (1998, p. 344), for a mathematical task to be a good task, it needs to be a task that has “potential to engage students in high-level thinking.” In this definition, the student’s prerequisite knowledge and skill to solving the task is a factor that needs to be considered for deciding if a task is good or not. So, for a teacher, grade level, age and earlier experiences are factors that needs to be taken into consideration when deciding a task is good or not. If the same task where to be given to students in the first year of elementary school and students of the first year of high school, the task could probably be considered a good task for one group of students, but not the other. This is because these two groups of students bring in different prerequisites to solve the task.

Furthermore, Smith and Stein (1998, p. 345) also point out the importance of looking at the task's cognitive demand to determine if the task is good or not. There are four stages of cognitive demands a task can have as suggested by Smith and Stein (1998): memorisation, procedures without connection, procedures with connections and doing mathematics. The first two levels, memorisation and procedures without connections are what constitute lower level of cognitive demand for tasks. Procedures with connections and doing mathematics are defined as the higher level of cognitive demands for tasks.

Tasks of the lowest level of cognitive demand, *memorization*, are non-ambiguous tasks. This entails tasks where previously learnt facts, rules, definitions, etc., are to be reproduced without necessarily having a conception to the concepts and meanings of the task. The time used to solve these types of tasks is short and no procedures are required (Smith & Stein, 1998, p. 348).

Procedures without connections are also considered lower-level demand tasks. These tasks are to be solved in a given way, either by the question specifically asking for it, or prior instructions gives only one way to solve it, and they are in that manner algorithmically solved. The solver of the task has no connection as to the meanings or concepts to the procedures they are working with, and hence the answers are what is set in focus and not the understanding of the mathematics (Smith & Stein, 1998, p. 348).

Procedures with connections are considered tasks of higher-level of cognitive demands. These tasks are meant to develop the students deeper understanding of mathematical concepts and ideas and therefor the procedures used to solve the task is meant to be in focus. Seeing the connection of various representations does also make these tasks help students develop meanings of the mathematics they are working with. The tasks require some cognitive effort from the person solving the task because procedures cannot be followed without thought. The solver needs to have some conception of the procedures he/she chooses to use to be able to solve the task. Through this, an understanding starts to develop (Smith & Stein, 1998, p. 348).

The highest level of cognitive demand in tasks is *doing mathematics*. Tasks of this cognitive demand require the students to have a good understanding of mathematical concepts, processes, and/or relationships. The task cannot be solved in an algorithmic way. The solver needs to examine the task and needs to access useful earlier experiences to find a way to be able to solve it. One need to reflect upon one's own cognitive processes, and the cognitive effort to solve it is high. As a result of the unpredictability of the task, students may feel some level of anxiety when working with a task like this. (Smith & Stein, 1998, p. 348).

In short, tasks of low level of cognitive demand are related to "know that" and "know how" when solving it, where else tasks of high level of cognitive demand are related to "know why" when solving it (Skott, Skott, Jess & Hans, 2019, p. 227-228).

2.2.2 The zone of proximal development

As mentioned in the previous part, student's prerequisite knowledge and skill to solving the task is a factor that needs to be considered for deciding if a task is good or not. A "good" task is not too easy so that no high-level thinking is included, but it cannot be too hard either because then the students will not be able to solve it. The level of the task needs to be within range of what the students can do.

To describe a person's potential of what it is possible for him/her to learn Vygotsky (1978) describes what he calls *the zone of proximal development*. He describes this zone as "*the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers*". This distance represents the functions in a person that is on its way to be matured, but for the moment has not been fully matured (Vygotsky & Cole, 1978, p. 86).

The actual development level of a person is thus what that person can manage to do on its own. The zone of proximal development implies what this person will learn in the nearest future (Vygotsky & Cole, 1978, p 86). The level of development that lays outside of a person's zone of proximal development is considered too difficult for the person to attain at this moment, even with help. The challenge of a teacher is to find tasks and such that utilize the student's zone of proximal development in a way that challenges the students to actively collaborate with each other and to help them when needed so that they can achieve this potential they have of learning. By considering the students way of learning in the light of the zone of proximal development, it is the potential of learning that is in focus and not the momentarily competence of a person (Imsen, 2018, p. 192-3).

2.3 Constructivism, Personal Construct Psychology and Repertory Grid Technique

Constructivism is a theory based on the thought that knowledge does not exist "in itself", but that it is the product of humans trying to understand the world. Through one's own exploration and experiences one learns and gains knowledge, and through new experiences develops and changes this knowledge. In constructivism, learning is viewed as an active process where one constructs and reconstructs perceptions based of the explorations one does. Work methods in the classroom that follow the constructivist view are often referred to as inquiry-based learning (Imsen, 2018, s.145-146).

The psychologist George Kelly published in 1955 his work on *Personal Construct Psychology*. In this psychology he introduces the proposition that humans construct their own bipolar meanings to explain their experiences. The purpose of creating such constructs is so that people can make predicaments and assumptions on the world and the behaviour of others, and this way be able to organize psychological experiences. Through every event a person experiences, that person gets to test her/his predictions and assumptions and from there alter or create new meanings (Raskin, 2002).

To determine a person's construct system, the Repertory Grid Technique is being used. This technique connects three main components: elements, constructs and links. The elements are objects within the domain of what is being studies. The constructs are the interpretations of the elements from the participants of the research. The links show connections between the elements and the constructs, similarities and differences (Tan & Hunter, 2002, p. 43).

3 Methodology

3.1 Repertory Grid Technique

To gather data, I have used the Repertory Grid Technique. The field of study for this thesis is "teacher's mathematical task selection" so the elements for the repertory grid technique are mathematical tasks. The elements, the tasks, were organized into three sets with each set containing three tasks. In total there were nine tasks. We call each set a triad. All the triads had their own mathematical topic, but the tasks within a triad varied in setup and mathematical power. (See the following section for more information on the construction of the triads)

The construct part of the repertory grid technique came from the interviews. In the interviews the participants were presented the triads, one at a time, and asked to compare the tasks in the triads. This way they came up with constructs about the tasks. These constructs were then used as a base for a Likert-scale the teachers where to fill out. The scale asked that the teacher rate all the nine tasks by the various constructs they had come up with. Each teacher got their own Likert-scale because they said different things. (See part 3.4 for more information on how the constructs came about).

Lastly is the link in the repertory grid. This link is the connection between the tasks and constructs of the teachers. The link was mainly constructed through analysing and comparing the constructs the teachers made with the tasks they had gotten based on a table. The table showed the constructs in the rows and the elements in the columns. Some things that were said during the interview was also a part of creating the link.

3.2 Constructing the triads

For the data collection, 1T teachers were to be presented sets of three tasks, called triads, and were then asked to first read, then analyse and then compare the tasks in each triad. I decided three triads was an ideal number of triads to make them analyse. Two triads might give to little data to give a reliable result, and four or more triads might be too time consuming and the repetitive process of comparing three tasks tiresome for the person being interviewed.

The first step in making the triads was looking, mostly online, for mathematical tasks and sorting which of them seemed fit for a 1T class's mathematical curriculum. After having found some suitable tasks, I divided these tasks into categories where tasks of similar mathematical key concepts or themes were placed together in groups.

Then came the time for choosing which of the groups I should evolve to become the finished triads. This selection was partially made by personal preference, so I chose groups that included tasks that I would consider "fun tasks." However, I also chose groups of tasks that included at least one theoretically good or more advanced task. After having settled on three groups of tasks, with the topics linear functions, geometry (area), and volume, I started "polishing" the groups to make them into triads.

In the "polishing" session I removed and added tasks to the groups, such that each group had three tasks with similar mathematical key concepts, but so that each task in the groups varied in cognitive demand, structure, visual appearance, use of mathematical language and such. The groups had now become the final triads the teachers were to analyse and compare.

3.3 Selecting the participants

To find participants for the data collection, I reached out to some mathematics teachers I knew in the southern region of Norway and asked if some of them wanted to participate in being an interview subject. I also asked if they would ask their 1T colleagues if some of them would like to participate. Three teachers volunteered to join me for interviews. Two of them were 1T teachers from the same school, and the third one was a 1T teacher from another school.

3.4 The interviews

Some days prior to the interviews, the person who was to be interviewed got sent, by e-mail, the triads so that he could look through them before the interview if interested. The teachers had been informed that they were to compare three and three tasks together. By sending the tasks earlier, the teachers got a chance to prepare for the interviews.

During the interview, the teacher I interviewed received one triad at a time. For each triad the teacher was presented, he compared the three tasks with each other and made constructs about the tasks. My role as interviewer was mostly to listen and take notes. If the teacher was out of things to say I asked the question “Is there any other way to compare the tasks?”, or something similar, to make the teacher think thoroughly thorough if he had more to say. When the teacher did not have any more comments to the tasks, we moved on to the next triad and did the same routine there. The interviews were audio recorded. I was done with all three first interviews within the same week.

One week after the first interview I had with a teacher, we met up again to perform the second interview. In this interview the teacher received a form with a Likert-scale to fill out. The assertions of the Likert-scale were based on the constructs the teacher himself had made. Each participant therefor got his own Likert-scale. For this round of interviews, I was present to answer questions the teacher might have about the constructs. Also here, I recorded the audio.

Since two of the teachers worked in the same school, and the interviews were to be rather quiet, I decided it would be OK to have them answer their Likert-scales at the same time in the same room without it impacting the outcome of their rating. The third teacher got a sick child the day we had planned to meet, so it was decided he could answer the Likert-scale on his own and send the form back to me when it was filled out.

3.5 Plan for analysis

For the analysis I will start with a qualitative content analysis. For each teacher I will compare the ratings of the constructs for their most liked and least liked tasks. What I will look for are constructs that have a rating with a noticeable difference between the most and least liked tasks, either by the most liked tasks being rated noticeably higher or lower than the least liked tasks. These constructs will give implications as to what perceptions the various teachers have for “good” tasks. Constructs where there is no noticeable difference between the ratings of the most liked and disliked tasks does not give explicit information on task features that separate what tasks the teacher like the most or the least.

The relevant constructs, those with a noticeable difference in rating between the most and least liked tasks, were then grouped together based off the theme they were about. The themes of the selected constructs were used to deduce the themes and factors that define the teacher's perception of a "good" mathematical task.

An example for a construct that I would take notice off is Teacher A's fifth construct, "The students will not understand this task." The rating for this construct is low for the most liked tasks and high for the most liked tasks. Since the liked tasks all score low on not being understandable it means that the liked tasks are tasks the students are likely to understand. The least liked tasks are rated as being tasks the students will not understand. This variation in the ratings give the implication that a "good" task should be one the students are able to understand. The construct is grouped in the theme complexity of a task.

When done with the qualitative content analysis of the teachers separately, I started comparing the teachers with each other. In the first part of the discussion, I compared the teachers' perception with each other. I looked at similarities and variations in their perceptions and tried to reflect how reasonable the perceptions were and how they align with the curriculum and theory on good tasks.

For the second part of the discussion, I compared the two most frequently liked and disliked tasks from all the teachers combined. I looked at how all three teachers had rated these four tasks and how their ratings varied.

