Developing flexibility in solving arithmetic word problems:

comparing strategies through semantic analysis.

Calliste SCHEIBLING-SEVE^(a), Elena Pasquinelli^(b), Emmanuel SANDER^(a)

(a) University of Geneva, IDEA Lab, 40 Bd du Pont-d'Arve, 1205 Geneva
calliste.scheibling-seve@unige.ch ORCID Identifier: 0000-0001-6559-3845
emmanuel.sander@unige.ch ORCID Identifier: 0000-0003-0044-3437

(b) Foundation La main à la pâte, 43 rue de Rennes, 75006, Paris
elena.pasquinelli@fondation-lamap.org

Abstract:

This study develops and evaluates an intervention aimed at improving flexibility in arithmetic word problems. 9 classes (4th-5th Grades) were divided between experimental, active and passive control groups, and followed this protocol: pre-test, 5 learning sessions for experimental and active control groups, post-test. During learning sessions, both groups learned to solve arithmetic word problems with two strategies — expansion and factorization — by comparing them. But comparing strategies was based on a semantic analysis and choice in point of view for the experimental group whereas comparing strategies was based on identifying keywords and calculations. The experimental condition was expected to promote flexibility in strategy for new transfer problems. At posttest, the active control group and the experimental group got similar scores for learning problems, higher than the passive control group. But for transfer problems, the experimental group was more flexible than the two other groups. Educational entailments of our finding are discussed.

Keywords: flexibility, evidence-based education, word arithmetical problem, teaching strategies

Highlights :

- The goal is to develop flexibility in strategies to solve distributive problems.
- Experimental and active control groups learned by comparing multiple solutions.
- The active control group focused on relevant information in keywords and calculations.
- The experimental group focused on semantic relations and change in points of view.
- The experimental group became more flexible and less dependent on semantic context.

1. Introduction

In mathematics, proposing flexible and adaptive representations and strategies reflects higher problem solving skills (Heinze, Star & Verschaffel 2009; Newton, Lange, & Booth, 2019). One recommended pedagogical tool for developing flexibility in problem solving is comparing multiple solutions to a same

problem (Kilpatrick, Swafford, & Findell, 2001; National Mathematics Panel, 2008; Common Core Curriculum, 2013). Yet, the only comparison does not lead necessarily to conceptual understanding. Students often fail to identify the underlying structure: it appears that expliciting cues supports the comparison between underlying structure even though identification of the conceptual structure is still tough (Alfieri, Nokes-Malach, & Schunn, 2013; Gick & Holyoak, 1980, 1983; Gentner, Loewenstein, & Thompson, 2003; Ross, 1989; Schwartz & Bransford, 1998). Indeed students' prior knowledge can lead students to intuitively induce a superficial structure, triggering a misleading categorization of the situation (Chi, 2008). The present study aimed at improving flexibility in problems solving in order to develop an expert representation on problems that reflects a better mastering of the underlying mathematical notions. To compare strategies, students are encouraged to reelaborate their first representation through a semantic analysis.

1.1. Flexibility in problem solving

Cognitive flexibility refers to the ability to select adaptively among multiple representations of an object, one relevant perspective or strategy in order to adjust to the demands of a situation (Cragg & Chevalier, 2012; Diamond, 2013). Some authors distinguish flexibility from adaptivity that emphasis the ability to select the most appropriate strategy (Verschaffel, Luwel, Torbeyns & Van Dooren, 2009). In equation-solving problems, the role of procedural flexibility defined as “knowing more than one type of procedure for solving a particular type of problem and applying them adaptively to a range of situations” (Rittle-Johnson, 2017, p.1; Schneider, Rittle-Johnson, & Star, 2011; Star & Newton, 2009) has been studied. Flexibility can therefore be measured through the mastering of multiple strategies (Rittle-Johnson & Star, 2007; Star & Seifert, 2006), the ability to identify the most efficient strategy (Verschaffel, Luwel, Torbeyns, & Van Dooren, 2011) and the ability to identify a novel method related to a taught strategy (Begolli and Richland, 2016). Students with high flexibility in problem solving are more likely to adapt strategies when faced with unfamiliar transfer problems, testifying a better conceptual understanding (Hiebert & Wearne, 1996; Rittle-Johnson & Star, 2007). And gains in flexibility appear to be linked to improvements in conceptual and procedural knowledge (Rittle-Johnson & Star, 2009; Rittle-Johnson, Star, & Durkin, 2009; Star & Rittle-Johnson, 2009; Scheibling-Sève, Pasquinelli, & Sander, 2020). Flexibility is opposed to routine that could derive from the traditional « mechanistic approach » (Treffers, 1987) that leads to impose on students one strategy with little or no alternative solutions in elementary mathematics. Some authors go even further by talking about a *rigidifying effect* due to years of imposed and diligent practice and overgeneralization (Feltovich, Spiro & Coulson, 1997; Van Dooren, De Bock, Hessels, Janssens & Verschaffel, 2005). Therefore, flexibility is more a disposition than a skill and therefore teaching flexibility in strategies cannot be postponed after students master routine in some strategies.

Flexibility should be the first goal of teaching (Baroody, 2003; Gravemeijer, 1994; Selter, 1998; Warner et al., 2002; Wittmann & Müller, 1990, 1992; Verschaffel, Luwel, Torbeyns, Van Dooren, 2009). That is why, to promote cognitive flexibility, we propose to test a teaching method using word problems and based on comparing strategies and on recoding semantic relationships.

