



Does this Answer Make Sense? Primary School Students and Adults Judge the Reasonableness of Computational Results in Context-Based and Context-Free Mathematical Tasks

Dimitrios Deslis¹ · Despoina Desli²

Received: 11 May 2021 / Accepted: 16 January 2022 / Published online: 5 February 2022
© The Author(s) 2022

Abstract

This study investigates students' and adults' performance in judging reasonableness of computational results, namely reflecting on whether these results qualify as acceptable answers to mathematical tasks. Data was gathered via task-based questionnaires from 160 participants, evenly divided between fifth-graders and adults. Their responses to a systematically varied collection of context-free and context-based items shed light on their performance and strategies in relation to two inter-related aspects of reasonableness: number relationships and the effect of operations (*internal reasonableness*) and the practicality of answers (*external reasonableness*). The results revealed that students were better at utilising the former criterion compared to the latter, while for adults, the opposite was found. However, adults clearly outperformed students in both aspects. Task characteristics such as the involved numbers and operations, and the correctness or incorrectness of computation results, were found to be associated with fluctuations in performance, although the effect differed between context-free and context-based items. Algorithm- and rule-based strategies were more popular than sense-making ones especially among students, although only the latter type was correlated with increased correct response rates. Students often failed to obtain results that made sense considering real-life situations or the involved numbers and operations. Instead, they blindly trusted algorithms and rules, even when their use was inappropriate or insufficient. The findings highlight that to help students refine their ability to judge reasonableness, mathematics classrooms should prioritise the development of number sense and promote the search for connections between school mathematics and everyday life.

Keywords Computational estimation · Context-based and context-free tasks · Judging reasonableness · Mental computation · Routine and adaptive expertise · Sense-making in primary school mathematics

✉ Dimitrios Deslis
dd544@cam.ac.uk

Extended author information available on the last page of the article

Introduction

The *judgement of reasonableness* in school mathematics usually refers to an evaluation of a computational result based on a critical judgement as to whether it can qualify as a sensible and acceptable answer to a mathematical task. It is an advanced metacognitive ability that requires critical reflection upon a result and the steps of the process through which this has been obtained (Schoenfeld, 1985). The ability to judge reasonableness can be conducive to problem solving, since solvers can utilise it to avoid mistakes, refine strategies, and adjust answers (Dougherty & Crites, 1989; Pólya, 1973). The ability to reflect critically on results and judge their reasonableness is not only associated with several key concepts in school mathematics, including number sense (e.g., Reys et al., 1999), sense making (Bonotto, 2005), mental calculation (e.g., Thompson, 1999), and computational estimation (e.g., Charles & Carmel, 2005), but can also have wide application in a variety of everyday life situations (Alajmi & Reys, 2007): “Did I get the correct change back?” “Did I set my alarm clock early enough to make sure I’ll be on time?”. In past years, researchers have shown a growing interest in the investigation of this concept in school contexts (e.g., Alajmi & Reys, 2010; Yang, 2017) which has also been acknowledged as a pivotal element in students’ learning of mathematics in school (National Council of Teachers of Mathematics [NCTM], 2000).

Theoretical Background

Drawing on previous theoretical discussions about the facets of reasonableness (Alajmi & Reys, 2007; Dougherty & Crites, 1989), we base our theoretical framework on the widely accepted distinction between two interrelated aspects of judging reasonableness (Fig. 1). These two aspects can also serve as criteria on which a solver can rely to evaluate and refine his/her answers.

The first criterion refers to the consistency of a computational result with expectations relevant to the size and properties of the involved numbers, the relationships between them, and the effects of operations on them. We name this criterion *internal reasonableness* as it answers the question “Does the result make sense considering the characteristics of the involved numbers and operations?”. Judgements based on this criterion are applicable even to context-free tasks. Imagine that after the execution of the algorithm for the multiplication 10.1×0.9 , the decimal point is misplaced, and the obtained result is 90.9. Mature solvers can reflect on whether this result can be correct and spot the mistake following various paths. Mental computations can be used to cross-check the correctness of the result. For example, after transforming the original expression into $(10 + 0.1) \times 9/10$ and applying the distributivity property, it becomes clear that the correct result is 9.09. Mathematical properties can also be used to judge reasonableness. For instance, one could expect that multiplying a positive number (like

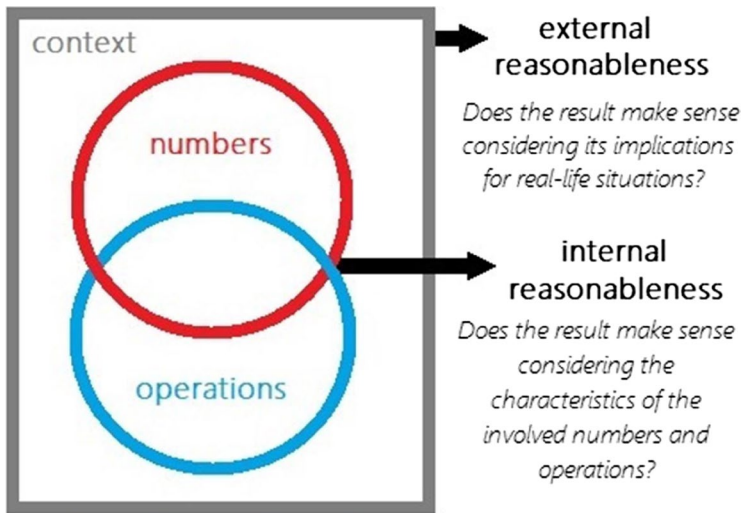


Fig. 1 The aspects of judging reasonableness: internal and external reasonableness

10.1) by a factor less than 1 (like 0.9) should decrease its value, and thus, only results less than 10.1 may qualify as acceptable. Moreover, rounding off 10.1 and 0.9 to 10 and 1, respectively, one can estimate that an acceptable result should be around 10. In fact, there is a two-way link between computational estimation and the judgement of reasonableness. Computational estimates may begin with an approximation of the result that can be obtained from an operation and conclude with the evaluation of the estimation as to whether it is close enough to the starting approximation. This may confirm the success of the estimation or lead to its refinement (Bonotto, 2005; Sowder & Wheeler, 1989). Additionally, computational estimation can be used to examine whether the results produced by calculators or by applying algorithms are reasonable (Charles & Carmel, 2005), like in the previous scenario. As shown by the preceding illustrations, the effective use of mental calculation, computational estimation, or properties in order to judge reasonableness requires advanced number sense (Alajmi & Reys, 2007; National Council of Teachers of Mathematics [NCTM], 2000), creativity, and flexibility (McMullen et al., 2020; McIntosh et al., 1997; Reys et al., 1999).