4 Results

4.1 Teacher A

Teacher A has rated his three most and least liked tasks as shown in Table 1.

Table 1: List of Teacher A's most and least liked tasks

	Top 3 (1 is best)	Bottom 3 (1 is worst)
1.	6 – Overlapping squares	8 – Can you can the can?
2.	2 – Stacked dices	5 – Pythagorean Theorem with circles
3.	7 - Popcorn	4 – Area between two circles

The three most and least liked tasks are rated as seen in Table 2. The constructs the teacher contributed with in the interviews are in the left column. The three most liked tasks are the first three tasks in the first row and are placed in the same order as the teacher ranked them. The most liked task is first. The most liked tasks are marked in green. The least liked tasks are the three last tasks in the column, where the least liked task is first. The least liked tasks are marked in yellow. The last three constructs are marked in orange. These are generated constructs all the teachers have gotten and will be looked at separately than the constructs the teachers provided.

The teacher has rated each of the tasks for each construct. The rating is as follows: 1-Do not agree at all, 2-Little agreement, 3-Some agreement, 4-Agree and 5-Very agreed. The constructs that have a noticeable difference in the ratings between the most and least liked tasks are in italic.

Table 2: Rating of Teacher A's most and least liked tasks

	Constructs	**Task 6**	**Task 2**	**Task 7**	**Task 8**	**Task 5**	**Task 4**
1	*This is a demanding task for the students*	4	3	4	5	5	4
2	The task is nice to use as a group task	5	5	5	3	5	5
3	*Most of the students would be able to do this task*	3	4	3	1	2	3
4	The task is explorative	3	4	5	5	4	3
5	*The students will not understand this task*	2	1	3	4	4	4
6	*The text is difficult and needs to be read several times*	1	1	1	4	1	5
7	The task should have a drawing	1	1	1	1	1	5

8	Bad reading abilities could be a hindrance for solving this task	1	1	1	1	1	5
9	*The task is not a task the students are used to*	1	1	4	5	4	3
10	*The students probably need some help to solve the task*	1	1	4	5	5	5
11	*The task is difficult*	3	1	3	5	5	4
12	It is easy to get "follow errors" in this task	5	3	3	3	3	3
13	*The task will take a lot of time to solve in the classroom*	2	2	2	5	5	3
14	*The task consists of difficult mathematics*	1	1	3	5	4	1
15	*This task will be too abstract for the students*	1	1	1	4	4	1
16	The illustration in the task is helpful for the students	5	5	5	5	1	5
17	The task is about a known topic for the students	3	4	4	4	3	3
18	The task is a "standard" task	3	4	4	4	3	3
19	*The task demands a lot of reflection*	3	2	3	5	5	5
20	The task is difficult to integrate into 1T	1	1	1	4	1	1
21	One can use the illustration to "assume" the answer	4	3	5	5	5	1
22	The task has a connection to the real world	3	4	4	5	3	3
23	*It is desired to use more of this type of task*	4	3	5	3	4	4
24	*I would use this task in my 1T education*	5	5	5	2	5	3
25	*Overall, I think this is a good task*	5	5	5	2	3	3

For teacher A, the constructs that give a noticeable difference in the ratings between the most liked and the least liked tasks are constructs 3, 5, 6, 9, 10, 11, 13, 14, 15 and 19. In addition to this, construct 1 also does imply a difference in the ratings, though not as big a difference as the other constructs. These constructs are divided into the topics level of challenge, complexity of the task, and mathematical content.

Level of challenge: From the constructs we are looking at there are 8 constructs classified under the theme level of challenge. These are the constructs 1, 3, 9, 10, 11, 13, 14 and 15, "This is a demanding task for the students," "Most of the students would be able to do this task," "The task is not a task the students are used to," "The students probably need some help to solve the task," "The task is difficult," "The task will take a lot of time to solve in the classroom," "The task consists of difficult mathematics" and "This task will be too abstract for the students."

Construct 1, "This is a demanding task for the students", did not have that big of a difference in the ratings between the liked and disliked tasks. The ratings of construct 1 vary between 3 and 4 for the most liked tasks, mode 4 and between 4 and 5 for the least liked tasks with mode 5. Both the most liked and disliked tasks give high average ratings, but what I deduce from this is that according to Teacher A, a task may be quite demanding for a student and be a "good" task, but if the task is too demanding it ceases to be good.

Construct 3, "Most students would be able to do this task", and Construct 11, "The task is difficult" have a high negative correlation at –0.89. The trend in the ratings of the construct is that the most liked tasks score high on Construct 3 and low on Construct 11, whereas the least liked tasks score low on Construct 3 and high on Construct 11.

The ratings of Construct 11 show that the most difficult of the most liked tasks is Task 6 and 7. They both have a difficulty level of 3 out of 5. This makes the tasks "middle of the road" difficult. For the least liked task, the easiest task is Task 4. This task has a rating of 4 out of 5 for being difficult, which means the task is in fact quite difficult. The most difficult task of the most liked tasks is easier than the easiest task among the least liked tasks. What I deduce from this is that a "good" task can be somewhat difficult, but not too difficult.

In Construct 3, "Most students would be able to do this task," there is some overlapping in the ratings and both the most liked and the disliked tasks have one or two tasks rated 3. However, the rating 3 is the lowest rating given to the most liked tasks and the highest rating for the least liked tasks. 3 is the middle rating so a rating of 3 on this construct implies that about half of the students are able to do the task." The most liked tasks have an average rating of over 3, so it implies an expectancy that half or more of the students will be able to do the tasks. For the least liked tasks, the average rating is 2 so the expectancy is that half or less of the students would be able to do those tasks. From this I deduce that a "good" task, according to Teacher A, can be difficult, but not so difficult that less than half of the students are expected to be able to solve it.

The constructs 9, 10, 14 and 15 all have a high, positive correlation with each other, ranging from [0.59, 0.91] as seen in Table 3. These constructs are "The task is not a task the students are used to," "The students probably need some help to solve the task," "The task consists of difficult mathematics" and "This task will be too abstract for the students." Generally, the liked tasks rates low in all four constructs and the disliked tasks rate high. However, Task 4 and Task 7 are outliers to this generalization.

For the most liked tasks, tasks 6 and 2 rates a 1 on all the four constructs 9, 10, 14 and 15. This rating shows that Teacher A sees the two tasks as tasks the students are used to, that do not consist of difficult mathematics, are not abstract and where help from the teacher most likely is not needed. Task 7 on the other hand gets rated 4 on constructs 9 and 10, and a 3 on construct 14, which is noticeably higher than the ratings of the other two most liked tasks. This makes Task 7 a task the students are not used to, it consists of difficult mathematics and help from the teacher is expected. This difference makes the task an outlier among the most liked tasks.

For the least liked tasks, tasks 8 and 5 rates between 4 and 5 for all four constructs. Task 4 is rated similarly on constructs 9 and 10, but not for constructs 14 and 15. This tells that none of the tasks are of a type the students are used to and help from the teacher is expected. For the constructs 14 and 15, tasks 8 and 5 still rate between 4-5, but Task 4 rates 1. Tasks 8 and 5 are then abstract and consist of difficult mathematics, but not Task 4. This makes Task 4 an outsider for the least liked tasks.

These results imply that Teacher A finds tasks the students are used to, that do not consist of difficult mathematics, are not abstract, and are tasks where the demand for help is low, generally are "good" tasks. Nevertheless, since tasks 7 and 4 are outlier tasks for this generalization, these factors alone are not sufficient to determine if a task is "good" or not. Even though Task 7 does not follow the general trend for the liked tasks, there must be some other factor in it that makes the teacher like it anyway. As for Task 4, it follows the trend of the most liked tasks for being a task that is not abstract and does not consist of difficult mathematics, but since this is one of the least liked tasks it shows that these are not factors that automatically makes a task "good."

Table 3: Correlation table showing the correlation between Teacher A's constructs 9, 10, 14 and 15.

	Construct 9	*Construct 10*	*Construct 14*	*Construct 15*
Construct 9	1			
Construct 10	0,91	1,00		
Construct 14	0,88	0,66	1,00	
Construct 15	0,69	0,59	0,88	1,00

The last construct in the level of challenge category is Construct 13, "The task will take a lot of time to solve in the classroom." All the most liked tasks are rated a 2 on this construct and are therefore tasks that get solved relatively fast. The least liked tasks rate between 3-5, mostly 5, and take considerably longer to solve. According to this trend, a "good" task is one that can be solved relatively fast in the classroom.

Complexity of the task: The second category regards the complexity of the task. The constructs we look at that belong to this category are constructs 5 and 6, "The students will not understand this task" and "The text is difficult and needs to be read several times."

In construct 5, the most liked tasks vary in their rating from 1 to 3 with an average of 2, and all the least liked tasks are rated a 4. From this we get that the liked tasks are tasks most of the students will understand. Regarding Construct 6 and the difficulty of the text, the most liked tasks get rated low on having a difficult text, and the least liked tasks get rated high. However, the second least liked task, Task 5, is an exception to this generalization with its rating 1. The ratings of the language and setup of the tasks implies that for a task to be "good" the text needs to be formulated in a way that does not require several readthroughs, but a good text itself is not what makes the task “good”. In addition to a good text, most of the students must be able to understand the task for it to be “good.”

Mathematical content: The last theme is mathematical content and the only construct in this category is Construct 19, “The task demands a lot of reflection.” The most liked tasks rated between 2 and 3 on this construct. Some reflection is needed for these tasks, but not to any large extent. All the least liked tasks are given the rating 5 and do therefore require a lot of reflection. From this I deduce that Teacher A considers tasks that require some reflections to be “good” tasks, but tasks that require a lot of reflection are not “good” tasks.

The three last constructs: For constructs 24 and 25 “I would use this task in my 1T education” and “Overall, I think this is a good task,” Teacher A have rated all the most liked tasks a 5. This implies that the teacher finds all these tasks as “overall good,” and they are all tasks he would use in his education. The least liked tasks are generally not as highly rated for these two constructs, but Task 5 is an outlier for Construct 24 with its rating of 5. Even though this is one of the tasks the teachers liked the least, it is still a task he would use.

For Construct 23, “It is desired to use more of this type of task,” there is not that big a difference between the most and least liked tasks, and all tasks rate relatively high from 3-5. This implies that all the tasks are types of tasks the teacher thinks should be used more.

Task 8 is the task that gets rated the lowest of all the tasks and is seem like a task the teacher really does not like that much.

Summary of Teacher A’s results: Teacher A ‘s results imply that the themes level of challenge, complexity of the task and mathematical content have an impact on if a task can be considered “good.” The teacher finds that “good” tasks can be relatively demanding and difficult, but not too demanding or difficult. A “good” task can also be solved fast in the classroom. A “good” task should be one the students are used to, that does not consist of difficult mathematics, is not abstract and requires little help from the teacher, but these factors do not automatically make a task good. The same goes for the text in the task. A “good” task has a text that does not require several readthroughs, but this criterion alone is not what makes the task “good.” Most students should be able to do the task for it to be “good” and some reflection should be required from the task, though not too much. All the tasks the teacher likes the most are tasks he finds as “overall good tasks” and are tasks he would use in his education.