1.2. Word arithmetic problems to measure conceptual development

Most studies whose aim is to analyze or develop flexibility in mathematics strategies use linear equations (Newton & Star, 2013; Rittle-Johnson & Star, 2007, 2009; Star & Rittle-Johnson, 2009; Hickendorff, 2020). In this study, we propose to use arithmetic word problems that can be used as a tool to identify students' conception on a mathematic notion (Humberstone and Reeve, 2018; Scheibling-Sève, Pasquinelli & Sander, 2020). Phrasing of mathematical word problems can influence the induced representation of the problem by students (Vergnaud, 1982; Hudson, 1983; De Corte, Verschaffel & De Win, 1985). But in addition to linguistic features, semantic effects also rely on semantic relations or scenario depicted in the problem: when solving a word problem, students build a mathematical representation based on semantic relations inferred from real-world objects (Bassok & Olseth, 1995; Bassok, Chase & Martin, 1998). For instance, word problems describing a situation with symmetric pairs such as apples and oranges spontaneously encourages to add the two quantities whereas word problems describing a situation with asymmetric pairs such as apples and baskets induces to divide apples among basket (Experiment 2, Bassok et al., 1998). Therefore alignment between the semantic and mathematical relations can influence the difficulty of mathematical problems.

A possibility to study this effect of semantic context on the spontaneous encoding is to use problems solvable by two strategies. Indeed the semantic context influences the encoding of the problem and thus leads to a preference for one of the two strategies (Thevenot & Oakhill, 2005; Gamo, Sander, & Richard, 2010). Distributive problems such as *“The florist prepares for each of the 14 candidates a bouquet made up of 5 roses and 7 tulips. How many flowers does the florist use in total?”* offer this possibility. These problems admit two distinct strategies: an expansion ($k \times a + k \times b$) and a factorization ($k \times (a+b)$). Coquin-Viennot and Moreau (2003) showed that students' choice of the solving strategy is influenced by the presence of a word semantically grouping the addends in the wording of the problem: the mere presence of the word “bouquet” increased the rate of the factorization by 20% among 3rd to 5th grade students. Wording therefore influences the perception of the problem: either it favors the representation of each addend as separate part, or the representation of the addends as a whole. Indeed, a resolution with expansion amounts to following a step-by-step method by independently treating the different homogenous parts (tulips and roses). In this case, students failed to group the separate parts (tulips and

roses) to form a super-ordinate whole (a bouquet). Reversely, a factorization requires grouping parts (tulips and roses) through categorizing them as members of a super-ordinate whole (a bouquet), and then multiplying the resulting quantity by a factor (number of candidates). Similarly, a problem whose parts does not form a relevant whole - such as “For back-to-school time, Claire buys in a shop: pencils for 2€, pens for 7€, markers for 6€. She buys 4 of each. How much did Claire spend in this shop in all?” for which $2€ + 7€ + 6€$ does not correspond to a price of any entity - does not favor factorization (12% for 4th and 5th graders). By comparison, problem whose parts form a relevant whole - such as “For back-to-school time, Claire buys in a shop: 2 pencils, 7 pens, 6 markers. In this shop, every item costs 4€. How much did Claire spend in this shop in all?” for which $2 \text{ pencils} + 7 \text{ pens} + 6 \text{ markers} = \text{the purchased items}$ - leads to twice as many factorization (28%) (Sander, 2008; Scheibling-Sève, Pasquinelli & Sander, 2020). Semantic context promotes a one-sided perspective, constraining the strategies used. Therefore conceptual knowledge can be characterized as the ability to get over dependence to semantic context, and being able to solve one problem with two strategies. However, very few students (7%) are able to propose the two strategies and thus be flexible strategies for this type of distributive problems (Scheibling-Sève, Pasquinelli & Sander, 2020).

1.3. Comparing multiple strategies

A way to develop flexibility in problem solving is to ask students to compare multiple strategies to solve a same problem (Rittle-Johnson & Star, 2007, 2011; Alfieri, Nokes-Malach & Schunn, 2013). Comparing solutions to the same problem is more profitable than comparing solutions of different problems sharing the same mathematical structure (Rittle-Johnson & Star, 2007; Rittle-Johnson, Star, & Durkin, 2009; Rittle-Johnson & Star, 2011): students better identify deep structure of the problem rather than surface element. Indeed, studies on learning by analogy (Gentner, 1983; Hummel & Holyoak, 1997) explain that comparing a new solution to a familiar one allow students to identify similarities or differences between the known solution and the new one, and therefore their ability to use several strategies in algebra (Rittle-Johnson, Star & Rittle-Johnson, 2009) and with word problems (Begolli & Richland, 2016). But the way to present the analog strategies is important (Durkin, Star & Rittle-Johnson, 2017). Rittle-Johnson and Star (2007) have showed that students who saw the different strategies – to solve multistep linear equations - on the same page with a side-by-side presentation outperformed students who saw the different strategies on different pages. However comparing multiple strategies appears to have some limits: huge difficulties to transfer a new strategy learned in one situation to another analog situation remain (Gick et Holyoak, 1983; Novick, 1988; Hayes & Simon, 1977; Reed, Ernst, & Banerji, 1974; Klahr et Chen, 2011; Goldin-Meadow & Alibali, 2002; Rittle-Johnson, 2006; Siegler & Opfer, 2003). In mathematics, one limit of the comparing strategies’ method is not to take into account students’ preconceptions (Vosniadou &

Verschaffel, 2004) contrary to most studies in science (Brown & Clement, 1989; Brown, 2014; Chi, 2013; Chinn & Brewer, 1993; di Sessa, 2004; Vosniadou, 2013).

