



Who-Is-Right tasks as a means for supporting collective looking-back practices

Boris Koichu¹ · Reut Parasha¹ · Michal Tabach²

Accepted: 3 April 2021 / Published online: 17 April 2021
© FIZ Karlsruhe 2021

Abstract

The looking-back stage is rarely observed in students' problem solving in spite of its recognized importance. The importance of this stage is attributed to practices of engagement with queries on verification of the obtained solution(s), comparative consideration of alternative solutions, and formulation of implications for future problem solving. We refer to such practices as *looking-back practices*. In the present study we explored the hypothesis that the looking-back practices can be evoked in small-group classroom discussions of controversial worked-out solutions to word problems. Such tasks are known as Who-Is-Right tasks. The data consisted of audio- and videotapes of six small groups of high-school students working on a Who-Is-Right task in the context of percentage. The data analysis, informed by a discursively-oriented perspective on problem solving, attended to strategies, dialogical moves and mathematical resources enacted by the students towards attempted agreement as to which of the solutions should be endorsed and why. The findings imply that Who-Is-Right tasks have undeniable potential for supporting collective looking-back practices. In addition, the study contributes to the literature on enactment of mathematical resources in problem-solving discourse and on patterns of students' dialogic participation in small-group problem solving.

Keywords Dialogical moves · Looking back · Resources · Small-group problem solving · Verification · Worked-out solutions

1 Introduction

Of the four stages comprising Pólya's (1945/1973) model of mathematical problem solving—understanding a problem, devising a plan, carrying out the plan, and looking back—the *looking back* stage is known to be particularly troublesome in mathematics education research and practice. Pólya (1945/1973) introduced *looking back* as a stage when a solution to a problem has already been produced but needs to be verified and examined for possible improvements and alternatives. At this stage, argued Pólya, the solvers are expected to address the following questions: (i) “Can you check the result?” (ii) “Can you check the argument?” (iii) “Can you derive the result differently?” (iv) “Can you see it

at a glance?” (v) “Can you use the result, or the method, for some other problems?” (p. 27).

Along with arguing for the importance of the looking back stage, Pólya (1945/1973) observed that this stage rarely appears in student problem solving. In his words, “even fairly good students, when they have obtained the solution to the problem and written down neatly the argument, shut their books and look for something else” (p. 27). This observation was reinforced by empirical evidence in several studies (e.g., Goos & Galbraith, 1996; Kantowski, 1977; Mashiach-Eisenberg & Zaslavsky, 2004; Malloy & Jones, 1998; Papadopoulos & Dagdilelis, 2008; Stillman & Galbraith, 1998).

A number of reasons for the rarity of the looking back stage in student problem solving have been pointed out. Kantowski (1977) explained this rarity by strong emotional satisfaction that emerges when a student finds a solution that feels correct. He suggested that this satisfaction makes the looking back stage intellectually unnecessary for a regular student. Stillman and Galbraith (1998) asserted that many students

✉ Boris Koichu
boris.koichu@weizmann.ac.il

¹ Weizmann Institute of Science, Rehovot, Israel

² Tel Aviv University, Tel Aviv, Israel

develop trust in their written work, and do not attempt to check their solutions even when they have enough time for this. Nonetheless, past research (Malloy & Jones, 1998; Mashiach-Eizenberg & Zaslavsky, 2004; Papadopoulos & Dagdilelis, 2008; Papadopoulos & Sekeroglou, 2018) has also shown that students can use a variety of verification strategies, as a rule, when properly prompted.

For example, Mashiach-Eizenberg and Zaslavsky (2004) described different occasions where students attempted to verify their solutions while solving combinatorial problems in a task-based interview setting, working individually or in pairs. In their study, the majority of students who had solved the problems incorrectly, were eager to reconsider after the interviewer's explicit or implicit prompts, the most influential of which was the interviewer's disclosure of her opinion about the students' solutions. We learn from this study that the looking-back stage in student problem solving is likely to appear when stimulated by a prompt that is beyond the independent reach of the students.

Being informed by past research that spontaneous engagement in the looking-back stage can hardly be expected from regular students, we recognized the need to understand better the principled design of realistic classroom situations that would be rich in opportunities for the students to engage in the *looking-back practices* if not in a looking-back stage. By looking-back practices we mean solvers' engagement with some of the questions (i)–(v) posed by Pólya (see above) in the course of solving problems. We suggested that looking-back practices can be evoked not only after independently solving a problem, but in the course of collaborative coping with specially designed tasks. Specifically, we relied on studies by Swan (2007) and by Rittle-Johnson and Star (2011) on the use of tasks requiring students to compare the validity of two or more worked-out solutions to a problem. Tabach and Koichu (2019) referred to such tasks as Who-Is-Right or WIR tasks. Our wish to examine empirically the potential of WIR tasks for evoking looking-back practices in realistic classroom situations constituted the rationale of our study. Hence, the study aimed to characterize small-group discussions of WIR tasks as a means to support looking back practices.

2 Theoretical framework

The design and conceptual apparatus of the present study were informed by three themes, namely, (1) problem solving as a discursive activity, (2) WIR tasks, and (3) collaborative argumentation and justification.

2.1 Discursively-oriented perspective on problem solving

In this section, we explicate theoretical premises underlying the use of such terms as 'problem solving', 'solution' and 'correct/incorrect' in our study. Generally, we adhered to a spectrum of socio-cultural perspectives on problem solving (Kilpatrick, 1985). In these perspectives a problem is considered as a task given and received in a social situation jointly constructed by the participants involved in it. In particular, the socio-cultural perspectives recognize that problem solvers are continuously involved in interpreting each other's actions and intentions while conforming to the existing social learning practices, and while contributing to the evolution of these practices (Cobb, 2000). Specifically, we relied on a discursively-oriented conceptualization of mathematical problem solving in instructional situations proposed by Koichu (2019), as follows:

Problem solving in mathematics instruction is a socio-culturally shaped process of not-immediate achieving a chosen or imposed goal, in which the involved individuals enact, through private or public exploratory discourse,¹ individual or shared resources² that they interpret, even for a short while, as appropriate for achieving the goal... [A] *solution to a problem* is a public narrative, which becomes endorsed, by the problem solvers or the problem proposers, as the one that achieves the goal. (p. 49, footnotes added).

This conceptualization helps us to balance analysis of individual contributions to a group discussion and analysis of decisions made by a group as a whole. In addition, the conceptualization theoretically backs up our decision to equate the solution *verification* with its endorsement by a group of students, and the solution's *correctness* with its endorsement by experts.

2.2 Who-Is-Right tasks and looking-back practices of problem solving

Tabach and Koichu (2019) described Who-Is-Right or WIR tasks as consisting of two parts. The first part is a narrative introducing a problematic situation. The second part consists of different solutions to the problematic situation, which rely

¹ Sfard and Kieran (2001) referred to *discourse* as "any specific instance of communicating, whether diachronic or synchronic, whether with others or with oneself, whether predominantly verbal or with the help of any other symbolic system" (p. 47).

² In line with Schoenfeld (1985), we refer to problem-solving resources as discursively-enacted pieces of knowledge held by a student as true at least temporary and which serve as anchors for constructing arguments.

on intuitively appealing but contradictory interpretations of the narrative. Each interpretation and corresponding solution consists of a solution-narrative. The solution-narratives are explicitly given and can represent full or partial solutions. The solvers of a WIR task are required to decide which solution-narrative should be endorsed and to support their decision by an argument that might lead to endorsement of the solution also by their peers.

In comparison to regular problem-solving tasks aligned with Pólya's four-stage model or to regular verification tasks (i.e., solve a problem and verify your solution, hereafter, Am-I-Right or AIR tasks), we see in WIR tasks two advantages for evoking and exploring the looking-back practices. First, AIR tasks and WIR tasks have different referential bases for verification. In an AIR task, the solver's own solution is usually the only available referent for the solver. If the solver is confident in his or her solution, this confidence can hinder the need to look back, as argued by Kantowski (1977) and Stillman and Galbraith (1998). If a solver is not confident, the referent—some 'gold-standard solution'—is usually unavailable to her without an intervention by an authority, as in the case described by Mashiach-Eisenberg and Zaslavsky (2004). Therefore, the request to look back at a self-produced solution may feel either artificial or unfeasible to students. In contrast, solution-narratives provided in a WIR task can by themselves serve as reciprocal referents. Second, a WIR task operates with solution-narratives produced by others. Accordingly, the solvers of the WIR task may be more emotionally detached from the solutions than those who produced the solutions first-handedly, and thus they may be more open-minded when verifying them. Hence, we suggest that the request to compare worked-out solutions can be perceived by a WIR task's solvers as intellectually appealing and feasible.