4.2 Teacher B

Teacher B has rated his top three most liked and disliked tasks as seen in Table 4. The ratings of this tasks are shown in Table 5.

Table 4: List of Teacher B's three most and least liked tasks

	Top 3 (1 is best)	Bottom 3 (1 is worst)
1.	5 – Pythagorean Theorem with circles	8 – Can you can the can?
2.	6 – Overlapping squares	9 – Drill out this problem
3.	3 – Squares in a frame	7 - Popcorn

Table 5: Rating of Teacher B's most and least liked tasks.

	Constructs	**Task 5**	**Task 6**	**Task 3**	**Task 8**	**Task 9**	**Task 7**
1	*This is a demanding task for the students*	5	2	1	5	1	5
2	The task requires prior knowledge	5	3	1	5	4	5
3	This is a “classical textbook” task	3	4	5	3	5	3
4	*All the students should be able to mean something about the task*	1	3	5	1	3	1
5	The task is increasing in difficulty	1	5	5	1	1	1
6	*The task can be solved regardless of prior knowledge*	1	3	5	1	3	1
7	*The picture helps to find a formula*	5	5	5	2	5	3
8	The students can think in various ways whilst solving the task	4	2	5	3	1	3
9	It is easy to get started on the task	1	3	5	1	5	3
10	It is easy to finish the task	1	3	5	1	5	1
11	The task is complicated	5	3	1	5	1	4
12	The text needs to be read several times	5	4	1	5	2	5
13	*The task has a too high level for a general 1T student*	3	1	1	4	1	3
14	*The task would have needed a helping figure*	3	1	1	5	1	3
15	*It demands a high level to be able to start the task*	5	1	1	5	1	4
16	*It demands a high level to be able to finish the task*	5	1	1	5	1	4

17	*The students are used to this type of task*	3	5	5	1	5	3
18	*The task is difficult*	5	1	1	5	1	4
19	*The task is abstract for the students*	3	1	1	5	1	3
20	*The drawing should have helping lines*	5	1	1	5	1	3
21	One can use the “guess and test” method to progress with the task	1	5	5	3	1	5
22	The students should be able to do this task	1	5	5	1	5	3
23	The task has an intuitive thought	2	1	1	3	1	5
24	The task can be solved by testing it in real life	3	3	5	5	1	5
26	Nothing in the task is known to the students	5	1	1	5	1	3
27	The task is concrete	5	5	5	3	5	5
28	One can use the visual in the task to think	5	5	5	3	5	5
29	The task is manageable	3	5	5	1	5	3
30	It is desired to use more of this type of task	5	5	5	3	5	5
31	I would use this task in my 1T education	5	5	5	1	5	5
32	*Overall, I think this is a good task*	5	5	5	4	4	4

Finding constructs that give a noticeable difference between the most liked and disliked tasks proved to be more of a challenging for Teacher B than for Teacher A. The first thing I noticed was that Task 5 was chosen as the most liked task. From the interview I did not get the impression that the teacher thought highly about Task 5. From looking at the ratings in the Likert-scale, Task 5 also tended to differ in rating from the other two most liked tasks and have a more similar rating to Task 8, the least liked task. I found it a bit odd that a task the teacher had expressed no interest in and that had a very similar rating as the least liked task was rated as the most liked task, so I contacted the teacher to get confirmation that his rankings were correct. The teacher responded that the ranking was correct, and though the rating of the tasks might have seemed odd, Task 5 was the task the teacher felt had the most potential.

After confirming the ranking of the tasks, I went back to the Likert-scale to look for patterns. What I found was that in the constructs where Task 5 was more like the least liked tasks than the other two most liked task, Task 9 also stood out from the other two least liked tasks and was rated more like the most liked task. This makes both task 5 and 9 outlier tasks. If looking at the remaining tasks, there is a clear difference in the rating of the most and least liked tasks for the 12 constructs 1, 4, 6, 7, 13, 14, 15, 16 17, 18, 19 and 20. These constructs are the same constructs as where tasks 5 and 9 mostly are outliers. The 12 constructs are divided into the two categories: level of challenge, and complexity of the task.

Level of challenge: Most of the 12 constructs belong to the theme level of challenge. This are the 9 constructs 1, 4, 6, 13, 15, 16, 17, 18 and 19. “This is a demanding task for the students,” “All the students should be able to mean something about the task,” “The task can be solved regardless of prior knowledge,” "The task has too high a level for a general 1T student,” “It demands a high level to be able to start the task, “It demands a too high a level to be able to finish the task,” “The students are used to this type of task,” “The task is difficult” and “The task is abstract for the students.” The correlation chart (Table 6) show that all these constructs have either a high negative or a high positive correlation that range from [-1, -0.8] and [0.8, 1].

Table 6: Correlation table between Teacher B’s constructs 1, 4, 6, 13, 15, 16, 17, 18 and 19. C is short for construct.

	C 1	*C 4*	*C 6*	*C 13*	*C 15*	*C 16*	*C 17*	*C 18*	*C 19*
C 1	1								
C 4	-0,92	1,00							
C 6	-0,92	1,00	1,00						
C 13	0,95	-0,86	-0,86	1,00					
C15	0,97	-0,88	-0,88	0,97	1,00				
C 16	0,97	-0,88	-0,88	0,97	1,00	1,00			
C 17	-0,88	0,80	0,80	-0,98	-0,92	-0,92	1,00		
C 18	0,97	-0,88	-0,88	0,97	1,00	1,00	-0,92	1,00	
C 19	0,88	-0,80	-0,80	0,98	0,92	0,92	-1,00	0,92	1

For all these constructs, tasks 5 and 7 are outliers. Task 5 is the most liked task, but is rated more like the least liked tasks, and Task 7, one of the least liked tasks, is rated more similarly like the most liked tasks.

If we ignore the two outlier tasks for a moment, the general trend is that for constructs 1, 13, 15, 16, 18 and 19, is that the ratings for the most liked tasks are lower than that of the least liked tasks. Tasks 3 and 6 all rate 1 on all these constructs, with the small exception of Task 6 rating a 2 on being a demanding task. This rating is however still low, so saying that the liked tasks score low on these constructs in not an understatement. These ratings make tasks 3 and 6 task that are not demanding for the students, it is not required a high mathematical level to start or finish the task, and the tasks are not abstract. The least liked tasks are for these same constructs mainly rated between 3-5, though mostly between 4-5. This is a noticeable difference from the most liked tasks and makes the least liked tasks demanding, abstract and in need of a high mathematical level to start and finish.

For constructs 4, 6 and 17, the trend is the opposite. The most liked tasks have high ratings between 3-5, and the least liked tasks have lower ratings of mostly 1. The most liked tasks are from this tasks that the students are used to, that most students should be able to mean something about, and that are solvable regardless of prior knowledge. This is not the case for the least liked tasks.

From the general pattern between the most and least liked tasks for the constructs 1, 4, 6, 13, 15, 16, 17 ,18 and 19 it seems like Teacher B have the perception that a "good" task is a type of task the students are used to, that does not have too high a level to start or finish. "Good" task does not require prior knowledge to be solved, and most of the students should be able to say something about the task. Tasks that are not particularly demanding, difficult or abstract are "good" tasks.

If the outlier tasks, tasks 5 and 9, were included again, one can see that the generalization does not hold in all cases. If we look at Task 5, the most liked task, the Likert-scale show that the task is rated more similarly like the least liked tasks than the most liked ones. In the same manner Task 5 differs from the rest of the most liked tasks, Task 9 differ from the least liked tasks. Because of these outliers, the generalization of the theme level of challenge will be more of a guideline feature to understand what considerations Teacher B have for "good" tasks but not definitive answer. A task can have features like the most liked task and not be liked that much, and vice versa. The theme level of challenge alone is therefore not enough to determine if a task is "good" or not. There must be other factors that also need consideration to decide if a task is "good."

Visuals of the task: Teacher B made four constructs regarding the visuals of the tasks and three of them followed the criteria of having a noticeable difference in the ratings between the most and least liked tasks. These were constructs 7, 14, 20, "The picture helps to find a formula", "The task would have needed a helping figure" and "The drawing should have helping lines." All three constructs are about the visuals of the task.

In Construct 7, "The picture helps to find a formula", all the most liked tasks rate a 5. The least liked tasks vary between 2 and 5. For the other two constructs, "The task would have needed a helping figure" and "The drawing should have helping lines," the teacher has rated several of the tasks with a 1 here. However, when the teacher did this s/he commented that the rating 1 was not given because he did not at all agree to the constructs, but rather that he felt the constructs were not relevant to these tasks because they already had an illustration that the teacher found satisfactory. The general trend is that the most liked tasks already have satisfactory visuals, and the least liked ones does not, but again are tasks 5 and 9 outliers. Task 5, even though the picture already there is helpful to find a formula, should have a more helpful illustration with helping lines. Task 9 has a fine illustration, but it is not a task the teacher favours.

The deduction I make of this is that for Teacher B the visual is an essential part of a task and a "good" task should have a good visual, but a good visual itself is not what make a task "good.". A good visual is one that helps the students to solve the task.

The three last constructs: All the Most liked tasks to Teacher B rates 5 on constructs 30, 31 and 32, "It is desired to use more of this type of task," "I would use this task in my 1T education" and "Overall, I think this is a good task." The least liked tasks mostly score high on these three constructs as well, but the general rating is somewhat lower. The rating for all the least liked tasks for being "overall good" is 4. The teachers here show that the tasks he likes the most are in fact the tasks he also finds to be overall good tasks.

For constructs 30 and 31, Task 8 is an outlier. The rating show that the teacher do think it is a type of task that could be used more of, but it is not a task he would use himself.

Summary of Teacher B's results: The themes that impact Teacher B's considerations for "good" tasks are level of challenge and visuals of the task. For Teacher B, a "good" task usually is one the students are used to that does not require prior knowledge or too high a mathematical level to solve and are not too demanding or difficult. A "good" task should also have a good visual that is helpful. However, there are two outliers for these perceptions, so these factors cannot be looked at individually to determine of a task is "good." The task the teacher liked the most was the task he thought had the most potential, not one that followed most of the general features the teacher thought a "good" task should have. The tasks the teacher have rates the highest on being "overall good" are also the task he like the most.

4.3 Teacher C

Teacher C has rated his most and least liked tasks as shown in Table 7 and the belonging ratings to these tasks is shown in Table 8.

Table 7: List of Teacher C's most and least liked tasks

	Top 3 (1 is best)	Bottom 3 (1 is worst)
1.	6 – Overlapping squares	2 – Stacked dices
2.	7 - Popcorn	9 – Drill out this problem
3.	8 – Can you can the can?	1 – Parallel lines

I did not meet up with Teacher C when he rated the Likert-scale and ranked his top three most and least liked tasks. Instead, I sent him the Likert-scale digitally and said I would be available via mail if any questions were to arise. He filled out the Likert-scale and the ranking list alone and sent it back to me. When I received the filled-out Likert-scale, the teacher also included some explanations for why he had rated the tasks as he had.