1.4. Semantic analysis and recoding

For a given problem solving by multiple strategies, comparing solutions is certainly a powerful teaching method. But a challenging prerequisite is to identify the preconceptions attached to the notion at stake (Richland, Begolli, Simms, Frausel & Lyons, 2017). Indeed previous knowledge interacts with the ability to identify deep structure (Holyoak & Thagard, 1989; Holyoak, 2002). For instance, in a classic multiplicative problem, the problem can be solved by a repeated addition ($3 + 3 + 3 + 3$) or by a multiplication (3×4). If a student fails to identify the similarities between the two strategies (repeated addition and multiplication), it is informative: the student's conception corresponds to the preconception of the multiplication, namely the repeated addition (Fischbein, 1985, 1989). In this example, a same situation – multiplicative problem - can be coded at different levels of abstraction – situation of repeated addition or situation of multiplication. The underlying assumption is that an expert can adopt the different encodings of the situation. Therefore training how to adopt multiple representations could be the principle of a learning method. Gamo et al. (2010) proposed such a training method in order to develop mathematical concepts through the semantic recoding of the word problem. The principle is to recode the semantically induced structure into a more apt mathematical structure. By recoding the problem, the student adopts a new point of view, which leads them to develop a representation of the problem that corresponds to the mathematical structure and succeed to use more expert solving strategies. In the present study, we proposed to focus on semantic analysis in order to recode a representation and transfer the adapted strategies to new contexts.

1.5. The present study: a training method based on semantic recoding

We tested a training method to improve flexibility in arithmetic problems. Contrary to previous work that mainly studied number calculations (Carpenter et al., 1997; Klein, Beishuizen, & Treffers, 1998, Rittle-Johnson & Star, 2007), the studied problems are word problems involving the distributive property. This type of problems - listed in the French curricula in 4th and 5th grade - has the methodological interest to be solvable by two strategies - expansion or factorization - whose prevalence depends on the semantic context involved in the problems (Moreau & Coquin-Viennot, 2003; Sander, 2008; Scheibling-Seve, Pasquinelli & Sander, 2020). The main goal of this method is to allow students to flexibly use the two strategies. Both active control and experimental groups had to compare solving strategies with a side-by-side presentation (Rittle-Johnson and Star, 2007). But in addition to comparing strategies, the experimental teaching method conducted a semantic analysis of the relationship between entities to code

them as parts or as whole. Vicente, Orrantia and Verschaffel (2007) pointed out the difficulty for students to develop conceptual relations between the entities of a problem. Thus, in their study, the success rate of problems whose rewording shed light on "part-whole" relationships was higher than problems with additional information about the problem's situation. The properties and relations of the entities or objects depicted in a problem are therefore key in the choice of strategies. In our study, the experimental method consisted in orienting students to establish these relations through an explicit labeling of "points of view". For instance, a training problem was the following: *For Christmas, Leo bought 5 boxes of chocolates: in each box, there are 4 black chocolates, 6 white chocolates and 7 milk chocolates. How many chocolates are there in total?* First, the semantic analysis consisted in identifying the parts (there are 3 parts: dark chocolates, white chocolate, milk chocolates) and then identifying the whole (chocolates in one box). The semantic relation between entities is: 4 dark chocolates + 6 white chocolates + 7 milk chocolates = 17 chocolates in one box. Second, students can adopt the point of view of the different parts: they calculate the total number of dark chocolates (5×4), the total number of white chocolates (5×6), the total number of milk chocolates (5×7) and then addition these sub-totals to find the total number of chocolate ($5 \times 4 + 5 \times 6 + 5 \times 7$). Adopting the point of view of the parts leads to an expansion. Third, students can adopt the point of view of the whole: they calculate the whole number of chocolates in one box ($4+6+7$), and then multiply the whole by the number of boxes ($5 \times (4+6+7)$). Adopting the point of view of the whole leads to do a factorization. Switching from one point of view to the other for a same problem allows adopting the two strategies. The two types of points of view were explicitly labelled with students (« the point of view of parts » and « the point of view of the whole »). Indeed using general labelling (Namy & Gentner, 2002) helps students identifying deep structure. Students were prompted to switch between the two conceivable points of view on the same situation and choose their preferred one. This choice of point of view by students is related to a reflexive level and is consistent with previous work (Siegler, 1999; Blöte, 2001) that stressed the flexible use of strategies and encouraged students to think about the value of different procedures for solving a given problem. The different steps of the experimental training are detailed in *Method, Training sessions*.

2. Method

2.1. Hypothesis

The experimental training method should favor students' cognitive flexibility on a mathematical concept - the distributive property - involved in arithmetic word problems. Students should be able to adopt the two points of view on the problem and use the two strategies (expansion and factorization) without depending on the semantic context.

First, we hypothesize experimental and active control groups improve more in using flexible strategies than the passive control group, since they receive specific training to solve the studied problems (Hypothesis 1). Then, for problems similar to training problems, no significant difference in strategy use between the experimental and active control groups should be observed, since they are both trained to solve this kind of problems (Hypothesis 2). Yet, the experimental method based on recategorization should favor transfer. Thus the experimental group should propose significantly more flexible strategies than the active control group at the post-test for the transfer problems (Hypothesis 3).

