Students’ perceived difficulty of mathematical tasks: an investigation on influencing factors

DIDACTICA MATHEMATICAE
Vol. 44, 2022, pp. 59–79

Marta Saccoletto (Turin, Italy)
Camilla Spagnolo (Bolzano, Italy)

Students' perceived difficulty of mathematical tasks: an investigation on influencing factors

Abstract: The paper shows the main results of a qualitative survey focusing on students' perceived difficulties after solving mathematical tasks (grade 9 and 10 students). The aim is to identify factors that influence students' perceived difficulty. Although factors contributing to an increasing or decreasing task difficulty (in an absolute sense) are widely discussed in the literature, students' perceived difficulty regarding a mathematical task is not. We believe that the analysis of the questionnaire and focus group conducted with students highlight some important reflections on the influence that metacognitive, affective and task factors have on students.

1 Introduction

Difficulty in mathematics is an extremely broad and fundamental issue in mathematics education research. It concerns several aspects already widely discussed in the literature, such as aspects related to mathematical content (Radmehr & Drake, 2017) or text comprehension (Spagnolo et al., 2021a).

Many studies aim to understand the possible causes of students' difficulty in mathematics, particularly in relation to mathematical tasks resolution. For example, Bolondi et al. (2018) investigate how text variations influence students' performance, highlighting that the difficulty of the task can depend on the wording of the task text. Task formulation is not necessarily better or worse

Key words: perceived difficulty, affect, grounded theory, large scale assessment, argumentative competence.

for everyone, but it does seem to influence students' performance. Complexity of tasks can be related also with the content of the question, such as numerical magnitude complexity (e.g., De Corte et al., 1988; Thevenot & Oakhill, 2005). Rewording the problem text has proved to be useful for improving children's performance, especially among younger children, and the difficulties related to rewording do not depend on the length of the resulting text (Vicente et al., 2007).

In addition to these studies, affective factors are taken into account in studying and interpreting students' behaviours and difficulties (Zan et al., 2006), also regarding mathematical problem solving (McLeod, 1998). Both adults and children often proclaim their lack of skills of mathematics without embarrassment, treating this absence of accomplishment in mathematics as a permanent state over which they have little control (McLeod, 1992).

The idea of the study is to take more into consideration the students' perspective by investigating their perception of difficulty of a mathematical task. A small number of studies have explored these factors from the point of view of students and there is no single definition of "perceived difficulty" in the field of mathematics education. In order to explain the focus of our study, we can give an example of a specific situation: the time when a student faces a mathematics task. In that moment, the student could run into multiple *difficulties* that may depend both on the student's own characteristics, such as his/her skills and knowledge, his/her beliefs and attitudes; or by task peculiarities, such as the text or the mathematical content involved. On the other hand, these latter item characteristics might influence the student's idea of the task, and they might help to set up his/her *perceived difficulties*. Consequently, *difficulties* and *perceived difficulties* are two different – but closely related – aspects. In particular, we reckon that *perceived difficulties* affect the student's behaviour in addressing the task. This study wants to investigate these aspects, asking students to evaluate the difficulty of a mathematical task, after they solved it.

2 Theoretical and context background

In this section, the theoretical background and the Italian context are presented. The first is crucial for interpreting the analysis of the results, while the second helps to understand our task choice.

2.1 Theoretical background

Taking into consideration a student facing a mathematical task, we strongly believe that, in addition to task characteristics (Bolondi et al., 2018), the affective factors could highly influence his/her idea of the tasks. McLeod (1992) considers the three constructs of beliefs, emotions, and attitudes to describe the general term of affect. Among the works that address the need to develop theoretical frameworks on affect, we refer particularly to the study of Di Martino and Zan (2010) on attitude, as we recognized some similarities with their study in reading and analysing our students' answers. In their work, Di Martino and Zan read and analysed 1,600 essays, in which Italian students from first to thirteenth grade recounted their experience with mathematics. From their study a three-dimensional model about attitude in mathematics emerged. The model involves three strictly interconnected dimensions: Emotional Dimension, Vision of Mathematics, Perceived Competence.

According to their work, Emotional Dimensions deal with the liking/disliking of mathematics, but it also encompasses the essays in which students write explicitly about emotions such as love, anger, etc. This dimension thus refers to the emotional disposition of students in respect to mathematics, and can be characterised as being *positive* or *negative*.

The second category (Perceived Competence) is "marked by utterances like 'I succeed/fail in mathematics', 'I understand/don't understand mathematics', 'I get good/bad marks in mathematics'" (Di Martino & Zan, 2010, p. 38). This dimension could be labelled as *high* or *low*.

Thanks to the third category, called Vision of Mathematics "[...] some indications emerge, often through the writers' theories of success (Nicholls et al., 1990), that is their beliefs about what needs to be done to be successful in mathematics. In particular, an instrumental view can be spotted in theories of success which emphasise the role of memory and recall a vision of mathematics as a set of rules to be memorised" (Di Martino & Zan, 2010, p. 38).

This model was useful for better interpreting our results, as will emerge from the Discussion.