> "*The reason that the tower of dices [Task 2] scores relatively low on the Liker-scale is because I feel that there some ambiguity in the text that needs clearance. […]*
>
> *"In the ranking of the tasks I liked least exists the drilling [Task 9] and parallel lines [Task 1]. There is nothing wrong with the tasks, but I feel they get done pretty fast, and have little problem-solving in them. That the dices [Task 2] gets on top is because one task needed to get that place, and I let the ambiguity in the text weigh heavily in this pall-placement. The task can be very good if some important things get clarified.*"

In these comments Teacher C implies that a "good" task should not include ambiguity and it should not be a task that gets solved fast with little problem-solving aspects.

Table 8: Rating of Teacher C's most and least liked tasks

	Construct	**Task 6**	**Task 7**	**Task 8**	**Task 2**	**Task 9**	**Task 1**
1	*The task repeats prior knowledge*	2	2	2	1	4	4
2	*The task test understanding*	4	4	4	2	3	3
3	*The task is solved fast*	2	2	2	1	4	5
4	The task uses good mathematical language	4	4	3	3	4	4
5	The task contains ambiguity	1	1	2	4	1	2
6	I would modify the task if I were to use it in my education	1	1	2	4	1	1
7	*I would give this task as a problem-solving task*	4	4	4	4	1	1
8	I like the construction of this task	4	4	3	2	3	4
9	The task is "too deep water" for the students	1	2	2	2	1	1
10	*This is an "algebra heavy" task*	3	3	3	2	1	1
11	A general 1T student probably does not have enough competence to solve this task	1	2	2	2	1	1
12	The task is insurmountable for the students	1	1	3	2	1	1
13	Measurements in the task are given, but one must look for them	1	1	3	2	1	1
14	The students can test out this task in real life	4	4	2	4	1	3
15	*I would not have to give hints in this task*	2	2	2	2	4	4
16	You can use your guts to start/continue this task	3	2	2	3	3	3
17	It is desired to use more of this type of task	4	4	4	2	2	3

18	I would use this task in my 1T education	4	4	4	3	1	4
19	*Overall, I think this is a good task*	5	5	5	3	4	4

From doing the qualitative analysis of the ratings of the Likert-scale, I have found 6 constructs where there is a noticeable difference in high and low ratings between the most and least liked tasks. These are constructs 1, 2, 3, 7, 10 and 15. The constructs I have divided into three categories: level of challenge, mathematical content and implementation of the task.

In addition to some of the constructs having a noticeable difference in the rating between the most liked and disliked tasks, I also saw that Task 2 was rated noticeably different in constructs 5, 6 and 7 than all the other tasks. These three constructs all consider the complexity of a task. Since Task 2 is such an outlier for the rating of these constructs and the constructs all regard the same topic, I will include an analysis of this topic too.

Level of challenge: From the constructs I am looking at, constructs 1, 3, 15, "The task repeats prior knowledge," "The task is solved fast" and "I would have to give hints in this task" are related to the theme level of challenge. The constructs are highly correlated with the range [0.94, 0.98]. The most liked tasks all get rated a 2 on all the constructs, which is a low rating. A rating of 2 on these constructs implies that the tasks do not repeat prior knowledge to any large degree, the tasks do take some time to solve, and the teacher would have to give aid to the students with hints.

Regarding the least liked tasks, tasks 9 and 1 rate high on these constructs, so they are the opposite of the liked tasks, which means they do repeat prior knowledge, are fast to solve and don't require hints from the teacher. Task 2 on the other hand is rated more like the most liked tasks than the least liked. This makes Task 2 an outlier to the general trend between the most and least liked tasks.

The results imply that the teacher thinks a task that takes some time to solve, does not repeat prior knowledge, and requires help from the teacher are "good" tasks. Though, because of the outlier in the ratings, I cannot conclude that this always is the case. The impression given by the ratings is that a "good" task must have these features, but these features alone are not enough to make a task "good", there must be other factors taken into consideration too to determine if a task is "good."

Mathematical content: From the constructs I am looking at, there are two constructs under the theme mathematical content. These are construct 2 and 10, "The task tests understanding" and "This is an 'algebra heavy' task" The constructs have a correlation of 0.66.

In Construct 2, the differences in ratings are not that extreme between the most and least liked tasks, but it is there. The most liked tasks all get rated a 4 on testing understanding. The least liked tasks have two ratings of 3 and one rating of 2. The rating 2 is given to the least liked task. The more a task tests understanding, the more it seems to be a task Teacher C finds to be "good.".

The difference in rating for Construct 10 is a somewhat bigger one. The most liked tasks have a higher rating than the least liked tasks. However, the most liked tasks are only rated a 3, so they are not the most “algebra heavy” tasks.

My deduction is that Teacher C thinks tasks that test understanding are “good” tasks. A “good” task also needs some degree of being “algebra heavy,” but this degree does not have to be too high.

Implementation of the task: There is one construct regarding implementation of the task that shows a big difference in the ratings between the most liked and disliked tasks and that is Construct 7, “I would give this task as a problem-solving task.” In this construct the generality is that the most liked tasks rate high and are therefore considered well suited as problem-solving tasks, and the least liked tasks rate low and are not good problem-solving tasks. However, Task 2 is again an outlier to the other least liked tasks and rates high on this construct, even though it is amongst the least liked tasks.

From this I interpret that Teacher C thinks a “good” task is one that can be used for problem-solving tasks. But like with the level of challenge, this one feature alone is not enough to determine if the task is “good” or not.

Complexity of the task: The constructs 5, 6 and 8, “The task contains ambiguity,” “I would modify the task if I were to use it in my education” and “I like the construction of this task” all regard the complexity of the task. For these constructs there was no general variation among the most and least liked tasks but rather there was one outlier task, Task 2. For all three constructs Task 2 is the only task that gets rated high on constructs 5 and 6, and low on construct 8. This makes it a task that contains ambiguity with at construction the teacher does not like and strongly feels the need to manipulate if he were to use it in class.

Generally, in the 9 constructs where there is a noticeable difference between the most and least liked tasks, Task 2 is often more similarly rated as the most liked tasks rather than the least liked ones. For these three constructs regarding task complexity, the task is rated significantly different from both the most liked and least liked tasks. This implies that the value of the complexity of a task is an important factor for Teacher C. Even though a task has potential to be a “good” task it cannot be considered a “good” task if the text contain ambiguity, and the construction of the task should be one the teacher likes sufficiently.

If we compare this deduction to the comments Teacher C gave with his ratings, one can see a clear alignment. The teacher himself said that the ambiguity of Task 2 weighed heavily, and this was what gave the task the lowest ranking. If not for the ambiguity the teacher says the task could be a good one. This may help explain why Task 2 often was more similarly rated as the liked tasks. The task may have had potential to be a “good” task, but the complexity of the task reduced the “goodness” of the task’s quality significantly.

The three last constructs: For constructs 17 and 18 “It is desired to use more of this type of task” and “I would use this task in my 1T education,” Teacher C have rated all the most liked tasks as 4. The general trend for the least liked tasks is that they get a lower rating and makes the teacher generally less enthusiastic about using these tasks in his own 1T education, and generally do not think it is desired to use more of such types of tasks.

In construct 19, “Overall, I think this is a good task,” the teacher has rated the most liked tasks a 5 and the least liked tasks rate between 3 and 4. The tasks the teacher likes the best are the ones he thinks are overall good.

Summary of Teacher C’s results: For Teacher C, the themes level of challenge, mathematical content, implementation of the task and complexity of the task are what impact if the teacher find a task “good.” According to the results, Teacher C considers a task “good” if it does not repeat knowledge, takes some time to solve and requires help from the teacher, but it is not given that a task with these features is “good.” A “good” task should test understanding and be somewhat algebra heavy. The text should be well formulated, and that construction of the task should be good. Problem-solving is also a feature a “good” task has, but only because a task can be used as a problem-solving task does not mean it is “good.” All the most liked tasks to Teacher C are ones he thinks are overall good tasks.

5 Discussion

In the discussion I will start by comparing the teachers' conceptions for a good task with each other to look for similarities and differences. Afterwards I will be comparing how the teachers ranked their most and least liked tasks by looking at the two most frequently liked and disliked tasks.

When I am done with the comparisons, I will contribute with some thoughts on how these results can be used, and I will come with a critical overview of my work.

5.1 Comparisons of the teachers' perceptions for "good" tasks

For this part of the discussion, I will be comparing the teachers' perceptions for a "good" task with each other. The perceptions are the ones deduced in the results. To add to the discussion, I will also include instances from the interviews to reflect upon the teachers' perceptions. I will do the comparisons of one topic at a time in the order level of challenge, complexity of the task, mathematical content, visuals of the task and implementation of the task.

Level of challenge: Level of challenge was a theme all the teachers offered constructs about, and everyone rated several of these constructs in a way that noticeably separated their most and least liked tasks. This led to all the teachers having perceptions of how the level of challenge impact if a task is "good."

The first deduction done was for Teacher A and this regarded his perception on how the difficulty and demandingness of a task play a role in deciding if a task is "good." According to the results, Teacher A finds that tasks can be quite demanding and somewhat difficult and still be "good" tasks, it is first if they get too difficult or too demanding that they are not considered "good" anymore. Generally, Teacher B seem to be equal in the thought that too difficult or too demanding tasks are not "good," but the results for this teacher are still different. The least liked tasks they have both rated as having about the same level of difficulty and demandingness, but the general trend for Teacher B's most liked tasks is that these tasks are rated as much easier and less demanding that the most liked tasks of Teacher A. Teacher B also have exceptions to his general trend, both for the most liked and least liked tasks. Teacher B's most liked task is one that is rated as being very demanding and difficult, but he still likes it so there must be some other feature to the task that thumps how much the difficulty pulls the task down. Teacher A is consistent in his rating that too difficult tasks are not "good." Teacher B show a looser view than Teacher A on how much the difficulty and demandingness impact if the task is "good" or not.

From the theory about cognitive challenge and the zone of proximal development, it is emphasised that for a task to be good, it needs to be within the range of what the students are able to do (Imsen, 2019; Smith & Stein, 1998; Vygotsky & Cole, 1978). Teacher A and B's perceptions seem to align with this theory. The tasks they felt were too difficult was rated among the least liked and are probably seen as being outside of the students' zone of proximal development.

None of Teacher C's constructs or ratings gave information that could be used to deduce if he had perceptions about difficulty or demandingness impacting if a task is seen as "good." However, in the interview Teacher C made a comment to one of the tasks that "There might be a point of throwing the students into deep water [referring to a task being difficult], but this might be too deep." Though the construct that came from this statement did not rate differently between the most and least liked tasks, it does showcase that the teacher has reflected about the difficulty of the tasks while analysing them. The statement gives the impression that Teacher C also agrees with the theory that task that are too difficult for the students should be avoided.

Other perceptions Teacher A and B have in common regarding level of challenge, is that they both mostly seem to preferer types of tasks the students are used to, and that they do not like tasks that are abstract. Teacher A also prefers tasks that do not consist of difficult mathematics and Teacher B tasks that do not require too high a mathematical level, which are somewhat interchangeable perceptions.