The second criterion refers to the practicality of a computational result, namely to its *external reasonableness*, as it answers the question “Does the result make sense considering its implications on real-life situations?”. Judgements based on it draw upon the context of a problem and thus are associated with context-based tasks. Consider these two problems: “How many buses with a capacity of 30 passengers are needed to transport 105 people?” “We have 105 candies and want to divide them fairly to 30 children. How many candies will each child get?”. In both cases, the application of the division algorithm gives 3.5 as a result, although it is not considered an acceptable answer to either of them. Since it is impossible

to rent half a bus, we would need 4 buses, even if one remains half-empty. However, as we cannot afford giving 4 candies to each child, each child would get 3, and inevitably there will be some candies left over. Although these problems involved the same numbers and operations, and the algorithm produced identical results, these needed to be adapted considering the context of the problem resulting in a different outcome in each case. Generally, in context-based problems, solvers should not only evaluate the appropriateness of answers based on the first criterion (internal reasonableness), but they should also reflect on whether the answers make sense in real-life contexts (external reasonableness). The degree to which solvers consider this aspect in their judgements highly depends on the richness of their real-life experiences as well as their ability to rely upon them to make appropriate judgements (Alajmi & Reys, 2010; Masingila et al., 1996). Although contextualised tasks are not necessarily more interesting or engaging for students than decontextualised ones (Beswick, 2011), nor always associated with increased performance (Can & Özdemir, 2020), the presence of context can potentially support connections with real life (Meyer et al., 2001; Sowder & Schappelle 1989), and thus enable reflections on the external reasonableness of their answers.

The internal and external reasonableness distinction fits well with that between routine and adaptive expertise (Hatano, 1988; McMullen et al., 2020). *Routine expertise* is associated with basic knowledge in mathematics which does not necessarily presuppose a deep understanding of the concepts, although it is sufficient for tackling familiar tasks. In contrast, *adaptive expertise*, which is more advanced and harder to reach, reflects the rich conceptual knowledge and in-depth understanding of concepts that are usable even in novel situations. Although routine expertise may suffice for low-demanding tasks, one needs to acquire and hone adaptive expertise in order to break free from routines when necessary (see Table 1).

Past studies with a specific focus on students' abilities for judging reasonableness are limited. However, there exists evidence suggesting that students at all school levels typically lack many of the competencies required for obtaining reasonable results. For instance, Alajmi and Reys (2010) used written tests and interviews to examine 200 eighth-graders' ability to recognise reasonable answers. Participants performed consistently low across the items on practicality, and number relationships and the effect of operations. Most students tended to over rely on algorithmic strategies (over 60%) and were generally unable to pursue connections with real-life situations. In a more recent paper, Yang and Sianturi (2019) reported that the mean score of the 942 sixth-graders from Hong-Kong who joined their study was very low (3.45/8). Their answers reflected various misconceptions relevant to both aspects of reasonableness. Most participants used ineffective strategies, and only one-fifth chose to employ strategies based on number sense. Consequently, more than 60% were unable to obtain reasonable results. An earlier study (Yang, 2017) with 790 Taiwanese fourth-graders led to similar findings. The average correct response rate was about 50%, and only one-fourth of them applied number-sense-based methods to recognise reasonable answers. Misconceptions regarding the meaning of responses in real-life contexts were also widespread. For instance, a total of one-third considered the sentences "I can fit 5,000 textbooks into my school bag" and "I

Table 1 Routine and adaptive expertise in judging reasonableness

| | Routine expertise | Adaptive expertise |
|-------------------------|--|--|
| Internal reasonableness | <ul style="list-style-type: none"> -Ability to apply known algorithm- and rule-based strategies -Sufficient for obtaining reasonable results in conventional tasks that usually aim at the practice of previously taught calculation techniques | <ul style="list-style-type: none"> -Ability to flexibly use knowledge about numbers and operations where the characteristics of the involved numbers and operations play an integral role in determining which results may qualify as acceptable answers and may require critical reflection upon the known calculation techniques |
| External reasonableness | <ul style="list-style-type: none"> -Ability to apply known problem-solving routines or algorithm-based procedures -Sufficient for obtaining reasonable results in conventional tasks that usually aim at the practice of previously taught problem-solving techniques, and in which the presence of context is typically immaterial to determine which responses may qualify as acceptable | <ul style="list-style-type: none"> -Ability to flexibly use problem-solving skills and develop novel strategies, when necessary -Required for obtaining reasonable results in unconventional tasks where the context plays an integral role in determining which results may qualify as acceptable answers and may require critical reflection upon the known problem-solving techniques |

can fit 5,000 M&Ms into my mouth” as more reasonable than the sentence “I can lift a pig that weighs 5,000 g”.

Additional enlightening findings come from studies that concentrated on concepts closely related to the judgement of reasonableness, albeit without explicitly focusing on it. For example, Menon (2004) studied 750 students’ (aged from 13 to 17) ability to use number sense. In many items, strikingly unreasonable responses were alarmingly popular among students of all ages. As an illustration, only 41% placed the decimal point correctly into the result of 15.24×4.5 displayed as 6858 in a “broken” calculator, while only half of them chose computational estimation over algorithms to do so. As illustrated by this example, students’ difficulties in achieving reasonable results often stem from their weakness in using mental computations and computational estimation flawlessly or their reluctance to do so altogether (Desli & Lioliou, 2020; Heirdsfield & Cooper, 2002). In another item of Menon’s (2004) study, more than half of the participants (56%) considered $2\frac{1}{4}$ as the correct answer to the number of vehicles required for the transportation of 9 passengers. Such responses are indicative of how a lack of consideration of the effects of a result on real-life contexts may lead to unreasonable answers. An earlier study conducted by Masingila et al. (1996), which focused on the differences between the use of mathematics in school and everyday life, also points towards a similar conclusion. The findings suggest that even at the age of 17, students fail to consider whether an answer to a problem makes sense in the real world. To illustrate, in this study, a pair of students and a restaurant manager were asked to change a cooking recipe initially made for 6 portions into a recipe for 10 or 20 portions. Unlike the professional who approached the problem creatively aiming at a practical solution, students’ reasoning prioritised the accuracy of measurements generating unpractical results (e.g., $\frac{2}{3}$ eggs, $\frac{5}{6}$ cups of carrots) which they did not attempt to adjust in the context of an everyday situation. This probably results from the great emphasis that is usually placed on the teaching and learning of standard algorithms in conjunction with the limited attention paid to seeking connections between the mathematical content and life outside the classroom (Mann, 2006; Verschaffel et al., 2007).

Overall, previous findings suggest that students tend to encounter numerous challenges in distinguishing between reasonable and unreasonable results, which span both internal and external reasonableness. Among the factors that contribute to their unsatisfactory performance appears to be their tendency to prefer unproductive algorithmic strategies over number-sense-based ones, as well as their failure to make connections between mathematics and the real world.