Moreover, from the students' answers and considerations the influence of the metacognitive aspects emerges. Metacognition is instrumental in building an appropriate representation of a given problem and monitoring the solution processes for solving it (Garofalo & Lester, 1985; Schoenfeld, 2016). Metacognition is also related to the decisions that a problem solver makes between different cognitive strategies when finding the solution, decisions which relate to their personal beliefs and values (Radmehr & Drake, 2017). Beliefs and values about learning, and problem solving are important in the encoding and retrieval of content knowledge (Radmehr & Drake, 2017).

In particular, metacognitive experience is "what the person is aware of and what she or he feels when coming across a task and processing the information related to it" (Efklides, 2008, pp. 279). Metacognitive experiences also include judgement of learning, estimation about effort and time that is needed and spent on the task, as well as estimating the correctness of the solution. Metacognitive experiences have an effect on decisions, which students make in learning situations regarding effort allocation, time investment or strategy use (Efklides, 2006).

2.2 Italian context

In Italy there are two cycles of education (www.miur.gov.it). The first cycle of education consists of two consecutive and mandatory school courses: primary school (lasting five years, for students aged 6 to 11 that corresponds to grade 1 to 5), and middle school (lasting three years, for students aged 11 to 14 that corresponds to grade 6 to 8). After finishing middle school, students have access to the second cycle of education, which ends at age 19 (grade 9 to 13). Mandatory education lasts 10 years (from age 6 to 16), and includes the eight years of the first cycle of education and the first two years of the second cycle.

In addition, in the Italian context, we have the possibility to track some students' difficulties over time thanks to INVALSI tests (tests with the purpose of measuring students' levels of competence in relation to the Italian curricular Guidelines) which were administered since 2008 in grades 2, 5, 8, 10 and 13 from the National Institute for the Evaluation of the Educational System (from 2009 to 2013, the tests also covered grade 6).

Up to now, the Italian Ministry of Public Education has established the standardised assessment of the Italian educational system, and commissioned the INVALSI (www.invalsi.it) to carry out annual surveys nationwide to all students in the second and fifth classes of primary school (grades 2 and 5), middle school third class (grade 8), and high school (grades 10 and 13). The INVALSI Institute carries out periodic and systematic checks on students' knowledge and skills (about reading comprehension, grammatical knowledge and mathematical competency), and on the overall quality of the educational outcomes from schools and vocational training institutions; in particular, it runs the National Evaluation System (SNV). The INVALSI standardised tests were created for system evaluation, and this is their primary purpose. The tests are administered every year at census level and student results are provided to each school institution. Results and questions of the INVALSI tests are considered as a resource also for researchers in the field of mathematics education (Garuti & Martignone, 2015) and are used in national and international research (e.g., Spagnolo et al., 2021b).

The SNV Framework is designed taking into consideration the Italian National Guidelines, in which argumentation is considered a competence goal for every school grade (Garuti & Martignone, 2019). For this reason, it is possible to find mathematical tasks that require the recognition or production of a justification, as the tasks we selected for our questionnaire design.

In particular, the tasks chosen in this study are INVALSI tasks; this choice has been made because it allows us to have some extra information that provides a further background for our study, such as the performance of Italian students (which is related to task peculiarities). In our study we focus on INVALSI tasks of grade 8 and grade 10, which in Italy are the transition years from middle school to high school.

3 Research question

The present study aims to outline some of the aspects that characterise the perceived difficulty of a mathematical task by students. This is a preliminary study with an exploratory function, and for this reason the research question (RQ) is broad.

> *RQ: What factors influencing students' perceived difficulty of mathematical tasks emerge from students' reflections?*

To answer this question, the study was divided into two phases that are explained in the next section.

4 Methodology

The study is qualitative and was carried out in two phases, both qualitative. Both phases included a first part of protocol collection (by protocols we refer to the students' answers given to a questionnaire that will be presented in the next section) and a subsequent phase of focus groups. The second phase was carried out with the specific aim of deeper investigating some interesting features that emerged during the first phase. In the following we present the descriptions of the sample, of the questionnaire, and, finally, in paragraph 4.3 we present the methods of analysis.

4.1 Sample description

The first experimentation (*phase 1*, carried out in October 2020) involved 79 students: two grade 9 classes and two grade 10 classes, from the same Italian school. The second experimentation (*phase 2*, carried out in October 2021) involved 69 students: three grade 9 classes from the same Italian school. Students involved in both phases of the study are the same age, but attend two different types of school: in phase 1 students attend a Humanistic school curriculum (called “Scienze Umane” in Italy), while in phase 2 students attend a Scientific school curriculum (called “Istituto tecnico” in Italy).

4.2 Questionnaire description

In both phase 1 and phase 2, students completed an online questionnaire which was followed by focus groups in each class. The questionnaire was administered during regular school activities. Students were not required to fill out the questionnaire, and numbers and analysis were done on the number of students who chose to answer the questionnaire. The questionnaire was administered in-person to students and was filled out through Google forms. The focus group was conducted remotely through Google Classroom (applet Meet) with the idea of helping categorise some of the answers given by the students.