What it means that a task is abstract is not specified by the teachers. My first thought is that abstractness of a task can regard either the topic, topic is unclear to the students, or the formulation/procedure, it is unclear for the students what needs to be done. Teacher A have a construct regarding the topic of tasks being known [Construct 17]. The ratings for this construct have no correlation with the constructs regarding tasks being abstract [Construct 15]. Teacher B does not have a construct about the topic of a task being known, but from looking at the tasks the teacher has rated as abstract, I would say some of them have topics the students should know, like popcorn [Task 7] and tomato can [Task 8], and some have not. It seems unlikely the teacher refers to the topic when deeming a task abstract.

The constructs regarding tasks being abstract and students not being used to the task on the other hand does have high correlations for both the Teachers. For Teacher A, the constructs "This task is not a task the students are used to" and "This task will be too abstract for the students" have a correlation at 0.69. Teacher B's constructs "The students are used to this type of task" and "The task is abstract for the students" have a –1 correlation. These correlations show that abstract tasks tend to be task the students are not used to. If students are not used to a type of task, it is most likely the formulation or the procedure of the task they are not familiar with.

These two observations regarding tasks being abstract make it plausible to believe that the teachers do not refer to the topic when deeming a task abstract but the rather the formulation/procedure is abstract for the students. If the students do not understand the formulation to a question because the task is not one they are used to, it is likely time will be wasted on figuring out the question rather than focusing on the mathematics. To reduce the amount of time waisted and to get the most out of class as possible, liking tasks the students are used to seem reasonable in this regard.

Both teachers do however also seem to prefer tasks that do not consists of difficult mathematics. A task that consists of easy mathematics does not automatically make the task easy. If the goal of a task is to learn a new procedure, then using easy mathematics can be a smart choice. Learning a new procedure can be challenging. Using easier mathematics for tasks with such purposes will facilitate for learning

Combining the features easy mathematics and types of tasks the students are used to does however seem like a task with little demand for cognitive challenge. I would assume the more a student is used to a type of task, the more possible it is they start to solve them algorithmically. If the mathematics in the task also is easy, then no challenge is given either. Finding tasks the students are used to and that consists of easy mathematics seem contrasting to what theory says about a “good” task.

Between teacher A and C there are some perceptions that are in contrast with each other. Teacher A find tasks that are solved fast and do not require much help from the teacher to be “good.”. Teacher C on the other hand has, according to the results, the opposite opinion and find tasks that do take time to solve and do require help from the students are good.

Having different perceptions on how long a “good” task take to solve is understandable. From the interview, Teacher A expressed that several of the tasks in the triads would take two class hours to finish if they were to be implemented correctly, and that would be time they did not have to spare on one task. Time in the mathematics classroom can feel tight because of a large pensum and little time to get through it. That Teacher A finds tasks that are shorter to be “good” makes sense.

Another opinion about task length, is that students learn better from one longer task they work more in depth with, rather than several smaller ones where they just learn the routine to solve the task, and not necessarily why. The perception to Teacher C seems to agree to this thought. The new Norwegian curricula is also made to facilitate for such longer tasks, so taking the time to solve tasks that take time should not be that big of a hindrance as it was with the old curriculum. Teacher C might have had this in mind when considering how long a “good” task should take to solve.

From what I have gathered of experiences the goal of mathematics education is to make the student as independent as possible in their solving process so the perception from Teacher C that tasks that need help from the teacher are “good” tasks seems like a weird one. The implication might however be the result of cause and effect. The tasks where the teacher expect s/he would have to give hints to the students are also rated as tasks that do not repeat prior knowledge and the teacher has the perception that task that do not repeat prior knowledge are “good” tasks. The two constructs have high correlations. My hypothesis is that the teacher not directly prefers tasks that demand a lot of help, but rather that s/he do prefer tasks that do not repeat prior knowledge and that these tasks usually require more help from the students.

Teacher B also had a construction regarding tasks needing prior knowledge. There can seem like the teacher has a slight preference towards tasks that so not repeat prior knowledge, but the ratings were quite varies, so no conclusion was made for Teachers B regarding prior knowledge in tasks. Though no perception is made for Teacher B regarding prior knowledge in tasks, what the construct does show is that the teacher has considered it.

The last perception to mention is that Teacher B usually likes tasks where all the students can mean something about the task. What he implies with this construct is that he likes task where all the students can say something about the task even though they might not be able to solve the whole task. Tasks like this will give everyone a chance to do something with the task. The lowest achieving student might not be able to do the whole task, but they can contribute with something and get a feeling of some accomplishment. We say that tasks where everyone can do something have a low threshold.

Complexity of the task: From the results, Teacher A and C were the teachers that had perceptions for the complexity of the task.

Teacher A finds that a "good" task is one most students will understand and that does not require several readthroughs. These constructs have a correlation of 0.6. My hypothesis is that this correlation has its base in tasks being difficult to understand and as a result the text needs several readthroughs. I imagine a task that is difficult to understand usually need several readthroughs.

When Teacher A rated the construct regarding understanding, it came to him that understanding a task can have two meanings and explains that "Understanding a task can mean understanding what the task is asking, or it can mean understanding how to solve it. Understanding what the task is asking does not automatically mean that the students will understand how they should solve it." The construct came from the teacher expressing that one of the tasks had a text the students would not understand, and the teacher was therefor asked to rate the construct based on the expectancy of the students being able to understand what the task is asking. Thus, the teacher's perception is that a task is "good" if the students understand what the task is asking.

Like Teacher A commented, understanding what a task asks do not entail that the students understand how to solve it. However, understanding the question is the first step in the solving process. If the students understand what the task is asking, they have a base for starting their mathematical thought process on their own. If they do not understand what the task is asking, valuable time that could be spent thinking mathematically must be spent to figure out what the task is asking, likely by reading the text more times or by getting the teacher to clarify what is being asked. To avoid "wasting" time on several readthroughs or listening to the teacher explain the task, tasks should have a text that is well formulated and easy to understand. Preferring tasks that the students understand does sound like a reasonable preference.

Teacher B does not have a construct on understanding the task but have one about reading the text several times. According to him, most of the tasks need to be read several times unless it feels like a "classical textbook" task, which are tasks students tend to be highly used to and have learned to understand. This rating matches the conjectures I gathered from Teacher A's result that tasks the students do not understand, are likely to need more than one readthrough.

Teacher C does not mention understanding the text as explicitly as Teacher A, but he does mention ambiguity in the text. Ambiguity in the text entails unclearness of certain parts of the text and this can lead to confusion as to what it is the task is asking and thereof lack of understanding the task. Both from the interview and the ratings, Teacher C heavily emphasises that ambiguity in a text weakens a task and makes it less "good." According to Teacher C, the text in a task should be formulated in a clear way where there is no doubt what the meanings in the task are and making the task possible to understand.

The only task rated as containing much ambiguity according to Teacher C is Task 2. However, it is only Teacher C that claims this task contains ambiguity. Teacher A even thinks this is a task the students will have no problem understanding and even has it as his second most liked task. That one teacher see ambiguity but not others enlighten the fact that something that makes full sense to someone, might not do so for others. This is something that is important to reflect upon. If a teacher has made or found a task and wants to use it in class, it can be smart to get a second opinion on the task to make sure the text makes sense to others and not just oneself.

In addition to the text being well formulated, Teacher C also find that the construction of the task is essential. In the interview, the teacher compares the constructions of Tasks 2 and 3 and says he much prefers the construction of Task 3 because it starts with an example and builds up to a generalization, where else Task 2 both starts and ends with general results. Teacher B has a construct that reads "The task is increasing in difficulty." This construct did not have an impact on the teacher's perceptions for a "good" task because he did not think this construct was relevant for most of the tasks, so most constructs got rated 1. Tasks 2 and 3 on the other hand got rated a 5 on this construct and the teacher thinks both tasks have an equal construction that increases in difficulty. Here we see a contrast in the opinions to Teacher B and C. No one can say that the opinion of one is more correct than that of the other. This contrast show that there is some level of subjectivity in deciding if a task is "good," and again emphasises that having more than one teacher evaluate a task can be smart to check its usability.

Mathematical content: Like with the previous theme, Teacher A and C are the only teachers having perception on the mathematical content impacting if the task is "good."

For Teacher A, a "good" task is one that require some reflection from the students, but not too much. He also mentioned tasks being explorative, but these ratings were similar for both the most and least liked tasks, so no conclusion was drawn from them. The shift in the new curricula leans towards having more inquiry-based learning where students spend more time exploring tasks, rather than just learning procedures. When exploring a task, the goal is to better understand the process that leads to an answer. To understand the process, one need to put some reflection upon it. In addition, tasks that require reflection and understanding are tasks with high cognitive demand. Preferring tasks that require reflection suit the new curriculum and it fits in with earlier theory on what a good task is (Smith & Stein, 1998, p. 348; Utdanningsdirektoratet, 2020).

What then needs to be evaluated with Teacher A's perception is that the preference is for tasks that require "some reflection, but not too much." For the most liked tasks the rating for needing a lot of reflection is between 2-3 and the least liked tasks rate a 5. The focus on incorporating tasks that are explorative and need reflections is a new one in the curriculum, so there is a chance that neither teachers nor students are used to such tasks yet. That the teacher does not prefer tasks that require too much reflection is probably connected to the teacher and students gradually need to get used to such tasks. When the students are used to tasks that require some reflection, they can get tasks that require more reflection. A rating of 2 on needing a lot of reflection is however not that high a rating. It seems the teacher would rather have closer to no reflection in the tasks rather than too much. This does not fit too well with the curriculum and taking away the reflection of a task decreases its level of cognitive demand.

Teacher C did not directly make constructs on tasks needing reflection but find that tasks that test understanding are "good." Tasks that test understanding imply that the students understand why they use the method they use. This understanding will most likely come from the students reflecting on the task. In this way, Teacher A and C have similar opinions for "good" tasks. Teacher B did not make any such constructs as Teacher A and C did. This imply that he did not consider the task up towards the curriculum the same way the other two teachers seem to have done.

Another perception Teacher C came with regarding the mathematical content, is that he likes tasks that are a bit "algebra heavy," but not too "algebra heavy." In the interview he expressed that he himself usually turned to algebra when solving tasks, even for geometry tasks. He pointed out that his preference for algebra might impact which tasks he chose to use in class. Here he shows that personal preference an impact the teachers choice. However, he was aware that even though he preferred algebraic tasks, it did not mean that his students needed to prefer algebraic tasks. He also pointed out that students often are not that fluent is using algebra, so a task that is too “algebra heavy” will be difficult for them to solve. So even though the teacher preferred algebraic tasks himself, he tried to vary the tasks he chose for class so that the students did not just receive algebraic tasks.

Visuals of the task: Both Teacher A and B made constructs about the visuals in the tasks, but the only teacher that has a clear perception for “good” tasks that regard the visuals of the task is Teacher B. Teacher C did not mention visuals at all.

As mentioned earlier, Teacher B made four constructs regarding the visuals, and three of these were rated differently between the most and least liked tasks. Emphasising the visuals value seems quite important to Teacher B who commented on the visuals of the tasks from the first triad. Teacher A did not comment on the visuals before the second triad where there was a task that did not include an illustration. When the Teacher A saw Task 4, he commented that “This task should have a drawing” and then proceeded to commented that the visuals of the other tasks were helpful. Other than this, the teacher did not focus much on the visuals in the interviews and the two constructs did have different ratings between the most and least liked tasks. It was almost as if Teacher A took it for granted that certain tasks had visual and did not see it as a positive thing that most tasks had a visual, but rather a negative thing if the task did not.

