TITLE:

Increasing the use of conceptually-derived strategies in arithmetic: using inversion problems to promote the use of associativity shortcuts

AUTHOR NAMES AND AFFILIATIONS:

Joanne Eaves[a]

Nina Attridge[a]

Camilla Gilmore[a]

[a]Mathematics Education Centre, Loughborough University, LE11 3TU

TO APPEAR IN: Learning and Instruction.

CORRESPONDING AUTHOR:

Joanne Eaves[a]

Email: J.Eaves@lboro.ac.uk

[a]Mathematics Education Centre, Loughborough University, LE11 3TU

Declarations of interest: None

ACKNOWLEDGEMENTS: This work was supported by a PhD Studentship from Loughborough University Doctoral college. We would like to thank Jayne Pickering who helped to dual-score the data. C.G. is supported by a Royal Society Dorothy Hodgkin Fellowship.

Highlights

- Associativity is an arithmetic concept that children and adults struggle to apply
- Inversion is a simpler concept that older children and adults easily apply
- Solving inversion problems (‘a + b – b’) increased subsequent associativity use
- ‘a + b – b’ inversion, but not ‘a + b – a’ inversion, increased associativity use
- ‘a + b – b’ problems may direct attention or validate a ‘right-to-left’ strategy

Abstract

Conceptual knowledge of key principles underlying arithmetic is an important precursor to understanding algebra and later success in mathematics. One such principle is associativity, which allows individuals to solve problems in different ways by decomposing and recombining sub-expressions (e.g. ‘a + b – c’ = ‘b – c + a’). More than any other principle, children and adults alike have difficulty understanding it, and educators have called for this to change. We report three intervention studies that were conducted in university classrooms to investigate whether adults’ use of associativity could be improved. In all three studies, it was found that those who first solved inversion problems (e.g. ‘a + b – b’) were more likely than controls to then use associativity on ‘a + b – c’ problems. We suggest that ‘a + b – b’ inversion problems may either direct spatial attention to the location of ‘b – c’ on associativity problems, or implicitly communicate the validity and efficiency of a right-to-left strategy. These findings may be helpful for those designing brief activities that aim to aid the understanding of arithmetic principles and algebra.

Keywords: Associativity; inversion; conceptual; arithmetic; strategy

1. Introduction

1.1. Arithmetic and algebra

The transition from arithmetic to algebra is difficult for many children (Stacey & MacGregor, 1999). However, it represents an important milestone in mathematics education, because algebra acts as a gateway to higher-level mathematical competencies (National Mathematics Advisory Panel, 2008), and predicts educational and employment success (Ladson-Billings, 1997; Knuth et al., 2006, p297). Conceptual knowledge of key principles underlying arithmetic is thought to be an important precursor to understanding algebra (e.g. Booth & Koedinger, 2008). In particular, understanding equivalence (Knuth et al., 2005; Alibali et al., 2007), commutativity and associativity (Warren, 2003) may help children with the transition, however, many children and adults struggle to understand these principles (Robinson & Dube, 2017; Robinson et al., 2018). Here we investigate ways to promote adults' use of associativity via brief interventions, in order to shed light on the reasons why individuals find it difficult.

Our research focused on associativity. This principle is relevant to algebra because it permits sub-expressions within problems to be solved in a different order to that in which they are presented, a process often required for solving algebra problems. Understanding associativity could also encourage individuals to use brackets appropriately, that is, only when they resolve the ambiguity of operation order. For example, an individual who understands associativity would know that the order of operations does not matter on problems such as 'a + b – c', but that it does on problems such as 'a – b – c'. Superfluous use of brackets may have negative consequences on later algebra learning (Gunnarsson et al., 2016) and thus a good understanding of the associativity principle may mitigate against their overuse. Below we discuss what associativity is and how it has been measured, followed by describing the difficulties many have with the principle.

1.2. Defining and measuring associativity

The associativity principle (hereafter 'associativity') states that (a + b) + c = a + (b + c). This refers to the fact that because some operations are related, problems can be decomposed and recombined in different ways to produce a correct answer (Canobi et al., 1998). In the psychology literature, this principle has been applied in different situations, such as with problems that contain only addition (e.g. 'a + b + c = c + b + a'), addition and subtraction (e.g. 'a + b – c = b – c + a'), only multiplication (e.g. 'a x b x c = b x c x a'), or multiplication and division (e.g. 'a x b ÷ c = b ÷ c x a') (Canobi, 2005; Robinson & Ninowski, 2003). Problems that contain opposing operations (addition and subtraction, multiplication and division) are the dominant paradigm used to investigate associativity, which our research adhered to by focusing specifically on addition and subtraction.

For those aged 7 years and above, knowledge of associativity has typically been measured using three-term arithmetic problems, e.g. 'a + b – c' (Klein & Bisanz, 2000). If an individual explicitly reports solving the problem by performing the subtraction before the addition, it may be inferred that they have applied knowledge of associativity. That is, they may have used their understanding of the principle to select a strategy for solving the problem. Here, we make explicit a distinction between two strategies that may be used on 'a + b – c' problems, a) 'left to right', where an individual solves the problem by performing the addition before the subtraction, and b) 'right to left' where an individual solves the problem by performing the subtraction before the addition.

While self-reports can provide evidence for the use of a strategy, they also require conscious awareness and verbal skills to describe the strategy used (Crooks & Alibali, 2014), making a reliance on self-reports alone sub-optimal. Implicit techniques, which infer strategy use from solution accuracy and response times, are therefore often used in conjunction with verbal reports (e.g. Robinson & Ninowski, 2003; Robinson et al., 2006; Dubé & Robinson, 2010; Robinson & Dubé, 2009, 2013; Dubé, 2014; Robinson & Dube, 2017). For associativity, performance can be compared on problems that are 'conducive' to the principle, with those that are not (Edwards, 2013). Conducive problems (e.g. '16 + 47 – 45') encourage the use of associativity because the subtraction yields a small positive number, which makes the 'right-to-left' procedure less computationally demanding. In other words, conducive problems contain a 'shortcut'. In contrast, non-conducive problems (e.g. '36 + 27 – 45') do not contain shortcuts and are designed to encourage a left-to-right procedure. If accuracy and response times are better on conducive than non-conducive problems, an individual is likely to have used the shortcut.

The inferences that can be drawn from both verbal report and implicit performance measures however are not clear-cut, because strategy use and knowledge of arithmetic principles are not perfectly related (Crooks & Alibali, 2014; Schneider & Stern, 2010). Use of shortcut strategies does not necessarily imply a 'deep' understanding of the principle (Star, 2005; Baroody et al., 2007), i.e. an understanding that it is associativity which justifies the shortcut procedure. Instead, shortcut use could reflect 'superficial' understanding, where individuals follow a right-to-left approach due to previous experience or memorised procedures: these individuals may erroneously solve any three-term problem in a different order (e.g. '8 – 4 + 2' as 8 – (4 + 2)). The reverse case is also plausible, where individuals have a deep understanding of associativity, but fail to use a shortcut (see section 1.3). Solutions to 'a + b – c' problems do not therefore perfectly measure knowledge of associativity. Rather, they capture an individual's ability and willingness to apply their knowledge, which may be superficial or deep.

1.3. Difficulties using associativity shortcuts

Associativity is often compared to inversion, the principle that addition and subtraction, and multiplication and division, have an opposite relation (Piaget, 1952; Bryant et al., 1999; Baroody, 2003). Knowledge of inversion is often measured through 'a + b – b' problems, where individuals who understand the principle know that the addition and subtraction cancel out, and that they can simply respond with 'a' (Starkey & Gelman, 1989; Bisanz & LeFevre, 1990).

Compared to inversion, children dislike associativity, and prefer to operate left-to-right rather than using a shortcut (Robinson & Dubé, 2012; Robinson et al., 2016). Indeed, compared to other arithmetic principles, associativity shortcuts are used least (Robinson & Dube, 2017; Robinson et al., 2018): only 15 – 25% of children aged 6 – 10 years use associativity shortcuts (Robinson & Dubé, 2009), a rate that remains low (approximately 30%) in early adolescence (11 – 13 years) (Robinson et al., 2006; Dubé, 2014). Even in adulthood (aged 18 years and over), there is substantial room for improvement, with use hovering at approximately 50% (Yarlas & Sloutsky, 2000; Robinson & Ninowski, 2003; Dubé & Robinson, 2010). Education practitioners have called for this situation to change, in order to ease the transition to algebra (National Mathematics Advisory Panel, 2008). However, to achieve this, we must first understand why associativity shortcuts are rarely used, a topic that we now address.

First, it should be noted that there are many factors that may influence shortcut use. These may be domain-general factors that apply to a range of tasks (Fuchs et al., 2010) or domain-specific factors that apply only to arithmetic. For example, the domain-general skills of attention, working memory and inhibition are likely to be required for using shortcuts, as they enable individuals to change from using familiar (e.g. left-to-right) to unfamiliar (e.g. shortcut) strategies (Luchins, 1942; Lemaire & Lecacheur, 2010; 2011). Deficiencies in any of these domain-general skills may therefore hinder shortcut use. From a domain-specific perspective, some individuals may have a poor understanding of associativity, or not understand the principle at all. Even if they do understand the principle, some may still choose to operate left-to-right, because they are proficient in calculating (Newton et al., 2010), or dislike the process of re-ordering operations (Robinson & Dubé, 2012). Such domain-specific factors may therefore also hinder shortcut use. Herein, we focus on one domain-general and one domain-specific factor, spatial attention and poor understanding of associativity respectively, because they are theoretically relevant (Siegler & Araya, 2005) and potentially malleable (Rittle-Johnson et al., 2016).

Spatial attention (hereafter, attention) is a domain-general process, referring to the prioritised processing of information at a relevant location (Kim & Cave, 1995), for example, looking to the left or right in response to a sound being presented from that direction. Evidence for its role in the use of associativity shortcuts stems from both theory and empirical investigation. Theoretically, SCADS*, the Strategy Change and Discovery Simulation Model* (Siegler & Araya, 2005) is most relevant. SCADS* is a model that was designed to predict when arithmetic strategies are discovered. It was primarily concerned with how domain-general cognitive mechanisms such as priming, forgetting and attention are involved in strategy discovery, and was in part developed from patterns of performance on 'a + b – b' inversion problems. The model proposes that on each trial, familiar strategies race against alternative strategies, a race that familiar strategies initially 'win' because they are well-rehearsed. Over trials, alternative strategies become familiar and gain strength, eventually leading to their application. Attention is proposed to be the first process required for shortcut discovery: an individual must direct their attention to the right-hand side ('b – b') to discover the shortcut. Some have applied the model to associativity (Robinson & LeFevre, 2012), suggesting that the same attentional mechanisms may be required.

Empirically, two studies provide preliminary evidence for the role of attention, one of which was correlational (Watchorn et al., 2014) and the other experimental (Dubé & Robinson, 2010). Watchorn et al., (2014) found that scores on a 'colour-trails' task, a task that partly measures attention skills, predicted inversion shortcut use in 7 – 10 year olds. The second study (Dubé & Robinson, 2010) used a priming paradigm, a technique used in psychology to activate the mental representation of a target before it is presented (Posner & Snyder, 1975). In Dubé & Robinson's study (2010), individuals were primed either to look to the left or to the right of three-term problems by presenting the left or right subexpression for 250ms before the whole problem. On large inversion problems (problems where a left-to-right procedure resulted in a large interim value), reaction time was faster for those in the right-prime condition, suggesting that attention may have facilitated shortcut use. They did not find an effect on associativity problems. However, their associativity problems were multiplication-division problems, a form of associativity that is known to be harder than addition-subtraction (Dubé & Robinson, 2017). Knowledge of multiplication-division associativity may simply be too poor for attention manipulations to aid the identification and use of the shortcut. However, because understanding of addition-subtraction associativity is better, attentional manipulations may be sufficient to facilitate shortcut use.