Importance of Study and Research Questions

Reasonableness is being increasingly considered by researchers in the field. However, research with an explicit focus on its underlying demands, such as the present study, remains scarce. Instead, reasonableness has been typically treated as one aspect among many others in studies elaborating on various topics, including number sense, sense making, and computational estimation (e.g., LeFevre et al., 1993; Markovits & Sowder, 1994; Menon, 2004; Yang, 2005). Thus, our knowledge about

which factors or task characteristics impede or facilitate one's ability to give reasonable answers to mathematical tasks is limited. Additionally, despite the intriguing results generated by a few studies with primary school students (Yang, 2017; Yang & Sianturi, 2019), relevant studies usually concentrate on school levels other than primary school (e.g., Alajmi & Reys, 2007, 2010). Furthermore, previous studies have typically concentrated on the study of either students (e.g., Yang & Sianturi, 2019) or teachers (Alajmi & Reys, 2007). However, none of the previous studies has pursued comparisons between the strategies and performance of students and the general population, particularly adults who are not necessarily teachers. The study of a group of students in parallel with a group of adults may enable inferences not only about how the judgement of reasonableness alters with age, but also whether the engagement with mathematics in the context of school environment and norms, or lack thereof, and in the context of everyday life situations, or lack thereof, differentiates participants' performance and strategies. Another novel contribution is the examination of the strategies used to judge reasonableness through the lens of routine and adaptive expertise. Finally, the reasonableness of results is often considered in relation to either the effect of operations and number relationships (Yang, 2005) or their practicality (Masingila et al., 1996).

The present study aspires to contribute to the body of literature that investigates the intersection of the two topics (e.g., Alajmi & Reys, 2010), as we believe that they are interrelated and equally important. Considering the above, the present study seeks to address the following research questions: (a) To what extent are fifth-graders and adults capable of judging the reasonableness of computational results? What are the competencies and difficulties of each age group in relation to the different aspects of reasonableness? (b) How is the performance of fifth-graders and adults associated with the characteristics of tasks, including the involved numbers and operations, and the sufficiency or insufficiency of algorithms to generate correct responses? (c) What strategies do fifth-graders and adults apply to judge the reasonableness of computational results and how are these strategies correlated with their performance?

Method

Participants

A total of 160 participants were selected to participate in this study through a combination of convenience and purposive sampling: accessible participants were recruited, ensuring that the sample was evenly divided in terms of age (80 fifth-graders, coded as S1-S80; 80 adults, coded as A81-A160) and gender (80 males; 80 females). The fifth-graders (average age: 10 years 10 months) were studying at two primary schools in Thessaloniki, Greece; their selection was conditional upon achieving a varied sample in terms of socioeconomic background and academic performance. The selection of the adult participants (age range: 18–64 years, 25% over 46 years of age) also aimed at a varied educational background: the majority had graduated from middle school (98.7%), and significant proportions also held higher

education qualifications (72.4%), including master's (10%) and doctoral degrees (2.5%).

These age groups were chosen for two reasons. Firstly, in contrast to very young learners, fifth-graders have sufficient mastery of numbers and operations since they normally receive instruction on both natural and decimal numbers from first and third grades onward, respectively. However, instruction is in line with typical mathematics practices in Greece and focuses on fluent application and computation of the algorithms for the operations, while judging reasonableness is not included in the most recent primary school mathematics curriculum (Greek Ministry of Education [MINEDU], 2003). Secondly, although they are not taught mathematics formally anymore, older participants frequently practise computations in their everyday activities.

Design and Instrument

Data were collected through task-based questionnaires which included demographic questions and two tasks, each consisting of eight items, following a cross-sectional design. In all items, the participants were asked to provide an answer alongside an explanation of their thinking. The reason why we asked participants to do both is because reasonableness and correctness are interrelated, but not equivalent terms. In fact, a result may be reasonable despite being erroneous. For example, 6364 is clearly an unreasonable result for 7×9092 ; in contrast, although erroneous, 63,654 might appear as reasonable at first glance. In our design, we decided to avoid boundary cases like this: all correct results were also reasonable, whereas all the erroneous results were also unreasonable for at least one apparent reason (e.g. in the previous example: $6364 < 9092$). Moreover, the requirement for the justification of answers enabled us to know whether participants made appropriate judgements about the *correctness* of results accidentally or based on appropriate judgements about their *reasonableness*, which was the focus of our study. Our pilot study showed that the trial items were interpreted correctly by the participants and from their responses it was clear if (and how) they attempted to judge reasonableness.

The tasks were designed for the needs of the study, and each was targeted towards a different aspect of reasonableness. In Task 1 (internal reasonableness), participants were asked to decide whether eight computational results of horizontal multiplications and divisions were true or false and justify their answers (e.g., $74.8: 3 = 26.2$ "Can this result be correct? Choose *True* or *False* and justify your answer."). In Task 2 (external reasonableness), participants were asked to give justified answers to eight items which were (or on the surface appeared to be) multiplication or division word problems (e.g. *Nick is 10 years old and 1.30-m tall. How tall will Nick be at the age of 20?* "Solve the problem and justify your answer."). In order to respond correctly to Task 1, the participants had to take into account the characteristics of the numbers and operations involved, while in Task 2, they had to consider whether the results made sense in the real world.

Both tasks included a systematic variation of items, based upon three binary classification conditions: (1) number type: *natural* (N) or *decimal* (D) numbers; (2)

Table 2 The 16 items of the questionnaire

| Correct computation result | | Erroneous computation result | |
|----------------------------|---|--|---|
| | Multiplication | Division | Multiplication |
| Task 1 | Natural numbers $709 \times 50 = 35,450$ | Division $696 : 4 = 174$ | Multiplication $7 \times 9,092 = 6364$ |
| | Decimal numbers $9.85 \times 11.04 = 108.744$ | $486.2 : 1.1 = 442$ | $25.3 \times 5 = 107.3$ |
| Task 2 | Natural numbers 4 balls cost 80€. How much do 8 such balls cost? | 4 cereal bars cost 2€. How much does 1 bar cost? | 2 shirts need 4 hours to fully air-dry. How long does it take for 6 shirts to air-dry? |
| | Decimal numbers 1 coffee bag weighs 0.4 kg. How much do 3 such coffee bags weigh? | 12.4 l oil is equally distributed among 4 containers. How many litres will each one of them contain? | At the age of 10 Nick is 1.30 m tall. How tall will Nick be at the age of 20? |
| | | | How many 30-seat buses are needed to transport a group of 315 travelers? 1 tissue pack costs 0.40€. How many packs can we get with 5€? |

The descriptions of the items in Task 2 have been abridged.

arithmetic operation: *multiplication* (\times) or *division* (\div); and (3) computation result: *correct* (\checkmark) or *erroneous* (\times) (i.e., whether applying the multiplication/division algorithm leads to a correct or erroneous answer). Each of the eight possible combinations of these characteristics was represented by one item in each task, resulting in a total of 16 systematically varied items (see Table 2). From now on, the items will be referred to with the use of codes, composed of the task number followed by three symbols representing their characteristics in relation to the three conditions (e.g., $2\text{ N}\div\checkmark$ translates as Task 2, natural numbers, division, correct result). A high degree of internal consistency was demonstrated for the total of task items (Cronbach's $\alpha=0.865$) as well as for each task separately (0.897 and 0.834, for Task 1 and Task 2, respectively).