2.2. Participants

The experiment was conducted with nine classes from two elementary schools located in high-priority education network in Paris region during regular classwork school hours. Students who missed the intervention were omitted. In the end, 161 students took part in the study ($M=10$ years and 3 months, $SD=6$ months, 88 boys, 73 girls). In each school, two classes from the same grade (4th or 5th grade) were randomly attributed to the experimental or active control group. In one of the school, three classes (one 4th, one 5th and one 4th-5th grades) were also attributed to a passive control group. The experimental group included 57 students (37 5th graders and 20 4th graders). The active control group included 56 students (37 5th graders and 19 4th graders). The passive control group included 48 students (30 5th graders and 18 4th graders). Teachers were blind to the conditions of their class.

2.3. Design

A priori power analysis determined that 156 participants were required to detect a medium effect size of .25 with a power of .80 in F -test between the three group conditions (52 students per group). The experiment included three phases: pretest at week 1, weekly training sessions during 5 weeks, post-tests at week 7. Pre- and posttest were strictly identical. Both the experimental group and the active control group followed 5 training sessions taught by the experimenter in the presence of the teacher. Within experimental and active control group, training sessions were identical in their duration, organization and

Table 1

Problems at pre and post-tests

Type of problems	Learning problems	Near-transfer problems		
	Number-of-elements	Duration	Weight	Price
Context	X wants to prepare a treat cone per child. There are y children. Making a treat cone requires the following items:	X has made a list of purchase for y years. Each year, X's purchases are:	X wants to fill his/her pencil case with items that weight y grams each. In the pencil case, there are:	X is at the checkout of a supermarket. He bought some items. For each item, he took y in his basket:
Version 1	a Lollipops, b Candies, c Chocolates	a Printers, b Computers, c Scanners	a Pens, b Gums, c Markers	Buns for a€ each, Cakes for b€ each, Pies for c€ each,
Version 2	a Balloons, b Cookies, c Figurines	a Microscopes, b Desks, c Hamsters	a Shells, b Key-rings, c Candies	Ices for a€ each, Plants for b€ each, Plates for c€ each
Question	In all, how many objects does X need?	In all, how many purchases did X buy?	In all, how much does these items weight?	In all, how much did X spend?

problems statements. The passive control group received no intervention. They followed their usual daily mathematics sessions by their own teacher.

2.4. Material

2.4.1. Pre and post-tests

The material was composed by 8 isomorphic distributive word problems (Table 1) and 5 filler problems. Filler problems were proposed between distributive problems, in order to make the structural similarities between the different distributive problems less salient for the students.

Each distributive problem describes a situation involving one factor and three summands. The final question whose structure is “How much/many ... in all?” was placed at the end of the text. Two main solving strategies make it possible to reach the solution: Expansion strategy (sum of each part multiplied by the factor: $4 \times 6 + 4 \times 7 + 4 \times 8$) and Factorization strategy (sum of the parts, then multiplied by the factor: $(6+7+8) \times 4$). Four different variables – Number-of-elements, Duration, Price, Weight - were selected to create eight problems with two versions per variable (Table 1). During the learning sessions, only

The statements were constructed by controlling the familiarity of the vocabulary and the numerical values at stake. Numerical values had different features in common in order to limit numbers’ effects. Factor value was between 4 and 8 (5 was excluded since the associated multiplication table is easier). The three summands were between 2 and 8 (5 is excluded for the same reason). Their sum lied between 11 and 21. And the result was inferior to 100 (between 72 and 98) in order to control the level of calculation difficulty. Therefore, 8 booklets were constructed by controlling the numerical values, the order of presentation of the problems variable and problems versions. On each page of the booklet, the problem was presented in written form with two sections in order to propose two strategies. The instruction for solving the problem with two strategies was both orally given by the experimenter and written on each page: “Write all your calculations and the result in the following section:” and “Do you see any other method to come up with the same result? If yes, write it down while writing all calculations you performed to find the result.”. Tests lasted 45 minutes. Students were given 3 minutes to solve each isomorphic problems with Number-of-elements variable were used. So we propose to name the two problems with Number-of-elements variable present at pre and post-test as *learning problems* and problems with the other variables as *transfer problems*. Students were instructed that they could ask the experimenter or the teacher to read aloud the problem in order to bypass reading difficulties, and that they

had to write down all calculations. When the time was over, students had to turn the page and begin the new problem when the experimenter gave the instruction. They could not modify their answer once they turned the page.

Training sessions

	Active control Group	Experimental Group
Session 1	A problem: a question, useful data and operations *	A problem: way of seeing a situation, parts and whole
Session 2	Multiplication and commutativity by repeated addition*	Multiplication and commutativity by recategorization
Session 3	Distributivity: useful data, choice in operations	Distributivity: semantic relations and choice in point of view
Session 4	Dual strategies by equivalency of procedures	Dual strategies by change in point of view
Session 5	Distributivity problems: useful data, operations	Distributivity problems: choice in point of view

* Pedagogic materials were extracted from usual French textbooks (*Vive Les Maths* and *Companion Math*)
 Note. Students in the passive control group followed no intervention.