From a domain-specific perspective, inadequate and conflicting knowledge may limit associativity shortcut use. In the worst-case scenario, some individuals may not understand associativity, i.e. they may not know that operations can be performed in a different order from that in which they are presented. Or, they may understand the principle in a simple context (e.g. with words or objects), but not in an abstract context with digits (Gilmore & Bryant, 2006). Given that some pre-schoolers make correct judgements of associativity with concrete objects (Klein & Bisanz, 2000; Asghari & Khosroshahi, 2016), a complete absence of knowledge is unlikely. It may be more likely that they do not know that the shortcut is valid and efficient on digit-based problems. Alternatively, some individuals could have conflicting knowledge: For example, the acronym BODMAS (Brackets, Order, Multiplication, Division, Addition, Subtraction) and the equivalents BIDMAS and PEMDAS are typically introduced around the age of 11 years in the UK and USA to help individuals remember operator precedence. However, the acronyms are often misinterpreted (Glidden, 2008; Zakis & Rouleau, 2017), with some individuals incorrectly ascribing precedence to addition over subtraction. For 'a + b – c' problems, this would hinder shortcut use.

1.4. Interventions to improve associativity

With the reasons for limited shortcut use in mind, we now address the ways that it could be improved. We begin by discussing previous intervention studies, followed by outlining the logic of our intervention studies. We then discuss the mechanisms through which our proposed interventions may encourage the use of associativity shortcuts.

A handful of interventions have attempted to improve the understanding of different arithmetic principles (e.g. Rittle-Johnson & Star, 2007, 2009), but only two have addressed associativity (Robinson & Dubé, 2012; Robinson & Dubé, 2013). These interventions focused on both inversion and associativity, where children aged 7 – 10 years received explicit demonstrations of the shortcuts and their equivalent left-to-right strategies. For each principle, the children evaluated each strategy by indicating whether they thought it was 'good' and then selected the one they preferred. In both studies, self-reported use of associativity shortcuts increased in an immediate post-test (by about 14 and 22% for those aged 8 and 9 years respectively). However, in a test one week later (Robinson & Dubé, 2012), shortcut use was no higher than before the intervention. Thus, even after an explicit demonstration, children are still reluctant to use associativity shortcuts.

Separately, demonstrations may carry a risk of compromising strategy flexibility (Luchins, 1942; Silver, 1986; Star, 2005), i.e. the ability to generate new strategies and to switch between them on different problems (Verschaffel et al., 2009). So, an individual instructed to perform '38 – 35' before the addition on '6 + 38 – 35' may then apply the same right-to-left procedure on problems where it is less efficient (e.g. '32 + 38 – 69'), or not valid (e.g. '38 – 5 + 6'). Interventions that teach arithmetic principles and the strategies they permit must therefore be designed to encourage their use *only* when they are appropriate and efficient, and to switch to an alternative when they are not. Methods that do not prescribe strategies but instead increase awareness of those available may be more likely to achieve this (Alfieri et al., 2011).

One potential solution arises when one revisits a definition of conceptual knowledge ". . . knowledge of the core principles, and their interrelations" (Schneider & Stern, 2010, p179): knowledge of one shortcut could therefore potentially facilitate the use of another shortcut if the principles from which they are derived are associated or related in memory. To date, only one study has investigated this possibility (Godau, 2014; Experiment 2), where 7 – 8 year olds were either exposed to inversion problems or 'ten-strategy' problems. The 'ten-strategy' is derived from commutativity, the arithmetic principle that permits operands to be added in any order, and the strategy itself refers to problems such as ‘4 + 7 + 6’ being solved through an interim value of 10 (e.g. ‘4 + 6 = 10’ and then ‘10 + 7 = 17’). In a post-test the children solved ‘addends-compare’ problems, problems that also require knowledge of commutativity, and the application of that knowledge across sequential trials (e.g. solving ‘14 + 6 + 8’ quicker after solving ‘6 + 8 + 14’). It was found that the ten-strategy problems led to better performance on addends-compare problems, however, inversion problems did not. The authors surmised that shortcuts facilitate each other only when they are derived from the same principle.

However, in the case of associativity, it may be possible that the shortcuts derived from a different principle (inversion) help, and there are three reasons why. First, in Godau's (2014) study, the low level of transfer found between inversion and commutativity problems was logical, as they are very different principles. Commutativity is concerned with the order of the same operation, while inversion is concerned with the relation between different operations (Canobi et al., 2003). Addition-subtraction associativity however, is similar to inversion because both refer to knowledge of opposing operations. This is consistent with the finding that knowledge of commutativity and inversion are not correlated, while inversion and associativity are (Robinson & Dube, 2017). Inversion and associativity are more aligned and may be associated in memory, making knowledge transfer between the principles a possibility.

Separately, the physical appearance of arithmetic problems, and the way they are perceived by individuals, have been found to influence which components of the problem individuals attend to (e.g. Landy & Goldstone, 2007). Inversion and commutativity problems have lower perceptual similarity than inversion and associativity problems (Robinson & Dubé, 2012). Inversion (a + b – b) and associativity (a + b – c) problems both contain two opposing relations, most often with a shortcut on the right-hand side (‘b – b’ or ‘b – c’). The perceptual resemblance could help individuals notice the shortcut strategy on the associativity problems.

Finally, and our third reason for why Godau's (2014) results may not apply to associativity, relates to the repertoire of strategies that individuals know to be valid for solving a given problem. In Godau's (2014) study, the strategy required on inversion problems (to operate right-to-left) was quite different to the strategy required on addends-compare problems (to look back at previous problems), so it would be unlikely that the former strategy would make individuals aware of the legitimacy of the latter strategy. 'a + b – b' inversion problems and 'a + b – c' associativity problems, however, can be solved through the same strategy, of operating right-to-left. Using a shortcut on 'a + b – b' problems may therefore make individuals aware that a right-to-left strategy produces a correct answer on 'a + b – c' problems.

Indeed, it has been suggested that when an individual is presented with a novel problem, they might think of a strategy they used on an analogous problem to help them solve it. When this strategy leads to a correct result, it is referred to as 'positive transfer'. Studies have investigated the factors that encourage positive transfer between analogous problems, and one that is particularly relevant to our work is the perceptual features within the problem (Sloutsky & Yarlas, 2000; Yarlas & Sloutsky, 2000; Bassok, 1990; Bassok & Novick, 2012). For example, Yarlas & Sloutsky (2000) presented mathematicians and novices with a target problem such as '6 + 3 + 4 = 3 + 4 + 6' and asked them to judge which of two subsequent problems most closely matched it. One of the problems matched the target based on the arithmetic principle of commutativity (e.g. '7 + 2 + 8 = 8 + 2 + 7'), while the other problem matched the target based on perceptual features such as similar numbers (e.g. '6 + 3 + 8 = 3 + 4 + 10'). One finding was that novices more often chose the problem with similar perceptual features, while experts made judgements based on principles. In our studies, participants were non-mathematics students and might therefore be deemed 'relative novices'. When presented with an associativity problem, they might therefore be susceptible to transferring solution procedures based on perceptual features in analogous problems (i.e. 'a + b – b' inversion problems).

1.5. Possible mechanisms for promoting associativity through inversion

From the aforementioned literature, we propose that solving inversion problems may encourage the use of associativity shortcuts, and suggest three possible (not mutually exclusive) mechanisms through which this could operate, a) conceptual knowledge, b) spatial attention, and c) validation of a strategy. We note that the mechanisms are likely to interact. For example, attention is likely to be required for strategy validation, and what an individual attends to may be modulated by their conceptual knowledge (Gibson, 1969). For clarity and simplicity however, these mechanisms are now discussed separately.

First, conceptual knowledge of the two principles may be associated in memory. Solving inversion problems could therefore activate this association, facilitating retrieval of the associativity principle and subsequent use of the shortcut. For this mechanism to work, an individual must have at least some understanding of both principles, stored in memory.

Alternatively, it could be that the mechanism is purely attentional. By this we mean that 'a + b – b' problems could increase associativity shortcut use through a domain-general process that operates without awareness and with little conscious thought. The salient 'b – b' feature in inversion problems may simply direct spatial attention to the right-hand side, where the shortcut is located on associativity problems. Specifically, the novelty of 'a + b – b', and the perceptual salience of the b – b component may encourage individuals to deploy their implicit attention towards to the right-hand side on subsequent problems. This mechanism is akin to directing attention in a 'bottom-up' manner and is domain-general in the sense that it could be activated using a wide variety of stimuli, e.g. by briefly presenting an arrow or spot of light toward the location of the shortcut. Individuals with either a superficial or deep understanding could benefit via this mechanism, as attention merely increases awareness of a strategy that they have prior knowledge of.

Finally, it could be that the mechanism is more domain-specific and that inversion implicitly communicates the legitimacy of a right-to-left strategy. While solving inversion problems, individuals who would otherwise be unaware or uncertain of whether a right-to-left procedure can be applied become more aware, and more certain, that it can. This mechanism is domain-specific in the sense that activating it requires more than a simple, general stimulus such as a spot of light. Rather, it requires top-down manipulation of arithmetic knowledge, such as instructions that challenge the social norms about the ways in which arithmetic problems can be solved. Any perceptual similarity between the problems could aid this mechanism; according to SCADS* (Siegler & Araya, 2005) 'dynamic feature selection' is a process by which salient features in problems become associated with strategy efficiency and allow that strategy to be deployed quickly. For inversion and associativity this association is identical: The salient feature ('b – b' or 'b – c') is similar, and the strategy (right-to-left) is the same. Thus, 'a + b – b' problems may do more than just direct spatial attention, they may inform an individual that a right-to-left strategy can result in a correct answer. After solving 'a + b – b' inversion problems, an individual may be more aware of the validity of a right-to-left strategy, and deploy it within the first few associativity problems presented. Individuals with poor/conflicting knowledge could benefit via this mechanism, as it validates a strategy that they may otherwise not know.

1.6. The present research

We investigated whether the use of associativity shortcuts could be improved. More specifically, we aimed to a) investigate whether exposure to inversion shortcuts could encourage subsequent use of associativity shortcuts and b) if inversion did increase the use of associativity shortcuts, investigate the possible mechanisms through which this occurred.

2. Study 1

In Study 1, we investigated whether exposure to inversion problems would increase adults' self-reported use of a shortcut on an associativity problem.

2.1 Method

All of the studies were approved by the University's Ethics Approvals (Human Participants) Sub-Committee. Before the data were collected the study hypotheses, design, sample size, exclusion criteria and analysis plan were pre-registered at https://aspredicted.org. The pre-registration is available at https://aspredicted.org/ib8ti.pdf.

2.1.1 Participants

109 first-year Psychology undergraduates aged 18 – 30 years (*M* = 19.09, *SD* = 1.54, 88 female, 21 male) participated. This sample provides 87.94% power to detect a medium-size effect in a chi-square analysis of two conditions.

2.1.2 Design

A between-subjects design was used whereby participants were allocated to one of two conditions: inversion or two-term arithmetic (control condition). We chose two-term problems for the control condition because three-term problems would have been substantially more difficult than inversion problems, despite being more visually similar. We judged that two-term problems provided a good comparison, in a first study, to see whether there was any evidence for any of our proposed mechanisms.

Participants were assigned to the conditions in a random order: Before the study, a randomly ordered list of the numbers 1 and 2 (representing each of the conditions) was created. The books that contained the stimuli for the conditions were then arranged in this order and distributed to participants after they were sat at their desks.

2.1.3 Materials and procedure

The stimuli used in Study 1 can be found here https://figshare.com/s/230ee9491e0b67d9d3f3. Each participant had one A4 book (https://figshare.com/s/230ee9491e0b67d9d3f3) that contained the stimuli. Within each book, the stimuli used during different stages of the study (the intervention problems, the self-report questions) were separated into different sections. Participants did not turn over the page from one section to the next without instruction.

Participants either solved 20 inversion problems (inversion condition) or 20 two-term arithmetic problems (control condition). Problems were presented on A4 paper, and participants were asked to mentally solve each problem and then write down their answer, without conferring with others. There was no set time-limit, and the task ended when all participants had put down their pens, which was after approximately 4 minutes and 30 seconds.

After solving the inversion or two-term problems, a test question (33 + 9 – 5) was projected onto a whiteboard at the front of the classroom for 10 seconds using PowerPoint. For all of the studies reported here, the time for the test question was chosen based on what we judged to be sufficient for people to solve it through either a left-to-right or a shortcut procedure, without it being too long such that participants could perform and then compare multiple strategies. '33 + 9 – 5' was chosen as a test question because we judged it to be not too conducive to a shortcut as to produce a ceiling effect, and not too non-conducive to a shortcut as to produce a floor effect. Participants wrote down their answer to the problem and were asked to write 'no answer' if they did not solve it in time. On a separate page, they answered the self-report questions, and were instructed to report the strategy that they were trying to use, even if they did not compute an answer.