Furthermore, we hypothesize that WIR tasks can evoke practices germane to the 'idealized' looking-back stage, as introduced by Pólya (1945/1973). These practices are not explicitly introduced in WIR tasks, but the tasks allude to them. To this end, we capitalize upon the Rittle-Johnson and Star (2011) analysis, in which they argued that comparison of alternative worked-out solutions to a given problem can involve considerations of correctness, clarity, simplicity, generalizability and applicability.

2.3 Collaborative argumentation

When a WIR task is approached in small groups, the process of endorsing or rejecting a particular solution can be considered as an argumentative activity, that is, an activity in which interlocutors engage in a dialogue characterized by a critical consideration of each other's arguments. Accordingly, the processes of endorsing or rejecting

worked-out solutions can be analyzed by means of conceptual tools developed in prior research on argumentation, justification and socially-mediated metacognition. We consider three studies, which we found particularly relevant to our study.

Goos et al. (2002) analyzed patterns of peer-to-peer social interactions in cases of successful and unsuccessful solving of a word problem in the context of motion. The main data-analysis procedures consisted of characterizing the student *conversation moves*, particularly concentrating on how the students were engaged in each other's thinking. The collaborative nature of the interaction was analyzed in terms of categories named *self-disclosure* (attending to one's own thinking), *feedback requests* (an invitation to a partner to attend to one's own thinking) and *other-monitoring* (attending to a partner's thinking). This analysis enabled the scholars to distinguish between discursive patterns characterizing different problem-solving cases. It was found that in successful cases, the students challenged and discarded unhelpful ideas and endorsed useful problem-solving strategies, whereas unsuccessful cases were characterized by the lack of critical engagement with partners' contributions to the discussion.

Ayalon and Even (2014) explored whole-class discussions of tasks requiring students to determine equivalence of algebraic expressions, for factors involved in shaping opportunities for the students to engage in argumentation. Videotaped discussions were analyzed by means of a scheme of dialogical moves developed by Asterhan and Schwarz (2009). This comprehensive scheme included categories reminiscent of those developed in Goos et al.'s (2002) study, along with additional categories for denoting moves that serve different dialogical functions (e.g., opposing or agreeing with the claim). Analysis in terms of dialogical moves enabled Ayalon and Even (2014) to reveal contributions of different actors to the justification of mathematical claims as well as the structure of the discussions. The scholars also attended to the content of the dialogical moves, and distinguished between student justifications based on mathematical rules, and justifications based on examples.

The dual focus on structure and content of argumentative activity, which can be observed in both Goos et al.'s (2002) and Ayalon and Even's (2014) studies, was appropriate also for our study. Furthermore, Ayalon and Even's (2014) study suggested how Asterhan and Schwarz's (2009) coding scheme, originally developed for analyzing argumentative activity in a non-mathematical context, could be adapted for analyzing discourse in mathematical contexts. The categories developed in all the aforementioned studies served as departure points for developing the data-analysis procedures in our study.

Fig. 1 The WIR task

Dan bought 2 kg of apples and 5 kg of pears. The price of 1 kg of apples is 25% higher than the price of 1 kg of pears. Overall, Dan paid 23 shekels. What is the cost of 1 kg of apples? What is the cost of 1 kg of pears?	
Hila wrote: The price of 1 kg of pears = x The price of 1 kg of apples = $1.25x$ Equation: $2 \cdot 1.25x + 5x = 23$	Sofia wrote: The price of 1 kg of apples = x The price of 1 kg of pears = $0.75x$ Equation: $2x + 5 \cdot 0.75x = 23$
<ol style="list-style-type: none"> 1. Explain Hila's solution. 2. Explain Sofia's solution. 3. Will Hila and Sofia obtain the same final results? If not, who is right? Explain. 	

3 Research questions

In the above-explained terms, our study pursues three research questions.

1. What strategies do high-school students use, and how, in small-group discussions of a WIR task in the context of percentages? What are the discursive products of the use of these strategies?

We answered this question in the spirit of studies by Papadopoulos and Sekeroglou (2018), Malloy and Jones (1998) and Mashiach-Eisenberg and Zaslavsky (2004). Furthermore, the intention of this question is to identify structures of the discussions, by inferring sub-questions attended to by students on the way to addressing the main WIR question, and by considering these sub-questions in light of the looking-back questions described by Pólya (1945/1973).

2. What are the characteristics of dialogical moves in the small-group discussions of students working on WIR task?

This question was answered by characterizing the extent to which students capitalize on each other's thinking, in the spirit of Asterhan and Schwarz (2009) and Goos et al. (2002).

3. What mathematical resources are discursively enacted, and how, in the small-group discussions of students working on a WIR task?

Here we attended to the specific content of the arguments, in the spirit of Asterhan and Schwarz's (2009) and Ayalon and Even's (2014) studies, and investigated, in particular, what the students relied on while constructing their arguments.

4 Methods

4.1 Participants and settings

A class of 16 11th grade students (16–17 years old) in a school for girls took part in the study. The second-named author of this paper was the teacher of the class. The class studied mathematics in accordance with the highest-level (out of three levels) of the Israeli mathematics curriculum (cf. Movshovitz-Hadar, 2019, for details).

To better understand the context, two comments are in order. First, these students often used different internet resources where solutions to textbook problems are published. Accordingly, they were accustomed to comparing their solutions with the published ones. Second, in Israel, the notion of percentage is first introduced in the 6th grade. From then on, word problems involving percentages are regularly considered in all grades up to high school. From her ongoing work with the class, the teacher was aware of some of her students' misconceptions with percentage problems and wished to create an opportunity for them to obtain a more nuanced understanding of the topic.

4.2 The WIR task

The WIR task used in our study is presented in Fig. 1.

The central task condition is 'the price of 1 kg of apples is 25% higher than the price of 1 kg of pears'. Hila and Sofia interpret this condition in two different ways.

Hila denotes the price of 1 kg of pears as x , and the price of 1 kg of apples as $1.25x$. In this approach, the price of 1 kg of pears corresponds to 100% and the price of 1 kg of apples to 125% of the price of the pears, which is a legitimate

interpretation of ‘25% higher’. Therefore, the corresponding equation $2 \cdot 1.25x + 5x = 23$ leads to the correct solution to the problem: the price of 1 kg of pears is $3\frac{1}{15}$ shekels, and the price of 1 kg of apples is $3\frac{5}{6}$ shekels.

Sofia denotes the price of 1 kg of apples as x and the price of 1 kg of pears as $0.75x$. In this approach, the price of 1 kg of apples corresponds to 100% and the price of 1 kg of pears is interpreted as 75% of the price of the apples. This interpretation does not meet the given condition. Indeed, in this way the price of 1 kg of apples is not 125% of the price of 1 kg of pears but $x/0.75x = 1\frac{1}{3}$ or 133.33%. Accordingly, the corresponding equation $2x + 5 \cdot 0.75x = 23$ leads to an incorrect solution to the given problem: the price of 1 kg of apples ‘is’ 4 shekels, and the price of 1 kg of pears ‘is’ 3 shekels.

In fact, Sofia’s equation solves another problem based on an alternative condition ‘the price of 1 kg of pears is 25% lower than the price of 1 kg of apples’. Parker and Leinhardt (1995) showed that the frequent confusion between the original and alternative conditions can be attributed to a variety of factors. These include the ambiguity involved in inferring a reference base for 100% from the problem narrative, and the interference of additive and multiplicative reasoning structures in students’ perception of problems having a ‘differ by percent’ structure (Parker & Leinhardt, 1995, p. 442). Indeed, the two conditions are not equivalent. However, two apparently similar conditions, for example, ‘the price of a book is 25 shekels more than the price of a pen’ and ‘the price of a pen is 25 shekels less than the price of a book’ are equivalent. Accordingly, ‘differ by percent’ problems and ‘differ by quantity’ problems can easily be confused.