Why Teacher C did not comment on the visuals I cannot say. It might be that the teachers saw the visual as part of the task such that if the visual was there it belonged there, but if there was no visual the task was not supposed to have a visual. The teacher might not have thought about the fact that the visual was something relevant to comment on.

Seeing how differently the teachers acted towards the visuals, one finding it important, one just mentioning it and one not commenting on it at all, shows how various teachers set their focus on different aspects. Something that one teacher might see as an important aspect of a task, might not even be considered by another. This again shows that deciding if a task is “good” includes a subjective aspect by the teacher.

Implementation of the task: Both Teacher A and C made constructs about the implementation of the task, but it was only Teacher C’s ratings that gave an implication on how implementation of a task can impact if the task is “good”. Teacher C have the perception that “good” tasks should be tasks that can be given as problem-solving tasks. Teacher A have a construct that reads “The task is nice to use as a group task”. The ratings on this construct does not vary between the most and least liked tasks so no implication can be made about this fact impacting if a task is “good,” but the construct show that the teacher have thought about implementation of the tasks.

According to the Documentational Approach to Didactics, even though a task is considered good, the teacher needs a scheme of usage on the task to get out the task’s potential (Trouche et al., 2020). Teacher A and C show some thought around the scheme of usage, but Teacher B does not. Why this is one can only suspect. One possibility is that the teacher did think about implementation of the tasks but did not utter them. Another possibility is that he did not think about implementation because he did not see it as important. If it is the last case, that he did not see it as important, his opinions will be in contrast with what theory says about a good task.

In addition to Teacher C mentioning implementation of the task through using it as a problem-solving task, the concept of problem-solving itself is also one to take note of. The new curriculum is highly arranged to focus more on problem-solving in the mathematics classroom and is part of one of the core-elements. Since problem-solving have become such a big part of the curriculum, I would have thought the teachers emphasised it more as well. The results I got are in misalignments between the curriculum and the practise of teachers regarding problem-solving, since only one of three teachers mentioned it when analysing the tasks.

The three last constructs: All the teachers had generally higher rating for their most liked tasks than their least liked tasks for the constructs "It is desired to use more of this type of task," "I would use this task in my 1T education" and "Overall, I think this is a good task." Most of the tasks were types of tasks the teachers thought should be used more of, and all the teachers expressed that they would like to use several of the tasks in their own education. All the teachers rated their most liked tasks as a 5 on being "overall good tasks." The implication here is that to decide if a task is "good" it is important to get an overall view of the task, but even though the overall view is not the best, the tasks can be desired to used anyway.

5.2 Compare the tasks the teachers chose as their favourites

In part 5.1 I compared all the teachers' perceptions to see how they varied in their definitions of a "good" task. In this part I will be comparing the two most and least liked tasks to all the teachers combined. I will look at how the teachers' ratings of these tasks are similar and how they differ from each other. I will also see if it is possible to see how the teachers' perceptions for "good" tasks may have impacted their choices.

In total, all the tasks that have been rated as one of the three most liked tasks are tasks 2, 3, 5, 6, 7 and 8 and the tasks that have been rated as least liked tasks are 1, 2, 4, 5, 7, 8 and 9. All the tasks have been mentioned from one or more teachers as being either most or least liked. There is only one task all the teachers have rated as one of their most liked tasks, and that is Task 6. Other than that, the agreement that a task is most or least liked only goes between two of the teachers at a time if even that. Some of the tasks have been selected as both most and least liked by different teachers. In Table 9. is a list of the two most frequently liked and disliked tasks by the teachers.

Table 9: The two most frequently liked and disliked tasks

The two most frequently liked tasks	1. 6 – Overlapping squares 2. 7 – Popcorn
The two least frequently liked tasks	1. 8 – Can you can the can? 2. 9 – Drill out this problem

Task 6 – Overlapping squares: There was only one task all the teachers had on their top three list of most liked tasks, and that was Task 6, "overlapping squares". Both teacher A and C had this as their most favourite task, and Teacher B had it at second place.

According to Teacher A, Task 6 is a demanding task, but not to the extent that no students would be able to solve it. The text is easy to read, the task is not abstract, and it is expected that most of the students will understand the task. It is a task the students are used to, and the teacher do not think the students would need much help to solve it, though it is a somewhat difficult task. The mathematics of the task is however not seen as difficult. Solving the task is not expected to take much time, but some reflection is required in the task.

Teacher B does not find that the task is particularly demanding or difficult for the students and it is not abstract. A high mathematical level is not required to start or finish the task. The figure in the task is helpful and good as it is. This is a type of task the students are used to. Some prior knowledge is required to solve the task, but it is expected that several students should be able to mean something about it.

Teacher C does not think this task repeats prior knowledge to any large extent. The task would be good as a problem-solving task that tests understanding. There is some level of algebra in the task, but it is not too heavy. Solving the task would take some time and the teacher would have to give hints to the students.

All the teachers' ratings of the tasks suit their individual perceptions for a “good” task. Still there are some perceptions the teachers have that contradict each other.

The first thing I notice from the teachers’ ratings of the task is that teacher A and B have contrasting thoughts regarding the task being demanding. Teacher A has rated it a 4 on being difficult and fits his description for “good” tasks being quite demanding but not too demanding. Teacher B has rated the task a 2 on being demanding and fits his perception that “good” tasks should not be very demanding. From the comparison of the teachers’ perception for a “good” task, Teacher A seemed to like tasks that are more demanding and difficult that Teacher B. However, looking at their rating for this one task makes me question whether they have a different perception of how difficult a task should be or have a different definition of what “difficult” is. It is also possible they relate the tasks as being demanding relative to their own students. This would explain why both teachers like the task but think differently regarding it being demanding.

Another contrasting rating for this task the expectancy of how long it will take to solve. Teacher A consider the task more difficult that Teacher C does, but still, he thinks it will take less time to solve. The deduction I make, based on the teachers’ ratings of the task, is that this probably have to do with how the teachers plan to implement the task. Teacher A simply say it can fit as a group task. Solving a task in a group can be done fast. Teacher C would implement this as a problem-solving task. When using a task for problem-solving, the goal is to focus on understanding the task. Getting understanding of a task cannot be fast obtained. Focusing on understanding in a task will take more time as opposed to just solving it. The consideration the teachers have for how the task is to be used seems to impact the perceptions the teacher has for how long it will take to solve the task.

Even with the different perceptions the teachers had, the task is still considered “good” by all the teachers.

Task 7 – Popcorn: The second most frequently liked task was Task 7. This task got a second place from Teacher C and a third place from Teacher A. However, Teacher B rated this task as his third least favourite task.

Teacher C have rated Task 7 almost exactly as he has rated Task 6. There are only three constructs [construct 9, 11 and 16] that have a slight difference in the ratings for these two tasks and the result is that Task 7 is seen as needing slightly more competence to be able to solve than Task 6. These small variations in the ratings can be why the teacher prefers Task 6 over Task 7. Regardless of this little change the task still fits the teachers' perceptions for a "good" task and makes sense that this is among the most liked tasks to Teacher C.

Teacher A have also rated Task 7 similarly to Task 6 for many of the constructs, and so for the most part the task fits the teacher's perception for a "good" task. However, Task 7 is not a task the students are too used to, the task consists of more difficult mathematics and, probably because of this, is a task where the students probably need some help. All these are factors that go against what the teacher generally finds "good" tasks should include, but it is still the third most liked task to the teacher. Why this is so is not possible to know, but some reflection might supply a possible reason. All the factors in the task that do go against the general thought to Teacher A regard the level of challenge, and the main focus is that the task consists of somewhat difficult mathematics and that the students are not that used to this type of task. Other that the higher level of difficulty of the task, it is similarly rated as the other "good" task. What it seems like is that this is a task the teacher generally likes and that the extra challenge the task will give from its mathematics level and un-common type is appreciated.

Teacher B did not rate Task 7 similarly as he rated Task 6. For Teacher B, Task 7 is a very demanding and quite difficult task. Prior knowledge is required for this task, and it is not expected that the students will be able to say something about the task. The level required to do this task is relatively high. It is not a task the students are particularly used to, and it is somewhat abstract. The visual in the task give some help to the students but could have had more helping lines. This drastic variation in ratings between Task 6 and 7 excuses why Task 7 is not among Teacher B's most liked task, but rather with the least liked ones.

There is not as much agreement between the teachers of how good of a task Task 7 is. If comparing the teachers' ratings of Task 6 and 7, Teacher C rates the tasks very similarly, Teacher A rates the tasks somewhat similar and Teacher B does not rate them similarly at all. It is not clear why the teachers have such various opinions on this task, neither from the Likert-scale not from the interviews.

Task 9 – Drill out this problem: One of the most frequently least liked tasks is Task 9. Both Teacher B and C have this task as their second to last least liked task. Teacher A did not rate this task as either most or least liked.

The ratings of Task 9, seen in the light of Teacher C's perceptions for "good" tasks, is a task that repeats too much knowledge and tests too little understanding. It is not a task that would be used as a problem-solving task, and it is not an "algebra heavy" task. Overall, the task seems to be rated as too easy and therefore it is not liked that much. This finding can be backed up with the comments the teacher gave along with the ratings. He does not think anything is wrong with the task, but it is done fast and do not have much problem-solving in it. This explanation the teacher gives match the description of tasks with low cognitive demands (Smith & Stein, 1998, p. 348).

In Teacher B's ratings this is clear outlier among the least liked task. The task is not demanding or difficult, nor abstract. The students are used to this type of task. There is not required a high level to be able to solve this task. Some prior knowledge is required to solve the task, but a lot of the students are expected to be able to say something about the task. The rating of the task is similar to the most liked tasks, and this highly suggest that the task should be one the teacher find as "good."

What probably is the case for Teacher B is the same as for Teacher C that the task is seen as too easy. As Stein and Smith (1998) have said, a "good" task should have the potential to engage the students in high-level thinking. A task that is too easy, like task 9, is not a part of the zone of proximal development, it is part of what the students' actual development zone (Vygotsky & Cole, 1978), and does not have potential to engage in high-level thinking and is therefore not considered a "good" task.

Something that is not possible to deduce from the ratings of the constructs, but that Teacher A mentioned in the interview, is that even a task that is as "straight forward" as Task 9, it can have its function, as for example a test yourself task. This is an important point he comes with; different tasks can have different purposes depending on the goal with the task. Even if a task is not considered particularly "good," it can still have a useful purpose.

Task 8 – Can you can the can?: Task 8 is a task all the teachers had an opinion on but only Teacher A and B agree that this is the worst task of them all. Teacher C like it quite well and have it as his third most liked task.