Intervention problems

For the inversion condition, 20 inversion problems were created, half in the format 'a + b – b' (e.g. 18 + 7 – 7), and half in the format 'a + b – a' (e.g. 7 + 18 – 7). Problems were presented in an intermixed order, and no more than two problems of the same format were presented consecutively. Half of the problems were large (all double-digit operands), and half were small (at least one single-digit operand). None of the operands contained the unit 0 or contained the unit 1, and answers to the problems were equally distributed among the 10s, 20s, 30s, 40s and 50s, none of which equalled 0, a negative number or a decade boundary.

For the control condition, 20 two-term arithmetic problems were created from the ‘a’ and ‘b’ terms of the inversion problems. Half were addition and half were subtraction problems to mirror the inversion problems, which contained both operations. They were presented in an intermixed order, where no more than two problems of the same type (addition, subtraction) were presented consecutively.

Self-report questions

Participants answered two self-report questions in response to the test problem, 1. “How did you solve the problem?” (open-ended) and 2. “When you solved the problem, what was the first calculation that you did?” (closed). The closed question had three response options, 1. “I did the addition 33 + 9”, 2. “I did the subtraction 9 – 5”, 3. “I did the subtraction 33 – 5”.

2.2 Results and discussion

First, we present our pre-registered analyses, including pre-processing, scoring method, descriptive statistics, and inferential statistics. Second, we present our exploratory analyses, which were not pre-registered. Finally, we interpret the results through our three proposed mechanisms (section 1.5). We hypothesised that exposure to inversion problems would increase the frequency with which associativity shortcuts were identified, and the data from Study 1 can be found here https://figshare.com/s/4191d061536fab1b936c.

2.2.1 Pre-registered analyses

Three participants did not attempt at least 50% of the inversion or two-term arithmetic problems (and thus had not completed the intervention), and one did not answer either of the self-report questions. In line with our pre-registration, these four participants were therefore excluded from all analyses. In addition, one person did not answer the open-ended question, and one person gave an ambiguous response (“I did no calculation”). These two participants were excluded from the analysis of responses to this question only. One person did not answer the categorical question and was excluded from the analysis of responses to this question.

For the open-ended question, participants were classed as a user of the shortcut if their response contained any indication that they had performed the subtraction before the addition (e.g. “I did 9 – 5 = 4, then 4 + 33 = 37” or “I did the subtraction first and then added the remainder”). If participants reported using both procedures (left-to-right and right-to-left), they were classed as a user. Ten of the responses were scored by another researcher, and classifications were 100% consistent. For the closed question, all participants either picked “33 + 9” or “9 – 5”, no participant chose “33 – 5”. Those who picked “9 – 5” were classed as a user of the shortcut.

Table 1 summarises the results and Figure 1 displays the result for the open-ended question. Self-reported use of the associativity shortcut was low overall, at 28% (n = 29/103) on the open-ended question and 22% (n = 23/104) on the closed-question. There were 6 people who were categorised as users of the shortcut on the open-ended question but not on the closed-question. These were people who reported using both a left-to-right and a right-to-left procedure on the open-ended question and reported using a left-to-right procedure on the closed question. On the closed question they may have therefore chosen the response that matched the first (left-to-right) strategy they used.

Table 1. *Number of associativity shortcut users and non-users in each condition for the a) open-ended and b) closed question. Parentheses contain the percent of users or non-users in each condition.*

a)

Measure	Condition	User	Frequency Non-User	Total
Open-ended	Inversion	19 (43%)	25 (57%)	44 (100%)
	Two-term arithmetic	10 (17%)	49 (83%)	59 (100%)
	Total	29 (28%)	74 (72%)	103 (100%)

b)

Measure	Condition	User	Frequency Non-User	Total
Closed	Inversion	15 (33%)	31 (67%)	46 (100%)
	Two-term arithmetic	8 (14%)	50 (86%)	58 (100%)
	Total	23 (22%)	81 (78%)	104 (100%)

Two chi-square tests were performed, one for each outcome measure. For the open-ended question, the frequencies of users to non-users was significantly different between the two conditions, $X^2(1, N = 103) = 8.57$, $p = 0.003$, phi = 0.29, with more shortcut users in the inversion condition. This difference was also significant for the responses to the closed question, $X^2(1, N = 104) = 5.27$, $p = 0.022$, phi = 0.23, with more shortcut users in the inversion condition.

2.2.2 Exploratory analyses

On the test question, 97 participants answered correctly, 3 answered incorrectly, and 5 did not solve it.

We performed exploratory Bayesian analyses to quantify the evidence for our hypothesis. Bayesian 2*2 chi-square tests were conducted in JASP, for which we used the default, uninformative prior concentration of 1 and manually checked for robustness (Jamil et al., 2016). We used the default distributions because we knew of no literature that could help us establish a more informative prior. For the open-ended question, the Bayes Factor (BF_{10}) was 14.73, which according to Jeffreys (1961) provides strong evidence in favour of the alternative hypothesis that there was a difference between the conditions. For the closed question, the BF_{10} was 2.65, which provides anecdotal evidence in support of the alternative hypothesis. Both Bayes factors were deemed to be robust, as they approached 1 (no evidence in favour of either hypothesis) only when the prior changed markedly (to a concentration of approximately 500).

We also analysed the accuracy scores on the intervention problems (inversion, two-term arithmetic). In the inversion condition, the mean number of inversion problems solved correctly was 19.93 (SD = 0.25). In the two-term condition, the mean number of problems correctly solved was 19.31 (SD = 0.97). The difference between the conditions was significant, $t(103) = 4.29$, $p < 0.001$, $d = 0.84$.

Interim conclusion

Solving a mixture of 'a + b – b' inversion problems and 'a + b – a' inversion problems increased self-reported use of the associativity shortcut on the problem '33 + 9 – 5', supporting our hypothesis. We propose three possible explanations of this finding. First, it may be that solving inversion problems activated knowledge not only of inversion, but also of principles that are associated with it in memory, such as associativity. Alternatively, it could be that the perceptual feature 'b – b', present in half of the inversion problems, directed participants' spatial attention to the two right-most digits in the problem, encouraging identification of the shortcut. Finally, a third explanation is that 'a + b – b' inversion problems communicated the legitimacy of a right-to-left strategy. As individuals solve 'a + b – b' problems using the shortcut, they may become aware that a right-to-left approach is a valid way to solve some three-term problems.

In Study 1, we used the self-reported solution strategy on one associativity problem to infer whether solving inversion problems encouraged the subsequent use of associativity shortcuts. Although the intervention did increase self-reported shortcut use, we do not know whether it would have led to improvements in problem-solving performance. In other words, we do not know whether, and if so by how much, solving inversion problems improves the accuracy and speed with which associativity problems are solved. In Study 2, we aimed to replicate the finding of Study 1 and extend it using an additional, implicit measure of shortcut use.

3. Study 2

In Study 1, we relied on one self-report response to infer strategy use, which has limitations. First, the strategy used on any single problem may depend more on the characteristics of the problem (e.g. the number and size of the digits) or the temporary state of the individual (e.g. their attention and concentration), potentially making it less reliable and valid than measuring strategy use across multiple problems. Second, some individuals reported using multiple strategies on the open-ended question, leading us to impose the criterion of 'any indication of shortcut use' for categorising them as a user. We did this to avoid missing instances of shortcut use, but it could be argued that if the shortcut was not the first strategy they wrote down, it was not their predominant strategy and therefore they might not use it on subsequent problems. Finally, self-report measures rely on an awareness of the strategy used, and verbal skills to express it (Crooks & Alibali, 2014). This could be difficult for some, or lead to ambiguous answers, e.g. "I used my head".

To address this, in Study 2, we sought to replicate the effect of the previous intervention and to extend it by measuring the use of associativity shortcuts through both an explicit self-report measure and through an additional, implicit problem-solving measure. The problem-solving measure required individuals to solve as many conducive problems as possible, and separately, as many non-conducive problems as possible, in a restricted timeframe. This measure is akin to those used in laboratory experiments, where strategy use is inferred from accuracy and response times. For example, in a laboratory experiment, if individuals solve 'a + b – b' inversion problems quicker than problems of similar difficulty without shortcuts, use of that shortcut may be inferred. This logic can be extended to develop a measure that is suitable for use in classroom settings (Godau, 2014): if an individual correctly solves more problems that contain shortcuts, than problems that lack shortcuts, in a restricted timeframe, shortcut use may be inferred.

3.1 Method

Before the data were collected the study hypotheses, design, sample size, exclusion criteria and analysis plan were pre-registered at https://aspredicted.org. The pre-registration is available at https://aspredicted.org/cv6dz.pdf.

3.1.1 Participants

52 undergraduates on an introduction to management-related statistics course, aged 18 – 22 years (M = 19.22, SD = 0.92, 24 female, 38 male) participated. This sample provides 58.06% power to detect a medium-size effect in a chi-square analysis of two conditions. We knew the sample size of the classroom would be smaller in advance and anticipated power to be lower than Study 1. We therefore made some changes to the inversion stimuli (section 3.1.3) to try to increase the strength of the manipulation and hence increase power.

3.1.2 Design

As per Study 1, a between-subjects design was used whereby participants were allocated to one of two conditions: inversion or two-term arithmetic (control condition). Participants were assigned to the conditions in a random order: Before the study, a randomly ordered list of the numbers 1 and 2 (representing each of the conditions) was created. The books that contained the stimuli for the conditions were then arranged in this order and distributed to participants after they were sat at their desks.

3.1.3 Materials and procedure

The stimuli used in Study 2 can be found here https://figshare.com/s/3b59db4ee96f411cf9e7. As per Study 1, each participant had one A4 book (https://figshare.com/s/c2b2f174271c7c18e1e6) that contained the stimuli. Within each book, the stimuli used during different phases of the study (the intervention problems, the self-report question, the problem-solving questions) were separated into different sections. Participants did not turn over the page from one section to the next until instructed.

Study 2 consisted of 3 phases, depicted in Figure 2. In the first phase, participants solved 20 intervention problems, which were either inversion or two-term arithmetic problems. Problems were presented on A4 paper, and participants were asked to mentally solve each problem and then write down their answer, without conferring with others. There was no set time-limit, and the task ended when all participants had put down their pens, which was after approximately 4 minutes and 30 seconds.

In the second phase, participants completed the self-report measure: a test question (33 + 9 – 5) was projected onto a whiteboard at the front of the classroom using PowerPoint for 10 seconds and participants were asked to mentally solve the problem, write down their answer and then complete the open-ended self-report question.

In the final phase, participants completed the problem-solving measure. In this task, they solved as many conducive associativity problems as possible in 35 seconds and then solved as many non-conducive problems as possible in the same timeframe (35 seconds was chosen to avoid ceiling effects). The order of the problems within each set was randomised for each person but all participants completed the conducive problems before the non-conducive problems. Non-conducive problems were always solved last, to avoid attenuating any effect of the intervention: If non-conducive problems had been solved first, they could have reinforced a left-to-right strategy and potentially negated any benefit from the inversion problems. If the order had been counterbalanced, this effect would create two groups whose performance on conducive problems would not be comparable: one group would have solved inversion problems, and the other, inversion problems plus non-conducive problems, before the conducive problems were presented.

Intervention problems

20 'a + b – b' inversion problems were used. These were the same problems used in Study 1, however the 'a + b – a' problems were reordered to the format 'a + b – b'. We thought 'a + b – b' problems might be a more powerful manipulation because they could capitalise on all three of the

potential conceptual, attentional and validation mechanisms (section 1.5). For the control condition, the two-term arithmetic problems described in Study 1 were used.

Self-report question

Participants answered one self-report question in response to the test problem, "How did you solve the problem?"