The task narrative includes two additional potential traps. First, the word ‘apples’ appears before the word ‘pears’ in the formulation. Here we capitalize on the well-known misconception in solving word problems, namely, *denote by x the first unknown quantity mentioned in the problem formulation* (e.g., Clement, 1982), as is done in Sofia’s solution.³ Also, Sofia’s solution-narrative leads to two ‘nice’ integers as the answers, while the correct solutions are fractions.

Yet, we assumed that the students in our study might be in a position to correctly solve the task, based on their past experiences with such problems as the following one (hereafter, the market-up/market-down problem).

When first produced, a fancy t-shirt cost 100 shekels. This price was increased by 25% at the beginning of the high season and then decreased by 25% at the end of the season. What was the price of the t-shirt at the end of the season?

³ Of course, choosing ‘apples’ as a referent for x in Sophia’s solution is legitimate, the confusion is in the interpretation of the price of pears, as explained above.

We hoped that the well-known to the students solution to the above problem (i.e., $100 \cdot 1.25 \cdot 0.75 = 93.75$) would serve as a resource for solving the current problem, by reminding the students that subsequent adding and subtracting of 25% does not preserve the initial price because the referents for 100% are different. However, the analogy between the WIR task and the market-up/market-down one is limited. Namely, the market-up/market-down problems presume that the first change (i.e., market-up) is applied to some initial quantity, and the second change (i.e., market-down) to the modified or dependant quantity. In our WIR task, the initial quantities in Hila and Sofia’s solutions are independent and acted upon only once.

An additional clue was provided by means of a follow-up task, as follows:

The price of a notebook is 3 shekels less than the price of a pen. The price of three pens is 25% less than the price of 7 notebooks. What is the price of one pen? What is the price of one notebook?

The plan was that during the lesson the groups might return to the WIR task after considering the follow-up task. In the lesson under investigation, the first 15 min were devoted to solving the WIR task in small groups, and the rest of the lesson (which was not part of the current study) was devoted to solving the follow-up task and to the whole-class discussion.

4.3 Data collection and data analysis

The students were working on the WIR task in six small groups, that is, four groups of three and two groups of two; the small-group discussions were audio- and video-taped and fully transcribed (overall, about 12,000 words). The students’ written work and the teacher’s reflective notes served as complementary data sources.

We implemented inductive analysis with partially pre-defined categories (Dey, 1999; Strauss & Corbin, 1990) as the main analytical approach. The transcripts were first split into *episodes* consisting of several mutually related consecutive conversational turns. For example, a part of the transcript was called ‘episode’ when the students collaboratively explored a particular argument. The episodes were assigned F-type codes (strategies of exploring the task formulation; see Table 1). In addition, each episode was aggregately summarized (Thomas, 2006) for the purpose of identifying its role in the flow of the discussion. The summaries attended to the central questions discussed and to the main discursive products of the episodes. Examples of such questions are ‘who is right?’, ‘why is Sofia wrong?’, ‘are the solutions indeed different?’ Examples of the discursive products are ‘two students agreed that Hila is right’, ‘the argument about different reference bases for 100% was asserted by one student and ignored by the rest’.

Table 1 Strategies of coping with the WIR task formulation (F-codes)

Code F (formulation)	Description
F-ESN (exploring the solution-narratives)	Coping with the task by exploring the given solution-narratives
F-SEQ (solving the equations)	Coping with the task by solving the equations presented in the solution-narratives up to the final answers to the problem
F-ISP (independently solving the problem)	Coping with the task by independently solving the problem while temporarily skipping the solution-narratives

Table 2 Dialogical moves (D-codes)

Code D (Dialogical moves)	Descriptions
D-SD (self-disclosure)	Self-disclosing utterances that clarify, elaborate, evaluate, or justify one's own ideas
D-RR (requests for response), including two sub-categories D-RR-RQ (rhetorical questions)	An utterance that requires partners to respond on one's own ideas Requests for response asserted in the middle of a self-disclosing conversational turn, but without giving the partners a fair chance to respond (e.g., "OK?" "Right?" asked without stopping the speech)
D-RR-IN (invitations to respond)	Requests for response at the end of a self-disclosing assertion (e.g., "Do you understand me?")
D-OO (other-oriented utterances), including 5 sub-categories D-OO-SA (simple agreement)	An utterance that represents an explicit response to a partner's self-disclosing assertion or to a request to respond An utterance that seems to express support for a previous assertion, but does not contain any explanation (e.g., "Right", "OK"). Such utterances are not necessarily expressions of understanding or a reasoned agreement. They may be interpreted as support but also as "go on", "I don't understand but don't let that interrupt you" or "leave me alone for now"
D-OO-RA (reasoned agreement)	An utterance that expresses support for a previous self-disclosing assertion that contains an elaboration (e.g., "Yes. The equations have different answers, so the solutions are not the same.")
D-OO-SO (simple opposition)	An utterance that expresses disagreement with a previous self-disclosing assertion, but without an elaborated explanation (e.g., "OK? No.")
D-OO-RO (reasoned opposition)	An utterance that expresses disagreement with a previous self-disclosing assertion that contains an elaboration (e.g., "OK? No, she decreased, so what?")
D-OO-LU (the lack of understanding)	An utterance that expresses lack of understanding of a previous self-disclosing assertion (e.g., "I still don't get the difference")
D-OTR (other)	A dialogical move, whose functionality is not captured by the previous codes

The transcripts were then analyzed according to *dialogical moves* (D-codes). In line with Asterhan and Schwarz's (2009) approach, we corresponded dialogical moves to conversational turns or parts of conversational turns having different dialogical functionality. Stimulated by Goos et al.'s (2002) study, we developed three types of D-codes: *self-disclosing utterances*, *requests for response* (with two subcategories) and *other-oriented utterances* (with five sub-categories), as shown in Table 2.

The third group of codes concerned the enacted mathematical resources (R-codes), see Table 3. In most of the cases, R-codes were applied to self-disclosing dialogical moves, and in some cases also to other-oriented utterances. Note that they are not mutually exclusive: more than one R-code can be applied to one dialogical move.

One full transcript was used to develop stable coding schemes. The rest of the transcripts were analyzed by the authors first separately and then together. Disagreements (overall, about 3% out of 1216 analytical decisions) were resolved in discussions among the authors. Most of the initial disagreements concerned the use of the R-AP group of codes (analogical problems). The R-codes were assigned (overall, 387 decisions) according to the following heuristics: while considering a dialogical move in the context of the discussion, we inferred which piece of knowledge could be held by the student as true in order to account for what she actually said. This heuristic conforms to a long-standing tradition of abductive analysis of inferring what is plausible from what is evident (Tavory & Timmermans, 2014). Illustrative examples of the data analysis are presented in the "Appendix".

