



Should we continue to teach standard written algorithms for the arithmetical operations? The example of subtraction

Jean-Paul Fischer, Bruno Vilette, Sophie Joffredo-Lebrun, Mireille Morellato, Céline Le Normand, Calliste Scheibling-Seve, Jean-François Richard

► To cite this version:

Jean-Paul Fischer, Bruno Vilette, Sophie Joffredo-Lebrun, Mireille Morellato, Céline Le Normand, et al.. Should we continue to teach standard written algorithms for the arithmetical operations? The example of subtraction. Educational Studies in Mathematics, 2019, 101 (1), pp.105-121. 10.1007/s10649-019-09884-9 . hal-02097469

HAL Id: hal-02097469

<https://hal.univ-lorraine.fr/hal-02097469>

Submitted on 12 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The final version of this paper was published as:

Fischer, J.-P., Vilette, B., Joffredo-Lebrun, S., Morellato, M., Le Normand, C., Scheibling-Seve, C., & Richard, J.-F. (2019). Should we continue to teach standard written algorithms for the arithmetical operations? The example of subtraction. *Educational Studies in Mathematics*, 101(1), 105-121. DOI: 10.1007/s10649-019-09884-9

Should we continue to teach standard written algorithms for the arithmetical operations? The example of subtraction

Jean-Paul FISCHER*, Bruno VILETTE^(a), Sophie JOFFREDO-LEBRUN^(b), Mireille MORELLATO^(c), Céline LE NORMAND^(d), Calliste SCHEIBLING-SEVE^(e) & Jean-François RICHARD^(f)

* Corresponding author, Professor Emeritus of Developmental Psychology,
University of Lorraine, 2LPN Laboratory, 23 Bd Albert Ier, BP 3397 54015 NANCY Cedex - FRANCE
E-mail: jean-paul.fischer@univ-lorraine.fr;

(a) Professor, Developmental Psychology, University of Lille, PSITEC Laboratory
bruno.vilette@univ-lille.fr

(b) ATER, University Nantes, Email: s.joffredolebrun@gmail.com

(c) Pedagogical Advisor, Marseille, E-mail: mireille@syrah.fr

(d) Pedagogical Advisor, Toulon, E-mail: Celine.Le-Normand@ac-nice.fr

(e) PhD Student, Department of Psychology, University Paris 8, Paragraphe Laboratory
Email: calliste.scheibling.seve@gmail.com

(f) Professor Emeritus of Psychology, University Paris 8, Paragraphe Laboratory
E-mail: jf.richard.paris8@gmail.com

Running head: Arithmetic Standard Written Algorithms

Funding: This research was supported by a grant from France's "Programme d'Investissements d'Avenir" (ANRU/DSDP/D14-2780).

Acknowledgments: We would like to thank all the children and teachers who took part in the experiment. We are also grateful to the experimenters, coders, and administrators for their invaluable help in organizing and conducting the experiment.

Abstract. The question of teaching a standard written algorithm for the four arithmetical operations in the first years of elementary school is increasingly raised by the possibility of using calculators, tablets and computers. This article first presents arguments both against and for maintaining this teaching in the first school-years. Then, it presents an analysis of the procedure—whether it implies the standard algorithm or not—used by 2nd grade students to solve a subtraction verbal problem. The 4720 students involved in the research were divided in a control group ($n = 2101$) and an experimental group ($n = 2619$) in whose the standard written algorithm (SWA) was taught, or not taught, respectively. The effectiveness of algorithm use was analyzed not only in terms of accuracy of the answers, but also for its influence on the choice of the correct arithmetical operation and for its choice by the students as a function of their ability. The SWA was strongly associated with the use of the false arithmetical operation (effect size $\phi = 0.79$)—addition instead of subtraction—but was chosen by students of all levels of ability. The latter result suggests that the SWA is well suited for the most advanced students but far less for the others who use it mechanically and without reflection.

Keywords: arithmetic teaching, standard algorithm, mental calculation, addition, subtraction, gender difference

1 Introduction

Each proposal of a new curriculum for the primary school raises the question of the need for continuing to teach a standard written algorithm (SWA) for the four classical arithmetical operations: addition, subtraction, multiplication, and division. The question is of particular concern in countries where this teaching is fundamental by its presence in the early grades, as in France where the SWA for addition and subtraction/multiplication are taught in first grade and second grade, respectively.

The changes in society, notably the numerical revolution, challenge the necessity of teaching an exact paper-pencil algorithm for computing each arithmetical operation. The numerous arguments that can be put forward against the teaching of the SWAs are here presented in eight categories, even if these categories can support some interactions (section 1.1). However, there are also some arguments that come in defense of the teaching of these ancestral algorithms (section 1.2).

1.1 Arguments against teaching SWAs

1.1.1 The SWAs harm the number sense

The SWAs, except that for division, impose beginning the computation with the smallest units. Therefore the magnitude of the numbers to add, subtract or multiply is masked. The persons engaged in the computation ignore the place value of the digit they manipulate. The consequences of such a practice are well summarized by Kamii and Dominick's (1997 p. 51) claim "that algorithms 'unteach' place value and hinder children's development of number sense".

1.1.2 The use of SWAs may lead to absurd results in arithmetical operations

Selter (2001 p. 157) reproduced the written subtraction SWA for $701 - 698$ from many third or fourth graders who found a result higher than 701 (e.g., $701 - 698 = 703$; $701 - 698 = 1903$). In the present research, many second graders found $72 + 37 = 19$. Whereas the latter result appears absurd, it may result rather logically from the columnar algorithm: In adding the tens, the students have $7 + 3 = 10$ and write the 10 as 1 in the column of the tens.

1.1.3 The SWAs run against the direction of writing

Beginning with the smallest units implies beginning on the right in our culture. In first grade, this seems not well come because teachers emphasize that we read and write from left to right. The reversal of the digits, omnipresent in Kindergarten but not totally fade-out in first grade (Fischer & Koch, 2016), shows that many students are in struggle when the left-right orientation is not respected.

1.1.4 The SWAs negatively influence mental calculation

The premature learning of a SWA suggests and allows a "paper-and-pencil mental analogue" strategy in mental computation. The French national annual assessments show that girls perform better in the written algorithm than boys, but that boys outperform girls in mental computation (Chesné & Fischer, 2015; Fischer, 2004). This twofold observation is consistent with a greater use of a paper-and-pencil mental analogue strategy in mental calculation by girls than by boys. Because of their low performance in the written algorithm boys could be less inclined, or even unable to repeat it mentally, whereas the contrary holds for girls. Insofar as this mental analogue strategy repeats the SWA, the observation is also consistent with the findings that girls apply more frequently standardized strategies, while boys prefer various approaches (Torbeyns, Hickendorff, & Verschaffel, 2017).

This interpretation predicts that boys will specifically perform better than girls in mental operations needing carrying, borrowing, bundling or unbundling because these complex transformations concern mainly the written algorithm. Indeed, Chesné and Fischer (2015, Appendix E) report many assessments, at the beginning of third grade and with a free choice of the method of computation, in which the $978 - 765$ subtraction problem was better solved by girls than boys, whereas the same girls performed lower than the boys on the $45 - 27$ subtraction problem.

1.1.5 The teaching of SWAs creates a bad didactical contract in problem solving

When teachers attach importance to unimportant things, when they convey an erroneous conception of the functioning of mathematics, they implicitly establish a bad didactical contract (Brousseau & Warfield, 2014). This is often the case in problem solving when the professors ask to present the response of a problem in the “Solution – Operation” format. This format emphasizes the computation of the arithmetical operation, which must be written to be presentable to the teacher, and thereby hides the heart of the problem solving activity, namely finding the operation that must be computed.

1.1.6 The SWAs are not always understandable by typical students

The subtraction SWA—let us write $a_2a_1a_0 - b_2b_1b_0$ the subtraction of two three-digit numbers—generally taught in second grade in France, needs borrowing or unbundling a higher order $(i+1)$ -unit when b_i cannot be taken away from a_i . In the latter case the higher order $(i+1)$ -unit is transformed in ten i -units through a sort of writing. Can a typically developing 7-year-old child understand this ingenious sort of writing? Even if the answer would be “yes”, it remains that the 1 corresponding to the $(i+1)$ -unit is misplaced in the i -unit column. This may be particularly troubling for students because teachers repeat at length that when an operation is put down in columns, the i -unit must imperatively be written in the column of the i -units. Other “harmful effects of carrying and borrowing in grades 1-4” can be found in Kamii and Dominick (2009).

Furthermore, the SWA for subtraction directly confronts the students with impossible $a - b$ computations, if $a < b$. Both by facility and because the difference between a and b is the same as that between b and a in the vernacular language, students are then rushed into the error of computing $b - a$ rather than $a - b$. This is, for example, supported by Csikos’ (2016) finding that 193 is the most frequent erroneous answer in the computation of $701 - 694$ (15.4%).

1.1.7 The learning of SWAs implies learning through authority

Because of the difficulty of understanding and the impossibility of reinventing the SWAs, the latter must be taught with authority. “If the classroom is governed by the authority of the teacher, children will be encouraged to conform to the teacher’s wishes, without separating sociomoral issues from purely intellectual ones”, noted Kamii (1994 p. 63).

1.1.8 The use of the standard algorithms in adulthood borders on ridicule in many cases

The premature initial learning of the SWA by the children (as young as 6 years of age for addition) leads to deeply buried memories. Not surprisingly, Dewi, Castel, Kerzel, Posada and Thevenot (2015) found that adult students prefer the columnar strategy taught at school rather than more flexible mental strategies. Such behavior in adult persons borders on ridicule when problem solving is obvious through mental computation, as is the case for many problems in the research by Dewi et al. (e.g., $57 + 99$, $25 + 25$).

1.2 Arguments in defense of teaching SWAs

1.2.1 Students use SWAs more effectively and faster than mental computation

Torbeyns and Verschaffel (2013, 2016) submitted two types of additive problems to 4th grade students in Belgium (Flanders). The MC items were easy to solve with mental computation (e.g., $498 + 263$, $963 - 499$), whereas the SA (standard algorithm) items were not (e.g., $456 + 266$, $952 - 474$). Thus, the SA items evoked computation with the SWA more strongly than the MC items. Not surprisingly, the SA items were better solved through using the SWA than through mental computation (88% vs. 63% correct for the subtractions tested in Torbeyns & Verschaffel, 2016); the first strategy was also faster than the second. However, very surprisingly, the MC items also were better solved through using the SWA than through mental computation (80% vs. 60% correct for the subtractions tested in Torbeyns & Verschaffel, 2016); the first strategy was faster than the second for the MC items also.

Dewi et al. (2015) not only found that adult students prefer the written algorithm of the addition of two two-digit numbers (see section 1.1.8), but also that this strategy was run faster than other strategies, without detrimental effect on accuracy. However, because there was a majority of female students in the sample by Dewi et al. (42 out of 58 participants), the results reported in section 1.1.4 may partially explain the strategy choice of Dewi et al.'s participants.

1.2.2 Students prefer SWA to mental computation

With the same sample of students as in section 1.2.1, Torbeyns and Verschaffel (2016) also used a choice condition in which students could choose between mental computation and SWA as a function of the item. The students clearly preferred the SWA (79% choices), not only for the SA items (82% choices) as expected, but also for the MC items (77% choices). These choices by the students can be explained with the cognitive and conative attractiveness of the SWA.

1.2.3 SWAs raise cognitive attractiveness

Knowing an algorithm that can be applied to all multi-digit numbers has the advantage of drastically limiting the choice of the students. If they do not retrieve the result of a computation in their declarative memory they immediately know what to do. In analyzing this phenomenon further, one sees that in computing the result with the SWA students identically repeat the same procedure. They have nothing new to learn. On the contrary, in genuine new mathematical learning the children must implement a reflective abstraction (Piaget, 1977). Such new learning unbalances the children's cognitive structures and needs efforts to restore the equilibrium. It is understandable that students—and not only students!—avoid such efforts.

1.2.4 SWAs raise conative attractiveness

A unique mean to calculate the result from a problem is not only cognitively economic but also affectively reassuring. This reassuring character of a unique algorithm or representation is a very general phenomenon. It was experimentally demonstrated with an algorithmic instruction for solving percentage problems on seventh grade students in Midwestern USA (Flores, Koontz, Inan & Alagic, 2015). The order of teaching this algorithmic instruction with respect to a multiple representation instruction was manipulated. Although the results indicated a very small superiority in abilities to solve percentage problems among the students who received multiple representation first, the students' perception of the algorithmic instruction was more favorable compared to that of the multiple representation instruction. Furthermore, the sole teacher involved in this study indicated that she also experienced difficulty with motivating the multiple representation instruction in the group of students who

had received the algorithmic instruction first. This may explain why the premature learning of the SWA hinders the learning of genuine mental computation.

Math anxiety—a feeling of tension, apprehension or fear that interferes with math performance—may be a conative consequence of mathematics learning. Ramirez, Chang, Maloney, Levine and Beilock's (2016) work confirms that math anxiety in first and second graders is negatively related to their use of more advanced strategy in the computations $3 + 6$, $9 + 8$, $13 + 5$, and $14 + 19$. Furthermore, and somewhat ironically, students who have the highest working memory capacity avoid using advanced problem-solving strategies when they are high in math anxiety. This can partially explain why students who spontaneously use the SWA perform better than students who use more advanced strategies, in the current observation as in others (e.g., Torbeyns & Verschaffel, 2013, 2016).

1.3 Goal of the study

A “battle” of arguments alone cannot solve the question of whether to teach arithmetical algorithms in the first school-years. The lighting of experimental work would be welcome. However, given the complexity and versatility of the question, an experimental approach can only investigate some specific aspects. The particular importance of finding the arithmetical operation to be performed led us to study arithmetical problem solving. The consequences of teaching or not teaching SWA on the solution procedures used by young students can then provide information to be included in an answer to our general research question about the effectiveness of SWA's teaching.

2 Method

2.1 The ACE program

The ACE (Arithmetic Comprehension at Elementary school) research program is a full arithmetic curriculum that was implemented during the first two years of France's school. In second grade, its efficiency was assessed with a pretest/posttest and control group design in 2015-2016 and 2016-2017. Here, the very similar data from the posttests 2016 and 2017 were merged. The posttest assessed mental computation, arithmetic writing, estimation, and mental arithmetic problem solving. These four arithmetical subdomains gave an arithmetic test score for each student. The posttest also included, at the end, a more complex arithmetic problem. The posttest data allowed then to relate the solving procedure used by the children for this problem not only with their response to the problem but also with their arithmetical level as measured by the arithmetic test (posttest without the problem).

Important for the present discussion, the ACE experimental program does not include the teaching of the SWA for computation. Hence, by comparison of the experimental and control groups, the ACE data are well suited for highlighting some of the consequences of teaching SWAs. The teaching of problem solving encouraged ACE students to examine the situation described in the problem and, if possible, to estimate an approximate result before trying to find the accurate answer.

2.2 Participants

The experiment involved 269 second-grade classes, which were divided between the experimental (147 classes) and control (122 classes) groups. Our analyses are based on 4720 students ($M_{\text{age}} = 7.95$, $SD = 0.35$), 2619 in the experimental group and 2101 in the control group, with an overall 48.6 % proportion of girls.

We recruited classes in the north, west, southeast, and center of France. Thanks to the support provided by France's Education Ministry, we were able to obtain all necessary administrative authorizations and to recruit a large number of classes. All the classes in the control group followed the French curriculum, which includes quasi-imperatively teaching an algorithm for addition, subtraction, and multiplication at this academic cycle. The algorithm is usually the classical computation in columns (see Fig. 1), which can be performed either through borrowing or unbundling for the subtraction. It is important to indicate that the curriculum at work in the control classes at the time of our tests emphasized the essential role of problem-solving in mathematical activity. Furthermore, not to favor the experimental group, our test-problem was extracted from the French National assessments (a former version), both in its wording and format of presentation.

2.3 The posttest problem

The problem was presented in writing, but nevertheless read aloud by the experimenter. The statement of the problem was as follows (in French): "At recess, Dimitri play marbles. At the beginning of the game, he has 37 marbles. At the end, he has 72 marbles. How much has he won? ". Students were given 5 minutes to solve the problem, using a pencil and an empty 8.5 x 16.5 cm rectangular frame (included in the students' booklet) for research and computation. They should indicate the response on a location outside of this frame.

Students who think that the essential thing in a verbal problem is to find the arithmetic operation—as a result of the emphasis on SWA (which leads to a bad didactical contract)—rush to find the operation. A quick way to find it, which often leads to the correct answer, is to superficially interpret the terms of the problem formulation. In our problem, the verb “won” (“gagner” in French) suggests an addition.

Our problem was selected for its special ability to test this behavior of the students taught with a focus on the SWAs, as is the case in many of the non-experimental classrooms. A more basic formulation of a subtraction problem, using the (conjugated) verb “lose” (“perdre” in French), is unable to test the real understanding. This is because the students who only resort to the verb “lose” for carry out a subtraction will successfully solve the problem (i.e., carry out the correct operation and probably give the correct answer), but possibly on the basis of an incorrect reasoning.

We focused on this question because a lot of research shows that this answer mode, with an arithmetical operation induced by the terms of the problem's formulation (rather than by the understanding of the situation), is the main problematic behavior in arithmetical problem-solving (at least in France; see, for example, Ehrlich, 1990). Including a problem-item in which the terms of the formulation induce the correct arithmetical operation would blur our comparison between experimental and classic teaching, and make difficult to distinguish between a correct mode of solving and an incorrect one. Thus, our Dimitri-problem was the sole written verbal problem.

2.4 Categorization of the problem-solving procedure

The 11 categories of procedure were defined in the aim to study the efficiency of the SWAs. They can be succinctly described as follow (see also Fig. 1):

1. *box*: rectangular diagram with representation of the total and the two parts

2. *column*: calculation with the SWA (in columns)
3. *line*: calculation completely in line
4. *nline*: representation on a number line (without unities or tens)
5. *numer*: computation base on numeration, with representation of tens and, possibly, units
6. *tree*: calculation first written in line, then reduced with a calculation tree
7. *unit*: computation with only representation of units
8. *mulcol*: multiple (at least two) representations including SWA
9. *muloth*: multiple (at least two) representations excluding SWA
10. *no*: there is no indication of the procedure (the rectangular frame of the posttest sheet is empty or only contains the response)
11. *other*: a procedure that cannot be categorized in anyone of the ten above mentioned procedures.

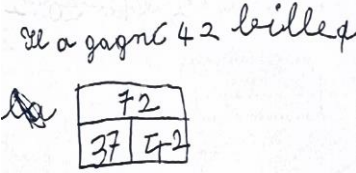
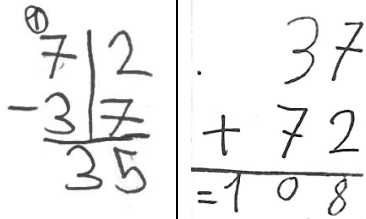
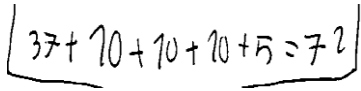
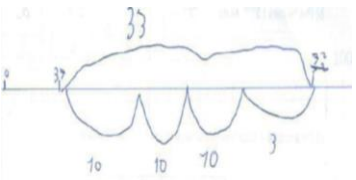
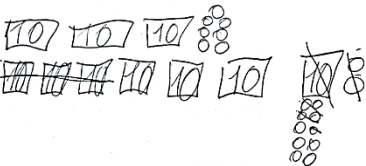
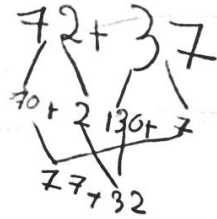
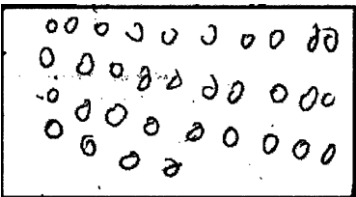
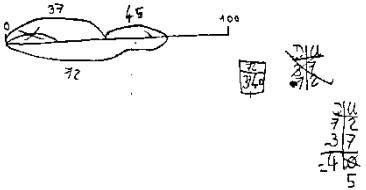
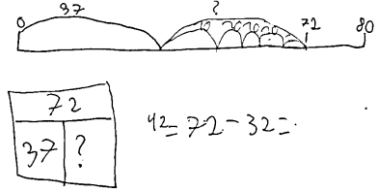
		
<i>box procedure</i>	<i>two column procedures</i>	<i>line procedure</i>
		
<i>nline procedure</i>	<i>numer procedure</i>	<i>tree procedure</i>
		
<i>unit procedure</i>	<i>mulcol procedure</i>	<i>muloth procedure</i>

Figure 1. A sample of the relevant procedures (the *no* and *other* procedures are not shown)

To examine coding reliability of the procedures by the first coders, other coders, independently of the firsts, have re-coded 20% randomly selected procedures from the 4720 students. Despite the large number of categories ($n = 11$), the agreement, estimated by Cohen's Kappa, is “substantial” (Landis & Koch, 1977): $K = .64$, 95% CI = [0.60, 0.67]. As a consequence, the procedures of the 4720 students were analyzed with their initial coding.

2.5 Specific research questions

The data collected allow us to investigate many questions about the SWAs, here identified as the *column* procedure. A first question results from the observation that children who rush into the algorithmic computation, perhaps through a bad didactical contract (see 1.1.5), often choose a false operation. Furthermore, they cannot correct this initial error because they do not have a sufficient operation sense (see 1.1.2) to predict that the outcome of the problem cannot exceed the greatest number of the difference. Thus, a specific hypothesis is that the *column* procedure would more lead to the answer 109 ($= 72 + 37$) than any other procedure.

A second question concerns the effectiveness of the *column* procedure. This major argument in support of teaching the algorithms (see 1.2.1) can be investigated directly by studying the accuracy of the answers by the children who used that procedure. However, with this way of investigation, a correct computation of the false operation will debatably be categorized as inaccurate. Another mean of investigation is more indirect. It is based on the postulate that the best performers in arithmetic will use the most effective procedure.

The ACE data are particularly interesting for discussing these questions because the answers can often be qualified with the separate analysis of the experimental and control groups. Even if the absence of algorithms teaching in the experimental group does not result in a total absence of the *column* procedure in this group, it should result in a considerably less use of this procedure. We would then underline that such data allowing the comparison in the same culture of two substantial samples of 7- to 8-year-old students who were submitted, or not submitted to the teaching of SWAs, are rare.

2.6 Statistical processing

Because the scale of grading of the posttests 2016 and 2017 slightly differ, the scores of the arithmetic tests were standardized, using the mean and standard deviation of the year of assessment. These z-scores permit to indicate the mean arithmetic achievement of the students in Tables 1 and 4, and, in Table 2 to distribute the students in 10 deciles of 472 students each in function of their arithmetic competence.

The subsequent analyses of procedures used by the students will be done at a participants' level. However, the observations of the students were not independent because there were many students in a class (18.7 students per class in mean). This data structure requires a multilevel analysis in order to avoid the statistical nuisance introduced by the dependency of students clustered in their classroom. As we are mainly interested in the use, or not use of a column algorithm, we defined a binary variable whose outcomes are 1 if the student was categorized in the *column* or *mulcol* procedure user, and 0 if not. Then, the data could be modelled with a two-level (level 1: student; level 2: class) logistic regression analysis carried out following the procedure described in Sommet and Morselli (2017), using the R computer statistical language and environment (R Core Team, 2015).

3 Results

3.0 A preliminary multilevel logistic regression analysis.

First, a null model with no other variable than the Class as random intercept was run. This null model allowed to quantify the intra-class coefficient (ICC), that is, the proportion of observed variance in the outcome that is attributable to the effect of clustering. The ICC would be equal to one if the same value for the outcome variable would be attributed to all students in a same class (e.g., 1 for all students in the control classes, in which the

column algorithm was taught; and 0 in the experimental classes in which the column algorithm was not taught). The ICC = .46 found indicates that there was substantial intra-class homogeneity in the use of a column algorithm, justifying this preliminary analysis, but also that this homogeneity is not sufficient for limiting the analysis to teachers' level.

Departing from the null model, we included Gender (level-1 variable) and Group (level-2 variable) as fixed effects, which improved significantly the model, $\chi^2(2) = 126.18, p < .001$. Then, neither authorizing random slope variation of Gender in the classes, nor adding the interaction between Gender and Group as fixed variable improved significantly the model.

The final model yields a significant fixed effect of Group, $z = -9.84, p < .001$; the odd-ratio OR = 0.11, 95%CI = [0.07, 0.16], indicates that students in the experimental group are about 9.5 times less likely to use the written column representation than students in the control group. The fixed effect of Gender is also significant, $z = -3.71, p < .001$; OR = 0.66, 95%CI = [0.52, 0.82], indicates that boys are 1.5 times less likely than girls to use the written column representation.

3.1 Accuracy of the answers as a function of the procedure

The effectiveness of a procedure was first assessed by reporting the number of correct answers to the number of uses of this procedure. The resulting percentages and their association with the arithmetic achievement level of the students who used the procedure are unambiguous (see Table 1): Whereas the two multiple procedures—*mulcol* and *muloth*—were the most effective they were also mostly used by the high achievement students.

Table 1. Number of uses, percentage of correct answers associated, and mean z-score of the students who used that procedure, for the 11 procedures

Procedure	Number of uses	Percentage of success	Mean z-score
<i>box</i>	205	12.68	-0.077
<i>column</i>	499	20.24	0.009
<i>line</i>	741	24.29	0.292
<i>nline</i>	426	20.19	0.154
<i>numer</i>	188	23.94	-0.015
<i>tree</i>	45	22.22	0.275
<i>unit</i>	326	21.78	-0.175
<i>mulcol</i>	104	30.77	0.462
<i>muloth</i>	519	27.75	0.348
<i>no</i>	1387	13.05	-0.272
<i>other</i>	280	12.86	-0.268

3.2 Procedure used as a function of arithmetical ability

Table 2 shows that the procedures are differently distributed in the decile groups, $\chi^2(90) = 408.20, p < .001$, Cramer's V = .098. The rank correlations between frequency of a procedure and decile differ considerably. The *line*, *mulcol*, and *muloth* procedures correlate positively, significantly and very strongly with the deciles, whereas the *no* and *other* procedures correlate negatively, significantly and very strongly with the deciles. In contrast, the

correlation for the *column* procedure is approximately null. Therefore, the *column* procedure seems not related to the level of arithmetical ability, whereas *line*, *mulcol*, and *muloth* procedures are increasingly (and the *no* and *other* procedures decreasingly) used as the level of ability increases.

Table 2. Number of uses of a procedure as a function of the arithmetic ability level of the students (ability levels are in deciles: decile 1 = low; decile 10 = high), with their rank correlations (r_s)

Decile	1	2	3	4	5	6	7	8	9	10	r_s	p
<i>Procedure</i>												
<i>box</i>	18	23	27	21	23	21	22	16	15	19	-.500	.141
<i>column</i>	52	51	52	58	43	40	44	53	55	51	-.006	.987
<i>line</i>	28	45	62	71	83	79	82	87	88	116	.963	<.001
<i>nline</i>	18	49	36	41	41	51	43	53	38	56	.584	.077
<i>numer</i>	18	15	20	20	15	21	25	27	15	12	-.037	.919
<i>tree</i>	2	2	3	5	6	7	5	4	6	5	.593	.071
<i>unit</i>	36	38	39	29	30	36	45	36	27	10	-.485	.156
<i>mulcol</i>	5	5	6	5	14	9	8	14	18	20	.899	<.001
<i>muloth</i>	22	27	49	45	35	54	54	82	76	75	.900	<.001
<i>no</i>	236	174	140	151	153	131	115	84	112	91	-.915	<.001
<i>other</i>	37	43	38	26	29	23	29	16	22	17	-.857	.002
z-score	-1.66	-1.11	-0.75	-0.45	-0.16	0.12	0.42	0.74	1.13	1.72	mean z = 0.00	

3.3 Types of response: *column* procedure versus all the other procedures

Table 3 shows the main types of final responses. The distribution of the students who used the sole *column* procedure and all the other students in the six types of response differs significantly $\chi^2(5) = 607.04$, $p < .001$, with a substantial effect size, Cramer's $V = .359$. For the *column* procedure, the "109" response percentage is strongly higher and the "None" response percentage strongly lower than these percentages for all other procedures together. Besides, the two main contributions to the chi-square statistics come from these "109" and

“None” responses. This double observation suggests that a great number of students applied blindly the *column* procedure.

Table 3. Percentage response types for the students who used the *column* procedure contrasted with the response types of all the other students

Procedure	Number students	Response					
		35 (correct)	109 (=72+37)	45	< 72 (\neq 35, 45)	\geq 72 (\neq 109)	None
<i>column</i> (sole)	499	20.24	35.67	9.82	21.64	9.02	3.61
all others (together)	4221	19.21	5.07	6.75	39.09	11.21	18.67
Total	4720	19.32	8.31	7.08	37.25	10.97	17.08

When these results are differentiated between the experimental and control groups, the data (not presented in Table 3)¹ show that the students in the experimental group produced in total slightly more the correct response “35” than the students in the control group, 20.20% versus 18.23%, $\chi^2(1) = 2.90$, $p = .089$, $\phi = 0.025$. However, they produced less the response “109”, which results from carrying out of the false operation, 5.08% versus 12.33%, $\chi^2(1) = 80.45$, $p < .001$, $\phi = 0.131$, and produced more a response below 72, which is consistent with the subtraction sense, 68.96% versus 57.02%, $\chi^2(1) = 71.79$, $p < .001$, $\phi = 0.123$.

3.4 Comparison for the *column* procedure: experimental versus control group

Table 4 shows the comparison on three variables of the experimental and control groups for the *column* procedure. The difference in the uses of the *column* procedure is not only highly significant, but the ϕ effect size of the group is impressive. Thus, the ACE experimental program reduces considerably the use of the *column* procedure. One can even wonder why this procedure, which was not taught, was nevertheless used at all in the experimental group. But students may learn the procedure at home, in first grade, or with teachers who taught it in addition to the experimental program (often under parental pressure). The significant higher z-score of the students in the experimental group who used the *column* procedure suggest that these students love arithmetic and have perhaps practiced the written algorithm at home, for themselves, with possibility to verify the answers on a computer or calculator.

¹ A complete table is available from the corresponding author of the article.

Table 4. Number of uses, percentage of correct answer associated, and mean z-score of the students who used the *column* procedure, as a function of the group: experimental vs. control.

Group	Number of uses	p $\phi^{(a)}$	Percentage of success	p $\phi^{(a)}$	Mean z-score	p $\sqrt{ges}^{(b)}$
Experimental	80	$p < .001$	22.50	$p = .301$	0.317	$p = .003$
Control	419	$\phi = .79$	19.81	$\phi = .02$	-0.050	$\sqrt{ges} = .132$
Total or significance	499	signif.	20.24	not signif.	0.009	signif.

(a) The significance p and the effect size ϕ of the difference between the two groups were approximated with the chi-square statistic.

(b) The significance p and the generalized effect size (ges) of the groups were obtained through an ANOVA (the square root of ges was used for direct comparability with ϕ).

3.5 Comparison between genders

Table 5 shows that the distribution of boys and girls in the different categories of procedures differs significantly, $\chi^2(10) = 42.58$, $p < .001$, even if the effect size of gender is small, Cramer's $V = 0.095$. In addition, we verified that the statistical significance of the chi-square test results mainly from the *column* data. Indeed, girls used the *column* procedure more than boys, $\chi^2(1) = 14.02$, $p < .001$, $\phi = 0.055$.

Table 5. Percentage of use of the procedures as a function of gender

Proc →	box	column	line	nline	numer	tree	unit	mulcol	muloth	empty	other
Girls	4.54	12.30	14.04	8.63	4.19	1.05	8.11	2.53	11.34	27.13	6.15
Boys	4.16	8.94	17.27	9.40	3.75	0.87	5.77	1.90	10.68	31.53	5.73
Total	4.34	10.57	15.70	9.03	3.96	0.95	6.91	2.20	11.00	29.39	5.93

4 Discussion

4.1 A limitation of the current study

Given the great proportion of unknown procedures it is possible that a preponderant internal procedure escapes our analysis. Students may also have used covertly one of the 9 relevant procedures, therefore invalidating the order of frequency of the procedures. However, the examination of the *no* procedure shows that it leads to the correct answer only in 13.05% of the students. Furthermore, these students performed below the mean at the arithmetical test (see Table 1) and were rather in the first deciles than in the last deciles (see Table 2). Therefore, a plausible hypothesis is that a majority of the students in the *no* procedure do not understand the problem.

Another limitation could arise from the statistical analysis mainly performed only at the students' level. However, the preliminary two-level logistic regression analysis yielded, for the differentiated use of columnar computation in the two groups and in the two genders, convergent results with the analysis at the students' level. Thus, we can think that the possible statistical nuisance induced by within-class homogeneity does not invalidate the results of our more complete analyses at the students' level.

4.2 Is the *column* procedure efficient in problem-solving?

Table 1 shows that the multiple procedures *mulcol* and *muloth* yielded the greatest percentage of success, 30.77 and 27.75% respectively. In comparison, the isolated *column* procedure, with 20.24% of successes (101 out of

499) holds only the 7th position out of 11, behind the multiple procedures and also the *line*, *numer*, *tree*, and *unit* procedures; it clearly outperformed the *no*, *other* and *box* procedures. This result is difficult to interpret for two reasons. First, in defining efficiency of a procedure by the accuracy of the final response, students who compute correctly the false operation in columns will lead to an underestimation of the effectiveness of the *column* procedure. However, as argued in section 1.1.5, the teaching of the *column* procedure often leads to a bad didactical contract. Therefore, the inflation of the “109” response (see section 3.3), which typically results from the correct computation of $72 + 37$, brings a negative support to the conceptual aspect of the *column* procedure.

Second, the procedure is not independent from the students’ ability. For example, if gifted students use a certain procedure, this procedure will be effective. However, a procedure that does not lead to the correct answer must be considered as ineffective regardless of the level of ability of the students who used it. Thus, it is not surprising that the procedures *no* and *other* are among the three most ineffective procedures. The *box* procedure—the schematic drawing for combine problems (Fuson & Willis, 1989)—was the most ineffective one (only 12.68% successes), and was almost exclusively used in the experimental group. Indeed, the teaching of this procedure is intended to find the correct arithmetic operation and does not provide a method for calculating this operation. Therefore only students who can calculate mentally $72 - 37$ in addition to correctly placing the numbers in the rectangular diagram can succeed.

The indirect mean to assess the relative efficiency of the procedures consists in looking at the procedures used by the most performing students in the arithmetical test. Table 1 shows that the most performing students, with a 0.462 mean z-score, were in the *mulcol* procedure. Table 2 is consistent with this finding because the frequency of use of the *mulcol* procedure increases when the level of arithmetic ability increases. Among the other procedures, the *muloth* and *line* procedures were also more chosen by the students with high performance in the arithmetic test. Therefore, the two ways of determining the most efficient procedure seem converge toward designating the multiple procedures, particularly the *mulcol* procedure, as the most efficient.

4.3 Other empirical observations on the effectiveness of SWA

Kamii and Dominick (1997) report the results of a computation given in 2nd, 3rd and 4th grade classes in which the conventional algorithm was, or was not taught. The children were asked to solve the horizontally written problem $6 + 53 + 185$ (in fact, $7 + 52 + 186$ in grade 2) without using any pencil or paper during individual interview at the end of the school year. The authors found that those who had not been taught any algorithms produced much more reasonable errors than those who were taught a SWA. They concluded that “the time has come to stop teaching the algorithms” (p. 60, see also section 1.1.1). However, the comparison seems unfair because the students who were taught the algorithm could not use it, at least in its normal paper and pencil form, whereas the compared students could use their normal procedure of computation. Thus, no clear conclusion can be drawn from this data.

Csikós (2016) studied the relationship between students’ performance in mental calculation and the strategies they use in a sample of 78 Hungarian fourth graders who solved three-digit addition problems (e.g., $342 + 235$, $498 + 256$) and two subtraction problems (e.g., $701 - 694$). Because the written algorithm is only introduced in the first half of grade 3 in Hungary, we can compare the frequency of use of the paper-and-pencil mental analogue strategy by the Hungarian students with that of students in countries where the written algorithm is introduced before grade 3. First, it is interesting to note that the paper-and-pencil mental analogue of the written algorithm seems a slower strategy than an indirect addition strategy (for subtraction): about 39 s

versus 12 s². Second, Csikos found only 25% Hungarian children using this written algorithm induced strategy, whereas Foxman and Beishuizen (2002) found 51% of 247 British (England, Wales and Northern Ireland) 11-year-old students using this strategy in a comparable 238 + 143 task.

Thus, the research by Csikos (2016) on students following a curriculum in which the algorithms are introduced somewhat later than usually does not confirm the speed of the written algorithm. Therefore, it qualifies the empirical results for this speed in section 1.2 (Dewi et al., 2015; Torbeyns & Verschaffel, 2013, 2016). Moreover, the comparison with Foxman and Beishuizen's (2002) finding suggests that the frequency of the paper-and-pencil mental analogue of the written algorithm strategy should decrease through delayed teaching of the written algorithm. This suggests that rather than “stop” the teaching as suggested by Kamii and Dominick (1997) a less radical but nevertheless effective option could be to delay this teaching.

4.4 A model of calculation learning without using or seeing the written algorithms

The ACE research provides some data that allow to anticipate what happens if students do not learn the classical algorithm in the first two years of school. However, how will older students or adult persons calculate, if they have never learned these algorithms despite living in our culture? Early blind individuals (EB) can provide a model to answer this question. Indeed, EB individuals have never seen the written algorithms and were probably not encouraged to use it. In consequence, they will not use a paper-and-pencil column mental analogue procedure in mental calculation. If the latter strategy really complicates the mental calculation through the complex operations of carrying and borrowing, early blind individuals should perform better than late blind individuals (LB) and sighted controls individuals (SC).

The research by Dormal, Crollen, Baumans, Lepore and Collignon (2016) suggests at least that our audacious hypothesis is not completely unrealistic. Indeed, the research by Dormal et al. allows to compare the three groups of individuals (EB, LB, and SC) on two types of addition and subtraction problems. In the first type of addition (e.g., 49 + 5) and subtraction (e.g., 61 – 6) problems, the second operand leads to carrying (or borrowing, unbundling) when using a written algorithm (or analogue), whereas in the second type it does not lead to these complications (e.g., 52+2, 57-2, 42+12, and 66-12). Comparing the performance of the three groups of individuals shows that the EB group performs better than the two other groups specifically in the first type of problems. The finding is reinforced by the fact that there were a majority of men in the EB group and a majority of women in the LB and SC groups. This because other findings (see sections 1.1.3, 3.5., and Fischer, 2004) suggest that girls should use the paper-and-pencil mental analogue strategy more than boys. Finally, Dormal et al. found no straightforward relation between working memory and arithmetic performance in the EB group. As an efficient working memory may explain the enhanced performance of the EB individuals, this latter negative result encourages finding another explanation. This explanation could be that the calculation strategy of the EB individuals differs from that of the LB and SC individuals, and, further, that the complex operation of carrying or borrowing makes the mental calculation harder for the LB and SC individuals, but not for the EB individuals.

² The author qualifies the differences in response time as “convincing”. However, because students spontaneously used their strategies, one cannot exclude that the slow students choose the (eventually) slow strategy and rapid students chose the (eventually) fast strategy.

5 Conclusions

In its empirical part, the present research used a sole subtraction word-problem. However, in practice, this problem allowed a representation of the described situation and the development of solution procedures. Insofar that the effectiveness of the SWA was not studied *per se*, our empirical approach allows then some generalizations to other arithmetical operations and conclusions about the appropriateness of teaching SWA for arithmetical problem-solving in the first school years. The conclusions are as follows.

Firstly, the multiple procedures, particularly the *mulcol* procedure—that is, at least one complementary procedure added to the columnar algorithm—, seems very effective. The complementary procedure (number line for example) provides the students with a way to find the arithmetical operation and, then, the columnar algorithm proves to be a mean of correctly carry out the operation. When the use of calculators and computers will replace the columnar algorithm, the emphasis should be on a procedure (e.g., *nline*, *box*) that leads to find the operation. The time saved through not teaching the algorithms could allow the teachers to reinforce the training in mental calculation. Thus, the students should higher perform in arithmetic problem solving and the negative effects of the algorithm, notably on mental calculation, will be avoided.

Secondly, our present research shows that students' use of the columnar algorithm increases the carrying out of the erroneous arithmetical operation, $72 + 37$ instead of $72 - 37$. Of course, the presence of the verb “win” in the statement of the problem can also explain this erroneous choice of the operation. However, the students in the experimental group had the same verbal problem to solve, and, therefore, had to inhibit this misleading verbal inductor as well. Nevertheless, they found less frequently a result equal or greater than 72 (16% of the students) than the students in the control group (24%), particularly the result 109 (5% vs. 12%). This lower proportion of students who found a result that violates the sense of subtraction in the experimental group, in addition to their slightly higher success rate (20% vs. 18%), is important because some neuropsychologists have attributed this type of error to a right hemispheric dysfunction (Rourke, 1982; Strang & Rourke, 1985). In showing that a change in teaching may be sufficient in many cases to remedy to this type of error, the present research contributes to eradicate this explanation through a putative neurological dysfunction.

Thirdly, this erroneous choice of the arithmetical operation, and the uniform use of the SWA according to the ability of the students, suggests that the SWA for subtraction should be well suited for the most advanced students in 2nd grade, but used without reflection or understanding by many others. For the latter students it is a blind practice, imposed from outwards, and partially or incorrectly automatized by its repetition.

More generally, deciding whether or not to teach a SWA for an arithmetic operation is a multi-faceted undertaking that depends especially on children's skills. However, the present empirical research provides information to be taken into account in the decision-making process as it suggests that this teaching may have a rather negative effect on solving arithmetic verbal problems.

6 References

- Brousseau, G., & Warfield, V. (2014). Didactical contract and the teaching and learning of science. In R. Gunstone (Ed.), *Encyclopedia of Science Education*. Dordrecht: Springer.
- Chesné, J. F. & Fischer, J. P. (2015). Ce que nous apportent les évaluations nationales sur les connaissances des élèves dans le domaine des nombres et du calcul à l'école primaire. <http://www.cnesco.fr/wp-content/uploads/2015/11/Acquis-des-%C3%A9l%C3%A8ves.pdf>. Accessed February 5, 2018.

- Csíkó, C. (2016). Strategies and performance in elementary students' three-digit mental addition. *Educational Studies in Mathematics*, 91, 123-139.
- Dewi, J. D. M., Castel, C., Kerzel, D., Posada, A., & Thevenot, C. (2015). Strategies for written additions in adults. *Journal of Cognitive Psychology*, 27, 979-991.
- Dormal, V., Crollen, V., Baumans, C., Lepore, F., & Collignon, O. (2016). Early but not late blindness leads to enhanced arithmetic and working memory abilities. *Cortex*, 83, 212-221.
- Ehrlich, S. (1990). *Sémantique et mathématique: Apprendre/Enseigner l'arithmétique simple*. Paris: Nathan.
- Fischer, J. P. (2004). Les différences cognitives entre sexes : une autre approche et d'autres observations. *Pratiques psychologiques*, 10, 401-413.
- Fischer, J. P., & Koch, A. M. (2016). Mirror writing in typically developing children: a first longitudinal study. *Cognitive Development*, 38, 114-124.
- Flores, R., Koontz, E., Inan, F. A., & Alagic, M. (2015). Multiple representation instruction first versus traditional algorithmic instruction first: Impact in middle school mathematics classrooms. *Educational Studies in Mathematics*, 89, 267-281.
- Foxman, D., & Beishuizen, M. (2002). Mental calculation methods used by 11-year-olds in different attainment bands: a reanalysis of data from the 1987 APU survey in the UK. *Educational Studies in Mathematics*, 51, 41-69.
- Fuson, K. C., & Willis, G. B. (1989). Second grader's use of schematic drawings in solve addition and subtraction word problems. *Journal of Educational Psychology*, 81, 514-520.
- Kamii, C. (1994). *Young children continue to reinvent arithmetic, 3rd grade*. New York (NY): Teachers College Press.
- Kamii, C., & Dominick, A. (1997). To teach or not to teach algorithms. *Journal of Mathematical Behavior*, 16(1), 51-61.
- Kamii, C., & Dominick, A. (2009). The harmful effects of "carrying" and "borrowing" in grades 1-4. [http://www.bradthiessen.com/html5/m340/11%20TheHarmfulEffectsof%3FCarrying%3Fand%3FBorrowing%3F\(2009\).pdf](http://www.bradthiessen.com/html5/m340/11%20TheHarmfulEffectsof%3FCarrying%3Fand%3FBorrowing%3F(2009).pdf). Accessed February 5, 2018.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Piaget, J. (1977). *Recherches sur l'abstraction réfléchissante: 1/L'abstraction des relations logico-arithmétiques*. Paris (France): PUF.
- R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. *Journal of Experimental Child Psychology*, 141, 83-100.
- Rourke, B. P. (1982). Central processing deficiencies in children: Toward a developmental neuropsychological model. *Journal of Clinical Neuropsychology*, 4, 1-18.
- Selter, C. (2001). Addition and subtraction of three-digit numbers: German elementary children's success, methods and strategies. *Educational Studies in Mathematics*, 47, 145-173.

- Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology*, 30(1), 203-218.
- Strang, J. D., & Rourke, B. P. (1985). Arithmetic disability subtypes: The neuropsychological significance of specific arithmetical impairment in childhood. In B. P. Rourke (Ed.), *Neuropsychology of learning disabilities: Essentials of subtype analysis* (pp. 167-193). New York (NY): Guilford Press.
- Torbeyns, J., Hickendorff, M., & Verschaffel, L. (2017). The use of number-based versus digit-based strategies on multi-digit subtractions: 9-12-year-olds' strategy use profiles and task performances. *Learning and Individual Differences*, 58, 64-74.
- Torbeyns, J., & Verschaffel, L. (2013). Efficient and flexible strategy use on multi-digit sums: a choice/no-choice study. *Research in Mathematics Education*, 15, 129-140.
- Torbeyns, J., & Verschaffel, L. (2016). Mental computation or standard algorithm? Children's strategy choices on multi-digit subtractions. *European Journal of Psychology of Education*. 31, 99-116.