Problem-solving measure

20 'a + b – c' problems were created that were deemed to be conducive to an associativity shortcut. The characteristics of the conducive stimuli were a) the right-hand side, 'b – c', resulted in a small positive integer (ranging between 1 and 5), b) 'b – c' did not involve a decade boundary cross or a borrow operation from the tens to the units, c) the left-hand side, 'a + b', resulted in a multiple-digit number (21 to 113) whose calculation involved a decade boundary cross and a carry operation from the units to the tens.

For each conducive problem, a non-conducive problem was created. Conducive and non-conducive problems were created to be of similar difficulty assuming that they were solved using a left-to-right procedure. For example, the counterpart for the conducive problem '23 + 29 – 27' was '16 + 36 – 27'. Assuming a left-to-right procedure on both problems, the product of the addition and the subtraction were identical, which was intended to make them similar in difficulty. The characteristics of the non-conducive stimuli were a) the result of the interim addition (a + b) and the value of the subtrahend (c) matched conducive stimuli, b) 'a + b' involved a decade boundary cross and a carry operation, c) 'b – c' involved a decade boundary cross, d) the result of 'b – c' ranged between +8 and +38 and between -4 and -42 and e) the result of 'a – c' ranged between +11 and +41 and between -2 and -39.

For each problem type (conducive, non-conducive), 10 were large (consisting of three double-digits) and 10 were small (containing one single-digit and two double-digits). None of the operands contained a unit with a 0 or unit with a 1, and none were identical (i.e. ties). Answers to the problems were equally distributed among the 10s, 20s, 30s, 40s and 50s, none of which equalled 0, a negative number or a decade boundary; interim solutions ('a + b' or 'b – c') also did not equal 0 or a decade boundary.

3.2 Results and discussion

First, we present the outcome of our pre-registered analyses to the self-report question and the problem-solving measure. We then report our exploratory analyses, which were not pre-registered, and then briefly discuss our findings. We hypothesised that the inversion condition would increase the frequency of self-reported use of an associativity shortcut, increase the number of associativity problems correctly solved, and increase the number of associativity problems attempted. The data from Study 2 can be found here https://figshare.com/s/c3c7ad3aece34ff71aa1.

3.2.1 Pre-registered analyses

Self-report question

For the self-report question, participants were classed as a user or non-user of a shortcut using the same criteria as Study 1. One person did not answer the question and so was excluded from the analysis of this question only. Ten of the responses were dual-scored by another researcher and were 100% consistent. Table 2 and Figure 1 display the summary statistics.

Table 2. *Number of users and non-users in each condition of Study 2. Parentheses contain the percent of shortcut users and non-users per condition.*

	Frequency		
Condition	User	Non-User	Total
Inversion	16 (62%)	10 (38%)	26 (100%)
Two-term arithmetic	8 (32%)	17 (68%)	25 (100%)
Total	24 (47%)	27 (53%)	51 (100%)

A chi-square test found that the frequencies of users to non-users was significantly different between the two conditions, $X^2(1, N = 51) = 4.46$, $p = 0.035$, phi = 0.30, with more shortcut users in the inversion condition.

Problem-solving measure

For the problem-solving measure, each participant's performance was indexed by the number of problems solved correctly, and the total number attempted, for each problem type (conducive, non-conducive). Table 3 displays the summary statistics.

Table 3. *The mean (SD) number of problems correctly solved and the total attempted in the inversion and two-term arithmetic (control) conditions*

Condition	Conducive correct	Conducive attempted	Non-conducive correct	Non-conducive attempted
Inversion	9.30 (6.20)	9.85 (5.83)	2.70 (2.00)	4.33 (2.53)
Two-term arithmetic	6.92 (6.19)	7.52 (5.85)	2.24 (1.56)	3.04 (1.49)
Overall	8.15 (6.25)	8.73 (5.90)	2.48 (1.80)	3.71 (2.17)

Between-subjects t-tests were conducted to see whether there was a difference between the inversion and two-term arithmetic conditions in the number of problems correctly solved. There was no significant difference between the conditions for either the conducive, $t(50) = 1.38$, $p = 0.173$, $d = 0.38$, or the non-conducive problems, $t(50) = 0.93$, $p = 0.358$, $d = 0.26$. The same analysis was performed on the number of problems attempted. For the conducive problems, there was no significant difference between the conditions, $t(50) = 1.44$, $p = 0.157$, $d = 0.40$. For the non-conducive problems, Levene's test was significant ($p < 0.05$) indicating that there was unequal variance between the conditions. A Mann-Whitney test was therefore conducted, which did not find a significant difference between the conditions, $W = 443.00$, $p = 0.051$, $r = 0.31$[1].

To assess the validity of the problem-solving measure, between-subjects t-tests were conducted to compare the problem-solving performance of those who were categorised as users on the self-report measure to those who were categorised as non-users. Users solved significantly more conducive problems correctly (M = 11.79, SD = 5.89) than non-users (M = 4.96, SD = 4.80), $t(49) = 4.56$, $p < 0.001$, $d = 1.28$, and attempted more conducive problems (M = 12.25, SD = 5.46) than non-users (M = 5.67, SD = 4.55), $t(49) = 4.70$, $p < 0.001$, $d = 1.32$. There was no significant difference between the users and non-users in the number of non-conducive problems solved correctly (M = 2.71, SD = 1.94 and M = 2.15, SD = 1.54, respectively), $t(49) = 1.15$, $p = 0.257$, $d = 0.32$, or in the number of non-conducive problems attempted (M = 3.96, SD = 2.33 and M = 3.37, SD = 1.96, respectively), $t(49) = 0.98$, $p = 0.333$, $d = 0.27$[2].

Finally, for each participant, a difference score on the problem-solving measure was calculated. The number of non-conducive problems correctly solved, and the number of non-conducive problems attempted were subtracted from the number of conducive problems correctly solved and attempted, respectively. From these scores, whether an individual has used the shortcut may be inferred: If an individual solved substantially more conducive than non-conducive problems, they are likely to have used a more efficient strategy (i.e. the shortcut) on at least some of the conducive problems. Between-subjects t-tests found no significant difference in the difference scores for participants in the inversion and control condition in terms of either the number of problems

[1] An exploratory 2*2 mixed ANOVA with condition (inversion, two-term arithmetic) and problem type (conducive, non-conducive) as the between- and within-subject factors respectively did not find a significant interaction for the number of problems correctly solved ($p = 0.251$), and the number attempted ($p = 0.499$).

[2] An exploratory 2*2 mixed ANOVA with classification type (user, non-user) and problem type (conducive, non-conducive) as the between- and within-subject factors respectively found a significant interaction between the factors for both the number of problems correctly solved ($p < 0.001$), and the number attempted ($p < 0.001$).

correctly solved, $t(50) = 1.16$, $p = 0.251$, $d = 0.32$, or the number attempted, $t(50) = 0.68$, $p = 0.499$, $d = 0.19$.

3.2.2 Exploratory analyses

On the test question, 50 participants answered correctly and 2 answered incorrectly.

We also analysed the accuracy scores on the intervention problems (inversion, two-term arithmetic) using the data from all participants. In the inversion condition, the mean number of problems solved correctly was 19.74 (SD = 0.81). In the two-term condition, the mean number of problems correctly solved was 19.36 (SD = 0.95). The difference between the conditions was not significant, $t(50) = 1.55$, $p = 0.126$, $d = 0.43$.

Bayesian analyses were performed to quantify the evidence for our hypotheses. A Bayesian 2*2 chi-square test on the frequencies of users to non-users in the two conditions returned a BF_{10} of 2.96, which provided anecdotal evidence in support of the alternative hypothesis and was robust to changes in the prior up to a concentration of 30. Bayesian between-subjects t-tests on the number of conducive problems solved correctly, conducive problems attempted, non-conducive problems solved correctly, and non-conducive problems attempted were conducted. These returned BF_{10}s of 0.61, 0.65, 0.40 and 2.05 respectively. There was therefore no evidence for our hypothesis on the problem-solving measure.

Interim conclusion

Study 2 replicated the findings from Study 1, with a group of participants who were studying a different course. As hypothesised, solving inversion problems increased the frequency of self-reported use of an associativity shortcut. Those who self-reported using the shortcut later solved more conducive problems than those who did not. However, we did not find that solving inversion problems in the intervention stage significantly increased the number of conducive or non-conducive problems correctly solved or attempted.

From Studies 1 and 2 we could not determine which, if any, of our three explanations were likely to account for the effect found on the self-report measure. Furthermore, because the control group in those studies always solved two-term arithmetic problems, it could just be that solving three-term problems in general (regardless of the operations or shortcuts they contain), improves performance on 'a + b – c' problems compared to two-term problems. Study 3 addressed these issues. Three conditions were used where individuals either solved 'a + b – b' inversion problems, 'a + b – a' inversion problems, or two-term arithmetic problems. By including an 'a + b – a' group as well as a two-term group, Study 3 helped to answer whether three-term problems in general increased the use of the associativity shortcut, and which of our three explanations was more likely: Did inversion problems encourage associativity use because of a conceptual, attentional or validation mechanism? Both 'a + b – b' and 'a + b – a' problems can be solved by using an inversion shortcut. However, if the mechanism was attentional or due to validation of a right-to-left strategy, only 'a + b – b' inversion problems would encourage shortcut use.

Study 3 also aimed to improve the problem-solving measure. Study 2 demonstrated that the method was valid because self-reported shortcut users solved significantly more conducive problems than non-users, but not significantly more non-conducive problems. However, those in the inversion condition did not solve more conducive problems than those in the two-term condition. This was unexpected: according to the self-report measure, shortcut use was higher in the inversion condition, which should translate into more conducive problems being solved in this condition than the control condition. Our interpretation is that the measure may have had limited sensitivity, given that the overall number of problems solved correctly was low, especially for the non-conducive problems. It may be that 35 seconds for problem-solving was insufficient to reveal a difference between the conditions. In Study 3 we therefore extended the time allowed to 65 seconds.

It should be noted that Studies 1 and 2 used the same test problem, '33 + 9 – 5'. Here, the difference in accuracy and efficiency between a shortcut and a left-to-right approach may be smaller than for a problem with a larger subtraction, e.g. ‘6 + 38 – 35’. It could be that inversion problems increase the use of associativity shortcuts only on problems such as ‘33 + 9 – 5’ because baseline associativity use (i.e. in the control conditions) is low, which offers more room for improvement. A different test problem was therefore used in Study 3 to assess the reliability of the effect found in the previous studies.

4. Study 3

Study 3 had three aims. First, we aimed to investigate the mechanism through which ‘a + b – b’ inversion problems increased associativity shortcut use (conceptual, attention, validation). Did ‘a + b – b’ inversion problems increase associativity shortcut use because they a) had three operands, b) shared conceptual similarity, c) directed attention or d) communicated the validity of a right-to-left strategy? This was done by creating three conditions, ‘a + b – b’ inversion, ‘a + b – a’ inversion[3] and two-term arithmetic (control) and comparing participants’ self-reports and problem-solving performance. Second, we aimed to improve the sensitivity of the problem-solving measure by extending the time allowed on the task. Finally, we aimed to investigate whether the effect found in Studies 1 and 2, where inversion problems increased associativity shortcut use, generalised to a different associativity problem, which we also judged to be not too conducive, and not too non-conducive to a shortcut.

4.1 Method

Before the data were collected the study hypotheses, design, sample size, exclusion criteria and analysis plan were pre-registered at https://aspredicted.org. The pre-registration is available at https://aspredicted.org/dv7n4.pdf.

[3] We note that ‘a + b – a’ are not widely used as inversion problems, but they are problems on which the inversion shortcut can be used.

4.1.1 Participants

257 students aged 18 – 55 years (*M* = 19.67, *SD* = 3.07, 186 female, 70 male, 1 not provided) participated. This sample provides 99.39% power to detect a medium-size effect in a chi-square analysis of three conditions. 188 were first-year Psychology students studying a qualitative research methods module and 69 were business and economics students studying an 'International corporate governance and firms' (ICGF) module.

4.1.2 Design

A between-subjects design was used whereby participants were allocated to one of three conditions: 'a + b – b' inversion, 'a + b – a' inversion, or two-term arithmetic (control condition). Students were allocated to the conditions through blocked random assignment: ten 30-item lists of the numbers 1, 2 and 3 were created, with each list containing 10 instances of each number. The lists were then shuffled to create a random order (Suresh, 2011). This ensured that the number of participants was approximately equal in each condition. Stimuli books were then arranged in the order dictated by the lists before the study and distributed to the students after they were sat at their desks.