Procedure

Participation in the study was voluntary and anonymity was guaranteed. Students were examined in their classroom during school time, while adults were examined at a place and time of their choice. All participants were examined individually and were not given a time limit to complete the questionnaire. The average time of completion was 40 minutes for students and 25 minutes for adults.

Results

Analysis of Participants' Performance

The participants' mean correct response in the total number of 16 items was 12.54 ($SD=2.72$). Their performance is analysed with regard to the type of task, number set, arithmetic operation, and computation result after normality assumption as well as random and independent selection of the participants from the population were checked to be satisfied. A repeated measures ANOVA was conducted to analyse the effects of age (5th grade children and adults) as the between-subjects factor, and the type of task (Task 1 and Task 2), number set (natural and decimal numbers), operation (multiplication and division), and computation result (correct and erroneous) as the within-subjects factors. There was a significant main effect of age ($F(1158)=130.565$, $p<0.001$), indicating that the adults reached greater level of success ($M=14.36$, $SD=1.58$) compared to the fifth-graders ($M=10.71$, $SD=2.38$). The main term of task was also significant ($F(1158)=12.304$, $p<0.01$), with participants performing significantly better in Task 1 than in Task 2. The interaction between age and type of task was significant ($F(1158)=38.609$, $p<0.001$). Further analyses showed that adults performed significantly better in Task 2 compared to Task 1 ($t(79)=-2.177$, $p<0.05$), whereas fifth-graders were significantly more successful with Task 1 than with Task 2 ($t(79)=6.204$, $p<0.001$). Figure 2 shows these findings.

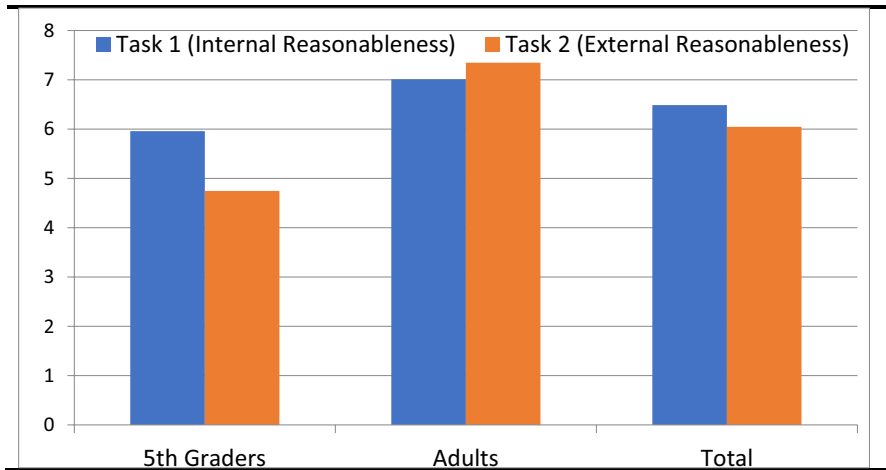


Fig. 2 Mean number of correct responses ($mx=8$) by task and age group

Number set The type of numbers used in the items had a significant effect on correct responses ($F(1158)=64.119$, $p<0.001$). The scores were higher in items with natural numbers than in items with decimal numbers. The interaction type of numbers by age was not significant ($F(1158)=3.010$, $p=0.085$), revealing that both age groups performed significantly better in natural than in decimal number items.

The two-way interaction between type of numbers and type of tasks was found significant ($F(1158)=114.714$, $p<0.001$). Further analyses showed that natural number items were significantly easier than decimal number items in Task 1 ($t(159)=11.121$, $p<0.001$), albeit with an even wider gap than that found for the total of the items. For Task 2, the opposite was found: participants performed significantly better in decimal numbers than in natural number items ($t(159)=-2.326$, $p<0.05$).

Last, the three-way interaction between type of numbers, type of tasks, and age was significant ($F(1158)=11.202$, $p<0.01$), confirming the previous inconsistencies between the tasks for both age groups. Children performed significantly better in natural numbers ($M=3.54$, $SD=0.73$) than in decimal numbers ($M=2.43$, $SD=0.85$) in Task 1 ($t(79)=9.221$, $p<0.001$), but this difference was reversed in Task 2 where their performance was significantly better in items with decimal numbers ($M=2.47$, $SD=0.99$) than in natural numbers ($M=2.28$, $SD=0.91$) ($t(79)=-2.231$, $p<0.01$). On the other hand, adults were significantly better with natural numbers in Task 1 ($M=3.82$, $SD=0.41$, $t(79)=6.743$, $p<0.001$), while no significant differences in their performance between natural and decimal numbers ($M=3.65$, $SD=0.64$ and $M=3.70$, $SD=0.60$, respectively) were spotted in Task 2 ($t(79)=-0.851$, $p=0.397$).

Arithmetic operation The main term of arithmetic operation was significant ($F(1158)=5.349$, $p<0.05$) with participants showing significantly better scores in division items compared to multiplication items. These differences were observed

for both age groups, since the interaction between arithmetic operation and age was not found significant ($F(1158)=0.237, p=0.627$).

The interaction between operation and type of tasks was found significant ($F(1158)=16.658, p<0.001$). Even though the operation involved did not affect participants' performance in Task 1 ($t(159)=1.078, p=0.283$), statistically significant differences were found in Task 2 ($t(159)=-4.399, p<0.001$) where the participants were more successful with division than with multiplication items. Additionally, the operation tended to interact with type of tasks and age ($F(1,158)=5.616, p<0.05$). Further analyses revealed that performance in division items in Task 2 was significantly better than in multiplication items for both children ($M=2.58, SD=1.03$ and $M=2.17, SD=0.99, t(79)=-3.445, p<0.01$) and adults ($M=3.79, SD=0.47$ and $M=3.56, SD=0.82, t(79)=-4.399, p<0.001$), whereas these differences were not spotted in Task 1 concerning either the children ($t(79)=1.850, p=0.068$) or the adult group ($t(79)=-0.820, p=0.415$). However, participants' greater success with division items in Task 2 might be attributed to the greater easiness of the numbers involved compared to the division items in Task 1. No other interactions were traced.