Table 3 Mathematical resources enacted (R-codes)

Code R (Resources)	Description
R-RPC (reference to a problematic condition of the task)	A student refers to the sentence “the price of 1 kg of apples is 25% higher than the price of 1 kg of pears”, in order to suggest, support or oppose a particular interpretation of Sofia or Hila’s solution
R-RB (reference base), with three sub-categories	A student attends to verbal clues in the task formulation in order to determine a reference base for 100% or 25%
R-RB-OF (“of”)	A student attends to “of” or “by” as a cue for what should be denoted as x
R-RB-OW (order of words)	A student refers to the order of words in the task formulation as an explanation for why “apples” or “pears” should be denoted as x . For example, Sofia’s choice of x is correct because “apples” is mentioned in the formulation before “pears”
R-RB-GS (greater and smaller)	A student attends to the words “greater” and “smaller” in order determine what should be denoted by x
R-DEF (definition)	A student refers to notion of percent or to its definition. For example, asserting the difference between “number” and “percentage” in order to explain why Sofia’s approach is wrong
R-AP (analogical problem), with three sub-categories	A student refers to an idea presumably taken from a previously solved problem, which she perceives as analogical
R-AP-MM (market-up/market-down)	A student refers to the market-up/market-down problem, for example, as an explanation for why Sofia’s and Hila’s approaches lead to different answers
R-AP-DQ (differ by quantity)	A student refers to a problem of the differ-by-quantity structure as justifying that a decrease by 25% and an increase by 25% in the context of the task are symmetrical or otherwise, in order to highlight the difference
R-AP-DP (differ by percent)	A student refers to previously solved problems of the differ-by-percent structure for arguing that the operation of multiplication (or addition, or both operations) should be used when translating the collocation “25% higher” and “25% lower” into the algebraic register
R-TS (true solution), with two subcategories	A student uses a particular solution to the given problem as a “gold-standard” reference solution for evaluating other solutions
R-TS-HS (Hila’s solution)	After concluding that Hila’s solution is correct, the student uses it as a reference solution for evaluating Sofia’s solution
R-TS-SP (self-produced solution)	A student uses a self-produced solution, which she believes to be true, as a reference solution for evaluating Hila’s and Sofia’s solutions
R-AF (algebraic/arithmetical facts)	A student attends to self-produced algebraic or arithmetic facts, which she holds as true, as a reference for evaluating Hila’s and Sofia’s solutions. For example, $\frac{0.75x}{0.25x} = \frac{1}{3} \neq \frac{1}{4}, \frac{1.25x}{0.25x} = 5 \neq 4, x + 1.25x \neq 0.75x + x$
R-EQ (equation), including three sub-categories	A student considers equations for evaluating the reasonableness of the approaches
R-EQ-SS (self-substitution)	A student solves the equations, substitutes the numerical answers in the equations, and, based on the fact that the answers fit the equations, concludes that the approaches are correct
R-EQ-DA (different answers)	A student refers to the fact that the answers to the equations are different in order to argue whether Sofia’s and Hila’s approaches are indeed different
R-EQ-WN (whole number)	A student refers to the fact that an answer to Sofia’s equation is a “nice” whole number, in order to argue that Sofia’s solution is reasonable
R-OTH (others)	A student refers to an additional resource, which has not been captured by means of the above codes

5 Findings

5.1 Strategies and discursive products

In this sub-section we address RQ1: What strategies do high-school students use, and how, in small-group discussions of a WIR task in the context of percentages? What are the discursive products of the use of these strategies?

Our first observation is that all the groups used more than one strategy, but employed them in different order. Three groups (G3, G4, and G5, see Table 4) used *independently solving the problem* (F-ISP) as their first strategy. The *exploring the solution-narratives* strategy (F-ESN) was used by all the groups, either as the first strategy (G1, G2 and G6) or as the second one (G3, G4 and G5). Accordingly, for G3, G4 and G5 experience of dealing with the WIR task was

Table 4 Summary by strategies used and pathways of sub-questions explored

Group	General strategies	Pathways of the sub-questions explored
G1	Started with ESN and moved to ISP	How did Hila and Sofia solve the problem? Who is wrong? Why is Sofia's solution wrong? Why are the solutions not the same? How is it possible that Sofia's equation is right but her solution is wrong? Why is Sofia's solution wrong? Why are the solutions different? How to formulate a final answer to the question "Why is Sofia's solution wrong?"
G2	Started with ESN and alternated between ESN and ISP	Who is right? Why is Sofia's solution wrong? How did Sofia solve the problem? How would we solve the problem? Why is Sofia's solution wrong, in light of previously solved problems? What class of percent problems does the given problem belong to? Why is Sofia's solution wrong as compared to Hila's solution, which we know is right? What is the role of language in the task formulation and in percent problems in general? Which percent problems can be solved by adding instead of multiplying?
G3	Started with ISP and moved to ESN	How to solve the problem? Why did one of the students in the group have a different solution? What is the solution to the problem? How did Hila and Sofia solve the problem? Why is Sofia's solution wrong?
G4	Started from using ISP and SEQ in parallel, and then moved to ESN with occasional use of SEQ	Who is right? Is Hila right? Who is right? Why is Sofia wrong? Can it be that Sofia is right? What is the reference base? Who is right? What is the reference base?
G5	Started with ISP, moved to ESN, then to SEQ and moved back to ESN	How to solve the problem? How did Hila and Sofia solve the problem? Will they get the same answer? What is the difference between the solutions? Are the solutions the same? After working on the follow-up task (see Section 4.2) Who is wrong and why?
G6	Started with ESN, switched to SEQ and ISP, and moved back to ESN	How did Hila and Sofia solve the problem? Will Sofia and Hila's equations return the same answers? After working on the follow-up task (see Section 4.2): Who is wrong and why? Why is Sofia's solution wrong? Who is right and why?

ESN exploring the solution narratives, *ISP* independently solving the problem, *SEQ* solving the equations (see Table 1 for details)

compatible with that in AIR tasks: first solve and then look back. The strategy of solving the equations (F-SEQ) was used by G4, G5 and G6 at the advanced stages of the discussions. For example, G6 started with F-ESN for describing and evaluating the provided solutions and concluded that they both were right. Only after they suspected that their conclusion might be wrong, they resorted to F-SEQ.

As for the 'how' part of RQ1, of interest is that the above strategies were used in order to address specific questions, which were continuously changing (see Table 4). Most of the time the discussions revolved not around the main WIR question but around more specific questions. As a rule, specific questions (e.g., 'why is Sofia's solution wrong?') re-appeared at different stages of the discussion. In G2, the tendency was to initially consider specific questions and then involve more general questions (e.g., 'But we've done exercises in this way many times, and it was correct... The question is, in which exercises is this way correct?'). The other five groups repeatedly came back to the previously asked specific questions. However, the discussions were not circular: the same specific questions were approached with the help of different resources and strategies. In addition,

in four groups (G1, G2, G4, and G6) 'strategic splits' were observed: the same questions were considered by different students in parallel but using different strategies.

Table 4 presents the sequences of questions considered in each group. Of note is that the questions are well-aligned to Pólya's questions for the (idealized) looking-back stage. Indeed, the first two of Pólya's questions were: (i) "Can you check the result?" (ii) "Can you check the argument?" These two questions were at the heart of all the discussions. The third of Pólya's questions was "Can you derive the result differently?". Arguably, this question underlined the use of the ISP strategy in all the groups. The fourth question, "Can you see it at a glance?" can be related to the 'what is the difference' questions at the advanced stages of the discussions in G1 and G5 as well as to such questions as 'what is the reference base' in G4. Finally, the fifth of Pólya's questions—"Can you use the result, or the method, for some other problems?"—connects to the 'what is the role of language' and 'what class of percentage problems the given problem belongs to' questions in G2.

As to the group discursive products produced by means of the above strategies, Table 5 shows that most of the

Table 5 Final products of the discussions

Groups	Final discursive products
G1	Hila's solution is endorsed. Sofia's solution is rejected based on the definition of percent and analogy to market up/down problems.
G2	Hila's solution is endorsed. Sofia's solution is different and therefore rejected.
G3	Hila's solution is endorsed. Sofia's solution is rejected based on the analogy to market up/down problems.
G4	Hila's solution is endorsed by one student, but the other two students endorsed both Hila and Sofia's solutions.
G5	None of the solutions is endorsed or rejected.
G6	Hila's solution is endorsed. Sofia's solution is rejected based on the definition of percent and analogy to differ-by-quantity problems.