4.1.3 Materials and procedure

The stimuli used in Study 3 can be found here https://figshare.com/s/6adad496590c6d713fec. As in Studies 1 and 2, each participant had one A4 book (https://figshare.com/s/7edc99daf2a7588d7d43) that contained the stimuli. Within each book, the stimuli used during different phases of the study (the intervention problems, the self-report question, the problem-solving questions) were separated into different sections. Participants did not turn over the page from one section to the next until instructed.

Study 3 consisted of the same 3 phases as Study 2, depicted in Figure 2. In the first phase, participants solved 20 intervention problems, which were either 'a + b – b' inversion problems, 'a + b – a' inversion problems or two-term arithmetic problems (control condition). Problems were presented on A4 paper, and participants were asked to mentally solve each problem and then write down their answer, without conferring with others. There was no set time-limit, and the task ended when all participants had put down their pens, which was after approximately 4 minutes and 30 seconds.

In the second phase, participants completed the self-report measure: a test question (6 + 38 – 35) was projected onto a whiteboard at the front of the classroom using PowerPoint for 15 seconds and participants were asked to mentally solve the problem, write down their answer and then complete the open-ended self-report question. The time limit was increased in Study 3 because the test question was changed to '6 + 38 – 35', which we judged to be more difficult than the test question used in Studies 1 and 2 ('33 + 9 – 5'). The extra 5 seconds allowed those using a left-to-right strategy to reach an answer.

In the final phase, participants completed the problem-solving measure. In this task, they solved as many conducive associativity problems as possible in 65 seconds and then solved as many non-conducive problems as possible in the same timeframe. 65 seconds was chosen because we judged it to be sufficient for detecting a difference between the conditions, without it being too long. Too long a timeframe might increase shortcut use in both conditions, through opportunity, and potentially mask any difference between them (Siegler & Stern, 1998). The order of the problems within each set was randomised for each person but all participants completed the conducive problems before the non-conducive problems.

Intervention problems

The inversion problems were similar to those in Studies 1 and 2. For the 'a + b – b' inversion condition, 20 stimuli were created, 10 of which had a single-digit for 'a' and a double-digit for 'b'. 5 of the stimuli had a double-digit for 'a' and a single-digit for 'b', while the remainder had double-digits for both 'a' and 'b'. Half of the inversion problems had answers <10, and the other half had answers that were equally distributed among the 10s, 20s, 30s, 40s and 50s. None of the answers equalled 0, a negative number or a decade boundary. Stimuli for the ‘a + b – a’ inversion condition were made by swapping the order of the first and second operands in the ‘a + b – b’ problems. ‘a + b – b’ inversion and ‘a + b – a’ inversion therefore only differed in the location of the shortcut; in the former the shortcut was on the right-hand side, while in the latter it was split across the first and third digits. The problems were presented in a different, randomised order for each participant.

For the control condition, 20 two-term arithmetic problems were created from the inversion problems. 10 were addition problems and 10 were subtraction problems, 15 of which were small (containing one single-digit operand) and 5 of which were large (both operands were double-digits). Answers to the problems were all multiple-digits in the range 11 – 102.

Self-report question

Participants answered one self-report question in response to the test problem, “How did you solve the problem?”

Problem solving measure

50 ‘a + b – c’ problems that were conducive to an associativity shortcut were created. The stimuli were similar, but different to Study 2 as more needed to be created due to the longer time limit on the problem-solving task (65 rather than 35 seconds), and the definition of the stimuli in the previous study (i.e. with one single digit) imposed a constraint on the maximum number of conducive and non-conducive stimuli pairs that could be created. To create more stimuli for Study 3, the main change was that ‘a’, ‘b’ and ‘c’ were all double digits. The conducive stimuli were defined by four features, a) ‘b’ and ‘c’ were distributed among the 10s, 20s, 30s, 40s, 50s and 60s, b) ‘b – c’ resulted in a small positive integer (2 to 5) which was never larger than either of the addends (‘a’ and ‘b’), c) ‘b – c’ did not involve a decade boundary cross or a borrow operation and d) ‘a + b’, resulted in a double-digit number whose calculation involved a decade boundary cross and a carry operation.

For each conducive stimulus, a non-conducive stimulus was created that was of similar difficulty assuming a left-to-right procedure. Non-conducive stimuli were defined by five features, a) the result of the interim addition (a + b) and the value of the subtrahend (c) matched conducive stimuli, b) 'a + b' involved a decade boundary cross and a carry operation, c) 'b – c' involved a decade boundary cross, d) the result of 'b – c' ranged between -7 to -52 and between +13 to +51 and e) the result of 'a – c' ranged between -6 to -49 and between +11 to +33.

4.2 Results and discussion

First, we present the outcome of our pre-registered analyses to the self-report question and the problem-solving data. We then report our pre-registered analysis of the data with 'inconsistent respondents' removed (participants who self-reported not identifying the shortcut but who solved many shortcut problems on the problem-solving task), and analyses for differences between the classes. Finally, we present our exploratory analysis and briefly discuss our findings. We hypothesised that there would be a significant difference among the three conditions in the frequency of self-reported use of an associativity shortcut, the number of problems solved in the time limit, and the number of problems attempted in the time limit. We also hypothesised that there would be a significant difference between the 'a + b – b' inversion and 'a + b – a' inversion conditions in the frequency of self-reported use of an associativity shortcut, the number of associativity problems solved in the time limit, and the number of associativity problems attempted in the time limit. The data from Study 3 can be found here https://figshare.com/s/d5c3fbb77763a9c53778.

4.2.1 Pre-registered analyses

In line with our pre-registration, 4 participants were excluded from all analyses because they had participated in other studies relating to the current research. 1 participant was excluded because they did not attempt at least 50% of the intervention problems. 14 people were excluded from the self-report measure only, because their responses could not be categorised: 5 participants did not answer the question, 3 people said they 'could not remember', and 6 people gave ambiguous responses ('I used my head', 'I nodded off', 'I split it into two equations', 'I did quick maths', 'I used my fingers', 'I did not see the question').

Another researcher scored the self-report responses, the number of problems correctly answered, and the number of problems attempted by 27 participants (9 per condition). The classification of participants as users or non-users of the shortcut were 100% consistent between the scorers. The number of problems answered correctly, and number of problems attempted were 95% consistent between the scorers (103/108 answers marked). Two inconsistencies were due to the scorers miscalculating the sum of their marks, which was resolved by recalculating the correct total. Two inconsistencies were due to the scorers accidentally marking an answer correct/incorrect when it was not, which was resolved by remarking. One inconsistency was for one participant who had unclear handwriting, and it was decided to score this as correct.

Self-report question

For the self-report data, responses were categorised in the same way as in Studies 1 and 2. Table 4 and Figure 1 display the result of the self-report data.

Table 4. *Frequencies of users and non-users in each of the conditions in Study 3*

	Frequency		
Condition	User	Non-User	Total
'a + b – b' inversion	50 (63%)	30 (37%)	80 (100%)
'a + b – a' inversion	34 (43%)	45 (57%)	79 (100%)
Two-term arithmetic	29 (37%)	50 (63%)	79 (100%)
Total	113 (47%)	125 (53%)	238

A 3*2 chi-square test found that the frequencies of users to non-users was significantly different among the conditions, X^2 (2, N = 238) = 11.54, p < 0.003, Cramer's V = 0.22. Subsequent 2*2 chi-square tests found that there were significantly more shortcut users in the 'a + b – b' inversion than the 'a + b – a' inversion condition, X^2 (1, N = 159) = 6.04, p = 0.014, phi= 0.20, more users in the 'a + b

– b’ inversion condition than the two-term condition, X^2 (1, N = 159) = 10.58, p = 0.001, phi= 0.26, but there was no significant difference between the ‘a + b – a’ inversion condition and two-term condition, X^2 (1, N = 158) = 0.66, p = 0.417, phi= 0.07.

Bayesian analyses indicated that for the 3*2 chi-square, the alternative hypothesis of a difference between the conditions in the proportion of users and non-users was supported by a BF_{10} of 10.41 (strong evidence), which was robust to an adjusted prior concentration up to 1000. For the 2*2 chi-square tests, the alternative hypotheses of a difference between the ‘a + b – b’ inversion and ‘a + b – a’ inversion conditions, and between the ‘a + b – b’ inversion and two-term conditions were supported by BF_{10}s of 3.94 (moderate evidence) and 38.50 (very strong evidence) respectively. For the ‘a + b – a’ inversion and two-term condition, the BF_{10} was 0.27, which provided moderate evidence in favour of the null hypothesis.

Problem-solving measure

For the problem-solving data, performance was scored as per Study 2, by the number of problems correctly solved, and the total number attempted. Table 5 displays the result.

Table 5. *The mean (SD) number of problems correctly solved and the total attempted in the ‘a + b – b’ inversion, ‘a + b – a’ inversion and the two-term arithmetic conditions.*

	Problems solved			
Condition	Conducive correct	Conducive attempted	Non-conducive correct	Non-conducive attempted
‘a + b – b’ inversion	14.28 (9.55)	14.87 (9.33)	3.02 (2.30)	4.02 (2.34)
‘a + b – a’ inversion	12.24 (8.72)	12.94 (8.62)	3.38 (3.14)	4.32 (3.15)
Two-term arithmetic	9.46 (8.69)	10.16 (8.57)	3.05 (2.40)	3.69 (2.32)
Overall	12.01 (9.18)	12.67 (9.02)	3.15 (2.63)	4.01 (2.63)

For the number of problems correctly solved, a 3*2 two-way mixed ANOVA was performed with condition (‘a + b – b’ inversion, ‘a + b – a’ inversion, two-term arithmetic) and problem type (conducive, non-conducive) as between- and within-subject factors respectively. There was a main effect of condition, F (2, 249) = 4.82, p = 0.009, η_p^2 = 0.04, a main effect of problem type, F (1, 249) = 271.11, p < 0.001, η_p^2 = 0.52, and a significant interaction between problem type and condition, F (2, 249) = 6.80, p = 0.001, η_p^2 = 0.05. Subsequent pairwise comparisons (Bonferroni corrected) for the main effect of condition indicated that the difference between 'a + b – b' inversion and the two-term arithmetic condition was significant (mean difference = 2.40, t = 3.06 p = 0.007). The difference between 'a + b – b' inversion and the 'a + b – a' inversion condition was not significant (mean difference = 0.84, t = 1.08, p = 0.844), and neither was the difference between 'a + b – a' inversion and the two-term arithmetic condition (mean difference = 1.56, t = 1.98, p = 0.146).

To interpret the interaction between condition and problem type for the number of problems correctly solved, simple main effects analyses were conducted. One way between-subjects ANOVAs indicated that there was a significant difference among the conditions on conducive problems, F (2, 249) = 6.08, p = 0.003, η_p^2 = 0.05, but not on non-conducive problems, F (2, 249) = 0.48, p = 0.619, η_p^2 = 0.00. Subsequent pairwise comparisons (Bonferroni corrected) on the conducive problems indicated that the difference between 'a + b – b' inversion and the two-term arithmetic condition was significant (mean difference = 4.83, t = 3.48, p =0.002). The difference between 'a + b – b' inversion and the 'a + b – a' inversion condition was not significant (mean difference = 2.04, t = 1.48, p = 0.423), and neither was the difference between 'a + b – a' inversion and the two-term arithmetic condition (mean difference = 2.78, t = 2.00 p = 0.141).

Another 3*2 two-way mixed ANOVA was performed to analyse the number of problems attempted, with condition ('a + b – b' inversion, 'a + b – a' inversion, two-term arithmetic) and problem type (conducive, non-conducive) as between and within- subject factors respectively. The same pattern of results emerged. There was a main effect of condition, F (2, 249) = 5.41, p = 0.005, η_p^2 = 0.04, a main effect of problem type, F (1,249) = 278.88, p < 0.001, η_p^2 = 0.53, and a significant interaction between problem type and condition, F (2, 249) = 5.96, p = 0.003, η_p^2 = 0.05. Subsequent pairwise comparisons (Bonferroni corrected) for the main effect of condition indicated that the difference between 'a + b – b' inversion and two-term arithmetic was significant (mean difference = 2.53, t = 3.23, p = 0.004). The difference between 'a + b – b' inversion and 'a + b – a' inversion was not significant (mean difference = 0.82, t = 1.05, p = 0.889), and neither was the difference between 'a + b – a' inversion and two-term arithmetic condition (mean difference = 1.71, t = 2.18, p = 0.091).

To interpret the interaction, simple main effects analyses were conducted. There was a significant difference among the conditions on conducive problems, F (2, 249) = 6.02, p = 0.03, η_p^2= 0.05, but not on non-conducive problems, F (2, 249) = 1.22, p = 0.298, η_p^2 = 0.01. Pairwise comparisons (Bonferroni corrected) on the conducive problems indicated that there was a significant difference between the 'a + b – b' inversion and the two-term arithmetic condition (mean difference = 4.71, t = 3.45, p = 0.002). There was no significant difference between the 'a + b – b' inversion and 'a + b – a' inversion condition (mean difference = 1.93, t = 1.42, p = 0.472), and no significant difference between the 'a + b – a' inversion and two-term arithmetic condition (mean difference = 2.78, t = 2.03, p = 0.129).

Bayesian analyses were also conducted on the problem-solving data. For the number of problems solved correctly, a Bayesian 3*2 ANOVA found that a model which included the interaction term had a BF_{10} that was 22.18 times larger than a model which just included the main effects (strong evidence in favour of the alternative hypothesis). For the number of problems attempted, a Bayesian 3*2 ANOVA had a BF_{10} that was 9.33 times larger (moderate evidence) for a model which included the interaction term, compared to a model which included just the two main effects. Both were robust to changes within the prior r-scale fixed effects range 0.1 – 1.0 (the default used was 0.5).

Inconsistent respondents

The problem-solving data were re-analysed with 'inconsistent respondents' removed (N = 28). Inconsistent respondents were defined in our pre-registration as those who self-reported not identifying the shortcut, but who then solved at least 12 conducive problems, and at least double the number of conducive than non-conducive problems in the problem-solving phase. Their removal did not change the result of the 3*2 ANOVAs. For the number of problems correctly solved, there was a main effect of problem type, F (1,221) = 196.60, $p < 0.001$, $\eta_p^2 = 0.47$, a main effect of condition, F (2, 221) = 5.59, $p = 0.004$, $\eta_p^2 = 0.05$, and a significant interaction between problem type and condition, F (2, 221) = 6.26, $p = 0.002$, $\eta_p^2 = 0.05$. For the number of problems attempted, there was a main effect of problem type, F (1,221) = 202.10, $p < 0.001$, $\eta_p^2 = 0.48$, a main effect of condition, F (2, 221) = 6.27, $p = 0.002$, $\eta_p^2 = 0.05$, and a significant interaction between problem type and condition, F (2, 221) = 5.68, $p = 0.004$, $\eta_p^2 = 0.05$.

Class type

The data were also re-analysed for differences between the classes (ICGF, Psychology). For the self-report data, a log-linear analysis was conducted using the three categorical variables of condition ('a + b – b' inversion, 'a + b – a' inversion, two-term), classification type (user, non-user) and class type (IGCF, Psychology). There was a significant three-way interaction among the variables, G^2 (7) = 14.82, $p = 0.04$. Subsequent 2*2 Chi-square tests identified that the proportion of users to non-users was significantly different across the three conditions for the Psychology class, X^2 (2, N = 177) = 12.42, $p < 0.002$, phi= 0.27, but not for the ICGF class, X^2 (2, N = 61) = 1.56, $p = 0.458$, phi= 0.16. For the problem-solving data, two 3*2*2 mixed ANOVAs were conducted, with condition ('a + b – b' inversion, 'a + b – a' inversion, two-term arithmetic), class type (ICGF, Psychology) and problem type (conducive, non-conducive) as the two between and one within-subjects factors respectively. There was no significant 3-way interaction between condition, problem type, and class type, F (2,246) = 0.88, $p = 0.416$, $\eta_p^2 = 0.01$, for the number of problems correctly solved. There was also no significant 3-way interaction for the total number of problems attempted, F (2,246) = 0.77, $p = 0.463$, $\eta_p^2 = 0.01$.

4.2.2 Exploratory analysis

We also explored the accuracy scores on a) the test question, and b) the intervention problems. For the test question, 212 participants answered correctly, 20 answered incorrectly, and 20 did not answer. Of those who answered incorrectly, 17 were users of the shortcut and 3 were non-users. 3 were in the 'a + b – b' condition, 4 were in the 'a + b – a' condition and 13 were in the two-term arithmetic condition.

For the intervention phase, the mean number of problems answered correctly in the 'a + b – b' inversion condition was 19.72 (SD = 2.18), in the 'a + b – a' inversion condition was 19.75 (SD = 0.74) and in the two-term arithmetic condition was 18.80 (SD = 2.04). A one-way between-subjects ANOVA found a significant difference among the conditions, F (2, 249) = 7.81, p = 0.001, η_p^2 = 0.06. Subsequent pairwise comparisons (Bonferroni corrected) indicated that the difference between the 'a + b – b' inversion condition and the two-term arithmetic condition was significantly different (mean difference = 0.92, t = 3.37, p = 0.003), the difference between the 'a + b – a' and two-term condition was significant (mean difference = 0.96, t = 3.48, p = 0.002), but the difference between the 'a + b – b' inversion condition and 'a + b – a' inversion condition was not significantly different (mean difference = 0.32, t = 0.12, p = 1.000). Our findings are therefore unlikely to be explained by differences in the use of inversion between the two inversion conditions.

Interim conclusion

Study 3 replicated the findings from Studies 1 and 2, where inversion problems increased the self-reported use of associativity shortcuts. As hypothesised, solving inversion problems increased the frequency of self-reported use of an associativity shortcut. A difference among the conditions was also found on the problem-solving measure, where solving inversion problems increased the number of conducive associativity problems solved in a restricted timeframe, but not the number of non-conducive problems. Study 3 extended the findings of Studies 1 and 2, by showing that 'a + b – b' inversion problems, but not 'a + b – a' inversion problems, increased self-reported shortcut use. However, contrary to our hypotheses, there was no significant difference between 'a + b – b' inversion and 'a + b – a' inversion on the problem-solving measure.

5. General discussion

Across three pre-registered intervention studies it was found that solving 20 inversion problems increased the subsequent use of associativity shortcuts. In Study 1, self-reported use of associativity was higher for those who had solved inversion problems than those who had solved two-term arithmetic problems. This finding was replicated in Study 2, which also introduced a problem-solving task to measure implicit shortcut use. In Study 3, those who had solved 'a + b – b' inversion problems, but not those who had solved 'a + b – a' inversion problems, had higher use of associativity on both the self-report and problem-solving measure than those who had solved two-term arithmetic problems. To our knowledge, these are the first studies to show improvements in adults' use of the associativity shortcut following a classroom intervention.

Previous studies have found that compared to other principles such as inversion and commutativity (Kilpatrick et al., 2002), associativity is used less often to solve arithmetic problems (Robinson & Dube, 2017). Our data are broadly consistent with the existing literature, which found that adults self-reported using addition-subtraction associativity shortcuts on 40 – 66% of trials (Robinson & Ninowski, 2003; Robinson & Beatch, 2016). In our studies, shortcut use was between 17 – 63% on the first problem that was presented, depending on the condition participants were in (e.g. inversion, two-term arithmetic) and the problem that was presented ('33 + 9 – 5' or '6 + 38 – 35').

Our results help researchers to understand why associativity shortcut use is low, even in well-educated adult samples (section 5.1). Developing this understanding is important if we are to build effective methods for teaching the principle and the strategies it permits. Given that associativity enables the development of more advanced mathematical competencies and helps to bridge the gap from arithmetic to algebra (Booth & Koedinger, 2008), knowledge of the principle, and the ability to apply that knowledge to arithmetic problems, is an important goal for education. Education practitioners have therefore called for greater research effort into improving the use of arithmetic principles (National Mathematics Advisory Panel, 2008). Our studies contributed to that goal by investigating whether the use of associativity shortcuts could be improved, and if so, through what mechanisms this improvement may occur. We now discuss the theoretical and practical significance of the results in further detail.

5.1 Theoretical contribution

Our findings make two important theoretical contributions. First, they indicate a possible mechanism through which inversion and associativity are linked, which in turn might help to reveal why associativity shortcut use is low. Second, they help to extend existing theories and models of strategy discovery and use. We discuss these contributions in turn, followed by the practical implications of our work. We note that these mechanisms may not be entirely independent (Gibson, 1969) and that our intervention might operate through an interaction among them. For clarity, we discuss each mechanism separately below.

5.1.1 A conceptual mechanism?

Recent cross-sectional studies have found that knowledge of some arithmetic principles are not related, such as commutativity and inversion (Robinson & Dube, 2017). This is backed by experimental work which found that inversion shortcuts did not encourage the use of commutativity shortcuts (Godau, 2014), leading to a suggestion that shortcut-to-shortcut transfer does not occur if those shortcuts are derived from different principles. Here, we show that this is not always the case. Solving 'a + b – b' inversion problems consistently increased the use of associativity shortcuts, in keeping with the positive correlation between inversion and associativity shortcut use (Robinson & Dube, 2017). Shortcut-to-shortcut transfer can therefore occur between different principles if they are closely related.

However, our studies suggest that the mechanism through which this occurred is unlikely to be purely conceptual (section 1.5), that is, through activation of an association between knowledge of inversion and associativity, or simply because both inversion and associativity stimuli have three terms. In Study 3, 'a + b – a' inversion problems did not increase associativity shortcut use, only 'a + b – b' inversion problems did. If the mechanism had been purely conceptual or because the stimuli both had three terms, ‘a + b – a’ inversion problems would have activated associativity knowledge and increased shortcut use: The fact that we found no evidence of this suggests that it is not simply the understanding of the underlying principles that aids performance. Rather, there is something specific to the ‘a + b – b’ inversion problems that enables them to drive the effect.

5.1.2 An attentional mechanism?

Another explanation of the results could be that ‘a + b – b’ problems (but not ‘a + b – a’ problems) directed spatial attention to the location of the shortcut in associativity problems. The perceptual salience of ‘b – b’ in inversion problems may direct attention in a domain-general, bottom-up manner, akin to the way any cue, such as an arrow or spot of light, can direct attention to different locations (Posner et al., 1980; Tipples, 2002). Two previous studies have investigated the role of attention in shortcut use, and provide some indication that this may be the case. Watchorn et al. (2014) reported a positive correlation between performance on a ‘colour-trails’ task and inversion shortcut use. However, as the colour-trails task measures a mixture of domain-general skills, including inhibition and switching, it may have been these, not attention, that were important for using the shortcut. Second, Dubé & Robinson (2010) found that those who were primed toward the right-hand side solved inversion problems quicker than those who were primed to the left. However, this effect was only found on inversion problems, not associativity problems. Attention may therefore be sufficient for encouraging inversion, but not associativity.

5.1.3 A strategy validation mechanism?

Another potential explanation for our findings is that ‘a + b – b’ inversion problems implicitly communicated the legitimacy of a right-to-left strategy. In Study 3, ‘a + b – b’ problems led to a consistent advantage over two-term arithmetic problems on both associativity outcome measures, and an advantage over ‘a + b – a’ problems on one of the outcome measures. We argue that for ‘a + b – b’ inversion problems, most adults know that a right-to-left strategy is valid (Robinson &

Ninowski, 2003). Solving them may therefore reinforce the validity of a right-to-left strategy in general, which leads to an increase in right-to-left strategy use on associativity problems.

This mechanism is likely to be aided by the similar features between the problem types. On inversion problems, an association is built between the right-to-left strategy and the perceptual feature ‘b – b’. On associativity problems, the association is similar, between a right-to-left strategy and the perceptual feature ‘b – c’. Prior activation of the association by the inversion problems could therefore allow the strategy to be deployed quickly on associativity problems (i.e. the first associativity problem presented).