2.4.2. Training sessions

Training sessions took place in one weekly session over 5 weeks (45 minute session each week) for each class (Table 2). A pedagogical advisor of the French National Education was involved in the conception of all sessions, in order to assure ecological material and ecological pedagogical acts. For sessions 1 and 2, usual French textbooks (*Vive Les Maths* and *Companion Math*) inspired the pedagogical method used by the active control group: focus on the choice of operations in a problem and multiplication as repeated addition. For the first two sessions, the experimental group focus on a problem as a way of seeing a situation, had to adopt several points of view on a same non-mathematic situation and worked on multiplication as a ratio between two numbers by recategorization. For the last 3 sessions, the training problems were the same in both groups and involved only two types of variables: Number-of-Elements and Distance. Therefore, isomorphic problems involving Number-of-Elements with a different statement are considered as *learning problems* in post-test. In both groups, they had to compare strategies with a side-by-side presentation for each problem. The two approaches did only differ in their treatment of the problem (Table 3).

Regarding all other aspects, session organization was similar between the two groups: students began by exercises on the slate in order to engage them in the task. Then the students had to answer on an individual exercise sheet, whose support was also projected on the blackboard. Finally, students ended the session by answering in writing the question "What have I learned today? " and then a general conclusion was proposed by the experimenter and was written by the students.

Table 3

Pedagogic approaches for experimental group and active control group

<p>Example of one training problem in session 4: For christmas, Leo bought 5 boxes of chocolates: in each box, there are 4 black chocolates, 6 white chocolates and 7 milk chocolates. How many chocolates are there in total?</p>		
	Active control group	Experimental Group
Expansion strategy	<p>The experimenter explicits to students that they need to identify useful information - 5 boxes, the word « each », 4 black / 6 white / 7 milk chocolates:</p> <p>- one strategy is to multiply the chocolates by 5 and then add the partial sums: $5 \times 4 = 20$; $5 \times 6 = 30$; $5 \times 7 = 35$; $20 + 30 + 35 = 85$ chocolates in total.</p>	<p>The experimenter explicits to students that they can adopt two points of view:</p> <p>- point of view of parts : the parts are the types of chocolate. Then we calculate the total number for each part, for each types of chocolate $5 \times 4 = 20$ black chocolates $5 \times 6 = 30$ white chocolates $5 \times 7 = 35$ milk chocolates $20 + 30 + 35 = 85$ chocolats in total.</p>
	<p>Factorization strategy</p> <p>- one strategy is to sum the chocolates and then multiplied by 5. $4 + 6 + 7 = 17$; $17 \times 5 = 85$ chocolates in total.</p>	<p>- point of view of the whole: the whole is the number of chocolates in one box: $4 + 6 + 7 = 17$ chocolates in one box. Then we calculate the total number of chocolates: $17 \times 5 = 85$ chocolates in total.</p>
Conclusion with students:	<p>We can do two strategies, that are similar. One requires 2 operations and the other 4 operations.</p>	<p>We can adopt two points of view to solve the problem: point of view of the parts or point of view of the whole; point of view of the chocolate's types or point of view of the box.</p>

2.5. Coding and scoring

For pre and post-test, we analysed students ability to propose factorization, expansion or dual-strategy for learning problems - whose the variable numbers-of-elements have been used during learning sessions, problem - and for transfer problems – whose variables have not been taught -. At pre and post-test, a global score was built: each problem per student was coded as 1 when it was solved by an expansion, by 2 when it was solved by a factorization, by 3 when it was solved by dual-strategies and 0 if otherwise. Each student got therefore two scores: one for the learning problems and one for the near-transfer problems.

Students' scores by condition

	PreTest		PostTest		Progress	
	Mean	SD	Mean	SD	Mean	SD
Experimental group						
Learning problems	0.44	0.38	0.74	0.29	0.31	0.42
Transfer problems	0.33	0.20	0.71	0.27	0.42	0.23
Active Control Group						
Learning problems	0.37	0.28	0.68	0.31	0.32	0.41
Transfer problems	0.31	0.14	0.57	0.30	0.28	0.26
Passive Control Group						
Learning problems	0.39	0.27	0.34	0.29	-0.05	0.26
Transfer problems	0.32	0.19	0.37	0.23	0.08	0.19

Note. Students in the passive control group did not study neither learning nor transfer problems.

3. Results

3.1. PreTest

Some preliminary analyses were conducted to test whether the three groups were comparable at the pretest moment. Three-way ANOVAs were conducted first to establish that there were no differences between the students' scores from the three groups ($F(2,161) = 0.67$, $p = .51$ for learning problems and $F(2,161) = 0.03$, $p = .97$ for transfer problems), the two grades ($F(1,161) = 0.16$, $p = .69$ for learning problems and $F(1,161) = 0.09$, $p = .77$ for transfer problems) and for the 9 teachers ($F(8,171) = 1.72$, $p = .12$, for learning problems and $F(8,161) = 1.66$, $p = .14$ for transfer problems). So, there were no difference between grades, teachers and groups. The score averages for learning and transfer problems are summarized in Table 4 by condition at pre and post-test.

Table 5Students mean (*SD*) per type of strategy per condition

	Learning Problems		Transfer problems	
	<i>Pre-test</i>	<i>Post-Test</i>	<i>Pre-test</i>	<i>Post-Test</i>
Expansion Strategy				
Experimental Group	0.55 (0.47)	0.74 (0.42)	0.62 (0.34)	0.78 (0.34)
Active Control Group	0.51(0.45)	0.71 (0.43)	0.62 (0.33)	0.72 (0.35)
Passive Control Group	0.51 (0.44)	0.57 (0.45)	0.64 (0.31)	0.64 (0.31)
Factorization Strategy				
Experimental Group	0.39 (0.47)	0.75 (0.40)	0.18 (0.27)	0.67 (0.36)
Active Control Group	0.29 (0.41)	0.65 (0.46)	0.15 (0.23)	0.49 (0.39)
Passive Control Group	0.33 (0.40)	0.21 (0.37)	0.16 (0.29)	0.23 (0.31)
Dual Strategy				
Experimental Group	0.19 (0.39)	0.52 (0.46)	0.08 (0.19)	0.52 (0.42)
Active Control Group	0.045 (0.17)	0.45 (0.48)	0.04 (0.10)	0.38 (0.41)
Passive Control Group	0.08 (0.26)	0.09 (0.27)	0.06 (0.21)	0.11 (0.25)

Note. Solving a problem by a factorization and an expansion was coded as correct in the three categories (expansion, factorization, dual strategy).

Kruskall-Wallis tests were conducted to establish there were no differences in rate of expansion, factorization and dual strategies between students. For learning problems, there were no effect of condition for expansion rate ($\chi^2(2) = 0.36, p > .05$), factorization rate ($\chi^2(2) = 0.92, p > .05$) and dual rate ($\chi^2(2) = 4.78, p > .05$). For transfer problems, there were also no effect of condition for expansion rate ($\chi^2(2) = 0.02, p > .05$), factorization rate ($\chi^2(2) = 0.63, p > .05$) and dual rate ($\chi^2(2) = 1.77, p > .05$). The mean students per type of strategies for learning and near-transfer problems are summarized in Table 5 by condition at pre and post-test.

3.1. PostTest

3.1.1. Score

We conducted separate ANOVAs with condition as a between-subjects factor. Levene's test of variance homogeneity was used to ensure that all measures were appropriate for use of the ANOVA statistic.

Learning problems: There was a main effect of condition with medium effect size ($F(2,161) = 26.41, p < .001, \eta^2 = 0.14$). Pairwise comparisons with Bonferroni correction reveal that students in the experimental group and in the control group scored significantly higher than students in the passive control group (Table 6). As predicted, there were no differences between students in the experimental and in the active control groups for training problems. This absence of difference between the two groups for the trained problems reflects the similarity in terms of learning method between the two groups: they both learned to use dual-strategies (expansion and factorization) for the distributive problems.

	Learning Problems		Near-transfer problems	
	<i>p</i> -values	<i>Cohen's d</i>	<i>p</i> -values	<i>Cohen's d</i>
Post-Test				
Experimental vs Active Control Groups	.3	0.19	.006	0.50
Experimental vs Passive Control Groups	< .001	1.47	< .001	1.38
Active vs Passive Control Groups	< .001	1.20	< .001	0.76

Note. The numbers in bold are *p*-values < .05.

The control group focused on relevant information in statements, digits and operations whereas the experimental group focused on semantic relations, words and point of view.

Transfer problems: There was a main effect of condition with medium effect size ($F(2,161) = 21.22$, $p < .001$, $\eta^2 = 0.12$). Groups' condition is a significant predictor for scores. Pairwise comparisons with Bonferroni correction reveal that students in the experimental and active control groups outperformed students in the passive control group. And students in the experimental group scored significantly higher than students in the active control group for transfer problems (Table 6 and Figure 1).

3.1.2. Dual-strategies and factorization

Learning problems: There was a main effect of condition ($\chi^2(2) = 25.427$, $p < .001$; $\chi^2(2) = 30.17$, $p < .001$, respectively). Mann–Whitney–Wilcoxon pairwise comparisons with a Holm correction show that students in the experimental group and in the active control group used more dual-strategies and factorization at post-test than students in the passive control group. No significative difference between experimental and active control group were found (Table 7).

Table 7

Mann-Whitney-Wilcoxon pairwise comparisons with Holm correction by condition

	Learning Problems		Transfer problems	
	<i>p</i> -values	<i>Cohen's d</i>	<i>p</i> -values	<i>Cohen's d</i>
Dual Strategy				
Experimental vs Active Control Groups	.042	0.15	.048	0.35
Experimental vs Passive Control Groups	< .001	1.12	< .001	1.17
Active vs Passive Control Groups	< .001	0.90	.002	0.78
Factorization Strategy				
Experimental vs Active Control Groups	0.23	0.24	.017	0.47
Experimental vs Passive Control Groups	< .001	1.41	< .001	1.31
Active vs Passive Control Groups	< .001	1.07	.002	0.74

Note. Numbers in bold are *p*-values < .05.

Transfer problems: When analysing dual-strategies and factorization, there was a main effect of condition ($\chi^2(2) = 25.53$, $p < .001$, $\chi^2(2) = 38.622$, $p < .001$, respectively). Mann–Whitney–Wilcoxon pairwise comparisons with a Holm correction show that students in the experimental group and in the active control group used more dual-strategies and factorization at post-test than students in the passive control group. And students in the experimental group used significantly more dual-strategies and factorization than active control group (Table 7).