Computational result The type of result was found to be a statistically significant variable in determining participants' success ($F(1158)=12.608, p<0.01$) with more successful responses being found with correct items than with erroneous ones. However, the participant children provided more successful responses in the items referring to correct than to incorrect results ($t(79)=5.962, p<0.001$), whereas neither such a difference or the opposite was true for the adult participants ($t(79)=-1.593, p=0.115$), as the interaction between type of result and age was found significant ($F(1158)=31.306, p<0.001$).

The ANOVA produced a significant difference for the two-term interaction between type of result and type of task ($F(1158)=305.149, p<0.001$). Further analyses showed that in Task 1 it was significantly easier for participants to reject an incorrect result than to verify a correct one ($t(159)=-10.673, p<0.001$). Yet, in Task 2, participants performed significantly better when the use of algorithms was sufficient to reach a correct outcome compared to when it was not ($t(159)=12.446, p<0.001$).

The type of result, type of task, and age interaction were significant ($F(1158)=51.006, p<0.001$). For the fifth-graders, the gap between the two types of result items in Task 2 was wide ($M=3.44, SD=0.79$ and $M=1.31, SD=1.29$) giving again a significant difference in favour of correct results ($t(79)=15.125, p<0.001$). However, in line with the average performance of the whole sample, in Task 1, the younger age group performed significantly better ($t(79)=-8.102, p<0.001$) in items with erroneous results than with correct ones ($M=3.51, SD=0.67$ and $M=2.45, SD=0.95$, respectively). The same was found for adults ($t(79)=-7.026, p<0.001$) whose performance in items of Task 1 with correct results ($M=3.11, SD=0.93$) was significantly lower than in those with erroneous results, in which they achieved an impressive mean score of 3.90 (out of 4, $SD=0.34$). However, in Task 2, adults performed significantly better

Table 3 Mean number of correct responses in both tasks by number set, operation, and computational result for the two age groups

| | <i>Task 1 – internal reasonableness</i> | | <i>Task 2 – external reasonableness</i> | |
|------------------------|---|-------------|---|-------------|
| | 5th graders | Adults | 5th graders | Adults |
| Natural numbers | 3.54 (0.73) | 3.82 (0.41) | 2.28 (0.91) | 3.65 (0.64) |
| Decimal numbers | 2.43 (0.85) | 3.19 (0.81) | 2.47 (0.99) | 3.70 (0.60) |
| Multiplication | 3.09 (0.62) | 3.48 (0.57) | 2.17 (0.99) | 3.56 (0.82) |
| Division | 2.87 (0.91) | 3.54 (0.61) | 2.58 (1.03) | 3.79 (0.47) |
| Correct | 2.45 (0.95) | 3.11 (0.93) | 3.44 (0.79) | 3.95 (0.22) |
| Erroneous | 3.51 (0.67) | 3.90 (0.34) | 1.31 (1.29) | 3.40 (1.02) |

Maximum correct score is 4; format: mean (SD).

($t(79) = 5.234$, $p < 0.001$) in items where the use of algorithms was sufficient to obtain correct results ($M = 3.95$, $SD = 0.22$) compared to the items where the use of algorithms alone was not sufficient ($M = 3.40$, $SD = 1.02$). No other interactions were found. Table 3 summarises the results.

Analysis of Participants' Strategies

The explanations accompanying participants' answers were analysed for themes revealing the strategies they used. For each Task, one set of themes was generated which varied in terms of sophistication, popularity, and efficacy.

Strategies used to judge the internal reasonableness (Task 1) Having excluded the *No Answer* category, participants based their responses on five different strategies (see Table 4) in Task 1. The *Algorithm* strategy referred to the use of standard multiplication or division algorithms, either for the given operation itself or for its inverse. Other participants invoked known (or made-up) *Rules and Properties*

Table 4 Mean strategy use (mx=8) and correlation between strategy use and performance by age in Task 1

| Strategies | <i>Frequency use</i> | | | <i>Correlation with performance</i> | |
|--------------------------|----------------------|-------------|--------|-------------------------------------|---------|
| | Overall | 5th graders | Adults | 5th graders | Adults |
| Without justification | 0.76 | 1.13 | 0.40 | -0.440** | -0.175 |
| Algorithm | 2.26 | 2.21 | 2.30 | -0.080 | -0.174 |
| Rules and properties | 2.03 | 2.41 | 1.65 | 0.236 | 0.001 |
| Split | 0.83 | 0.41 | 1.25 | 0.249* | 0.308** |
| Equivalent expression | 0.35 | 0.29 | 0.41 | 0.125 | -0.011 |
| Computational estimation | 1.77 | 1.55 | 1.99 | 0.206* | 0.194* |

*Significant correlation at the 0.05 level, **significant correlation at the 0.01 level.

relevant to the involved operations and/or numbers to justify their responses (e.g., S45 rejected the result 107.3 as incorrect in $1D \times \times$ recalling that multiples of 5 end in either 0 or 5). The *Split* strategy was a form of mental computation that was applied either to the given operation or to its inverse (e.g., S21 in $1N \times \checkmark$ reasoned the Eq. $709 \times 50 = 35,450$ by splitting 709 into 700 and 9 and then calculating $700 \times 50 + 9 \times 50$). *Equivalent Expressions* was another mental calculation strategy that was based on the reformulation of one part of the equation enabling multiplication or division in stages, usually involving doubles or halves (e.g., S76 in $1N \times \checkmark$ calculated $709 \times 100 : 2$ to check the result of 709×50). Finally, participants who employed *computational estimation* relied upon finding approximate results for either the given operation or its inverse (e.g., A104 in $1D \times \times$ recognised that 107.3 is too small to be the correct result because $25.3 \times 5 \approx 25 \times 5 = 125 > 107.3$).

Frequency of strategy use As shown in Table 4, strategies based on algorithms ($M = 2.26$) and rules or properties ($M = 2.03$) dominate over the number-sense-based ones (computational estimation: 1.77; split: 0.83; equivalent expressions: 0.35). Comparing the trends within the two age groups, children chose algorithms almost as much as adults did ($t(158) = -0.275$, $p = 0.784$), but they used rules and properties significantly more frequently than adults ($t(158) = 3.270$, $p < 0.01$). In contrast, adults relied on the split strategy to a significantly greater extent than children ($t(158) = -5.324$, $p < 0.001$). Regarding the use of equivalent expressions and computational estimation, the differences between the two age groups were not found significant. Finally, it is worth mentioning that the fifth-graders failed to give a justified response at all almost three times as frequently as the adults.