The phase 1 questionnaire was composed of four sections. The questions in Section 1 aimed to investigate metacognitive factors and factors related to students' attitudes and beliefs, such as negative or positive attitudes towards mathematics. The questions in Section 2 and 3 referred to two specific mathematics tasks (represented in Figure 1). We asked the students to solve the tasks and, for each one, to respond to specific questions related to the perceived difficulties.

We choose two mathematics INVALSI tasks, because INVALSI tasks are *statistically validated* (Lazersfeld, 1958). We paid attention to argumentative questions relating to the *Numbers* area. With the help of the teachers of the classes involved in the experimentation, we selected tasks whose content had already been dealt with. This decision made it possible to exclude that the perception of difficulty was influenced by the fact that the students did not know the topic. The two chosen items involved mathematical similarities and differences. From one hand, the task chosen for Section 2 was a multiple-choice task that required recognition of a correct argumentation, while the task chosen for Section 3 was an open-ended task that required to produce an argumentation. On the other hand, for both items, the content was related to literal calculation and both tasks could be solved using the same strategy: proving the falsity of a statement through a counterexample. Final questions in Section 2 and 3 are the same, but for Task 1 and Task 2, respectively. For example, in Section 2 students were asked to evaluate from 1 to 10 the difficulty of Task 1, and in Section 3 they were asked to evaluate from 1 to 10 the difficulty of Task 2. The purpose of these additional questions was to inquire students' ideas and to link them – in a strictly qualitative way – with students' attitudes, beliefs or peculiar INVALSI items elements. Finally, in Section 4 we asked the students which of the two INVALSI tasks they consider more difficult and the reason why. Specifically, the questions in Section 2, 3 and 4 that were analysed for this study are shown in paragraph 6.1.

The phase 2 questionnaire was the same as in phase 1, with two more questions (related to the tasks represented in Figure 1): we asked students to explain why they assigned a specific level of perceived difficulty to Task 1 and Task 2. Such questions in phase 1 were carried out during the focus group session.

In phase 2 of the study we chose to have the students' explanations regarding why they attributed a specific level of perceived difficulty to Task 1 and Task 2 in written form.

Task 1

***n* is a natural number.**

Anthony affirms that "4*n*-1 is always a multiple of 3".
Is Anthony right?
In the table below, mark <u>the only</u> argument that justifies the correct answer.

Anthony is right...		Anthony is not right...	
A. ☐	because 4n-1=3n	C. ☐	because 4n-1 is always odd
B. ☐	because if n=4 then 4n-1=15	D. ☐	because if n=3 then 4n-1=11

Task 2

Mark states that, for every natural number *n* greater than 0, n^2+n+1 is a prim
Is Mark right?

Choose one of the two answers and complete the sentence.

☐ **Mark is right, because** ....

....

☐ **Mark is not right, because** ....

Figure 1. Task 1 belongs to Section 2 administered to Grade 08 Italian students by INVALSI in 2017 and task 2 belongs to Section 3 administered to Grade 10 Italian students by INVALSI in 2014.

4.3 Method of analysis

In this section we highlight our methodological strategy for addressing data management in a grounded theory study of the collected students' protocols, in particular considering students' answers to open-ended questions of the questionnaire.

In particular, the method we used is inductive: the categories of analysis were constructed by reasoning from the specific to the whole and focusing on the particular rather than the general. We based our conclusions on the database of protocols (students' responses consisted of rich descriptive data). The analysis of phase 2 (presented in Section 6.2.) started with explicit planning from the results of phase 1. There are significant regularities in our data collection and data analysis procedures.

Bearing in mind that there are few qualitative studies about students' perceived difficulty in performing a mathematical task (as we clarified in the introduction), constructive grounded theory (Charmaz, 1994, 2003) was used as our method of analysis. The theory shares some characteristics with quantitative methods (Creswell, 2005; Greckhamer & Koro-Ljungberg, 2005) but is positioned in the qualitative tradition. Grounded theory analysis procedures have been well documented in the methodology literature (Charmaz, 1990; Creswell, 1998, 2005; Harry, Sturges, & Klinger, 2005), and highlight the validity (and, some would argue, the objectivist underpinnings) of this research method. For a full discussion on the terrain, evolution, and developments of grounded theory, see Bruce (2007) and Mills et al. (2006). Concerning our qualitative study, the method used is inductive: reasoning from the specific to a whole and focusing on the particulars rather than the general. Qualitative researchers are expected to gather rich descriptive data and ground conclusions and understandings in the data mined, not prior theories (Bruce, 2007). On closer examination, however, it becomes apparent that qualitative studies often involve overt planning before the researcher launches into a main analysis. There are significant regularities in data collection and analysis procedures (Mills et al., 2006).

The categories that emerged from the analysis were compared with the categories in the theoretical framework of Di Martino, Zan (2010), highlighting the differences relative to the construct of perceived difficulty.

5 Analysis of INVALSI tasks

In this section we present the response results with regard to the two INVALSI tasks chosen. The results are shown both at the Italian national level and at the level of our study. The data of the national sample are gathered, analysed and released by INVALSI. Data can be found at www.gestinv.it. The percentage of correct/incorrect answers is one of the indicators used by INVALSI to define the level of difficulty of the task. Although it is not the purpose of the article to relate the level of difficulty defined by INVALSI to the students' perceived difficulty, we believe it may be of interest to the reader to receive some more information about the tasks and some of their peculiarities.

Following, we report the national results of the Tasks shown in Figure 1 (page 65).

Task 1 aimed to assess the ability to manipulate algebraic expressions by recognizing their properties in the set of Natural numbers and the ability to argue by acknowledging the correct argumentation. The percentage of correct answers of the Italian national sample was 40.3% (answer D), while the percentage of incorrect answers was 50.2% and of unanswered questions 9.6%. Those who gave the incorrect answer included 21.1% of students who chose C, that is 21.1% of students chose a true statement that did not support the conclusion. 20.6% of students chose A, showing a lack of control in literal calculation. Finally, 8.5% of students chose B, considering that one true example is sufficient to justify the answer.

We also report the percentages relative to the 148 students involved in the two qualitative phases of the study. In order to respect the subdivision adopted during the presentation of the analysis results, the results are presented with respect to the phase (1 and 2) of which they are part. As far as phase 1 is concerned, the percentage of correct answers was 32% (answer D), while the percentage of incorrect answers was 68% and no question was left unanswered. Those who gave the incorrect answer included 22% of students who chose A, 5.1% of students who chose B, and 41% of students who chose C. Regarding phase 2, the percentage of correct answers was 42% (answer D), while the percentage of incorrect answers was 58% and no question was left unanswered. Those who gave the incorrect answer include 36.2% of students who chose A, 11.6% of students who chose B, and 10.14% of students who chose C.

Task 2 aimed to assess the ability to manipulate the fundamental elements of literal calculus, the ability to interpret algebraic expressions recognizing their properties in the set of Natural numbers and the ability to argue using counterexamples. The percentage of correct answers of the Italian national sample was 17.8%, while the percentage of incorrect answers was 55.3% and of unanswered questions 26.9%. Additionally, in this case we report the percentages relative to the 148 students involved in the two qualitative phases of the study. The percentage of correct answers was 9% in phase 1, while the percentage of incorrect answers was 91%. In phase 2, the percentage of correct answers was equal to 13%, while the percentage of incorrect answers was 87%. We note that the percentage of uncorrected answers was extremely high. Although approximately half of the students understood that Mark is not right, they were not able to produce a valid justification.

Finally, we would underline that the statistics for the 148 students involved in phase 1 and phase 2 of the study were used for an initial qualitative categorisation of their responses. As specified in the previous section, we used Grounded Theory for the qualitative analysis.

6 Qualitative analysis of results

In this section results are presented separately with respect to the two phases: phase 1 and phase 2.

6.1 Analysis of phase 1

Regarding the questionnaire related to the first phase, we focus in particular on three questions of the Section 2, 3 and 4 of the questionnaire:

- Section 2 – D1: On a scale 1 to 10, how difficult did you find this [first] task?
- Section 3 – D2: On a scale 1 to 10, how difficult did you find this [second] task?
- Section 4 – D3: Compare the two tasks you addressed during this test. Which of the two tasks did you find more difficult?

Concerning questions D1 and D2, students' answers were distributed among all choice options, and a particular preference did not emerge (also the averages were quite similar for the two questions). In answering question D3, we observed a difference between the difficulty perception of the two mathematics tasks, as almost fifty percent of the students stated that they had found the second mathematics task more difficult.

Consequently, we investigated the consistency of the students' answers to D1, D2 and D3. The result is represented in the graph below.

Every point corresponds to one or more students' answers: the size represents how many students answered that way. In other words, bubble size is directly proportional to answer frequency. The numerous little points correspond to frequency 1: each of these answers were selected by a student. The biggest bubble corresponds to a frequency equal to 4, i.e., the answer

(5; 6) was chosen by four different students. The medium bubble corresponds to a frequency equal to 2, for example the response (4; 2) was selected by two different students. There is no bubble related to frequency 3, as it never happened that three students proposed the same answer. The cartesian plane coordinates represent the answers to the question D1 and D2. The abscissa value is the difficulty level attributed to the first INVALSI task (Task 1) and the ordinate is the value attributed to the second INVALSI task (Task 2). The colours (blue, yellow and red) represent the answer to the third question (D3), which is related to task comparison. As shown in the legend of Figure 2, the blue colour is related with the answers "The two tasks were difficult alike", i.e., a blue bubble represents a student who considered the two INVALSI tasks were not one more difficult than the other. The yellow colour is for students who answered that the first task was more difficult than the second task. Finally, the red colour represents the students who stated that the second task was more difficult than the first one. For example, the blue point with coordinates (10; 3) represents a student who attributed a difficulty level equal to ten to the first INVALSI task (Task 1), equal to three to the second INVALSI task (Task 2) and considered the two tasks difficult alike.

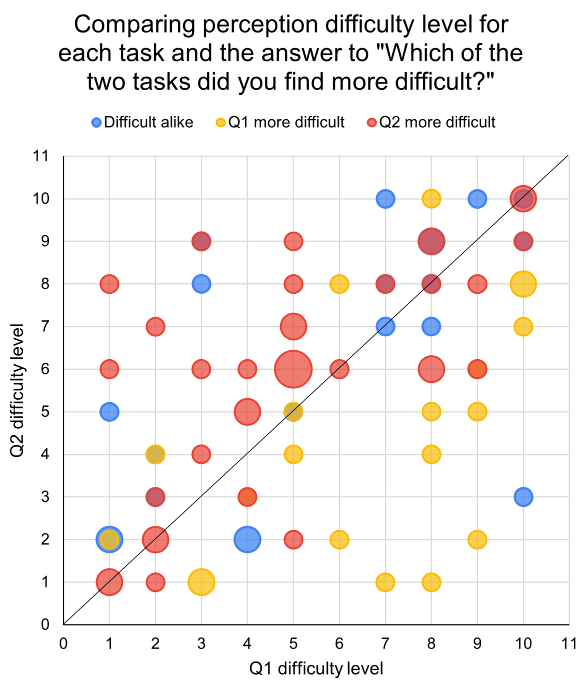

Figure 2. Comparison between phase 1-students' answers to D1, D2 and D3.

This graph confirms, at least for some students, the mismatch between the difficulty level chosen during the single item evaluations (i.e., the answers to D1 and D2) and the tasks' difficulty comparison (question D3). In fact, if, for each student, all three answers were consistent, all the blue points would belong to the bisector, all the yellow points to the x> y half plane and all the red points to the y> x half plane. Thanks to the graph, we can easily observe that this is not the case. Particularly, this erratic behaviour affects half of the students (50%). In our opinion, this could be the evidence of students' difficulty in evaluating a task, or it could indicate that students consider different factors during the individual tasks' evaluation or the comparison.

In phase 1, the reasons behind the students' choice of the difficulty level were discussed during whole classes focus groups. During the discussion different factors emerged, such as the students' previous experience with this kind of task, the students' difficulty regarding the mathematical content involved or consideration about text and task formulation. We believed that these observations were important in relation to the students' perception of difficulty, and for this reason we decided to make explicit the "Reason why" of their perception of difficulty in relation to the tasks in the second version of the questionnaire (administered in phase 2). These results will be discussed in the following section.

6.2 Analysis of phase 2

In order to investigate the previous discussed evidence that we found intriguing, we planned a second phase study (phase 2). The aim of the second phase was twofold. On one hand, we were interested to witness whether the mismatch would occurred again, on the other hand, we wanted to further inquire into the motivations for students to select one level of difficulty over another. In this section we present the results from our second study. They consist of two parts: the study of mismatch presented above for phase one, and the analysis of the answers given to the students to the questions that investigated their motivations. The questions taken into consideration for the second phase are the following:

- Section 2 – D1: On a scale from 1 to 10, how difficult did you find this [first] task?
  – D1b: Why?
- Section 3 – D2: On a scale from 1 to 10, how difficult did you find this [second] task?
  – D2b: Why?

- Section 4 – D3: Compare the two tasks you addressed during this test. Which of the two tasks did you find more difficult?

The analysis of questions D1, D2, and D3 is the same as the one for step one. The thematic analysis of the questions D1b, D2b was the novelty of this second phase.

As far as the inquiry of the mismatch between answers of D1, D2 and D3 is concerned, we found that approximately 43% of students (30 of 69 students) exhibited a conflicting behaviour. However, the graph built on the second phase results shows great differences from the one built on the results of the first phase. Figure 3 represents the graph concerning the second phase of the study.

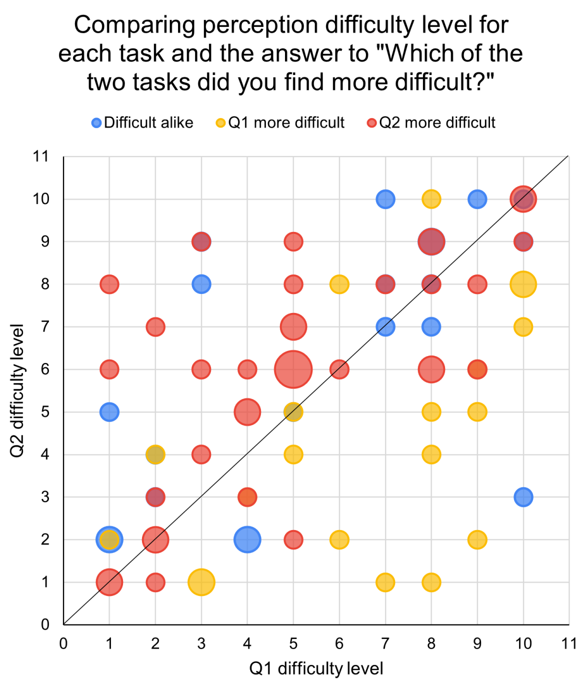

Figure 3. Comparison between phase 2-students' answers to D1, D2 and D3.

The graph reveals that this time, students' answers are less scattered across all responses. Moreover, the points that represent the mismatched answers are concentrated near the bisector. We see this as a sign that students were paying more attention (consciously or not) to consistently answering these questions.