Figure 1
Score at post-test for transfer problems by condition

4. Discussion

This intervention supported our hypothesis that strategy comparison through semantic analysis favors students' flexibility in strategy. As predicted, the passive control group did not adopt flexible points of view for any problems. The simple fact to redo the same problems at pre and post-tests does not lead to learning. Compared to the passive control group, the two intervention groups perform significantly better for learning and near-transfer problems. However, strategy comparison with the usual practice to identify the relevant information only allows students in the active control group to reapply taught strategies for problems similar to those seen during the training lessons. They were not able to transfer the flexible point of view to new problems involving other variables. Compared to the active control group,

the experimental method based on both strategy comparison and the explicit analysis of semantic relations showed its success in promoting transfer: it was more successful in transferring factorization strategy and dual strategies to new problems in the post-test. The use of isomorphic problems made it possible to identify precisely the robustness of the transfer effects from the learning method. Indeed as we studied the transfer problems, the greater progression for the experimental group shows that the experimental method was not superficial. This transfer reflects a semantic change by students that could adopt a double point of view on the problem: it became less dependent on the choice of the variable than the active control group. Experimental method succeeded in making the students move from a spontaneous representation influenced by problem's variables to flexible representation based on the mathematical structure. In sum, these results provide insights into the role of semantic analysis in strategy comparison. Focus on establishing the semantic relation between entities of the word problem and labelling the two possible points of view can help the identification of conceptual structure, and therefore adopt flexibility in strategy.

5. Conclusion

The results from this study have the potential to contribute to cognitive flexibility literatures and evidence-based education field as well as to inform teaching practices. The influence of preconceptions in conceptual change has been investigated in science education (Chi, 2008; Chinn & Brewer, 1993; di Sessa, 2004; Vosniadou, 2013) but its role in transferring mathematic strategies remains less studies (Vamvakoussi & Vosniadou, 2012; Begolli & Richland, 2016). This study has also a methodological interest in terms of evidence-based education. It extends previous results on learning with comparison strategy from laboratory to more ecologic school context. Even though we designed the study with active and passive control groups to control this bias, replication of the experimental method teaching by the teachers and not by an experimenter should be conducted. However, these first results are already informative for teaching word problems at school. The students from the experimental group became less dependent on semantic context. Their choice of strategy was less constrained by the nature of the variable. The substantial transfer of the flexibility in strategy illustrates the ability to adopt a new point of view on the situation. Thus students were able to change their encoding based on spontaneous representations to an encoding based on conceptual relations. We interpret this result to support the importance of conceptual rewording (Vicente, Orrantia & Verschaffel, 2007; Gvozdic & Sander, 2020). Students in the experimental group were taught to establish semantic relationship by their own, moving their attention from calculations to reasoning (Savard & Polotskaia, 2017). In addition to improve semantic analysis, students were encouraged to adopt a reflexive attitude thanks to the notion of point of view. Thus students

developed their flexibility: developing flexible strategies with the ability to transfer them to new isomorphic problems. Yet studying the extent of this transfer could be the goal of further research. The teaching method appears to be a useful framework to identify if cognitive flexibility is domain-general or domain-specific. Indeed, fostering cognitive flexibility takes part of a broader goal, namely promoting conceptual development.

Acknowledgments

This research was supported in part by a National Research Agency Grant ANR-06-APPR-015. We thank the pedagogical advisor Arbya Eichi and all the teachers who participated in the study.

References

1. Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, 48(2), 87-113.
2. Bassok, M., & Olseth, K. L. (1995). Object-based representations: Transfer between cases of continuous and discrete models of change. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6)
3. Bassok, M., Chase, V. M., & Martin S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive psychology*, vol. 35, no2, 99-134.
4. Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody & A. Dowker (Eds.), *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise Studies*. Mahwah, N.J. : Lawrence Erlbaum Associates, Inc.
5. Begolli, K. N., & Richland, L. E. (2016). Teaching mathematics by comparison: Analog visibility as a double-edged sword. *Journal of Educational Psychology*, 108(2), 194-213.
6. Blöte, A., Van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems. *Journal of Educational Psychology*, 93(3), 627.
7. Brown, D. E., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: Abstract transfer versus explanatory model construction. *Instructional science*, 18(4), 237-261.
8. Chi, M. T. (2013). Two kinds and four sub-types of misconceived knowledge, ways to change it, and the learning outcomes. In *International handbook of research on conceptual change* (pp. 61-82). Routledge.
9. Chi, M. T. (2008). Three types of conceptual change: *International handbook of research on conceptual change*, 61-82.
10. Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of educational research*, 63(1), 1-49.
11. Coquin-Viennot, D., & Moreau, S. (2003). Highlighting the role of the episodic situation model in the solving of arithmetical problems. *European Journal of Psychology of Education*, 18(3), 267-279.
12. Cragg, L., & Chevalier, N. (2012). The processes underlying flexibility in childhood. *The Quarterly Journal of Experimental Psychology*, 65(2), 209-232.
13. DeCorte, E. Verschaffel, L., & De Win, L. (1985) Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology*, 77. 460-470
14. Diamond, A. (2013). Executive functions. *Annual review of psychology*, 64, 135-168.
15. disessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive science*, 28(6), 843-900.
16. Durkin, K., Star, J. R., & Rittle-Johnson, B. (2017). Using comparison of multiple strategies in the mathematics classroom: Lessons learned and next steps. *ZDM*, 49(4), 585-597.
17. Feltovich, P.J., Spiro, R.J., & Coulson, R.L. (1997). Issues of expert flexibility in contexts characterized by complexity and change. In P.J. Feltovich, K.M. Ford, & R.R. Hoffman (Eds.), *Expertise in context: Human and machine* (pp. 125-146). Menio Park, CA: AAAI Press.
18. Fischbein, E. (1989). Tacit models and mathematical reasoning, *For the Learning of Mathematics*, 9, 9-14
19. Fischbein, E., Deri, M., Nello, M. S., & Marino, M. S. (1985). The role of implicit models in solving verbal problems in multiplication and division. *Journal for Research in Mathematics Education*, 16, 3-17