Correlations between strategy use and performance Looking at the efficacy of the different strategies (Table 4), the absence of positive correlations with the correct response rates revealed that the use of algorithms did not guarantee an appropriate response. In contrast, the split strategy (Pearson's $r = 0.249$, $p < 0.05$ and Pearson's $r = 0.308$, $p < 0.01$, for children and adults, respectively), albeit used less frequently, and computational estimation (Pearson's $r = 0.206$, $p < 0.05$ and Pearson's $r = 0.194$, $p < 0.05$, for children and adults, respectively) were both found to be associated with increased correct response rates in the items of Task 1.

Table 5 Mean strategy use ($mx = 8$) and correlation between strategy use and performance by age in Task 2

| Strategies | Frequency use | | | Correlation with performance | |
|-----------------------|---------------|-------------|--------|------------------------------|----------|
| | Overall | 5th graders | Adults | 5th graders | Adults |
| Without justification | 0.47 | 0.71 | 0.23 | -0.534** | -0.007 |
| Algorithm | 3.72 | 3.71 | 3.73 | -0.302* | -0.645** |
| Guess and check | 1.93 | 2.23 | 1.64 | 0.191 | 0.262 |
| Practicality | 1.88 | 1.35 | 2.41 | 0.420** | 0.559** |

*Significant correlation at the 0.05 level, **significant correlation at the 0.01 level.

Strategies used to judge the external reasonableness (Task 2) In Task 2, aside from the *Without Justification* category (unjustified responses), three strategies emerged from the thematic analysis (see Table 5). The *Algorithm* strategy represented routine expertise and included all responses that were based on the execution of long division and multiplication or known algorithmic procedures such as the rule of three (e.g., A151 in $2 N \times X$ explained as follows: 2 shirts \rightarrow 4h, 6 shirts \rightarrow 12h). Furthermore, two adaptive expertise strategies were identified. *Guess and Check* was a mental strategy that started with an estimation and gradually approached the result through appropriate adjustments (e.g., A125 in $2D \div X$ first calculated the cost of 10 packs, given that 1 pack costs 0.40€, and then added 2 more packs to get as close to 5€ as possible). Finally, participants who employed the *Practicality* strategy aimed at responses that make sense in the real world, either by relying solely on the context of the problem or by considering the context to adapt the results generated by algorithms accordingly. As an example of the former (which was mainly used in the two $\times X$ items), A124 in $2 N \times X$ answered that the number of shirts does not affect how long their air-drying will take, as long as that they are made from the same material and the weather conditions remain unchanged. To illustrate the latter (which was mainly used in the two $\div X$ items), M10 in $2 N \div X$ calculated that $315:30 = 10,5$ and answered 11 buses explaining that there is no such thing as half a bus. A few participants employed an interesting variation of this strategy: they used some of the numbers involved in the problem to perform operations that generated reasonable results, although these operations were irrelevant. For instance, in $2D \times X$ participant S19 multiplied the decimal part of 1.30 m, which was the 10-year-old Nick's height, by 2 to find that at the age of 20, he will be $1 + 2 \times 0,30 \text{ m} = 1.60 \text{ m}$ tall. Similarly, S41 added the 20 years to the decimal part of the current height to conclude that he will become 1.50-m tall.

Frequency of strategy use Algorithm-based strategies were by far the most commonly used ones for both age groups. Interestingly, children and adults based their solutions on algorithms with strikingly similar frequency (means: 3.71 and 3.73, respectively). However, further analyses showed that children clearly preferred conventional ways of executing and presenting algorithms as they used written algorithms significantly more frequently ($t(158) = 7.676, p < 0.001$) than adults. On the other hand, the mental execution of algorithms, typically assisted by note-taking, was significantly more popular ($t(158) = 11.310, p < 0.001$) within the adult than the student group. The two adaptive strategies (guess and check and practicality) accumulated comparable overall mean use (1.93 and 1.88), with statistically significant differences emerging from the comparisons between the two groups. Guess and check was more popular within the younger age group ($M = 2.23$) than the older group ($M = 1.64$), a difference that was found to be significant ($t(158) = 2.987, p < 0.01$). For practicality, the opposite was found: adults used this strategy to a significantly greater extent ($t(158) = -6.338, p < 0.001$) than students. The prevailing variation of this strategy was the adjustment of answers in light of the task context, a method that was again used more widely ($t(158) = -7.230, p < 0.001$) by adults in comparison with students. A small proportion of participants gave reasonable answers, albeit generated by operations that were not appropriate. This variation

was mainly used by students as it was almost non-existent in the adult group, with the difference between the two groups being found significant ($t(158)=4.194$, $p<0.001$). Finally, it is worth noticing that, similarly to Task 1, the student participants failed to give justifications three times as frequently as the adult group.

Correlations between strategy use and performance Despite the clear dominance of algorithm-based solutions, it was found that this strategy was negatively correlated with performance (Pearson's $r=-0.302$, $p<0.05$ and Pearson's $r=-0.645$, $p<0.01$, for children and adults, respectively). It seems that a frequent use of algorithms is not always the safest path, but in fact it might reveal that a person may be strongly based on rules due to their lack of other tools when checking the reasonableness of a response. In contrast, participants who aimed at responses that made sense in the real world had an increased chance of giving correct answers, as revealed by the positive correlation of practicality strategy with the correct response rates (Pearson's $r=0.420$, $p<0.01$ and Pearson's $r=0.559$, $p<0.01$, for children and adults, respectively). Finally, no statistically significant correlation was found for the guess and check strategy.

Discussion and Conclusions

In this study, we explored Greek fifth-graders' and adults' ability to judge the reasonableness of computational results in context-free and context-based tasks. Responding to the research questions of the study, results revealed three key findings.

First, the performance of adults (14.36/16) was substantially better than that of students (10.71/16) with the adults clearly outperforming students in both tasks. The relatively weak performance of students may be partially attributed to their limited chances for engagement with relevant mathematical activities prior to their participation in the study due to the lack of instructional attention to reasonableness in primary school (Greek Ministry of Education [MINEDU], 2003). However, it is worth noticing that despite the wide gap between the two age groups, the results concerning students' performance are more encouraging than past studies have revealed (e.g., Alajmi & Reys, 2010; Yang, 2017; Yang & Sianturi, 2019). Of course, due to differences in the research design, the findings might not be directly comparable with each other, but the relatively high performance of our student participants indicates that the concept of reasonableness may be less difficult for students than previously thought.