We will now present the process and the results of the analysis of questions D1b and D2b. In the following we present by way of example some students' answers[1]. The analysis followed a *grounded approach*, namely we categorised students' answers to questions D1b and D2b in relation to the main aspect(s) that were explicitly mentioned in these answers. The focus of the thematic analysis was the content of students' answers. Different aspects emerged, for example many students explicitly referred to the *time* factor, as in the answer "Because I was able to solve it quickly", or to the procedure used to solve the problem, as in the answer "Because it does not require complicated calculations". In addition, we found explicit references to emotions, for example "I get anxious even if there is no grade", or to the fact that they were not confident about the given answer, in fact some students stated that "I am not sure of the answer". Moreover, some students referred to their previous experience with similar questions, as in the answer "because it is not the first time that I have been asked questions like this". Answers have been read multiple times, and categories have been designed and modified gradually.

As a result of this first analysis phase, we had many categories related to the main aspect that students mentioned in their answers. We then considered these categories and reread the responses to unify, compare, and try to address the main aspects that these categories referred to. Four supercategories have emerged as a result of this second phase of the analysis:

1. Resolution strategy
2. Capability and experience
3. Emotions
4. Task Formulation

In addition, a fifth category grouped together all those responses in which the answer was missing or students stated "it was just hard" and no reason was provided.

In "Resolution strategy" we grouped together those answers in which students explicitly referred to the kind of strategy or process that, in students' view, was needed to solve the problem. The attention then was on *what students need to do to get the solution*. This category presented some nuances. Different aspects were in fact highlighted in students' texts. Many students referred to calculus, or to the fact that a *reasoning* was needed. For example, stating [the first task was difficult 1, because] "The calculus was easy", or [the second task was difficult 2, because] "There weren't so many calculations and it was enough to think a moment"; or again, [The second task was difficult 2, because] "It was purely logical". Other students referred explicitly to the fact

[1]The answers, collected in Italian, are translated by the authors.

that the tasks could be solved by means of example. For example, a student answered that the first task was difficult 3 and she/he stated that "It wasn't very difficult because it was enough to try several numbers and state whether the question was false or true". Another student claimed [the second task was difficult 5, because] "because I had to try several numbers before I found the answer".

The second category, "Capabilities and experience" is the most widespread and concerns answers that referred to students' perceived capabilities or competence and previous experience that affected their perceived difficulty of the task. This category includes students that referred explicitly to the fact that they were (not) familiar with this kind of task, as in the answer: [the first task is difficult 2] "because it is not the first time that I have been asked questions like this", or [the second task is difficult 10, because] "I never faced similar problems". This was somehow reinforcing the idea that a problem is easier if it is similar to something already known. Moreover, this category contains the answers which referred to what students were (not) able to do or what they (did not) know. The attention of these answers is on students' self-perception, in general, or in referring to these tasks. For example, a student stated that [the first task is difficult 9, because] "I am not a logical person" and [the second task is difficult 10, because] "I do not know how to do it, I have fundamental gaps". The category also includes students who were (not) *sure* about what they know or did, as [the second task is difficult 8] "Because I am not sure about the answer", or answers in which students presented a consideration about the correctness/incorrectness of their answer [the second task is difficult 9] "[...] because it's definitely wrong and I didn't fully understand the reasoning to be done". Finally, this category concerns answers in which students referred to the fact they did (not) solve the task easily or smoothly. These responses usually referred, more or less explicitly, to some obstacles students encountered in tackling the problems, or to the time students invested in solving the task. Generally, the task was perceived easier if students were engaged for a short time or if they reported that they had an insight and that they solved the problem on the first try. Answers of this kind were for example, [the first task is difficult 2] "Because I was able to solve it quickly" or [the second task was difficult 3, because] "I figured it out right away". Otherwise, some students answered [the first task was difficult 6, because] "Because it took me some time to think it through", or [the second task was difficult 4, because] "because it took me a long time to get the answer", or again [the second task was difficult 3] "Because at first I didn't understand how I had to start".

The category "Emotions" refers to the fact that students explicitly considered their emotions in motivating the difficulty level they chose. This category is smaller than the previous ones, and we can find only a few responses here. However, we decided to have this category because these responses presented some peculiar aspects that could hardly be included in the remaining categories. This category includes, for example, the answers:

- [The first task is difficult 5, because] "I get anxious even if there is no grade"
- [The first task is difficult 8, because] "I'm afraid I've made a mistake".

It is perhaps worth noting that only negative emotions are highlighted. Moreover, no answer to question 2b falls into this category.

The fourth category represents considerations about the formulation of the task, in particular with respect to the text. For example, a student stated that [the Task 1 is difficult 3, because] "it was a little tricky for me to understand the text, but once I understood it, it was easy to give an answer", or similar. Additionally, this category is very small and includes only answers to the question 1b. As for the previous one, we think however it shows characteristics that are peculiar and that need to be considered. In fact, despite the fact that in the students' written responses the category emerged marginally, during the focus groups conducted after the questionnaire, students themselves commented about the text and the formulation of the tasks. For example, students noticed that one task was a multiple choice one (Task 1) and the other was an open-ended question (Task 2). Moreover, some students claimed that Task 2 was more complex because they had to write their own answers instead of choosing among the different options provided.