Table 6 Summary by types of dialogical moves and resources

Groups	Dialogical moves (DM)	Essential resources
G1	Out of 160 DM, SD – 33%, RR – 11%, OO – 56%	Reliance on analogical problems (MM) arithmetic facts (AF) and Hila's solution (TS-HS). The final argument is based on the definition of percent (DEF)
G2	Out of 158 DM, SD – 53%, RR- 20% OO-28%.	Reliance on self-produced solution and Hila's solution (TS), analogical problems (DP, DQ, MM). These resources became shared and combined towards the end of the discussion
G3	Out of 63 DM, SD – 35%, RR – 10%, OO – 55%	Referring to the problematic condition (RPC), reliance on arithmetic facts (AF) and analogical problems (DP, DQ, MM)
G4	Out of 204 DM, SD – 45%, RR – 16%, OO – 39%	Reliance on: a self-produced solution (TS-SP), arithmetic facts (AF), definition of percent (DEF). Referring to the problematic condition (RPC) and to words in the formulation (OF/GS)
G5	Out of 94 DM, SD – 30%, RR – 8%, OO – 62%	Reliance on arithmetic facts (AP) and solutions believed to be true (TS) at the beginning. Then reliance on the fact that the answers to the equations are different (EQ-DA)
G6	Out of 83 DM, SD – 42%, RR – 8% OO – 47%	Reliance on arithmetic facts (AP), equations (EQ), and solutions believed to be true (TS) at the beginning. Then reliance on analogical problems (MM and DQ) and definition of percent (DEF)

DM dialogical moves, *SD* self-disclosure, *RR* request for response, *OO* other-oriented (see Table 2), *TS* true solution, *DP* differ by percent, *DQ* differ by quantity, *MM* market-up/market-down, *EQ* equations, *OF/GS* "of"/"greater-smaller", *RPS* referring to the problematic condition, *DEF* referring to definition of percent (see Table 3)

groups (correctly) endorsed Hila's solution, though the final endorsement and rejection decisions achieved different extents of agreement. Four groups (G1, G2, G3, and G6) reached a full agreement that Hila's solution is right and Sofia's solution is wrong. Of these, three groups (G1, G3 and G6) agreed on a justification as to why Hila's solution is right, and why Sofia's solution is wrong. The agreement in G4 was partial, and G5 did not achieve closure: they only agreed that one of the solutions must be wrong. These findings suggest that the task was quite challenging for all the groups, in spite of the considerably strong mathematical background of the students.

5.2 Dialogical moves

In this sub-section we answer RQ2: What are the characteristics of dialogical moves in the small-group discussions of students working on WIR task? As a reminder, dialogical moves were categorized as self-disclosing utterances, requests for response and other-oriented utterances (see Table 2). As can be seen in Table 6, self-disclosing utterances constituted about 1/3 to 1/2 of the overall number of dialogical moves, which means that, generally speaking, the discussions were not collections of students' monologues but had a dialogical nature. The general picture presented in the

Fig. 2 A coded transcript of a part of Episode 6 in G4

No	St.	Conversational turn	D-codes
122	M.	Listen to me for a second, ok? Listen to me.	D-RR-IN
123	Y.	Ok, I will listen to you and then you will listen to me?	D-OO D-RR-RQ
124	M.	The price...	D-SD
125	Y.	First will you truly listen?	D-RR-IN
126	M.	I have already listened to you	D-OO
127	Y.	No, but listen again, cause you listened while in your head you had your own answer, I will listen to you now without the answer in my head....	D-RR-IN

Legend: **D-codes**; SD: self-disclosure; RR: request for response; OO: other-oriented; IN: invitation to respond.

left column of Table 6 suggests that in G1, G3 and G5 the students were more responsive to each other's ideas than in G2, G4 and G6.

Of interest is an apparent mismatch between the percentages of the requests for response (D-RR) and other-oriented utterances (D-OO). This finding implies that students often responded to their groupmates when not being explicitly asked to do so, and sometimes did not respond to explicit requests for response.

Following the chains of D-codes, the complex structures of the dialogues is further revealed. In all six groups, there were episodes in which the students readily responded to each other. This observation is reflected in the frequent appearance of the codes of the *other-oriented* group of codes (D-OO) (e.g., see two consecutive episodes analyzed in the "Appendix"). In contrast, in all six groups there were episodes in which the students asserted their own ideas in parallel. This observation is reflected in the frequent appearance of the *self-disclosure* code (D-SD). Such episodes often involved a struggle for attention. In Fig. 2, we present part of Episode 6 in G4, in which the struggle for attention became the explicit topic of discussion.

This dialogue exemplifies the reciprocal nature of the struggle: listen now in order to be listened-to later. G4 developed an explicit norm of listening 'in turns' whereas in the rest of the groups the norms of listening remained unarticulated. The different patterns of talking and listening observed in the data seem to be connected to the use of mathematical resources, which brings us to the third research question.

5.3 Mathematical resources

In this sub-section we address RQ3: What mathematical resources are discursively enacted, and how, in the small-group discussions of students working on a WIR task? A wealth of mathematical resources was enacted (see Table 6). As a reminder, we refer to mathematical resources as inferences from the students' utterances which can reasonably

explain what the students might temporarily have held as true while constructing their arguments.

Resources that manifested themselves in all groups were related to the use of analogical problems (R-AP) and to someone's specific solutions to the problems (R-TS). The rest of the resources were used occasionally. Attending to analogical problems (AP-MM for G1 and G6 and AP-DP/Q for G3, see Table 3) helped three groups out of six to solve the task fully. Repeated reliance on verbal clues in order to determine a reference base for 100% (R-RB-OF) led G2 to address the general role of language in percentage problems. Referring to the definition of percent (R-DEF) was used as a final argument in G1 and G6. However, the nomenclatures of the enacted resources by themselves do not fully explain success or its lack thereof in solving the task (see Table 5). Indeed, there were groups that used similar sets of resources, but did not all succeed in reaching a (correct) agreement. The differences seem to be related to how the resources were enacted and responded to.

The enacted resources had different influence on the endorsement/rejection decisions. Some were ignored. Sometimes, an argument that did not immediately sound right to the interlocutors was put aside and another argument was put forward instead. Some of the resources were 'peripheral' in one episode, though became central in the next episodes (e.g., R-AP-DP in Episode 2 in G2, "Appendix"). Simultaneously, a resource that was central in one episode could be 'forgotten' in the next one (e.g., R-AF in Episode 2 in G2, "Appendix"). Sometimes the use of a particular resource was short-lived in the sense that it helped the students to make some progress and then ceased being useful. For example, relying on the fact that the solutions of the equations in Hila and Sofia's solutions were different (R-EQ-DA) enabled the students of G2, G5 and G6 to conclude that at least one of the solutions must be wrong. However, it was not enough for determining who was wrong and why. On this basis, student N. from G6 put into play a new resource, analogy with market-up/market-down problems (R-AP-MM),

which eventually led to their solving of the task. Also, some resources could subsume each other in the discussion. For example, a self-produced solution to the problem in G2 (R-TS-SP) was a resource that underlined the decision to endorse Hila's solution ('Let's say that we do the task. How would we do it? A kilo of the pears equals x . Correct?'). Then reliance on Hila's solution being true (R-TS-HS) became a resource on the way of evaluating Sofia's solution.

In many cases, a student who brought to the discussion a particular resource tended to keep on using it, with little attention to the resources brought by the other students, though, as illustrated, they sometimes struggled for each other's attention (see Fig. 2 and "Appendix").

In sum, the kaleidoscope of different resources and a limited readiness to work with others' resources for more than several moments in a row might explain the inconclusiveness of the discussions in G2 and the lack of agreement in G4 (see Table 5). Success of G1, G3 and G6 in solving the task seems to be related to moderate attention to a considerable number of each other's ideas. The story was different in G5, which ended the discussion without endorsing or rejecting any solution. In this group, when a student brought to the discussion a particular resource (see Table 6), her peer joined her in a non-critical way (i.e., a typical response consisted of *simple agreement*, D-OO-SA). Furthermore, G5 was the only group in which the number of the other-oriented dialogical moves (D-OO) was approximately twice the number of the self-disclosing utterances (D-SD). Accordingly, the inconclusiveness of the discussion in G5 can be attributed to over-collaboration based on the relatively shallow repertoire of enacted resources.

6 Discussion

The goal of the study was to characterize processes involved in small-group discussions of Who-Is-Right (WIR) tasks as a means to support, in realistic classroom situations, practices known as germane to the looking-back stage of problem solving. To address this goal, we designed a WIR task in the context of percentage and enacted it in a regular mathematics lesson, in which high-school students worked on the task in small groups.