Turning to the competencies and difficulties of each age group, the no answer rates and the justifications of answers indicated that adults found it easier to give sensible responses in Task 2, which required examining the meaning of numbers in the real world, compared to Task 1, which revolved around number relationships and the effect of operations. However, the opposite was true for students, a finding that is in opposition to Alajmi and Reys' (2010) results according to which the two criteria of reasonableness were equally difficult for eighth-graders. Still, a significant proportion of students gave inappropriate or unreasonable responses in both tasks, showing that students very often experience difficulties relevant to both, a finding

that is consistent with past studies (Menon, 2004; Yang, 2017). What might explain the different strengths and weaknesses of the two groups is that school mathematics often places great emphasis on developing fluency with operations, but students are given limited opportunities to connect the mathematical content they learn to real-life situations (Mann, 2006; Masingila et al., 1996). In contrast, adults have long been detached from school environment and have been using mathematics more often in the context of daily life than in that of school-like activities.

The second key finding centres upon factors that may facilitate or place obstacles on solvers' efforts to give reasonable answers. Unlike previous studies that reported consistently low performance across the different number domains (e.g., Alajmi & Reys, 2010), our participants performed better in natural than in decimal numbers in context-free items (Task 1), but not in context-based ones (Task 2) in which the average performance in decimals was relatively higher. This probably indicates that the presence of context can potentially allow solvers to overcome difficulties associated with the difficulty of specific mathematical concepts or topics, since it enables the consideration of the problem in the context of real life (Meyer et al., 2001; Sowder & Schappelle 1989). However, it is worth noting that the presence of context was more beneficial for adults than it was for students, who performed relatively better in context-free than in context-based items. This supports the growingly accepted argument that the presence of context is not necessarily associated with increased student performance (Beswick, 2011; Can & Özdemir, 2020). Students were found to be particularly weak in contextualised problems where algorithms did not suffice for a correct answer or when the results generated from them were incomplete and needed adaptation to make sense. This is in line with previous findings suggesting that students often rely too heavily on algorithms to solve mathematical tasks (Desli & Lioliou, 2020; Heirdsfield & Cooper, 2002), even when algorithms are insufficient or inappropriate to generate correct responses (Masingila et al., 1996; Menon, 2004).

The third key finding focuses on the range of the employed strategies alongside the frequency and the efficacy of each. Overall, two broad categories of strategies emerged: routine-based and sense-making strategies. Despite being less effective, the former clearly prevailed over the latter in terms of frequency. This result is consistent with findings of previous studies showing a general preference for the use of algorithmic techniques (e.g., Alajmi & Reys, 2010; Can & Özdemir, 2020). Interestingly, in either task, the use of algorithms was equally widespread within the two age groups, although the popularity of the other strategies varied between them.

In Task 1, students typically resorted to the use of rules and properties of the involved numbers and operations. These were often misinterpreted, or unrelated to the particular task and insufficient for its solution, highlighting that students often do not make sense of the rules and algorithms they learn, which results in the misunderstanding of their meaning or their limitations (Markovits & Sowder, 1994). In general, student participants were inclined to use unproductive routine strategies that indicated static and sparsely connected mathematical knowledge (McMullen et al., 2020). On the other hand, adults opted more frequently for sense-making adaptive strategies, such as estimates and mental computations, which were more effective and indicative of flexible mathematical thinking (Hatano, 1988; Markovits

& Sowder, 1994). In a school environment that prioritises the instruction and memorisation of algorithms and rules (Mann, 2006; Verschaffel et al., 2007), students learn to consider the use of routine-based strategies as the safest path and become reluctant to opt for non-standard computation strategies (Can & Özdemir, 2020; Heirdsfield & Cooper, 2002). On the other hand, as a result of using mathematics in real-life situations, adults become more familiar with estimating and calculating mentally (Northcote & McIntosh, 1999; Thompson, 1999) which, consistently with findings of past studies (Alajmi & Reys, 2010; Dougherty & Crites, 1989; Yang, 2017), were found to be instrumental in successfully judging reasonableness. Therefore, the effective incorporation of reasonableness into school mathematics may require the classroom community to shift its focus from the mere memorisation and practice of rules and routines towards the development of well-connected knowledge of number and operation characteristics that can foster their meaningful utilisation through sense-making strategies and the acquisition of adaptive expertise (Heirdsfield & Cooper, 2002; McMullen et al., 2020).

The strategies used in Task 2 offer similar insights. The use of algorithms tended to be associated with an increased risk of giving inappropriate responses. The very effective adaptive strategy of filtering results through considering the task context was again much more popular among adults than among students. In general, students very often did not pursue connections between mathematics and real life, which is consistent with findings from past studies (e.g., Yang & Sianturi, 2019). In contrast, adults' familiarity with everyday circumstances that are relevant to the context of the problems (e.g., financial transactions, housework etc.) might have given them a significant advantage here, enabling them to productively rely on the context to obtain reasonable results. This underlines again the importance of encouraging students to reflect critically upon the sufficiency of taught problem-solving routines and use the known techniques flexibly (McMullen et al., 2020) considering whether adaptations are required. Furthermore, it is worth mentioning that a few students felt the need to give dual responses in some items of Task 2, which is not only indicative of how their overreliance on algorithmic procedures restricts their thinking, but also shows how they view school mathematics as disconnected from reality. For instance, when asking how tall a person will be at the age of 20, given that now he is 10 years old and 1.30-m tall, a student responded that "according to mathematics" he will be 2.60-m tall, but "according to reality", none ever reaches this height. In our interpretation, this student probably considers mathematics in the classroom as constrained by the limits of routine expertise, while viewing the development and use of adaptive expertise as permissible only in out-of-school mathematics. Interestingly, such responses were given exclusively by students, which raises concerns about the role engagement with mathematics in school contexts may play in promoting the erroneous belief that mathematics and reality are incompatible with each other. In light of that and reflecting on how beneficial the experience of using mathematics outside the mathematics classroom appears to be for adults, we argue that it is crucial to step up efforts towards making the identification of connections between mathematics and real life an integral part of the mathematical activity in school.

This study was subject to at least two limitations. First, concerning the selected sampling technique, although the combination of purposive and convenience

sampling enabled a varied sample, it is unknown whether our participants are typical Greek adults and fifth-graders. Second, turning to the design of the task-based questionnaires, the two aspects of reasonableness are deeply interrelated, and thus, it is technically impossible for them to be completely distinguished from one another. To tackle this issue, we included only context-free items in Task 1; the absence of context eliminated the need for pursuing connections between the numbers and the real world, encouraging judgements solely based on the relationships between numbers and the effect of operations on them, namely the internal reasonableness. In contrast, since external reasonableness was the focus of Task 2, this included only context-based items that enabled reflection upon the meaning of results in real-life situations. The numbers and the operations involved in Task 2 were fairly easy to eliminate the need for judgements based on internal reasonableness. Additionally, sometimes different items within the same task may have favoured the use of different strategies. For example, in the \div items of Task 2, it was probably easier to adjust the result of the algorithm than to avoid the use of algorithms altogether, while for the \times items of the same task, the opposite was probably true. All these may have eliminated the gap that typically exists between the level of difficulty of multiplications and divisions as well as natural and decimal numbers. Thus, a selection of more complex items in terms of operations and numbers might shed light on this issue.