Finally, we note that these categories are not intended as exclusive, and some answers could be categorised referring to more than one category.

6.3 Discussion analysis of phase 2

Presenting our categories, it is impossible to discuss them without referring to a fundamental model for aptitude in mathematics proposed by Di Martino and Zan (2010) and presented in Section 2.1 of this work. In the following, we briefly underlying similarities and differences between their model and our findings.

Our work differs from the study carried out by Di Martino and Zan (2010). In particular, we focus on problem resolution, in respect to a particular mathematical content and to selected tasks that require a justification. In our study, students' answers were briefer, and referred to specific situations. Students were not free to talk about their experience with mathematics in general but focused on the task and on their experience in solving them. Despite this, the analysis of the students' responses to questions 1b and 2b of our questionnaire reflect, in our view, some of the results found by Di Martino and Zan. Firstly, an emotional dimension is presented, although in our case it was not the crucial aspect of the responses. Secondly, a certain vision about mathematics emerged from students' answers. In particular, students referred to methods that, in their perspective, were required to solve the problem, by paying attention to calculus and procedures, or to the importance of reasoning and reflecting. Thirdly, also in our case, students referred, more or less explicitly, to their perceived competences in solving the selected tasks, and some ideas about their perceived knowledge and abilities might be inferred. However, some differences between their model and our categories cannot be denied. In particular, the *time reference*, or the fact that students could solve the problem following their first idea are peculiar in defining the difficult perception. Such aspects we think may be related to metacognitive factors (Radmehr & Drake, 2017). Specifically, the estimation about effort and time that is needed and spent on the task, as well as estimating the correctness of the solution can be categorised as metacognitive experiences (Efklides, 2006). In addition, the *experiences* play an important role in defining what is a (not) easy task. Moreover, considering a specific task, the consideration about task formulation or specific characteristics of the text had emerged.

7 Discussion and concluding remarks

We have highlighted in previous research that a student's perceived difficulty seems to not be related to being able to correctly answer the question (Saccoletto & Spagnolo, *in press*).

In this paper we highlight a fundamental feature that emerges from the analysis of the qualitative questionnaire: when a student expresses his/her perceived difficulty in relation to a single task or in comparing several tasks, we could get different results (not necessarily consistent). This suggests that task characteristics alone are not sufficient to understand the students' perceived difficulty of mathematical tasks, as also the focus groups conducted in the first phase (Saccoletto & Spagnolo, *in press*) suggested. The categories that emerged from the analysis allow to clarify some of the fundamental aspects involved when a student expresses his/her perceived difficulty in relation to mathematical tasks. In particular, in assigning a level of perceived difficulty to a task, students seemed to be influenced by factors more closely related to the task (such as text elements), by factors related to their attitude or their emotions and by metacognitive aspects (such as lack of ability to judge their own skills, knowledge, and abilities). Further studies with a broader sample will help us to move towards two directions: finding new aspects that influence students' perceived difficulty, highlighting additional aspects of perceived difficulty and hence broadening the aspects taken into consideration; characterising in more depth the different aspects that emerged.

This preliminary study can be also developed in several directions. We believe, for example, that it may be interesting to investigate the perceived difficulty even before solving the task and relate it to the perceived difficulty after solving it. Furthermore, the analyses developed from the data represented in Figures 2 and 3 show that by comparing students' perceived difficulty between two tasks, the results are not consistent. So, we ask whether it is possible to classify tasks according to the students' perceived difficulty. Starting from the qualitative phase results, we could build an adaptive questionnaire, and we will inquire whether it is possible to arrange tasks (more than 2) in order of difficulty.

In addition, we are currently setting up a further study in order to examine the phenomenon quantitatively. Finally, we think that it would be interesting to understand how the perceived difficulty of students is related to perceived difficulty of the teachers and whether it matches.

References

Bolondi, G., Branchetti, L., & Giberti, C. (2018). A quantitative methodology for analyzing the impact of the formulation of a mathematical item on students learning assessment. *Studies in Educational Evaluation*, *58*, 37–50.

Bruce, C. (2007). Questions arising about emergence, data collection, and its interaction with analysis in a grounded theory study. *International Journal of Qualitative Methods*, *6*(1), 51–68.

Charmaz, K. (1990). Discovering chronic illness: Using grounded theory. *Social Science & Medicine*, *30*, 1161–1172.

Charmaz, K. (1994). Identity dilemmas of chronically ill men. *Sociology Quarterly*, *35*, 269–288.

Creswell, J. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Sage.

Creswell, J. (2005). *Educational research: Planning, conducting, and evaluating qualitative research* (2nd Edition). Pearson Education.

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: a review of linguistic and numerical factors contributing to their difficulty. *Frontiers in psychology*, *6*, 348.

De Corte, E., Verschaffel, L., & Van Coillie, V. (1988). Influence of number size, problem structure, and response mode on children's solutions of multiplication word problems. *Journal of Mathematical Behavior*, *7*(3), 197–216.