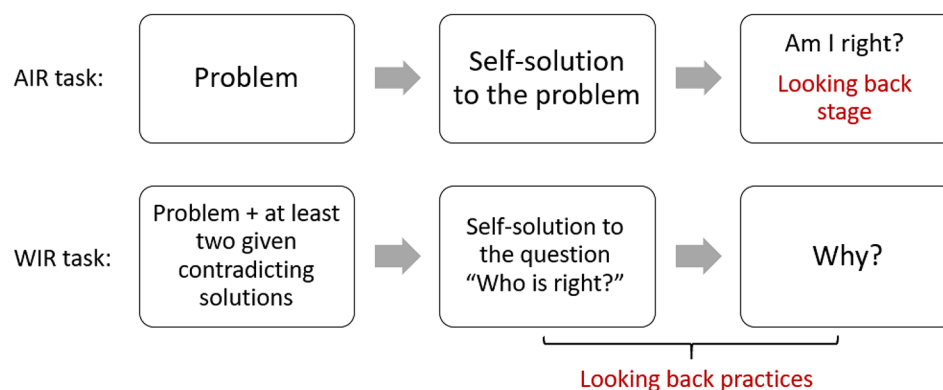
The study was driven by three research questions. In response to RQ1, concerning strategies of coping with the WIR task, we identified three strategies used in the groups in different orders. Those groups who began from independently solving the task and then moved, sometimes repeatedly, to exploring the solution-narratives and examining the algebraic parts of the provided solutions, arguably experienced the task as if it were a regular 'solve and then look back' task (referred to as 'Am I Right' or AIR task). In all cases of coping with the WIR task in our study, the use of the

above three strategies was driven by attempting to address specific sub-questions. Many of these questions appeared to be well-aligned with five questions reserved by Pólya for the (idealized) looking-back stage. Therefore, our central claim was that all the participants in our study were engaged in looking-back practices, if not always in a looking-back stage. In response to RQ2, concerning dialogical moves, we showed that some parts of the discussions were of a collaborative nature, whereas in some other parts the students tended to ignore each others' arguments. In response to RQ3, concerning mathematical resources enacted, we identified a wealth of mathematical resources, but also showed that success with the WIR task cannot be associated with mere use or non-use of certain resources, but rather with ways of enacting some of the resources (most notably, the use of analogical problems) in a dialogue.

We begin the discussion by elaborating on the above-formulated central finding of our study. It is based on the distinction between a looking-back *stage* in AIR tasks, and looking-back *practices* in WIR tasks. An initial comparative discussion of these two types of tasks was provided in Sect. 2.2. We are now in position to say more. First, as our data show, independent solving of a given problem takes place in both cases, though in AIR tasks it is necessarily the first stage and in WIR task it can be either the first or the intermediate stage. Second, as past research shows (see Sect. 1), the appearance or non-appearance of looking-back in AIR tasks frequently depends on prompts provided by an authority. In WIR tasks, even if there is no designated looking-back stage, the solvers engage in verifying solutions and arguments, comparing alternative solutions, and sometimes in considering how the considered means of solution corresponds to that used in solving other problems. These practices are in remarkable accordance with Pólya's (1945/1973) vision for the looking-back stage. Importantly, they were not solicited, at least not explicitly, in the context of the WIR task. Third, even when the looking back stage is present in students' coping with regular AIR tasks, it frequently ends with mere endorsement or rejection of the constructed solution (Malloy & Jones, 1998; Mashiach-Eisenberg & Zaslavsky, 2004). In our WIR task, endorsement of a particular solution came relatively early (e.g., from G1: 'Hila is right because this is how everyone does it'), but then the main challenge began. Beyond endorsement of a particular solution, a WIR task requires students to engage with a series of subtle 'why-questions' stemming from the need to compare the provided and the self-produced solutions. Of note is that AIR tasks, which as a rule operate with one self-produced solution, do not include this opportunity.

Due to the enriched referential basis embedded in the WIR task formulation, handling the 'why' questions naturally required considering the question 'why is the other solution wrong?', which is not the same as considering the

Fig. 3 Comparison of AIR tasks and WIR tasks



question ‘why is the chosen solution right?’ Answering the former question requires the students to use various looking-back practices (e.g., connecting problems by considering analogous problems, formulating implications for future problem solving, and comparing alternative mathematical points of view for explaining their endorsement and rejection decisions to the peers), whereas the answer to the latter question can be socially-based and not necessarily mathematically based (e.g., G1: ‘this is how everyone does it’). The differences between AIR and WIR tasks are schematically summarized in Fig. 3.

To recapitulate, WIR tasks and AIR tasks are essentially different with respect to the opportunities they provide for the use of looking-back practices. In particular, we argue that in addition to affordances inherent in comparing worked-out solutions (Rittle-Johnson & Star, 2011; Swan, 2007), the WIR task contains affordances provided by a regular AIR task. Our hypothesis that WIR tasks can be a valuable tool for engaging students in looking-back practices was partially based on the relative richness (in comparison with AIR tasks) of the space of referential bases for verification. This hypothesis is supported by our findings, which, in turn, calls for future research: it would be interesting to find out if students who systematically engage in WIR tasks tend to employ more verification strategies in regular AIR tasks than students who are exposed only to AIR tasks.

We now turn to discussing our findings in light of past research. The identified strategies of coping with the WIR task—*exploring the solution-narratives* (ESN), *independently solving the problem* (ISP) and *solving the given equations* (SEQ)—are in line with verification strategies identified in past studies. For example—referring to Malloy and Jones’ (1998) and Mashiach-Eizenberg and Zaslavsky (2004) lists of strategies—ESN corresponds to *rereading the problem*, *checking the plan* and *adding justifications to the solution*, ISP to *redoing the problem*, and SEQ to *checking calculations* and *comparing answers*. A novel (yet not surprising) finding is that all small groups in our study used several strategies in the context of the WIR task. As mentioned,

the looking-back stage in the context of AIR tasks is rarely observed, as well as the phenomenon of voluntarily using more than one verification strategy (Mashiach-Eisenberg & Zaslavsky, 2004). We also note that our study documented the student strategies in a realistic classroom situation, and not in an interview setting, as had been done in the prior studies (e.g., Malloy & Jones, 1998; Mashiach-Eisenberg & Zaslavsky, 2004; Papadopoulos & Dagdilelis, 2008).

In addition, our findings imply that comparing worked-out solutions in the WIR format raised the level of difficulty of the task for the students. Indeed, the task in the context of percentages, which could have been offered to middle-school students, evoked hot discussions among high-school students studying mathematics at the highest level of the Israeli curriculum. How so? This may seem particularly surprising in light of the fact that almost all students who adhered to the ISP strategy reiterated Hila’s (correct) solution. Furthermore, based on close familiarity of the second author with the students as their mathematics teacher, we suggest that most of the students could easily have answered the question included in the task-narrative (see Fig. 1) if it were given in a standard format. We deem that the student difficulty with the WIR task can be explained by pointing out that explaining why Sophia’s solution was wrong was a higher-challenge request than producing the correct solution. This is because producing the correct solution could have been solely based on procedures memorized while solving similar problems in the past, and understanding why Sophia’s solution was wrong required a nuanced understanding of the problem situation, as the task analysis provided in Sect. 4.2 shows.

Returning to Pólya’s argument concerning the importance of looking back, our next claim is about the potential of the WIR task for creating connections between problems with an eye to future problem solving. Evidence for this claim consists of the use of analogical problems in all groups and the tendency (observed only in G2) to consider questions of a more general nature than the original task question towards the end of the discussion. G2 discussed which class of percentage problems the given problem belongs to, and

compared the difficulties they encountered in the given task with difficulties in tasks from the past. To our knowledge, documentation of a gradual shift of focus in a problem-solving discussion towards unsolicited generalization (see G2 in Table 4) is a novelty in empirical research on problem solving.

Our next point is about enacting mathematical resources in the group problem-solving. We recall that mathematical resources were conceptualized in our study as discursively-enacted pieces of knowledge held by a student as temporarily true, and which served as anchors for constructing arguments. It was not the resources themselves, which were more or less the same in the most of the groups, but rather the students' patterns of responding to the enacted resources that seemed to account for success or failure with the task. Indeed, resources that could lead to a quick solution (e.g., analogy to differ-by-percent problems) were sometimes enacted but ignored, then repeated, being subsumed under the other resources, modified, forgotten or otherwise put forward.

Asterhan and Schwarz (2009) described a similar phenomenon when writing about shifts of the epistemic status of ideas in small-group discussions. Our study contributes to this line of research. In particular, we found that, as a rule, each student was an enactor of a small set of her 'favorite' resources that she repeatedly returned to throughout the discussion. Our data enable us to suggest that one's 'favorite' resource had a chance to influence further discussion (namely, to gain response from others) if and when it appeared at particular moments, namely, when it resonated with the resources enacted by the others. Abdu and Schwarz (2020) showed that cooperation in small-group problem solving can be as important and useful as collaboration. The above suggestion about how and why the particular resources can influence the discussion sheds light on how and why collaboration and cooperation in problem solving are interleaved.

We now discuss a related finding: *request for response* was the least frequent dialogical move in all the groups, and the *other-oriented dialogical moves* were relatively frequent, especially in G1, G3 and G5. To interpret this finding, let us recall that in written dialogues—for example, in classic Socratic or Lakatosian dialogues (Zazkis & Koichu, 2018) or in dialogues written for learning purposes (Koichu & Zazkis, 2018)—the number of requests for response is roughly equal to the number of responses. An idealized dialogue develops in a linear manner, from one idea to another. The lived dialogues reported in our study are much more complex, and include instances of simultaneous use of several strategies or ideas. A vivid dialogue among teenagers seems to be shaped by their free choice regarding what to respond to, when to respond, and how, in accordance with their own interests and lines of reasoning. To this end, our findings challenge

one of the conclusions of Goos et al.'s (2002) study. In that study, successful problem-solving cases were characterized as those in which the students discarded unhelpful ideas and actively endorsed useful problem-solving strategies, whereas unsuccessful cases were characterized by the lack of critical engagement with the partners' contributions to the discussion. Our findings reveal a more complex picture: it seems that, as a rule, the students were unaware of which idea should have been endorsed or rejected. Accordingly, success or failure in completing the task seems to be related to the extent to which one's ideas could be combined or contrasted with others' ideas, in the quid-pro-quo struggle for attention. In addition, our data corpus includes a case (G5) where the lack of success with the task may be attributed to over-collaboration.

The phenomena discussed in the preceding paragraphs were revealed through the dual coding of dialogical moves (D-codes) and of mathematical resources (R-codes). The former coding scheme is of generic nature, and the latter coding scheme is task-specific. We believe that the codes we developed, along with the way of using them in tandem, can be useful in future studies.

This being said, it is important to discuss limitations of our study. Obviously, the use of one WIR task in one classroom does not enable us to generalize our claims to other tasks and classrooms. Another limitation is related to the fact that our analysis, as complex as it was, did not attend to individual differences among the students, and only narrowly attended to social roles of the interlocutors. With more comprehensive analytical tools we might be able to refine some of the explanations of the observed phenomena. Next, the complexity of the analysis presumes that, in spite of our attempts to assure proper reliability of coding, some mistakes could occur. As a measure of dealing with this inevitable threat, we formulated our findings using qualitative rather than quantitative language whenever possible. Thus, we hope that our conclusions are immune to occasional mistakes in coding.

We conclude by pointing out implications of our study. First, the findings seem encouraging for practice, which gives us an opportunity to reinforce a previous call for the greater use of specially designed WIR tasks in school setting (e.g., Rittle-Johnson & Star, 2011; Swan, 2007; Tabach & Koichu, 2019). To this call we would like to add that WIR tasks have deeply entered the school reality even when not intentionally designed. For example, students encounter spontaneous WIR tasks in various internet resources, in which a wealth of unverified worked-out solutions are discussed (Koichu et al., 2018). Next, we propose that the study provides some ideas for how to increase the feasibility of regular AIR tasks for students. For example, an AIR task can be followed by a WIR task or be included in a sequence of mathematically related

Fig. 4 A coded transcript of Episode 2 in G2

No	St.	Conversational turn	D-codes	R-codes
24	B.	Now let's construct an example. Say, the price for a kilo [of pears] is 5 shekels. Right? Then the price of a kilo of apples is 6.25. OK? But what if it costs, say 2 shekels? Like we do a different number, OK? Not 5.	D-SD D-RR-RQ D-SD D-RR-RQ D-SD D-RR-RQ	R-AF
25	A.	Just a second, I'll bring a calculator.		
26	B.	Now, let's say that a kilo of apples costs 5 shekels.	D-SD	R-AF
27	M.	All this as if for the first time? From the beginning? This plus 25 of x Right?	D-OO-RQ D-SD D-RR-IN	R-AF
28	B.	Yes.	D-OO-SA	
29	M.	So, this is 1 plus 25	D-SD	R-AF
30	A.	As though, OK, let's say that a kilo of pears costs 5 shekels.	D-OO-RA	R-AF
31	B.	Oh, you see? If you do...OK, let's say that we concluded that this is 6.25, and if you take 0.75 of this, it is not 5. Do you understand? It is like, 0.75 of this, it does not say 1.25 multiply this.	D-RR-RQ D-SD D-RR-RQ D-SD	R-AF R-AP-DP
32	A.	This 6.25, it is now $x = 6.25$	D-SD	R-AF
33	B.	No, this is not the same.	D-OO-RA	
34	A.	Yes.	D-OO-SA	
35	B.	OK?	D-RR-IN	
36	A.	Right. This is incorrect.	D-OO-SA D-SD	
37	B.	So, how can we explain this? I mean, why is this [Sofia's solution] wrong?	D-RR-IN	

AIR tasks for the sake of enriching the space of possible reference bases for verification.

As to theoretical implications, our study creates grounds for re-thinking the looking back as the desirable (but hardly achievable) last stage of problem solving (Pólya, 1945/1973), into a set of collaborative looking-back practices that can be evoked at different stages of problem solving. In line with the discursively-oriented perception of problem solving (Koichu, 2019), we offer the following (tentative) conceptualization: the looking back practices in problem solving are socially-shaped processes of endorsing or rejecting a particular set of solution-narratives to a given problem, in which the solvers discursively enact resources available to them prior to or during solving the problem, and also resources developed in collaboration, cooperation or exposure to solutions, which do not belong to the initial set. We hope that this conceptualization

will support discussion of and research attention to the (still underexplored) phenomenon of collective looking back in mathematical problem solving.

Appendix: Illustration of the data analysis

We illustrate the coding processes based on two episodes (out of 11) from the work of Group 2, consisting of three students, A., B. and M.

In Episode 1, the students read the problem formulation and the solution narratives and attended to the central task question 'who is right?'. A. quickly suggested that Hila's solution was right because this was what she would have done herself ('That's it. I'd write, x is a kilogram of pears. As Hila wrote.'). B. and M. agreed with A. However, A.

Fig. 5 A coded transcript of part of Episode 3 in G2

No	St.	Conversational turn	D-codes	R-codes
41	M.	Oh! Here [in Sofia's solution], it is of the price of the apples, and here [in Sofia's solution] – of the price of the pears. This is wrong, this is not 25% of the same price.	D-SD	R-RB-OF
42	B.	This is right, but...	D-OO-SA D-OO-SO	
43	M.	Well, so what do you want to explain?	D-OO-IN	
44	B.	That this exercise [Sofia's solution] is, like, incorrect.	D-OO+ D-SD	
45	M.	OK. This is not of the same, it is not 25% of the same price. Here she took 25% of the price of the apples, and we want 25% of the price of the pears.	D-SD	R-RB-OF R-AP-DP
46	B.	No, this is, this is like...	D-OO-SO	
47	M.	Here it is correct [points to Hila's solution]	D-SD	R-TS-HS
48	B.	It is 25% more.	D-SD	R-RPC
49	M.	OK	D-OO-SA	
50	A.	Right, but it is...	D-OO-SA D-OO-SO	
51	B.	We can think as though it is less.	D-SD	R-AP-DQ

also suggested that Sofia's solution was wrong, and this assertion launched a discussion. A. argued that Sofia's approach was wrong by pointing out that it was based on 'decreasing' the initial price. This argument appeared to be unconvincing to M. A. attempted to formalize her argument by using algebraic tools. She asserted that the price of 1 kg of apples plus the price of 1 kg of pears in Hila's approach and in Sofia's approach were different ($2.25x$ for Hila and $1.75x$ for Sofia). B. opposed A. without entering deeply into what A. had said, and suggested her own line of exploration in Episode 2 (Fig. 4).

As for strategies, a change in strategy, from *exploring the solution narratives* (F-ESN) to *independently solving the problem* (F-ISP), was inferred for A. and B., whereas M. was likely to be still captured by *exploring the solution-narratives* (F-ESN) strategy (# 27). B.'s question at the end of the episode (# 37) suggests that her numerical example was aimed at addressing the question 'why is Sofia's solution wrong?' This question differs from Episode 1's central question 'who is right and why?'

In terms of dialogical moves, B. introduced her own ideas starting in #24. Note that she alternated between disclosing her line of thought (D-SD) and requesting her peers to respond by using rhetorical questions (D-RR-RQ). Both A. and M. followed B.'s argument, as reflected in the frequent appearance of the 'other-oriented' (D-OO) group of codes. Note the difference between a simple agreement move (D-OO-SA) by B. in #28, and a reasoned agreement move (D-OO-RA) by A. in #30.

Regarding mathematical resources enacted, an R-AF code (self-produced arithmetic/algebraic facts) was assigned to most of the conversational turns in Episode 2. An additional resource, an analogy to differ-by-percent problems (R-AP-DP, #31) was employed only by B.

In Episode 3 (Fig. 5), the group continued to explore the 'why is Sofia's solution wrong?' question. What had been a peripheral strategy in Episode 2 became a central strategy in Episode 3. M., who had taken the lead, continued *exploring the solution narratives* (F-ESN). She suggested that Hila's referent for 25% is correct and Sofia's referent is wrong, but B. opposed her. In response, M. modified her argument and, in addition to attending to the role of the word 'of' as a clue (R-RB-OF), resorted to an analogy with differ-by-percent problems (R-AP-DP).

The students were attentive to each other's ideas in Episodes 2–3 (operationally, the D-OO codes were relatively frequent).

The sequence of two episodes is illustrative of the following phenomenon: an opposition to an incomplete or not-understood argument results in an attempted elaboration of the argument by repetition and enactment of additional resources. In this way, a resource that had been peripheral for a while could be put forward again. Simultaneously, a resource, which was central in one episode (i.e., R-AF in Episode 2) could be 'forgotten'. Realization that none of the employed resources led to an unequivocal resolution of the question under exploration may lead to enactment of additional resources, but up to some

saturation point. At that point the ‘forgotten’ resources can come into play again, sometimes in a slightly modified way.

Acknowledgements The study is partially supported by the Israeli Scientific Foundation, Grant No. 2699-17. We thank Jason Cooper and anonymous reviewers for their valuable suggestions following early drafts of the paper.

References

- Abdu, R., & Schwarz, B. (2020). Split up, but stay together: Collaboration and cooperation in mathematical problem solving. *Instructional Science*. <https://doi.org/10.1007/s11251-020-09512-7>
- Asterhan, C. S. C., & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialogue. *Cognitive Science*, 33(3), 374–400
- Ayalon, M., & Even, R. (2014). Factors shaping students’ opportunities to engage in argumentative activity. *International Journal of Science and Mathematics Education*, 14(3), 575–601
- Clement, J. (1982). Algebra word problem solutions: Thought processes underlying a common misconception. *Journal for Research in Mathematics Education*, 13(1), 16–30
- Cobb, P. (2000). Conducting teaching experiments in collaboration with teachers. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education*. (pp. 307–333). Lawrence Erlbaum.
- Dey, I. (1999). *Grounding grounded theory: Guidelines for qualitative inquiry*. Academic Press.
- Goos, M., & Galbraith, P. (1996). Do it this way! Metacognitive strategies in collaborative mathematical problem solving. *Educational Studies in Mathematics*, 30(3), 229–260
- Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics*, 49(2), 193–223
- Kantowski, M. (1977). Processes involved in mathematical problem solving. *Journal for Research in Mathematics Education*, 8, 163–180
- Kilpatrick, J. (1985). A retrospective account of the past 25 years of research on teaching mathematical problem solving. In E. A. Silver (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives*. Hillsdale Lawrence Erlbaum Associates.
- Koichu, B. (2019). A discursively oriented conceptualization of mathematical problem solving. In P. Felmer, P. Liljedahl, & B. Koichu (Eds.), *Problem solving in mathematics instruction and teacher professional development*. (pp. 43–66). Springer.
- Koichu, B., Biton, Y., Gendler, O., & Keller, N. (2018). Problem-solving forums at social networks that accompany mathematics study in Israeli high school. In N. Movshovitz-Hadar (Ed.), *Israel mathematics education K–12*. (pp. 199–208). World Scientific.
- Koichu, B., & Zazkis, R. (2018). “I understand” talk in script writing. A case from Euclid’s Elements. In R. Zazkis & P. Herbst (Eds.), *Mathematical dialogue: Scripting approaches in mathematics education research and practice*. (pp. 163–184). Springer.
- Lester, F. (2013). Thoughts about research on mathematical problem-solving instruction. *The Mathematics Enthusiast*, 10(1–2), 245–278
- Malloy, C. E., & Jones, M. G. (1998). An investigation of African American students’ mathematical problem solving. *Journal for Research in Mathematics Education*, 29(2), 143–163
- Mashiach-Eizenberg, M., & Zaslavsky, O. (2004). Students’ verification strategies for combinatorial problems. *Mathematical Thinking and Learning*, 6(1), 15–36
- Movshovitz-Hadar, N. (Ed.). (2019). *Israel mathematics education K–12*. World Scientific.
- Papadopoulos, I., & Dagdilelis, V. (2008). Students’ use of technological tools for verification purposes in geometry problem solving. *Journal of Mathematical Behavior*, 27(4), 311–325
- Papadopoulos, I., & Sekeroglou, I. (2018). Types of control in collaborative problem solving. In A. Ambrus & E. Vasarhekyi (Eds.), *Problem solving in mathematics education. Proceedings of the 19th ProMath Conference*. (pp. 97–112). Eötvös Loránd University.
- Parker, M., & Leinhardt, G. (1995). Percent: A privileged proportion. *Review of Educational Research*, 65(4), 421–481
- Pólya, G. (1945/1973). *How to solve it*. Princeton: Princeton University Press.
- Rittle-Johnson, B., & Star, J. R. (2011). The power of comparison in learning and instruction: Learning outcomes supported by different types of comparisons. In: *Psychology of learning and motivation* (Vol. 55, pp. 199–225). Academic Press.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.
- Sfard, A., & Kieran, C. (2001). Cognition as communication: Rethinking learning-by-talking through multi-faceted analysis of students’ mathematical interactions. *Mind, Culture, and Activity*, 8(1), 42–76
- Stillman, G. A., & Galbraith, P. (1998). Applying mathematics with real world connections: Metacognitive characteristics of secondary students. *Educational Studies in Mathematics*, 36, 157–189
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage.
- Swan, M. (2007). The impact of task-based professional development on teachers’ practices and beliefs: A design research study. *Journal of Mathematics Teacher Education*, 10(4–6), 217–237
- Tabach, M., & Koichu, B. (2019). Who is right? Theoretical analysis of representational activities. *Proceedings of the 11th Congress of the European Society for Research in Mathematics Education*. Utrecht, the Netherlands. Available at <https://hal.archives-ouvertes.fr/hal-02435330/document>.
- Tavory, I., & Timmermans, S. (2014). *Abductive analysis: Theorizing qualitative research*. University of Chicago Press.
- Thomas, D. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237–246
- Zazkis, R., & Koichu, B. (2018). Dialogues on dialogues: The use of classical dialogues in mathematics teacher education. In R. Zazkis & P. Herbst (Eds.), *Mathematical dialogue: Scripting approaches in mathematics education research and practice*. (pp. 365–387). Springer.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.