Our data do not allow us to differentiate between the attention and validation mechanisms, and it is likely that both play a role in the use of the associativity shortcut. These mechanisms may explain the relatively low use of associativity shortcuts found in previous studies (e.g. Robinson & Ninowski, 2003; Robinson & Beatch, 2016; Robinson & Dube, 2017). Individuals may have paid insufficient attention to the location of the shortcut, or may have been unaware that a right-to-left approach was a valid strategy for solving the problems. We suggest that future research should investigate the relative explanatory power of both potential mechanisms (section 5.3).

5.1.4 Existing models of strategy discovery

The validation mechanism is broadly consistent with existing cognitive models in the strategy literature (Siegler & Shipley, 1995; Shrager & Siegler, 1998). For example, in SCADS* (Siegler & Araya, 2005), priming and dynamic feature detection are two processes involved in the discovery of inversion shortcuts. Priming refers to the finding that successful solution strategies used on previous trials encourage the use of that strategy on subsequent trials, while feature detection refers to a monitoring system that, over a series of trials, extracts the features that are relevant for solving problems (e.g. for inversion problems, the numbers that are identical). Feature extraction and their association with valid strategies (priming) are key aspects of our explanation. Thus, while some have commented that SCADS* may be applied to associativity problems (Lefevre & Robinson, 2010; Robinson & LeFevre, 2012), our studies provide the first indication that, to some extent, it can.

However, our explanation goes beyond these mechanisms. Implicit in SCADS* is that the feature and strategy on one problem need to be identical to the feature and strategy on another problem, for that strategy to be selected. Indeed, according to Siegler (1998, 2005), solving inversion problems could be *harmful* for strategy selection on associativity problems (2005, p.26). For example, with children, solving inversion problems could cause them to overgeneralise the strategy, e.g. after solving many 'a + b – b' inversion problems, an individual may erroneously apply the exact same strategy to the problem '18 + 7 – 5' and respond '18'. However, that was not the case in the current studies, no participant in the inversion conditions gave 'a' as their answer to the associativity problem. Our results show, for the first time, that 'a + b – b' inversion problems can increase the use of a shortcut on 'a + b – c' problems, and thus that the features and strategy on one problem need only be similar to the features and strategy in a different problem, to encourage strategy selection.

5.2 Practical implications

Our studies were not designed to uncover how associativity should be taught, and the findings should not be taken as such. Many factors are likely to contribute to associativity shortcut use (section 1.3) and solving inversion problems is not a remedy for all these issues. Furthermore, we cannot guarantee that our studies encouraged a deep understanding of associativity (Star, 2005), as strategies do not necessarily imply knowledge of the principle (Baroody et al., 2009; Siegler, 1988; Torbeyns et al., 2016; Hansen et al., 2015). Some individuals may have obtained superficial knowledge, undertaking a right-to-left approach without understanding the principle of associativity that justifies the procedure. This could lead them to erroneously solve any problem with multiple terms in a different order (e.g. '8 – 4 + 2' as 8 – (4 + 2)). The findings should not, therefore, be applied without further investigation.

Nevertheless, the findings are important because they uncover underlying mechanisms that could be used to fuel further research into the design of interventions for improving the understanding of arithmetic principles and performance on algebra problems. Here, we found that a quick activity, administered to a whole classroom at once, improved the use of associativity shortcuts. Future studies could therefore investigate whether these simple, efficient tasks can encourage a long-lasting, 'deep' understanding of associativity.

5.3 Future research

Our data do not allow us to differentiate between the attention and validation mechanisms, and it is plausible that both interact to play a role in the use of the associativity shortcut. For example, for 'a + b – b' problems to validate a right-to-left strategy it is likely that they must also, at least initially, direct spatial attention to the right-hand side of the problem. Future research could try to disentangle the contribution of these two mechanisms, and we suggest two possible ways in which this could be done. First, studies that directly manipulate spatial attention to different components of the problem could be conducted. One previous study (Dubé & Robinson, 2010) used a priming technique to achieve this manipulation, which did improve performance on inversion shortcut problems. However, due to the nature of the stimuli (Dubé & Robinson, 2010 were concerned with multiplication-division), the role of attention in addition-subtraction associativity problems is unknown. A second approach to disentangle the validation and attention mechanisms would be to include a control condition where the problems direct attention to the right, but do not validate a right-to-left strategy. For example, 'a + b + b' problems such as 3 + 47 + 47 may direct individuals' attention to the right-hand side because the two right-most digits are identical. However, 3 + 47 + 47 is unlikely to validate a right-to-left strategy, as adding 3 to 47 is easier than adding 47 to 47. Future research could explore this possibility: if solving both 'a + b + b' and 'a + b – b' problems increase the use of associativity shortcuts on subsequent problems, then this is likely to be due to increased attention directed to the location of the shortcut. However, if 'a + b + b' problems do not increase subsequent associativity shortcut identification as much as 'a + b – b' problems, then it is more likely that the effect is driven through strategy validation, because only ‘a + b – b’ problems promote a right-to-left solution approach.

We also do not know whether our findings generalise to other forms of associativity shortcuts, or whether they are limited to shortcuts on the right-hand side of ‘a + b – c’ problems. For example, it would be interesting to know whether ‘a + b – b’ problems could increase shortcut use on ‘a + b + c’ problems such as 48 + 7 + 3, ‘a + b – c’ problems such as ‘38 + 6 – 35’, and multiplication-division problems such as ‘7 x 6 ÷ 2’. These problems all contain shortcuts that are derived from the associativity principle. In theory, if ‘a + b – b’ inversion problems operate through a conceptual mechanism, shortcut use should increase on all three problems. However, the problems differ in the operations they contain and the location of the shortcut; different outcomes for each problem might therefore be expected and add further weight to our claim that our findings are better explained through a non-conceptual mechanism.

Lastly, our studies also do not tell us whether or how long-lived the effect of the intervention may be, or the depth of knowledge that may be encouraged. Future research should therefore include delayed post-tests and tasks that require participants to provide explanations and justifications, to investigate the extent to which inversion problems foster knowledge and use of the associativity principle.

6. Conclusion

Improving knowledge of associativity may ease the transition from arithmetic to algebra. In three studies, we found that the use of associativity shortcuts could be improved by solving problems that contained inversion shortcuts beforehand. Our findings suggest that this effect is unlikely to be driven through a link between the concepts stored in memory. Rather, it is more likely that inversion problems direct spatial attention to the location of the shortcut, or implicitly communicate the validity of a right-to-left strategy, or both. Further work investigating which of these two mechanisms is more likely, and the depth and longevity of knowledge that the interventions delivered, is warranted before the findings can be used to help counter the difficulties that many have with the associativity principle.

Figures

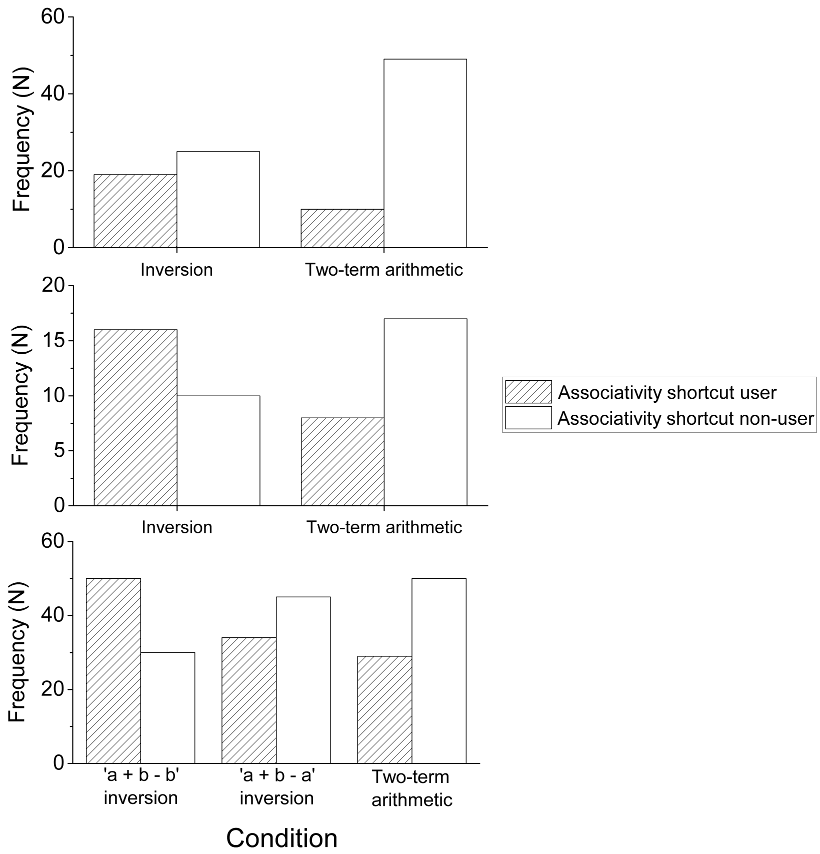

Figure 1: The number of participants who self-reported using an associativity shortcut in an open-ended question for each of the conditions in Study 1 (upper), 2 (middle) and 3 (lower)

Phase	Inversion condition	Two-term arithmetic condition
1 Intervention problems	20 inversion problems "18 + 7 – 7"	20 two-term arithmetic problems "18 + 7" and "18 – 7"
2 Self-report measure	Solve "33 + 9 – 5"	
	"How did you solve the problem?"	
3 Problem-solving measure	Conducive problems e.g. "53 + 58 – 56" (35 seconds)	
	Non-Conducive problems e.g. "67 + 44 – 56" (35 seconds)	

Figure 2: Overview of the procedure used in Study 2

References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *103*(1), 1–18. https://doi.org/10.1037/a0021017

Alibali, M. W., Knuth, E. J., Hattikudur, S., McNeil, N. M., & Stephens, A. C. (2007). A longitudinal examination of middle school students' understanding of the equal sign and equivalent equations. *Mathematical Thinking and Learning*, *9*(3), 221–247. https://doi.org/10.1080/10986060701360902

Asghari, A., & Khosroshahi, L. (2016). Making associativity operational. *International Journal of Science and Mathematics Education*, *15*(8), 1559–1577. https://doi.org/https://doi.org/10.1007/s10763-016-9759-1

Baroody, A. J. (2003). The development of arithmetic concepts and skills: Constructing adaptive expertise. In A. J. Baroody & A. Dowker (Eds.), *The Development of Arithmetic Concepts and Skills: Constructive Adaptive Expertise* (pp. 1–490). Mahwah, N.J.: Lawrence Erlbaum Associates, Inc. https://doi.org/10.4324/9781410607218

Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education*, *38*(2), 115–131. https://doi.org/10.2307/30034952

Baroody, Torbeyns, J., & Verschaffel, L. (2009). Young children's understanding and application of subtraction-related principles. *Mathematical Thinking and Learning*, *11*(1–2), 1–9. https://doi.org/https://doi.org/10.1080/10986060802583873

Bassok, M. (1990). Transfer of domain-specific problem-solving procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 522–533. https://doi.org/10.1037/0278-7393.16.3.522

Bassok, M., & Novick, L. R. (2012). Problem Solving. In K. Holyoak & R. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 321–349). New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0021

Bisanz, J., & LeFevre, J. (1990). Strategic and nonstrategic processing in the development of mathematical cognition. In D. Bjorklund (Ed.), *Children's Strategies: Contemporary Views of Cognitive Development* (pp. 213–244). New York: Psychology Press.

Booth, J. L., & Koedinger, K. R. (2008). Key misconceptions in algebraic problem solving. In B. Love, K. McRae, & V. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 571–576). Austin, USA: Cognitive Science Society.