20. Gamo, S., Sander, E., & Richard, J-F. (2010). Transfer of strategies by semantic recoding in arithmetic problem solving. *Learning and Instruction*, 20, 400-410.
21. Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, 95(2), 393-408.
22. Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2), 155-170.
23. Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306-355.
24. Goldin-Meadow, S., & Alibali, M. W. (2002). Looking at the hands through time: A microgenetic perspective on learning and instruction. *Microdevelopment: Transition processes in development and learning*, 80-105.
25. Gravemeijer, K. (1994). Educational development and developmental research in mathematics education. *Journal for research in Mathematics Education*, 443-471.
26. Gvozdic, K., & Sander, E. (2020). Learning to be an opportunistic word problem solver: going beyond informal solving strategies. *ZDM*, 1-13.
27. Hinsley, D. A., Hayes, J. R., & Simon, H. A. (1977). From words to equations: Meaning and representation in algebra word problems. *Cognitive processes in comprehension*, 329.
28. Heinze, A., Star, J. R., & Verschaffel, L. (2009). Flexible and adaptive use of strategies and representations in mathematics education.
29. Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and instruction*, 14(3), 251-283.
30. Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 234-259). New York, NY: Oxford University Press.
31. Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13, 295-355.
32. Humberstone, J., & Reeve, R. A. (2018). The conceptual overlap between arithmetic and algebraic referential mapping. *Learning and Instruction*, 54, 138-146.
33. Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological review*, 104(3), 427.
34. Klahr, D., & Chen, Z. (2011). Finding one's place in transfer space. *Child Development Perspectives*, 5(3), 196-204.
35. Klein, A. S., Beishuizen, M., & Treffers, A. (1998). The empty number line in Dutch second grades: Realistic versus gradual program design. *Journal for Research in Mathematics Education*, 443-464.
36. Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics* (Vol. 2101). National research council (Ed.). Washington, DC: National Academy Press.
37. Lautrey, J., Rémi-Giraud, S., Sander, E., & Tiberghien, A. (2008). *Les connaissances naïves*. Armand Colin.
Changer Sander 2008
38. Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, 131(1), 5-15.
39. National Mathematics Panel, 2008
40. Newton, K. J., & Star, J. R. (2013). Exploring the nature and impact of model teaching with worked example pairs. *Mathematics Teacher Educator*, 2(1), 86-102.
41. Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, memory, and cognition*, 14(3), 510-520.
42. Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive psychology*, 6(3), 436-450.
43. Richland, L. E., Begolli, K. N., Simms, N., Frausel, R. R., & Lyons, E. A. (2017). Supporting mathematical discussions: the roles of comparison and cognitive load. *Educational Psychology Review*, 29(1), 41-53.
44. Rittle-Johnson, B. (2017). Developing mathematics knowledge. *Child Development Perspectives*, 11(3), 184-190.
45. Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, 101(4), 836.
46. 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.

47. Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 629-639.
48. Scheibling-Sève, C., Pasquinelli, E., Sander, E. (2020) Assessing conceptual knowledge through solving arithmetic word problems. *Educational Studies in Mathematics*, 103, 293-311.
49. Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental psychology*, 47(6), 1525-1538.
50. Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and instruction*, 16(4), 475-522.
51. Selter, C. (1998). Building on children's mathematics-A teaching experiment in grade three. *Educational Studies in Mathematics*, 36(1), 1-27.
52. Siegler, R. S. (1999). Strategic development. *Trends in Cognitive Sciences*, 3(11), 430-435.
53. Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological science*, 14(3), 237-250.
54. Star, J. R., & Newton, K. J. (2009). The nature and development of experts' strategy flexibility for solving equations. *ZDM*, 41(5), 557-567.
55. Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology*, 31(3), 280-300.
56. Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, 18(6), 565-579.
57. Treffers, A. (1987). Integrated column arithmetic according to progressive schematisation. *Educational studies in Mathematics*, 18(2), 125-145.
58. Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representations in arithmetic word problem solving. *The Quarterly Journal of Experimental Psychology Section A*, 58(7), 1311-1323.
59. Vamvakoussi, X., & Vosniadou, S. (2012). Bridging the gap between the dense and the discrete: The number line and the "rubber line" bridging analogy. *Mathematical Thinking and Learning*, 14(4), 265-284.
60. Van Dooren, W., De Bock, D., Hessels, A., Janssens, D., & Verschaffel, L. (2005). Not everything is proportional: Effects of age and problem type on propensities for overgeneralization. *Cognition and Instruction*, 23(1), 57-86.
61. Vergnaud, G. (1982). A classification of cognitive tasks and operations of 1195 thought involved in addition and subtraction problems. In T. P. Carpenter, 1196 J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A 1197 cognitive perspective* (pp. 39e59). Hillsdale, NJ: Erlbaum.
62. Verschaffel, Luwel, Torbeyns & Van Dooren, 2009
63. Vicente, S., Orrantia, J., and Verschaffel, L., Influence of situational and conceptual rewording on word problem solving. *British journal of educational psychology* 77.4 (2007): 829-848.
64. Vosniadou, S., & Verschaffel, L. (2004). Extending the conceptual change approach to mathematics learning and teaching. *Learning and Instruction*, 14(5), 445-451.
65. Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator*, 8-19.