During the interviews, neither Teacher A nor B could see how this task would be solved in a mathematical way. The only possibility the teachers saw for somewhat solving the task was by looking at the illustration and comparing the task with what they knew from real life. Neither found this as a valid way to solve the task. Teacher A also find the task to be vague and from there arises the question of what is allowed to do when solving the task and what is not. The negativity the teachers had for the task is reflected in their ratings for the task where both teachers consider the task too demanding task for the students.

Teacher C have rated the task more positively than the other two teachers. The reason for this most likely lay in the teacher seeing a mathematical way to solve the task, as opposed to Teacher A and B who did not. When Teacher A saw the task in the interview, he commented that "if I just give them [the students] a special example, they should be able to find a solution."

As Teacher A points out, Task 8 is vague and knowing what is allowed to do to solve the task is unclear. To make a general solution to such a vague task, the concept of specializing is helpful. Mason and Davis (1991) explain specializing as such:

> "Specializing *means finding some thing, whether a physical object, a mental image, a diagram, a special case or example, that can be manipulated with more confidence than the words used in the question*" (Mason & Davis, 1991, p. 8).

When one is able to manipulate a task more confidently, it may be easier to understand how factors in the task impact each other. From here it is easier to look for patterns, and thereof get a better understanding of what is going on in the task. When this understanding is acquired, the path to make a generalization of the task is easily accessed (Mason & Davis, 1991).

That Teacher A sees a possibility in the task that the others do not see can be connected to instrumentalization (Trouche et al., 2020). Through experiences one gather knowledge and get inspired on how things can be done. For Teacher C it seems obvious that specializing is a helpful tool to solve this task, and this is most likely because he has used or reflected on such a way to solve tasks before and probably have earlier experiences in the field. If Teacher A and B do not have prior experience with specializing, it is less likely they will realize that this is a tool that can be incorporated in their education, and this lack of knowledge might be why the teachers did not like the task.

Specializing is not directly connected to the new curriculum, but generalizing is a part of the core elements, and is therefore something that should be considered by the teachers in their education. Of all three teachers, it is only Teacher C that somewhat touches the topic of generalization through mentioning that specializing as a tool to solve the task, which ultimately will lead to the making of a generalization. I do not know if the teacher actively and intentionally made connections to the curriculum when analysing the task, but there is some compliance between the two.

5.3 How can the results be used?

These results can be used to make mathematics teachers aware of how many considerations there are to be taken when selecting tasks for class and how complex task selection can be. The results give concrete examples of considerations that can be taken when analysing tasks. I hope these examples can inspire other teachers to reflect upon the tasks they choose to incorporate in their classroom.

5.4 Critical overview

In this thesis I have tried my best to be systematic and objective in the data analysis. However, my own perceptions of what may be happening can somewhat have led to my own biases. My deductions and hypothesis might have been impacted by my own experiences and thoughts.

6 Conclusion

The purpose of this thesis has been to study task selection to better understand what considerations mathematics teachers have for a “good” mathematical task. The three teachers that participated in the data collection had several perceptions of task features that impacted if a task could be considered “good.” Some perceptions were similar for all the teachers, but some were just perceptions of one or two of them. The totality of the perception a teacher had was used to deduce how that teacher considered if a task was “good.” The different teachers had various perceptions that lead to various considerations for what a “good” task is.

All the themes that have been considered by the teachers are level of challenge, complexity of the task, mathematical content, visuals of the task and implementation of the task. All the teachers made constructs for most of the themes, but each teacher had different focuses on what they found important. For example, did one teacher emphasise the importance of the visual in tasks, where else another teacher did not mention visuals at all in his considerations.

All the teachers had general perception on features they thought a task should have. The tasks the teacher consider “good” usually follow several, sometimes all, of the perceptions the teacher has for a “good” task. However, oftentimes these perceptions alone were not enough to tell if a teacher would consider a task “good.” A task that mostly was rated like a “good” task for a teachers could still be considered as not that good. The same way could a task that seemingly rated like a task the teacher would not like be considered “good.” This shows how the teachers weighted their perceptions differently. Some perceptions were more important for “good” tasks than others. Therefore, to consider if a task is “good” one need an overall view of the task. This makes evaluation of tasks and task selection a complex process.

There are also some individual features that can impact how "good” a teacher would find a task, such as subjectivity and bias. It can therefore be smart for a teacher to get a second opinion on tasks that is planned to use in class, to clear the task for ambiguity and such. Other factors the teachers seemed to consider when deciding if a task was “good” was the class's perquisite.

Some teachers also seem to consider the curriculum more than others when they consider tasks. The teachers’ opinions on what a “good” task is are often the same as what earlier research and theory proclaims “good” tasks to be, but sometimes there are some contrasts.

The consequences of instrumentalization are also evident in the teachers’ considerations. There was one task that two of the teachers expressed much dislike towards that the third teachers liked well, most likely because he saw a possibility in the task the others did not.

This study has provided some concrete examples of teachers’ perceptions regarding task features for instructional use. “Local” insights into more general theories of teachers reasoning have been provided. Hence it will add to our understanding of teachers’ needs. This may have ramifications for those curriculum specialists who are engaged in task design.

7 References

Alacaci, C., Prospective Elementary Teacher's Selection of Mathematical Tasks. Paper presented at The 12th Conference of European Mathematics Education Research, February 2-5th, 2022, Bolzano, Italy.

DeGraaf, B. E. (2015). *What Makes a Good Problem?: Perspectives of Students, Teachers and Mathematicians*, (Doctoral dissertation, Columbia University). Retrieved from https://academiccommons.columbia.edu/doi/10.7916/D8NV9HC3

Imsen, G. (2018). *Elevens verden: innføring i pedagogisk psykologi* (5th edition). Universitetsforlaget.

Mason, J. and Davis, J. (1991). Fostering and sustaining mathematics thinking through problem solving. Deakin University, Geelong, Victoria: Deakin University Press.

Nymann, R. (2016). What makes a mathematical task interesting? *Educational Research and Reviews*, *11*(16), 1509-1520. Retrieved from https://files.eric.ed.gov/fulltext/EJ1111486.pdf

Raskin, J. D. (2002). Constructivism in Psychology: Personal Construct Psychology, Radical Constructivism, and Social Constructivism. *American Communication Journal*, *5*(3).

Skott, J., Skott, C. K., Jess, K. & Hansen, H. C. (2019). *Matematikk for lærerstuderende: Delta 2.0 Fagdidaktikk, 1.-10. klasse* (2nd edition). Samfunds litteratur.

Smith, M. S & Stein, M. K. (1998). Selecting and Creating mathematical Tasks: From Research to Practice. *Mathematics Teaching in Middle School*, *3*(5), 344-350. Retrieved from https://www.jstor.org/stable/41180423?seq=5#metadata_info_tab_contents

Tan, F. B., & Hunter, M. G. (2002). The Repertory Grid Technique: A Method for the Study of Cognition in Information Systems. *MIS Quarterly*, *26*(1), p. 39-57. Retrieved from https://www.jstor.org/stable/4132340?seq=1&cid=pdf-reference#references_tab_contents

Trouche, L., Gueudet, G., and Pepin, B. (2020). Documentational Approach to Didactics In S. Lerman (Ed.). *Encyclopedia of Mathematics Education*, (pp. 237-247) Springer. Retrieved from https://doi.org/10.1007/978-3-030-15789-0_100011

Utdanningsdirektoratet. (2006). *Læreplan i matematikk fellesfag (MAT1-04)*. Retrieved from https://www.udir.no/kl06/MAT1-04/Hele/Kompetansemaal/kompetansemal-etter-1t-%E2%80%93-vg1-studieforebuande-utdanningsprogram

Utdanningsdirektoratet. (2019). *Hva er et kjerneelement?* Retrieved from https://www.udir.no/laring-og-trivsel/lareplanverket/stotte/hva-er-kjerneelementer/

Utdanningsdirektoratet. (2020). *Læreplan I matematikk fellesfag vg1 teoretisk (matematikk T) (MAT09-01)*. Retrieved from https://www.udir.no/lk20/mat09-01?lang=nob

Vygotsky, L. S. & Cole, M. (1978). *Mind in Society: Development of Higher Psychological Processes*. U. S.: Harvard University Press.

Appendix

i. Letter of consent

Are you interested in taking part in the research project?

"Teacher's mathematical task selection"

This is an inquiry about participation in a research project where the main purpose is to identify the factors that affect the perceptions of task quality in school mathematics. In this letter we will give you information about the purpose of the project and what your participation will involve.

Purpose of the project

This is an exploratory research project about how mathematics teachers perceive the quality of instructional tasks in school mathematics. The purpose is to identify the factors that affect the perception of task quality as perceived by teachers. Your personal information (e.g., your name or position) will not be linked to the data, and data will not be used for any other purpose than the purposes of this research project.

Who is responsible for the research project?

Department of Mathematical Sciences of the University of Agder is the institution responsible for the project.

Why are you being asked to participate?

The intended population of this research project is mathematics teachers. You are selected to participate in this project because you teach mathematics at a high school.

What does participation involve for you?

If you choose to take part in the project, this will involve that you will be interviewed in two separate sessions with the first one taking approximately 60 minutes and the second one about 60 minutes. During the first interview, you will be asked to compare and contrast pairs of instructional tasks. During the second interview, you will be asked to rate the tasks on a Likert-scale. I will take written notes of your responses during the interviews and I may make audio recordings of the interview as well.

Participation is voluntary

Participation in the project is voluntary. If you choose to participate, you can withdraw your consent at any time without giving a reason. All information about you will then be made anonymous. There will be no negative consequences for you if you chose not to participate or later decide to withdraw.

Your personal privacy – how we will store and use your personal data
We will only use your personal data for the purpose(s) specified in this information letter. We will process your personal data confidentially and in accordance with data protection legislation (the General Data Protection Regulation and Personal Data Act).

Kristine Moldestad will have access to the data to be collected. However, I will replace your name and contact details with a code. The list of names, contact details and respective codes will be stored separately from the rest of the data. I will store the data on a safe research server. Participants will not be recognizable in publications. In publications, no reference will be made to names or other personal information.

What will happen to your personal data at the end of the research project?
The project is scheduled to end in June 2022. Data will be kept until one year after the project ends, and then all the data including the electronic records will be deleted permanently.

Your rights
So long as you can be identified in the collected data, you have the right to:
access the personal data that is being processed about you
request that your personal data is deleted
request that incorrect personal data about you is corrected/rectified
receive a copy of your personal data (data portability), and
send a complaint to the Data Protection Officer or The Norwegian Data Protection Authority regarding the processing of your personal data

What gives us the right to process your personal data?
We will process your personal data based on your consent.

Based on an agreement with the Department of Mathematical Sciences of the University of Agder, *and* NSD – The Norwegian Centre for Research Data AS has assessed that the processing of personal data in this project is in accordance with data protection legislation.

Where can I find out more?
If you have questions about the project, or want to exercise your rights, contact:
Supervising professor Cengiz Alacaci, Department of Mathematical Sciences of the University of Agder, by email: (cengiza@uia.no) or by telephone: +47 45 91 96 92.
Person in charge of data protection, Johanne Warberg Lavold, by email (johanne.lavold@uia.no) or by: telephone: +47 41 21 20 48.
NSD – The Norwegian Centre for Research Data AS, by email: (personverntjenester@nsd.no) or by telephone: +47 55 58 21 17.