Di Martino, P., & Zan, R. (2010). 'Me and maths': Towards a definition of attitude grounded on students' narratives. *Journal of mathematics teacher education*, *13*(1), 27–48.

Efklides, A. (2006). Metacognition and affect: what can metacognitive experiences tell us about the learning process? *Educational Research Review*, *1*(1), 3–14.

Efklides, A. (2008). Metacognition: defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist*, *13*(4), 277–287.

Garofalo, J., & Lester, F. K. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for research in mathematics education*, *16*(3), 163–176.

Garuti, R., & Martignone, F. (2015). The SNV (INVALSI) experience. Teaching and learning mathematics: resources and obstacles. *Proceedings of CIEAEM 67 – Quaderni di ricerca diddattica*, *25*(2), 95–98.

Garuti, R., & Martignone, F. (2019). Assessment and argumentation: an analysis of mathematics standardized items. In U. T. Jankvist, M. Heuvel-Panhuizen, & M. Veldhuis (Eds.). *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education* (pp. 4075–4082). Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.

Greckhamer, T., & Koro-Ljungberg, M. (2005). The erosion of a method: Examples from grounded theory. *International Journal of Qualitative Studies in Education*, *18*(6), 729–750.

Harry, B., Sturges, K., & Klinger, J. (2005). Mapping the process: An exemplar of process and challenge in grounded theory analysis. *Educational Researcher*, *34*(2), 3–13.

Lazarsfeld, P. F. (1958). Evidence and inference in social research. *Daedalus*, *87*(4), 99–130.

McLeod, D. B. (1989). The Role of Affect in Mathematical Problem Solving. In D. McLeod & V. M. Adams (Eds.), *Affect and mathematical problem solving: A new perspective* (pp. 20–36). Springer.

McLeod, D. (1992). Research on affect in mathematics education: a reconceptualization. In D. Grows (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 575–596). McMillan.

Mills, J., Bonner, A., & Francis, K. (2006). The development of constructivist grounded theory. *International Journal of Qualitative Methods*, *5*(1), Article 3.

Nicholls, J., Cobb, P., Wood, T., Yackel, E., & Patashnick, M. (1990). Assessing students' theories of success in mathematics: individual and classroom difference. *Journal for Research in Mathematics Education*, *21*(2), 109–122.

Radmehr, F., & Drake, M. (2017). Exploring students' mathematical performance, metacognitive experiences and skills in relation to fundamental theorem of calculus. *International Journal of Mathematical Education in Science and Technology*, *48*(7), 1043–1071.

Schoenfeld, A. H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics (Reprint). *Journal of education*, *196*(2), 1–38.

Saccoletto, M., & Spagnolo, C. (in press). Difficulty perception in answering argumentative INVALSI tests: a qualitative study. Volume The school and its protagonists: the students. V Seminar "INVALSI data: a tool for teaching and scientific research". *FrancoAngeli*.

Spagnolo, C., Capone, R., & Gambini, A. (2021a). Where do students focus their attention on solving mathematical tasks? An eye tracker explorative study. In M. Inprasitha, N. Changsri, & N. Boonsena (Eds.), *Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 89–96). PME.

Spagnolo, C., Giglio, R., Tiralongo, S., & Bolondi, G. (2021b). Formative Assessment in LDL: A Teacher-training Experiment. In B. Csapó & J. Uhomoibhi (Eds.), *Proceedings of the 13th International Conference on Computer Supported Education* (Vol. 1, pp. 657–664). Scitepress.

Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representation in arithmetic word problem solving. *Quarterly Journal of Experimental Psychology*, *58*, 1311–1323.

Vicente, S., Orrantia, J., & Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British Journal of Educational Psychology*, *77*(4), 829–848.

Zan, R., Brown, L., Evans, J., & Hannula, M. (2006) Affect in mathematics education: an introduction. *Educational studies in mathematics*, *63*(2), 113–121.

Trudności zadań matematycznych, postrzegane przez uczniów: badanie czynników wpływających na odczuwaną trudność

Streszczenie

W artykule przedstawiono główne wyniki badania jakościowego koncentrującego się na postrzeganych przez uczniów trudnościach po rozwiązaniu zadań matematycznych (uczniowie klas 9 i 10). Celem badania było zidentyfikowanie czynników, które wpływają na postrzegane przez uczniów trudności. Chociaż czynniki przyczyniające się do rosnącej lub malejącej trudności zadania są szeroko omawiane w literaturze, trudności postrzegane przez uczniów dotyczące zadania matematycznego są rzadziej analizowane. Analiza kwestionariusza i dyskusji grupowej przeprowadzonej z uczniami podkreśla kilka ważnych refleksji na temat wpływu czynników metapoznawczych, afektywnych i zadaniowych na uczniów.

Marta Saccoletto
University of Turin
Turin, Italy
e-mail: *marta.saccoletto@unito.it*
ORCID: *0000-0002-7417-4383*

Camilla Spagnolo
Free University of Bozen-Bolzano
Brixen-Bressanone, Italy
e-mail: *Camilla.Spagnolo@unibz.it*
ORCID: *0000-0002-9133-7578*