Bryant, P., Christie, C., & Rendu, a. (1999). Children's understanding of the relation between addition and subtraction: inversion, identity, and decomposition. *Journal of Experimental Child Psychology*, *74*(3), 194–212. https://doi.org/10.1006/jecp.1999.2517

Canobi, K. H. (2005). Children's profiles of addition and subtraction understanding. *Journal of Experimental Child Psychology*, *92*(3), 220–246. https://doi.org/10.1016/j.jecp.2005.06.001

Canobi, K. H., Reeve, R. A., & Pattison, P. E. (1998). The role of conceptual understanding in children's addition problem solving. *Developmental Psychology*, *34*(5), 882–891. https://doi.org/10.1037/0012-1649.34.5.882

Canobi, K. H., Reeve, R. A., & Pattison, P. E. (2003). Patterns of knowledge in children's addition. *Developmental Psychology*, *39*(3), 521–534. https://doi.org/10.1037/0012-1649.39.3.521

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review*, *34*(4), 344–377. https://doi.org/10.1016/j.dr.2014.10.001

Dubé, A. K. (2014). Adolescents' understanding of inversion and associativity. *Learning and Individual*

Differences, *36*, 49–59. https://doi.org/10.1016/j.lindif.2014.09.002

Dubé, A. K., & Robinson, K. M. (2010). The relationship between adults' conceptual understanding of inversion and associativity. *Canadian Journal of Experimental Psychology*, *64*(1), 60–66. https://doi.org/10.1037/a0017756

Dubé, A. K., & Robinson, K. M. (2017). Children's understanding of multiplication and division: Insights from a pooled analysis of seven studies conducted across 7 years. *British Journal of Developmental Psychology*, *36*, 206–219. https://doi.org/10.1111/bjdp.12217

Edwards, W. (2013). *Underlying components and conceptual knowledge in arithmetic*. University of Regina, Saskatchewan. Retrieved from https://ourspace.uregina.ca/bitstream/handle/10294/5434/Edwards_William_200275843_MA_EAP_Spring2014.pdf?sequence=1

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., … Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology*, *46*(6), 1731–1746. https://doi.org/10.1037/a0020662

Gibson, E. J. (1969). Trends in perceptual development. In E. J. Gibson (Ed.), *An Odyssey in Learning and Perception* (pp. 450–472). Cambridge, Massachusetts: MIT Press.

Gilmore, C. K., & Bryant, P. (2006). Individual differences in children's understanding of inversion and arithmetical skill. *The British Journal of Educational Psychology*, *76*(Pt 2), 309–331. https://doi.org/10.1348/000709905X39125

Glidden, P. L. (2008). Prospective elementary teachers' understanding of order of operations. *School Science and Mathematics*, *108*(4), 130–136. Retrieved from https://doi.org/10.1111/j.1949-8594.2008.tb17819.x

Godau, C. (2014). Spontaneously spotting and applying shortcuts in arithmetic - a primary school perspective on expertise. *Frontiers in Psychology*, *5*, 1–11. https://doi.org/10.3389/fpsyg.2014.00556

Gunnarsson, R., Sönnerhed, W. W., & Hernell, B. (2016). Does it help to use mathematically superfluous brackets when teaching the rules for the order of operations? *Educational Studies in Mathematics*, *92*(1), 91–105. https://doi.org/10.1007/s10649-015-9667-2

Hansen, S. M., Haider, H., Eichler, A., Godau, C., Frensch, P. A., & Gaschler, R. (2015). Fostering formal commutativity knowledge with approximate arithmetic. *PLoS ONE*, *10*(11), 1–27. https://doi.org/10.1371/journal.pone.0142551

Jamil, T., Ly, A., Morey, R. D., Love, J., Marsman, M., & Wagenmakers, E.-J. (2016). Default "Gunel and Dickey" Bayes factors for contingency tables. *Behavior Research Methods*, *49*(2), 638–652. https://doi.org/10.3758/s13428-016-0739-8

Jeffreys, H. (1961). *Theory of Probability*. *Oxford Classic Texts in the Physical Sciences* (3rd ed.). Oxford: Oxford University Press, Clarendon press.

Kilpatrick, J., Swafford, J., & Findell, B. (2002). *Adding it Up: Helping children learn mathematics*. Washington, D.C: National Academic Press. https://doi.org/10.17226/9822

Kim, M. S., & Cave, K. R. (1995). Spatial attention in visual search for features and feature conjunctions. *Psychological Science*, *6*(6), 376–380. https://doi.org/10.1111/j.1467-9280.1995.tb00529.x

Klein, J. S., & Bisanz, J. (2000). Preschoolers doing arithmetic: the concepts are willing but the working memory is weak. *Canadian Journal of Experimental Psychology*, *54*(2), 105–116. https://doi.org/10.1037/h0087333

Knuth, E. J., Alibali, M. W., McNeil, N. M., Weinberg, A., & Stephens, A. C. (2005). Middle school students' understanding of core algebraic concepts: Equivalence & Variable. *Zentralblatt Für Didaktik Der Mathematik*, *37*(1), 68–76. https://doi.org/10.1007/BF02655899

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education*, *37*(4), 297–312. https://doi.org/10.2307/30034852

Ladson-Billings, G. (1997). It doesn't add up: African American students' mathematics achievement. *Journal for Research in Mathematics Education*, *28*(6), 697–708. https://doi.org/10.2307/749638

Landy, D., & Goldstone, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 720–733. https://doi.org/10.1037/0278-7393.33.4.720

Lefevre, J. A., & Robinson, K. M. (2010). Use of conceptual knowledge among adults: Experimental evidence. *Unpublished Manuscript*.

Lemaire, P., & Lecacheur, M. (2010). Strategy switch costs in arithmetic problem solving. *Memory & Cognition*, *38*(3), 322–32. https://doi.org/10.3758/MC.38.3.322

Lemaire, P., & Lecacheur, M. (2011). Age-related changes in children's executive functions and strategy selection: A study in computational estimation. *Cognitive Development*, *26*(3), 282–294. https://doi.org/10.1016/j.cogdev.2011.01.002

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, *54*(6), 1–95. https://doi.org/http://dx.doi.org/10.1037/h0093502

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel* (Vol. 37). Washington, D.C. https://doi.org/10.3102/0013189X08329195

Newton, K. J., Star, J. R., & Lynch, K. (2010). Understanding the development of flexibility in struggling algebra students. *Mathematical Thinking and Learning*, *12*(4), 282–305. https://doi.org/10.1080/10986065.2010.482150

Piaget, J. (1952). *The child's conception of number.* London: Routledge & Kegan Paul.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 669–682). London: Academic Press.

Rittle-Johnson, B., Fyfe, E. R., & Loehr, A. M. (2016). Improving conceptual and procedural knowledge: The impact of instructional content within a mathematics lesson. *British Journal of Educational Psychology*, *86*(4), 576–591. https://doi.org/10.1111/bjep.12124

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, *101*(3), 529–544. https://doi.org/10.1037/a0014224

Robinson, K. M. (2001). The validity of verbal reports in children' s subtraction. *Journal of Educational Psychology*, *93*(1), 211–222. https://doi.org/10.1037/0022-0663.93.1.211

Robinson, K. M., & Beatch, J.-A. (2016). Conceptual knowledge of arithmetic for Chinese- and Canadian-educated adults. *Canadian Journal of Experimental Psychology/Revue Canadienne de*

Psychologie Expérimentale, *70*(4), 335–342.

Robinson, K. M., & Dube, A. (2017). Children's understanding of additive concepts. *Journal of Experimental Psychology*, *156*, 16–28. https://doi.org/https://doi.org/10.1016/j.jecp.2016.11.009

Robinson, K. M., & Dubé, A. K. (2009). Children's understanding of addition and subtraction concepts. *Journal of Experimental Child Psychology*, *103*(4), 532–545. https://doi.org/10.1016/j.jecp.2008.12.002

Robinson, K. M., & Dubé, A. K. (2012). Children's use of arithmetic shortcuts: The role of attitudes in strategy choice. *Child Development Research*, *2012*, 1–10. https://doi.org/10.1155/2012/459385

Robinson, K. M., & Dubé, A. K. (2013). Children's additive concepts: Promoting understanding and the role of inhibition. *Learning and Individual Differences*, *23*(1), 101–107. https://doi.org/10.1016/j.lindif.2012.07.016

Robinson, K. M., Dubé, A. K., & Beatch, J. A. (2016). Children's multiplication and division shortcuts: Increasing shortcut use depends on how the shortcuts are evaluated. *Learning and Individual Differences*, *49*, 297–304. https://doi.org/10.1016/j.lindif.2016.06.014

Robinson, K. M., & LeFevre, J. A. (2012). The inverse relation between multiplication and division: Concepts, procedures, and a cognitive framework. *Educational Studies in Mathematics*, *79*(3), 409–428. https://doi.org/10.1007/s10649-011-9330-5

Robinson, K. M., & Ninowski, J. E. (2003). Adults' understanding of inversion concepts: how does performance on addition and subtraction inversion problems compare to performance on multiplication and division inversion problems? *Canadian Journal of Experimental Psychology*, *57*(4), 321–330.

Robinson, K. M., Ninowski, J. E., & Gray, M. L. (2006). Children's understanding of the arithmetic concepts of inversion and associativity. *Journal of Experimental Child Psychology*, *94*(4), 349–362. https://doi.org/10.1016/j.jecp.2006.03.004

Robinson, K. M., Price, J. A. B., & Demyen, B. (2018). Understanding arithmetic concepts: Does operation matter? *Journal of Experimental Child Psychology*, *166*, 421–436. https://doi.org/10.1016/j.jecp.2017.09.003

Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: A multimethod approach. *Developmental Psychology*, *46*(1), 178–192. https://doi.org/10.1037/a0016701

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*(5), 405–410. https://doi.org/10.1111/1467-9280.00076

Siegler, R., & Araya, R. (2005). A computational model of conscious and unconscious strategy discovery. *Advances in Child Development and Behavior*, *33*, 1–42. https://doi.org/10.1016/S0065-2407(05)80003-5

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*(3), 258–275. https://doi.org/10.1037/0096-3445.117.3.258

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. Simon & G. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Hillsdale, NJ: Erlbaum.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries : A microgenetic analysis. *Journal of Experimental Psychology: General*, *127*(4), 377–397. https://doi.org/10.1037/0096-3445.127.4.377

Silver, E. A. (1986). Using conceptual and procedural knowledge A focus on relationships. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 181–198). Hillsdale, NJ: Erlbaum.

Sloutsky, V. M., & Yarlas, A. S. (2000). Problem representation in experts and novices: Part 2. Underlying processing mechanisms. In *Proceedings of the XXII Annual Conference of the Cognitive Science Society* (pp. 475–480). Mahwah, NJ: Erlbaum.

Stacey, K., & MacGregor, M. (1999). Learning the algebraic method of solving problems. *The Journal of Mathematical Behavior*, *18*(2), 149–167. https://doi.org/10.1016/S0732-3123(99)00026-7

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, *36*(5), 404–411. https://doi.org/10.2307/30034943

Starkey, P., & Gelman, R. (1989). The development of addition and subtraction abilities prior to formal schooling in arithmetic. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and Subtraction: A cognitive perspective* (pp. 99–116). Hillsdale, NJ: Laurence Erlbaum Associates.

Suresh, K. (2011). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of Human Reproductive Sciences*, *4*(1), 8–11. https://doi.org/10.4103/0974-1208.82352

Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic Bulletin and Review*, *9*(2), 314–318. https://doi.org/10.3758/BF03196287

Torbeyns, J., Peters, G., De Smedt, B., Ghesquiere, P., & Verschaffel, L. (2016). Understanding the complementarity principle. *British Journal of Educational Psychology*, *86*(3), 382–396.

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of*

Psychology of Education, *24*(3), 335–359. https://doi.org/10.1007/BF03174765

Warren, E. (2003). The role of arithmetic structure in the transition from arithmetic to algebra. *Mathematics Education Research Journal*, *15*(2), 122–137. https://doi.org/10.1007/BF03217374

Watchorn, R. P. D., Bisanz, J., Fast, L., LeFevre, J.-A., Skwarchuk, S.-L., & Smith-Chant, B. L. (2014). Development of mathematical knowledge in young children: attentional skill and the use of inversion. *Journal of Cognition and Development*, *15*(1), 161–180. https://doi.org/10.1080/15248372.2012.742899

Yarlas, A. S., & Sloutsky, V. M. (2000). Problem representation in experts and novices: Part 1. Differences in the content of representation. In *Proceedings of the XXII Annual Conference of the Cognitive Science Society* (pp. 1006–1011). Mahwah, NJ: Erlbaum.

Zakis, R., & Rouleau, A. (2017). Order of operations: On convention and met-before acronyms. *Educational Studies in Mathematics*, *97*(2), 143–162. https://doi.org/https://doi.org/10.1007/s10649-017-9789-9