Yours sincerely,

Researcher

Kristine Moldestad

Consent form

I have received and understood information about the project *Perceptions of Mathematical Instruction Tasks* and have been given the opportunity to ask questions. I give consent:

- ❑ to participate in *individual interviews*.

I give consent for my personal data to be processed until the end date of the project, approx. *September 2022.*

(Signed by participant, date)

ii. Triads

Dear teacher,

With this, I am requesting you to consider the attached set of instructional tasks in high school mathematics. This is part of a study that examines teachers' selection of instructional tasks and the possible implications for task design. There are 6 tasks in this set.

In our first interview, we will consider and compare these tasks in any way and from the perspective of any characteristics you think of. The tasks are placed in triads (groups of three). The tasks in the same triads are placed consecutively in this document (as tasks 1-3, 4-6 and 7-9). These tasks can potentially <u>be used in the same grade levels,</u> and they are designed to <u>tap into similar or the same learning goals</u>. At this time before the first interview, you may consider the tasks so it will be easier to compare them in the first interview.

1. **First, please read the tasks. Think about a possible solution. You can use the back of the pages or extra paper to work on your solution if you think it is helpful. You can take notes in the spaces following the tasks if you like.**
2. **Think about how pupils would respond to the task. Can they understand the task? Can they solve it? Would pupils feel motivated to engage in the task? Why?**

In the second interview, I will ask you to rate on a Likert scale these same tasks based on the information I gather from the first interview.

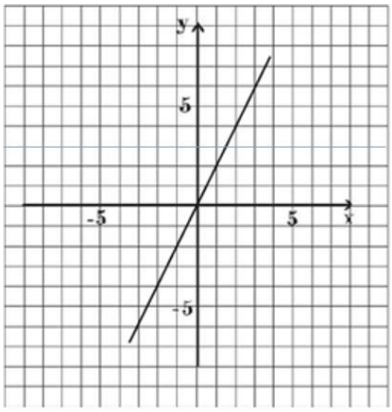
Thank you for your time and I appreciate your help in this research project.

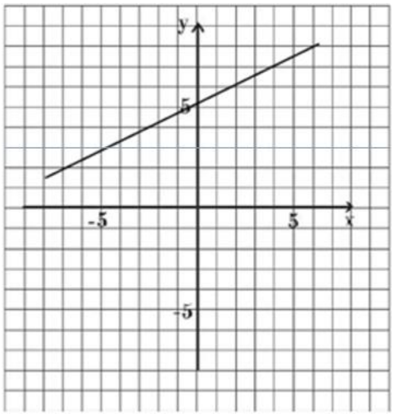
Set 1 – Linear functions

Task 1 - Parallel lines

Below are four straight lines. Two are in the form of equations, and two are in the form of graphs.

Part A:
Circle all those that are parallel to y = 2x + 5. There can be more than one answer.

1. y = 2x + 9
2. y = 5x + 2
3. 

4. 

Part B:
How do you know your answer is correct? Explain in as much detail and mathematical language as you can

……

……

……

Task 2 - Stacked dices

Build a tower by putting dices (or centicubes) on top of each other. Count the number of visible sides on all the dices in the tower. Build towers with different hights.

1. What is the connection between the number of dices and the number of visible sides?

   ……………………………………………..…………………………………………….

   ……………………………………………..…………………………………………….

   ……………………………………………..…………………………………………….

2. How many visible sides will there be in a tower of a 100 dices?

   ……………………………………………..…………………………………………….

   ……………………………………………..…………………………………………….

   ……………………………………………..…………………………………………….

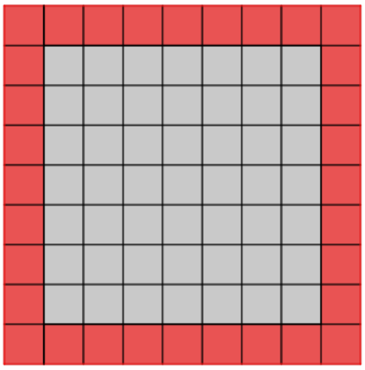
3. What is the connection between the number of dices and non-visible sides?

   ……………………………………………..…………………………………………….

   ……………………………………………..…………………………………………….

   ……………………………………………..…………………………………………….

Task 3 - Squares in a frame



The figure is a square put together of 81 smaller squares. We call the red smaller squares the frame.

1. How many smaller squares are in the frame?

   ……………………………………..……………………………………

   ……………………………………..……………………………………

   ……………………………………..……………………………………

2. How many smaller squares will there be in the frame of a square of 16, 25 and 36 smaller squares?

   ……………………………………..……………………………………

   ……………………………………..……………………………………

   ……………………………………..……………………………………

3. How many smaller squares will there be in the frame of a square with n^2 smaller squares?

   ……………………………………..……………………………………

   ……………………………………..……………………………………

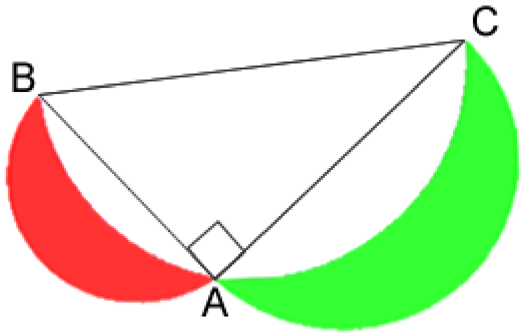
   ……………………………………..……………………………………

Set 2 – Geometry (area)

Task 4 - Area between two concentric circles

Given two concentric circles with radius R and r respectively, such that a chord tangent to the inner circle is 10 cm. Find the area between the circles (the circle ring).

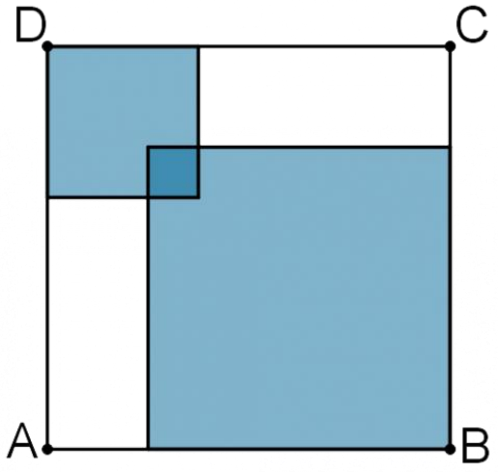
Task 5 - Pythagorean theorem with circles



The triangle ABC is right-angled with angle A right. There is drawn a halfmoon on each side of the triangle. There is formed two “quarter moons” as shown by the coloured area.

Show that the area of the quarter moons together is the area of the triangle.

Task 6 – Overlapping squares



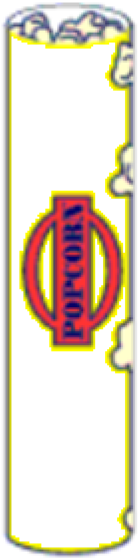
The square ABCD has an area of 196. It contains two overlapping squares. The biggest of these squares has an area that is four times bigger than the smallest one. The overlapping area is 1.

How big is the coloured area of the figure?

Set 3 – Volume

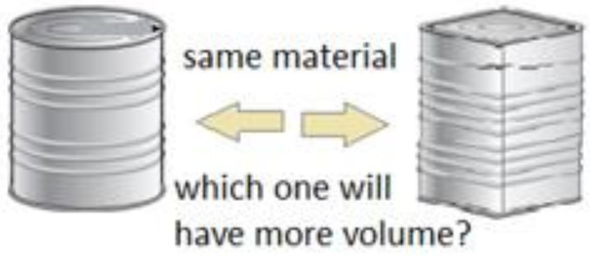
<u>Task 7 – Popcorn</u>

Which two cylinders made of A4 size paper, made by folding length-wise or hight-wise hold more popcorn?
How do you know?



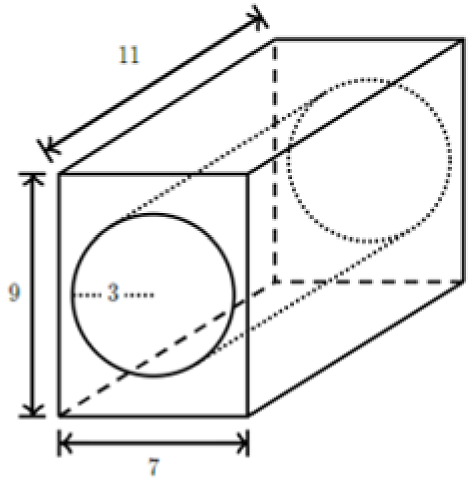
Task 8 – Can you can the can?

Assume you are designing a container for a tomato paste company. If you want to use the same material, is it more profitable to make cylindrical or rectangular-prism containers? Which one would hold more tomato paste?



Task 9 – Drill out this problem

A cylinder has been drilled out of this block to allow a rod to pass through. All measurements are in millimeters. Find the remaining volume of the block.



iii. Example of a filled-out Likert-scale and rating of the tasks

Rating of the tasks

Task 1	Parallel lines	**Task 4**	Area between two circles	**Task 7**	Popcorn
Task 2	Stacked dices	**Task 5**	Pyth. theorem w/circles	**Task 8**	Can you can the can?
Task 3	Squares in a frame	**Task 6**	Overlapping squares	**Task 9**	Drill out this problem

Please rate the constructs:

1: Don't agree at all 2: Little agreement 3: Some agreement 4: Agree 5: Verry agreed

	Construct	**Task 1**	**Task 2**	**Task 3**	**Task 4**	**Task 5**	**Task 6**	**Task 7**	**Task 8**	**Task 9**
1	The task repeats prior knowledge	4	1	1	2	2	2	2	2	4
2	The task test understanding	3	2	2	4	4	4	4	4	3
3	The task if solved fast	5	1	2	1	1	2	2	2	4
4	The task uses good mathematical language	4	3	3	4	4	4	4	3	4
5	The task contains ambiguity	2	4	2	1	1	1	1	2	1

6	I would modify the task if I were to use it in my education	1	4	2	4	4	1	1	2	1
7	I would give this task as a problem-solving task	1	4	4	3	3	4	4	4	1
8	I like the construction of this task	4	2	3	3	3	4	4	3	3
9	The task is "too deep water" for the students	1	2	2	4	4	1	2	2	1
10	This is an "algebra heavy" task	1	2	2	4	4	3	3	3	1
11	A general 1T student probably does not have enough competence to solve this task	1	2	1	4	4	1	2	2	1
12	The task is insurmountable for the students	1	2	1	2	2	1	1	3	1
13	Measurements in the task are given, but one must look for them	1	2	1	2	2	1	1	3	1
14	The students can test out this task in real life	3	4	3	3	2	4	4	2	1
15	I would not have to give hints in this task	4	2	4	1	2	2	2	2	4

16	You can use your guts to start/continue this task	3	3	4	2	2	3	2	2	3
17	It is desired to use more of this type of task	3	2	3	3	3	4	4	4	2
18	I would use this task in my 1T education	4	3	3	2	2	4	4	4	1
19	Overall, I think this is a good task	4	3	5	3	3	5	5	5	4