Given the previous limitation regarding the generalisability of our results, studies with a larger number of participants recruited through more refined sampling techniques could be conducted in the future. We also recommend exploring the performance and strategies of students in early primary grades which has largely remained understudied. Finally, all past studies have revealed severe weaknesses in students' understandings about the concept of reasonableness, which stresses the need for classroom-based intervention studies. This research direction can offer valuable insights into how students at different school levels can be appropriately introduced to the concept of judging reasonableness, as well as how they can best meet these multidimensional ability demands.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alajmi, A., & Reys, R. (2007). Reasonable and reasonableness of answers: Kuwaiti middle school teachers' perspectives. *Educational Studies in Mathematics*, 65(1), 77–94.

- Alajmi, A., & Reys, R. (2010). Examining eighth grade Kuwaiti students' recognition and interpretation of reasonable answers. *International Journal of Science and Mathematics Education*, 8(1), 117–139.
- Beswick, K. (2011). Putting context in context: An examination of the evidence for the benefits of 'contextualised' tasks. *International Journal of Science and Mathematics Education*, 9(2), 367–390.
- Bonotto, C. (2005). How informal out-of-school mathematics can help students make sense of formal in-school mathematics: The case of multiplying by decimal numbers. *Mathematical Thinking and Learning*, 7(4), 313–344.
- Can, D., & Özdemir, İE. Y. (2020). An examination of fourth-grade elementary school students' number sense in context-based and non-context-based problems. *International Journal of Science and Mathematics Education*, 18(7), 1333–1354.
- Charles, R. I., & Carmel, C. A. (2005). Big ideas and understandings as the foundation for elementary and middle school mathematics. *Journal of Mathematics Education*, 7(3), 9–24.
- Desli, D., & Lioliou, A. (2020). Relationship between computational estimation and problem solving. *International Electronic Journal of Mathematics Education*, 15(3), em0602.
- Dougherty, B. J., & Crites, T. (1989). Applying number sense to problem solving. *The Arithmetic Teacher*, 36(6), 22–25.
- Greek Ministry of Education [MINEDU]. (2003). Διαθεματικό Ενιαίο Πλαίσιο Προγραμμάτων Σπουδών Πρωτοβάθμιας Εκπαίδευσης και Αναλυτικά Προγράμματα Σπουδών Δημοτικού [Interdisciplinary Curriculum for Primary School], ΦΕΚ 303B/τεύχ.2ο, 13–03–2003. Retrieved from: <http://www.pi-schools.gr/programs/depps/>
- Hatano, G. (1988). Social and motivational bases for mathematical understanding. In G. B. Saxe & M. Gearhart (Eds.), *Children's mathematics* (pp. 55–70). Jossey-Bass.
- Heirdsfield, A. M., & Cooper, T. J. (2002). Flexibility and inflexibility in accurate mental computation: Two case studies. *Journal of Mathematical Behavior*, 21(1), 57–74.
- LeFevre, J. A., Greenham, S. L., & Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cognition and Instruction*, 11(2), 95–132.
- Mann, E. L. (2006). Creativity: The essence of mathematics. *Journal for the Education of the Gifted*, 30(2), 236–260.
- Markovits, Z., & Sowder, J. (1994). Developing number sense: An intervention study in Grade 7. *Journal for Research in Mathematics Education*, 25(1), 4–29.
- Masingila, J. O., Davidenko, S., & Prus-Wisniowska, E. (1996). Mathematics learning and practice in and out of school: A framework for connecting these experiences. *Educational Studies in Mathematics*, 31(1–2), 175–200.
- McMullen, J., Hannula-Sormunen, M. M., Lehtinen, E., & Siegler, R. S. (2020). Distinguishing adaptive from routine expertise with rational number arithmetic. *Learning and Instruction*, 68, 101347.
- McIntosh, A., Reys, B., Reys, R., Bana, J., & Farrell, B. (1997). *Number sense in school mathematics: Student performance in four countries*. Mathematics, Science & Technology Education Centre, Edith Cowan University.
- Menon, R. (2004). Elementary school children's number sense. *International Journal for Mathematics Teaching and Learning*, 57, 1–16.
- Meyer, M. R., Dekker, T., & Querelle, N. (2001). Context in mathematics curricula. *Mathematics Teaching in the Middle School*, 6(9), 522–527.
- National Council of Teachers of Mathematics [NCTM]. (2000). *Principles and standards for school mathematics*. Author.
- Northcote, M., & McIntosh, A. (1999). What mathematics do adults really do in everyday life? *Australian Primary Mathematics Classroom*, 4(1), 19–21.
- Pólya, G. (1973). *How to solve it*. Princeton University Press.
- Reys, R., Reys, B., McIntosh, A., Emanuelsson, G., Johansson, B., & Yang, D. C. (1999). Assessing number sense of students in Australia, Sweden, Taiwan, and the United States. *School Science and Mathematics*, 99(2), 61–70.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.
- Sowder, J. T., & Schappelle, B. P. (Eds.). (1989). *Establishing foundations for research on number sense and related topics: Report of a conference*. San Diego: San Diego State University Center for Research in Mathematics and Science Education.
- Sowder, J. T., & Wheeler, M. M. (1989). The development of concepts and strategies used in computational estimation. *Journal for Research in Mathematics Education*, 20(2), 130–146.

- Thompson, I. (1999). Getting your head around mental calculation. In I. Thompson (Ed.), *Issues in teaching numeracy in primary schools* (pp. 145–156). Open University Press.
- Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 557–628). NCTM.
- Yang, D. C. (2005). Number sense strategies used by 6th grade students in Taiwan. *Educational Studies*, 31(3), 317–333.
- Yang, D. C. (2017). Performance of fourth graders when judging the reasonableness of a computational result. *International Journal of Science and Mathematics Education*, 17(1), 197–215.
- Yang, D. C., & Sianturi, I. A. J. (2019). Sixth grade students' performance, misconceptions, and confidence when judging the reasonableness of computational results. *International Journal of Science and Mathematics Education*, 17(8), 1519–1540.

Authors and Affiliations

Dimitrios Deslis¹  · Despoina Desli²

Despoina Desli
ddesli@eled.auth.gr

¹ Faculty of Education, University of Cambridge, 164 Hills Road, Cambridge CB2 8PQ, UK

² Faculty of Education, School of Primary Education University Campus, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece