CrossMark

Can young students understand the mathematical concept of equality? A whole-year arithmetic teaching experiment in second grade

Jean-Paul Fischer[1] ⓘ · Emmanuel Sander[2] · Gérard Sensevy[3] ·
Bruno Vilette[4] · Jean-François Richard[5]

Abstract Ensuring students correctly understand the notion of equality is a fundamental problem in teaching mathematics. Is it possible to teach arithmetic in such a way that students do not misinterpret the equal sign "=" as indicating the result of an arithmetic operation and, consequently, viewing the arithmetic operation symbols (+, -, or x) as systematically meaning perform a computation? The present paper describes the implementation of an experimental

✉   Jean-Paul Fischer
     jean-paul.fischer@univ-lorraine.fr

     Emmanuel Sander
     emmanuel.sander@unige.ch

     Gérard Sensevy
     gerard.sensevy@espe-bretagne.fr

     Bruno Vilette
     bruno.vilette@univ-lille3.f

     Jean-François Richard
     jf.richard.paris8@gmail.com

[1]   Lorraine Laboratory of Psychology and Neuroscience (2LPN), University of Lorraine, 54015 Nancy,
      France

[2]   Developmental Psychology and Educational Sciences, IDEA Lab, University of Geneva, Geneva,
      Switzerland

[3]   Educational Sciences, CREAD Laboratory, University Bretagne Occidentale, Brest, France

[4]   Developmental Psychology, PSITEC Laboratory, University Lille, Villeneuve-d'Ascq, France

[5]   Paragraphe Laboratory, University Paris 8, Saint-Denis, France

arithmetic teaching program (called ACE) drawn up in conjunction with teachers and focusing on these issues. We assessed the program's impact via an experiment involving 1140 experimental group students and 1155 control group students, and using a pretest/posttest design. The experimental group students achieved higher composite arithmetic scores, combining all four subdomains (arithmetic writing, mental computation, word-problem solving, and estimation), than control group students. The effect size, computed using individuals' posttest scores adjusted for pretest scores was d = 0.559. The ACE program was effective in all four subdomains tested; however, it was particularly successful in the arithmetic writing subdomain, especially in writing equalities. The program's effect not only held one year later, it was cumulative, as the benefits produced by following the program in both first and second grade were almost double those of following the program in just one grade.

Keywords Arithmetic teaching · Arithmetic writing · Arithmetic comprehension · Second grade · Equal sign

Introduction

This article is based on a new teaching program, called ACE (Arithmetic Comprehension at Elementary school), which was experimented during the years 2012–2016 in France. Rather than assessing the global effectiveness of the program, the main purpose of the article is to test the possibility for teaching a correct concept of equality by training the students in arithmetic writing. However, other taught subdomains of arithmetic and some general characteristics of the program's effect that indirectly may influence arithmetic writing are also studied.

The mathematical concept of equality

For preschool children, the logic of exact equality means understanding that collections of objects have an exact numerical value and understanding how set manipulations affect that value (cf., Jara-Ettinger et al. 2017). However, by the time most children enter school, the most important marker of the notion of equality is the "=" sign. Mathematically, "$a = b$" means that the symbols a and b represent the same mathematical object. The equality is a reflexive, symmetric, and transitive relation. With this definition and properties, the most obvious equality should be "$a = a$." Unfortunately, as Ginsburg (1977) noted, such equalities are not easy for children to comprehend.

A major obstacle

Attempts to understand and overcome this difficulty have then generated a substantial literature in the field of didactics (Baroody and Ginsburg 1983; Fischbein 1989; Ginsburg 1977; Kieran 1981). A major obstacle to students' understanding of equality arises from the way arithmetical facts (addition, subtraction, multiplication) are learnt in the first school grades, as current teaching programs tend to induce the "operations = answer" format (McNeil et al. 2015). As is often relayed by textbooks (Powell 2012), this results in students understanding the equal sign as an operational sign rather than as a relational or equivalence sign.

The "operations = answer" format encapsulates a frequent complaint of mathematics educators that students understand an arithmetic operation sign (e.g., the "+" sign for addition)

as a signal for computing (adding) the two terms (addends), and subsequently interpret the equal sign as announcing the result of the computation. This partial understanding of the writing "a + b =" can have unfortunate consequences.

Some consequences of the "operations = answer" format

The "operations = answer" format has a number of direct, short-term consequences. First, it makes students reluctant to break down a number by, for example, writing it as a sum of two or more numbers. This issue was at the heart of the experiment conducted by McNeil et al. (2015) in second-grade classrooms. They found that presenting problems with operations on the right side, for example $\_ = 4 + 3$, promotes understanding of mathematical equivalence. Second, when second graders have to judge whether $825 + 57$, for example, is less or more than $825 + 66$, some students try to calculate both $825 + 57$ and $825 + 66$, and then compare the two results. Given the difficulty of doing these computations mentally and of memorizing and comparing their results, students are unlikely to obtain the correct answer by using this strategy. Third, compliance to this format limits the use of substitution. For example, as Chesney et al. (2014) showed, students will not spontaneously approach a problem such as $3 + 4$ by replacing $3 + 4$ with $3 + 3 + 1$, because they do not conceive of 4 as a way of writing $3 + 1$, even though the involvement of a double means that $3 + 3 + 1$ is easier to calculate than $3 + 4$.

A more indirect short-term consequence of the "operations = answer" format is that it makes it more difficult to associate two inverse operations, most notably addition and subtraction (Baroody 1999). For example, second graders who conceive $866 - 128$ as an operation to run (often taking away 128 from 866) cannot use the fact that $738 + 128 = 866$ to obtain the result by replacing 866 with $738 + 128$ in the operation $866 - 128$, and then using inversion to remove 128 from $738 + 128$.

In the long term, this limited conception of arithmetical signs, especially the equal sign, as a signal for computing from left to right may hinder algebraic problem-solving (Byrd et al. 2015; Knuth et al. 2006; McNeil et al. 2010) because correctly and completely understanding that the two sides of an equation represent the same quantity is "foundational for more advanced mathematics, particularly algebra" (Crooks and Alibali 2014, p. 350).

McNeil (2007) highlighted an unusual development curve in mathematics when computations are extended to three numbers. Contrary to the usual progression with increasing age, she found that performance on mathematical equivalence problems, such as $7 + 4 + 5 = 7 + \_$, declined between the ages of 7 and 9. Performance improved once again after the age of 9, thereby producing the unusual development curve. This decline may be due to 2 or 3 years of school teaching leading children to perceive the plus sign as a signal to add. Furthermore, the children in this study were raised in a culture that conditions them to read from left to right. However, before the age of 6 or 7 years, this perception is weaker and left-to-right reading is not so ingrained (as demonstrated by the ease with which 5- to 6-year-olds write from right to left: see Fischer and Tazouti 2012).

Enhancing the notion of equality in students

As Baroody and Ginsburg (1983 p. 210) noted: "it may be easier to teach a relational view of 'equals' if it is taught from the very beginning of formal math instruction." Accordingly, one of the aims of our ACE research program was to implement during the first 2 years of school a full arithmetic curriculum whose main purpose is to improve understanding of the

mathematical notion of equality. Drawing up a full curriculum allowed us to include activities that would reduce the weight of the "operations = answer" conception in students' arithmetic. For example, activities designed to enhance number sense (Dehaene 2011) should reduce students' tendencies to rush into attempting an exact calculation by teaching them to begin by predicting the approximate result.

Baroody and Ginsburg (1983) directly demonstrated that first- to third-grade students can be taught a correct conception of the plus sign. A more indirect demonstration is provided by DeCaro (2016), who showed that the context in which older students (fifth- and sixth-grade students, and undergraduates) solve problems can impact both their strategy and their understanding of the problems. Participants who solved complex problems, such as $7 + 5 + 9 = 3 + \_$, before solving problems with a repeated addend, such as $7 + 5 + 9 = 7 + \_$, were less likely to use the most efficient strategy (here adding 5 and 9) than participants who solved the problems after first solving problems with a repeated addend. Hence, early instruction can enable students to adapt their strategies. However, if early instruction is too complex, it may lead students to overlook important features of a problem, as in the example of an addend repeated on both sides of the equal sign. More generally, DeCaro underlines the important role played by the initial learning context, including in the writing of equalities.

Aim of the present study

Despite long awareness of children's difficulties in understanding the notion of equality, this issue has not yet been resolved. In fact, Byrd-Hornburg et al.'s (2017) integrative analysis of studies carried out after 2010 showed that children perform poorly on all measures of mathematical equivalence. They identified gender as a source of variation in understanding the concept of equality, but what is most important for educational psychologists is to create teaching tools and methods that can help children correctly grasp the concept of equality (e.g., McNeil et al. 2015, 2011).

Consequently, the present study was designed to determine whether it is possible to create a successful program for typical teaching. One obvious condition was that teachers would have to be trained so they did not implicitly teach any restricted or distorted conceptions of the mathematical notion of equality. Our ACE program fulfilled this condition, even if teacher training was limited to between 2 and 4 days at the end of the school year prior to the experiment or at the start of the school year in which the experiment began.

Research questions and hypotheses

The unusual U-shaped development curve reported by McNeil (2007) raises the question of whether the initial developmental decline is unavoidable or whether it is a by-product of inappropriate teaching methods. A more general question is whether an appropriate teaching program, which does not instill in students the "operations = answer" conception, can induce correct understanding of the notion of equality. Because this was the aim of the ACE program, we hypothesized this program would be more effective than a classic teaching program in the arithmetic writing subdomain.

However, we also examined two further questions. First, is learning arithmetic writing detrimental to other fields of mathematics learning? We investigated this issue by assessing three other major subdomains of arithmetic learnt in second grade: mental computation, word problem-solving, and estimation. In order to demonstrate a non-detrimental effect,

achievement scores in these subdomains had to be at least as high for students who followed the ACE program as for students who followed a classic teaching program. An additional hypothesis is that the other subdomains help students construct a correct notion of equality and vice versa. This hypothesis led us to predict that the ACE program would be globally more effective than classic teaching programs. This hypothesis is beyond the scope of short experiments with a narrow focus, concentrating on the arithmetic writing subdomain, such as Alibali et al.'s (2009) experiment. In that experiment, third- and fourth-grade students received a half-an-hour lesson about one (in each of two experimental groups) of two strategies for solving equations such as $3 + 4 + 6 = 3 + \_$. The teaching of the equalize strategy ($4 + 6 = 10$) proved to be effective, but not that of the add-subtract strategy ($3 + 4 + 6 = 13, 13 - 3 = 10$).

Second, is the predicted positive effect of the ACE program a transient phenomenon? Answering this question requires assessing the permanence of a student's learning. Although the present study was not longitudinal, some of the second-grade students had already followed an ACE program in first grade, but without being tested. Hence, we were able to examine the hypotheses (a) that students who follow the ACE program for 2 years will perform better on the final test than students who follow the ACE program for just 1 year, (b) that the latter will perform better than the students who follow a classic teaching program during the two first grades, and moreover (c) that the students who followed the ACE program only in first grade, thus being in the control group in second grade, will also benefit from this program.

Method

Teaching programs

Because the official program for teaching mathematics in French schools introduces multiplication during second grade, we could not restrict the experimental program and tests to additive problems and writings. However, in the interests of concision, all the examples included in the following presentation relate to the teaching of addition, which constitutes the majority of the second-grade arithmetic program, in so far as subtraction is considered an additive operation. Nevertheless, the ACE program used similar approaches for the other arithmetical operations (multiplication, and, to a lesser degree, division).

Students were initially taught to write sums without computation via a "Statement Game" (Joffredo-Le Brun et al. 2018), in which they had to note in full the outcome of throwing two dice. For example, if the dice showed five and three, they had to write $5 + 3$, not 8. If a second throw of the dice resulted in a 6 and a 4, the students could write $5 + 3 < 6 + 4$, justifying their writing without computation by observing that $5 < 6$ and $3 < 4$. Likewise, if the outcome of the second throw was 4 and 1, they could write $5 + 3 > 4 + 1$, because $5 > 4$ and $3 > 1$. When this procedure was not applicable, for example, when comparing $5 + 1$ and $6 + 3$, the students explicitly stated that the procedure did not apply, and used their spontaneous method of control, that is, counting the two sums. As shown by Hattikudur and Alibali's (2010) experiment involving third- and fourth-grade students, using the inequality symbols in this way facilitates learning about the equal sign. Many other challenging writings (e.g., $9 + 1 + 4 + 6 + 5 + 5 + 8 + 2$) that do not promote an automatic left-to-right computation—here computing $(9 + 1) + (4 + 6) + (5 + 5) + (8 + 2)$ when the students have memorized the complements to 10—were practiced during the school year. Moreover, each student had a personal notebook—a "Journal of Numbers"—in which he/she could invent challenging writings.

To show the utility of writing a sum, students were shown displays of, for example, four dots in a square, five dots in a quincunx, and three dots in a triangle. The display time was limited so students could perceive, but not count, the three groups of dots. They were then allowed to note the number they saw by writing "$4 + 5 + 3$." To favor the reading $a = b + c$, the rectangular representations shown in Fig. 1 were used to represent both "equalities" and "missing addend" problems (Brissiaud 1994).

The teacher training course included a hidden treasure game (see Appendix 1, see electronic supplementary material) to show that writing "$a + b$" sometimes provides more information than calculating the sum. In this game, the treasure is hidden in one of nine bouquets of tulips and roses. The teacher (or a student) indicates the hiding place via an additive writing, which allows the bouquet to be identified. For example, the statement "$5 + 4$ flowers" unambiguously indicates one of the nine bouquets, whereas the sum of the two numbers, 9 flowers, could indicate any of the nine bouquets. Furthermore, comparing a bouquet with a tulips on the left and b roses on the right and a bouquet with b tulips on the left and a roses on the right illustrates the difference between $a + b$ and $b + a$ and shows the need to further demonstrate that $a + b = b + a$.

The other major components of arithmetic were also taught in such a way as to avoid instilling the "operations = answer" format. For example, in the problem-solving and numerical estimation subdomains, students were taught not to rush into performing a computation. Instead, they were encouraged to examine the situation described in the problem and estimate an approximate result for a complex computation (e.g., $39 + 52$) before computing a precise answer. The program consisted in 150 h of arithmetic teaching, namely 36 weeks × 5 h per week of mathematics as planned in the French official program minus 30 h of geometry, which was not included in the program. This number of hours allows many other activities, although it is beyond the scope of the present paper to describe them in detail. Because it matches the number of hours recommended in the official curriculum, it should be equivalent to that practiced in the control classes.

All teachers of the control group elaborated the mathematical activities of their class by following the official curriculum of the National Education. This curriculum includes numbers (up to 1000), calculation with notably a written algorithm for addition and subtraction, and problem-solving. Because the District Inspectors control the application of the curriculum, the program taught in the control classes should present a certain homogeneity. Moreover, the official curriculum is also relayed by the textbooks. Consequently, we verified that some major peculiarities of the ACE program are not developed in the textbook mainly used in the control classes (Dussuc et al. 2009). For example, $a \times b$ is not defined as the number of cells in an $a \times b$ rectangle (see section "A mutual influence of two subdomains" of "Discussion") in this textbook. In the other direction, a written algorithm for addition and subtraction is present in the textbook, but not in the ACE program.

In addition, because we hypothesized that the experimental group would achieve better performance scores due to the way the notion of equality is taught, it should be noted that the

13	
7	6

13	
?	6

Fig. 1 Example of how children could represent additive equalities (left-hand box) and "missing addend" problems (right-hand box) in the ACE program

French official curriculum at work in 2015–2016 (the school year of the present experience) does not stress decomposition, in contrast to the French official curriculum carried out in 2016–2017. In fact, decomposing a number (i.e., writing c = a + b), rather than composing it (i.e., writing a + b = c), is a strong indication of the way the concept of equality can be taught correctly. This technique is omnipresent in the 2016–2017 curriculum but almost absent from the 2015–2016 curriculum. The technique of decomposition constitutes a major tool for add and subtract two- or three-digit numbers for the students in the experimental group, whereas the students in the control group are prompted to use the written algorithms for these computations.

Participants

The experiment involved 92 schools and 129 second-grade classes, which were divided approximately evenly between the experimental (64 classes) and control (65 classes) groups. All the classes contained approximately equal numbers of boys and girls. Only the students present at both the pre-test and post-test were included in our analyses. We also excluded students who were absent for more than 25% of the items on a test. For the few other children who did not complete all the items, we calculated scores by extrapolating from the results for the items they did complete. Consequently, our analyses are based on 1140 experimental group students ($M_{age} = 7.95$, $SD = 0.38$) and 1155 control group students ($M_{age} = 7.96$, $SD = 0.37$). We calculated ages at the time of the post-test, which was administered approximately 8 months after the pre-test.

We recruited schools in the north, west, southeast, and center of France. These regions were chosen simply for convenience. 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/schools. All the classes in the control group followed the 2015 French curriculum, under the supervision of their district inspector, and were a priori chosen to be comparable to the classes in the experimental group.

Procedure

The pre-test and post-test, each of which took 45 min, were administrated during the first and last months of the school year, respectively. The experimenters were graduate students or teaching advisers, never the teacher of the class being tested. Each experimenter alternated between testing experimental and control classes.

All the students in each class took the tests. The items were set out in a small booklet and read one after the other by the experimenter. We worked in conjunction with the teachers to ensure the booklet was as easy to follow as possible. Precautions were taken to prevent students copying and to avoid interference from the class's teacher. Any classroom illustrations capable of helping the students, such as addition and multiplication tables and graduated number lines, were taken down before the test.

Items

We begin by describing the post-test, which covered the four subdomains of arithmetic learning. The mental computation subdomain included 16 mental computation items. The experimenter read each problem aloud (e.g., 7 + 8, 62 – 10, 3 × 4, 30 ÷ 5) and then gave the

students 5 s to write their answers (e.g., 15, 52, 12, 6, for the examples) in their answer booklets before moving onto the next item.

The estimation subdomain section contained six number-line items. For each item, students were given 15 s to circle the mark on a horizontal line that most closely corresponded to the target, which was presented apart, to the left of the line. For example, for the target 68, the line had marks at 46, 54, 68, and 88. Separate lines were used for the six targets and each line ran from 0, on the left, to 100 on the right.

For the problem-solving subdomain, the experimenter read aloud eight verbal problems, allowing the students 30 s to solve each problem by mental computation and write the answer. A further three problems, presented in various forms, referred to quantities (weights) and had answer times of between 30 s and 3 min. The test also included a more complex problem that was presented in writing. Students were given 5 min to solve the problem, using a pencil and paper.

Four tasks were used to assess the arithmetic writing subdomain. The first task involved three items in which students had to insert the correct sign ($<$ or $>$) in a statement such as $200 + 70 + 5 \_\_ 200 + 40 + 5$ (15 s). In the second task, they had to complete two items with the form $866 - 128 = \_\_$, given $738 + 128 = 866$ (5 s to read the addition and 5 s to provide the answer). This task tested the "fundamentally important mathematical principle" of inversion (Verschaffel et al. 2012, p. 327). In the third task, students were given 1 min to generate as many equations equaling a target value of 21 as they could (maximum 5). Finally, they were given 30 s to complete the equation $7 \times 5 = 4 \times 5 + \_ \times \_$, with the help of a rectangle of 7×5 cells, of which 4×5 cells were shaded.

We also administered a pre-test, mostly in order to provide a relevant covariate (score out of 100). Because many arithmetical notions (e.g., multiplication and division) involved in the post-test are not taught in first grade, and in order to avoid a ceiling effect, only half of the items were the same as the items on the post-test.

Scoring method

Thirty of the 41 items were scored using a binary code (1 vs. 0). The 11 remaining items were assessed using either 5-point scales or, in the case of the four-choice number-line task, a more complex assessment method that provided a score based on how close an answer was to the correct choice. Then, as is standard practice for assessments in French schools, we applied a grading scale, which we drew up in conjunction with the teachers. This scale was designed to take into account both an item's "importance"—a simple mental computation carried out in 5 s had a lower weighting than generating a set of equations equating a target value or solving a written problem in 5 min—and to give a final score out of 100. Each of the four arithmetic subdomains were scored out of 20, as this is standard practice in French schools.

There were no outliers, mainly because scores were bounded—between 0 and 100 or 0 and 20. Consequently, we were able to include all the participants in our final sample. Point-biserial or Pearson correlation coefficients for each of the 41 items contributing to the total score were between .25 and .64, so we were also able to keep all the items in the analyses. Internal consistency of the results across items, measured with Cronbach's alpha, was 0.854.

Statistical analyses

As the students were nested within classes, multilevel modeling would appear to be the most appropriate analysis method, notably because it handles the problem of non-independence of

the students in a class. However, a factor can be significant even if small and unimportant, especially with large samples (Kreft and De Leeuw 1998). Therefore, rather than statistical significance, our main concern is estimation of the effect size of the program. Effect sizes or, more generally, comparisons of groups (see Zieffler et al. 2011) are often reported at the participants' level. Consequently, after a preliminary multilevel analysis, we relied mainly on descriptive statistics (M, SD, d) at this level. Only secondarily we used Student's parametric two-sample t test to compare two parameters of our populations. Welch's adaptation of this test t test is noted t^{W}.

A comparison of the pre-test scores for the two groups showed that the mean score for the control group ($M_C = 45.01$, $SD_C = 21.09$) was significantly higher than the mean score for the experimental group ($M_E = 41.95$, $SD_E = 21.14$): $t^{W}(2292.4) = 3.47$, $p = .0005$, 95% CI (1.33, 4.79). This is one of the reasons why we used adjusted (for pre-test scores) post-test scores in the comparisons of the two groups. A second reason was the strong correlation between the pre-test and post-test scores ($r_P = 0.722$): Adjusting the post-test scores leads to variance reduction and avoids the need to use Glass et al.'s (1981) correction for correlated data when estimating the effect size of the experimental teaching program. Consequently, we estimated the effect size by calculating a generic Cohen's d (Fritz et al. 2012) on the adjusted post-test scores, which we computed using the pooled standard deviation, calculated by weighting each of the two standard deviations according to the sample size.

Results

A preliminary multilevel analysis

We used R (R Core Team 2015) and *lme4* package (Bates et al. 2015) to perform a multilevel analysis of the influence of the scaled pre-test *zpre* score and *status* (experimental vs. control) on the scaled post-test score *zpost*. First, the null model, with only the class as random intercept (grouping factor), yields a .214 intra-class correlation. Thus, 21.4% of the variance in students' post-test score can be explained by their class membership. Adding *zpre* as fixed factor to the model improved considerably and significantly the fit ($\chi^2(1) = 1777.61$, $p < .001$). Allowing then random slope variation for the *zpre* factor improved slightly but significantly the model ($\chi^2(2) = 6.41$, $p = .041$). Further, adding *status* as fixed factor improved clearly and significantly the model ($\chi^2(1) = 37.00$, $p < .001$). However, adding the interaction between *zpre* and *status* to the latter model does not more improve it ($\chi^2(1) = 0.05$, $p = .82$). In addition, using the *nmle* package (Pinheiro et al. 2017) to compute the model showed a highly significant effect of the factor *status*, $t(127) = 6.51$, $p < .001$.

Analysis of the composite score at the students' level

Graphical examination of the data

Figure 2 suggests that the normality of the two post-test distributions is not strongly violated; Figs. 2 and 3 suggest that the data is not affected by any ceiling or floor effect. Furthermore, Fig. 3 shows that the regression lines of the composite post-test score on the pre-test score are almost parallel, which is a primary assumption for the computation of adjusted scores.
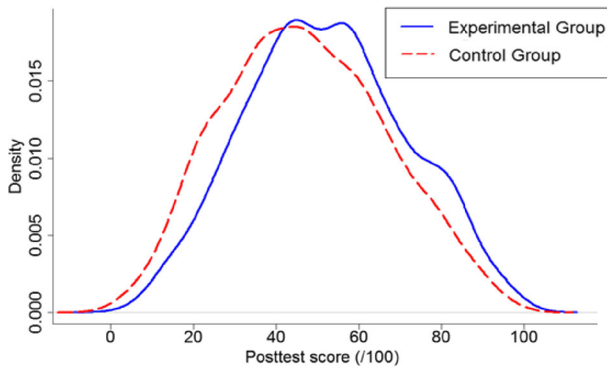
Fig. 2 Kernel density estimate of students' post-test scores as a function of group (densities were estimated using a smoothing parameter of 4.25)

Test for interaction

To statistically test the visual impressions generated in Fig. 3, including a possible interaction between pre-test and status (experimental vs. control), a multiple linear regression model was implemented. The model, with the pre-test, status, and their interaction as predictors, fits well and significantly the post-test scores, $F(3,2291) = 956$, $p < .001$, $R^2 = 0.556$. Moreover, the model confirms a highly significant effect both of pre-test and status: $t(2291) = 36.35$, $p < .001$, and $t(2291) = 6.90$, $p < .001$, respectively. But their interaction reaches not significance, $t(2291) = 1.17$, $p = .242$.

Adjusted and non-adjusted means, and effect size

We computed both simple individual post-test scores and individual post-test scores adjusted for pre-test scores. Table 1 shows the means for the two groups, differences, and standardized mean differences as measured by Cohen's *d*. The adjustment worked as expected. The control
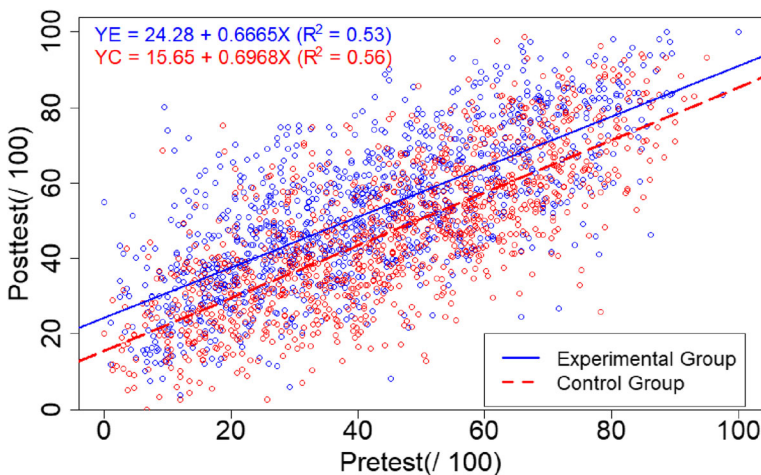

Fig. 3 Scatter diagram for composite scores on the pre-test and the post-test for the whole sample of 2295 students, with regression lines and equations for the two subgroups

Table 1 Comparison of non-adjusted and adjusted post-test means (*SD* in parentheses) for the experimental group with those for the control group, and effect sizes, *d*, of the experimental program computed using these statistics

Measure	Group E	Group C	E–C	*d*
Non-adjusted mean (/100)	52.24 (19.34)	47.01 (19.56)	5.23	0.269
Adjusted mean (/100)	53.29 (13.25)	45.97 (12.92)	7.32	0.559

group had higher scores on the pre-test than the experimental group, so adjusting the post-test scores reduced the mean for the control group, increased the mean for the experimental group, and reduced variance in both groups.

Mean composite scores for the whole sample were significantly higher for boys than they were for girls, both at pre-test (45.03 vs. 41.85, $t^W(2291.2) = 3.62$, $p < .001$, 95% CI [1.46, 4.91]) and post-test (51.30 vs. 47.81, $t^W(2289.0) = 4.29$, $p < .001$, 95% CI [1.90, 5.09]). But, importantly, effect sizes for boys ($d = 0.634$) and girls ($d = 0.474$) showed that the program was effective for both genders.

Analysis by subdomain

Adjusted mean scores (out of 20) and effect sizes for each arithmetic subdomain are shown in Table 2. As for Table 1, this table shows only descriptive statistics. The relative sizes of the experimental effects in each subdomain, as measured by *d*, reflect the experimental program's emphasis on arithmetic writing. Not only was the effect highest for arithmetic writing, it was significantly higher than the combined mean of the other three effects ($M = 0.355$), when compared using a unilateral, one-sample Student's *t* test: $t(2) = 6.22$, $p = .012$.

Analysis of long-term and cumulative effects

Because a substantial number of students who followed the experimental ACE program in second grade had already followed the program in first grade, we were able to divide our participants into four subgroups: ACE1&2 students, who followed the ACE program in both first and second grade ($n = 594$); ACE2 students, who followed the program only in second grade ($n = 544$); ACE1 students, who followed the program only in first grade ($n = 248$); and ACE0 students, who did not follow the program in either first or second grade ($n = 907$). Comparing all possible pairs of these subgroups allowed us to determine whether the ACE program had long-term and cumulative effects. The results from the six pairwise comparisons shown in Table 3 were established with simple t^W tests on composite post-test scores. We used the false discovery rate method, as implemented in R program (R Core Team 2015), to correct for multiple comparisons.

Table 2 Comparison of adjusted post-test means for the experimental group with those for the control group, and effect sizes, *d*, of the experimental program as a function of arithmetic subdomain

Domain	Mean E (/20)	Mean C (/20)	Difference	*d*
Mental computation	11.98	10.83	1.15	0.331
Problem-solving	8.42	7.14	1.28	0.346
Arithmetic writing	10.79	8.83	1.95	0.485
Number line	14.04	12.45	1.58	0.388

Table 3 Comparison of mean post-test composite scores as a function of the years following the ACE program. Mean differences and d values (given in parentheses) are shown above the empty diagonal; p values and false discovery rates (adjusted p values, given in parentheses) are shown below the empty diagonal

↱ minus	ACE1&2	ACE2	ACE1	ACE0
ACE1&2: in 1st and 2nd grade	–	Diff = 3.82 (d = 0.198)	Diff = 2.93 (d = 0.151)	Diff = 8.21 (d = 0.431)
ACE2: only in 2nd grade	p = .0009 (adj. = .0013)	–	Diff = −0.89 (d = −0.045)	Diff = 4.38 (d = 0.227)
ACE1: only in 1st grade	p = .0545 (adj. = .0654)	p = .5668 (adj. = .5668)	–	Diff = 5.28 (d = 0.271)
ACE0: never ACE	p < .0001 (adj. < .0001)	p < .0001 (adj. = .0001)	p = .0003 (adj. = .0007)	–

A post hoc observation

The experimental group students scored higher than the control group students on nearly all of the individual items. However, the problem-solving item "How many pens are there in 10 packs of 2 pens?" was an interesting exception. In this case, the control group students outperformed the experimental group students: 53 versus 42% correct answers, $\chi^2(1) = 29.73$, $p < .001$. The difficulty of the $2 \times 10 = 20$ computation does not explain this result because the experimental group was far better at mental computations involving multiplication by 10. For example, the item $\_ \times 4 = 40$ (read aloud: "How many times 4 equals 40?") was answered correctly by 62% of the experimental group students and by only 48% of the control group students ($\chi^2(1) = 45.12$, $p < .001$). The experimental group students also outperformed the control group students on the diagrammatic item $7 \times 5 = 4 \times 5 + \_ \times \_$, with 29% of the experimental group students providing the correct answer, compared with 23% of the control group students. This difference is significant, $\chi^2(1) = 9.57$, $p = .002$.

Discussion

The main aim of the present study was to determine whether it is possible to effectively teach a mathematically correct notion of equality in the first two grades of elementary school. We did this using a pre-test/post-test design to evaluate the impact of an experimental teaching program on arithmetic writing performance in second graders. Thus, the following discussion analyzes the effectiveness of the experimental program demonstrated by the results presented above. We then examine the important issue of the program's long-term and cumulative effect. Because the ACE program covered all arithmetic learning in first and second grade, we discuss its general effectiveness, based on the exploratory hypotheses that (1) sound learning of the notion of equality improves understanding in other subdomains of arithmetic learning, and that (2) other subdomains can also help improve arithmetic writing. We illustrate this hypothesized interaction between subdomains by briefly commenting on the post hoc observation noted in the

"Results" section. Finally, we note and discuss some limitations of our test of the efficacy of the ACE program.

The experimental program's effectiveness in teaching arithmetic writing

The experimental program's effectiveness in teaching arithmetic writing was moderate, as indicated by the effect size $d = 0.485$. Most importantly, because the program focused on arithmetic writing, its effect was greater in the arithmetic writing subdomain than in the other three subdomains.

This expected result supports the idea that it is possible to reduce many children's partial (and, in the long term, disadvantageous) interpretation of some mathematical writings. Although this result is encouraging, it cannot be considered conclusive, mainly because only 29% of the experimental group students answered the item $7 \times 5 = 4 \times 5 + \_ \times \_$ correctly, and only 35% used the writing $738 + 128 = 866$ to complete the equation $866 - 128 = \_$. Therefore, a fundamental question remains unanswered: What about the students who failed to answer these items correctly? Are they developing a deeper understanding of $a \times b$ or $a - b$, or have they simply replaced this more appropriate understanding with the partial signification that 7×5 means adding together 5 seven times and that $866 - 128$ means taking 128 from 866?

The literature does not include any results with which the effect size obtained for our experimental program can be compared directly. However, McNeil et al. (2015) showed that reversing the traditional format for presenting elementary addition problems (e.g., presenting $4 + 3 = \_$ as $\_ = 4 + 3$), organizing problems by equivalent sums, and replacing the equal sign by relational words, can lead to a better understanding of mathematics in second graders. The teaching component of McNeil et al.'s study had two important methodological features: (1) The experiment was of short duration; (2) interventions for both groups were through workbooks; therefore, children could be assigned randomly to the experimental and control conditions. McNeil et al. reported separate effect sizes for different types of item in their precisely targeted program. For example, for the equation encoding items (e.g., $2 + 3 + 6 = 2 + \_\_$), they obtained $d = 0.60$ at post-test and $d = 0.44$ at a follow-up test (5–6 months after the intervention). The $d = 0.485$ obtained for the arithmetic writing subdomain in the present study is of a similar order of magnitude.

In their quantitatively (in hours) substantial interventions, Fuchs et al. (2014) obtained a mean effect size of 0.465 for their calculation intervention compared to non-intervention (business-as-usual control). Because our study focused on arithmetic writing, the effect size of Fuchs et al.'s calculation intervention should be compared to the arithmetic writing subdomain (rather than to the mental computation subdomain). Again, the two effect sizes reported by Fuchs et al.—0.55 and 0.38—are of a similar order of magnitude to the effect size obtained for the arithmetic writing subdomain of our experimental program.

Because students in the ACE program regularly wrote in their personal notebook, we have qualitative evidence that children enjoy unusual writings. Appendix 2 (see electronic supplementary material) provides a sample of such writings. None of the writings were in the "operations = answer" format or were intended for computing a sum. Some of these writings have a long-term conceptual function in the ACE program. For example, writing a larger number as a series of repetitions of a smaller number (e.g., $23 = 5 + 5 + 5 + 5 + 3$) will help students later understand divisions with remainders. Other writings were mainly aimed at familiarizing the students with a simple method for writing more complex forms of an equality by adding a null sum. The null sum in these more complex equalities can be very apparent

(e.g., $6 = 6 + 0$; $6 = 6 + 8 − 8$), less apparent (e.g., $10 = 10 + 9−9 + 3 + 3−6$), difficult to detect (e.g., $50 = 100 − 40 − 10 − 20 + 40 − 20$), or very difficult to detect (e.g., $52 = 60 + 10 + 20 + 10 + 10 − 10 + 4 + 2 + 2 − 4 − 2 − 50$).

A long-term and cumulative effect

The comparison between the post-test performances of four subgroups of students in Table 3 clearly shows the program's long-term and cumulative effect. The pattern of results shows that the students who followed the ACE program in first grade still benefit from it in second grade, whether the ACE program is applied ($d = 0.43$) or not applied ($d = 0.27$) in this grade. It also shows that following the ACE program during 2 years accumulates its benefits ($d = 0.43$) compared with following the program only 1 year, whether this was in first ($d = 0.27$) or second grade ($d = 0.23$).

Thus, the ACE program overcomes, at least over a 1-year period, the fade-out problem associated with many early mathematics interventions (see Bailey et al. 2016, for a recent account). Resolving this problem is essential for any teaching program, as a program whose effect declines or disappears in the following school year has little, if any, value.

Moderate overall effectiveness of the experimental teaching program

According to Cohen's (1988) nomenclature, the effect size for the effectiveness of the experimental program, computed on the composite adjusted post-test scores, was medium ($d = 0.559$). To the best of our knowledge, this degree of effectiveness, measured by comparing scores covering all the main arithmetic subdomains studied during second grade, is unique in the psychological literature. In general, previous studies have found beneficial effects in some subdomains or tasks but not in others. For example, Fuchs et al. (2006) reported an effect size of 0.82 for improvement in addition, but effect sizes for improvement in subtraction and story problems were negative (although not significantly so).

A mutual influence of two subdomains

One of our hypotheses was that many very different learning activities can influence students' understanding of the "operations = answer" concept. The post hoc observation reported at the end of the "Results" section provides a good example of the local, negative influence of one learning activity on another. Probably because the experimental program emphasized an interpretation of $a \times b$ as the number of elements in a rectangular array of length a and width b (see Barmby et al. 2009), the experimental group students outperformed the control group students in the diagrammatic item $7 \times 5 = 4 \times 5 + \_ \times \_$. However, this interpretation of $a \times b$ is not appropriate for perceiving the multiplicative structure of the problem-solving item: "How many pens are there in 10 packs of 2 pens?" Consequently, the control group students outperformed the experimental group students on this item.

Limitations of the study

Our study had great ecological validity, but it also had weaknesses. First, the sample was not constituted completely randomly. Second, we used only a small number of items to directly test the notion of equivalence. However, by examining plots of the data to assess distributions and a major assumption—parallelism of the regression lines—of the score adjustment method, as

reported in the "Results" section, we were able to show that these weaknesses are not crippling (see Figs. 2 and 3). Third, we scored the tests using a scale that attributed different weights to different items as a function of their difficulty, length, and importance in the curriculum, as is standard practice when grading tests in French schools. This scoring system makes our tests difficult to replicate. In fact, the study's complexity means that it would be difficult, if not impossible, to replicate it exactly.

Given the importance of randomization for comparison between groups (Zieffler et al. 2011), we would note, about the first weakness, that a lot of our observations show that teacher self-selection could not be, alone, their explanation. For example, if self-selection led to having more effective teachers in the experimental group than in the control group, it would be difficult to explain why the students of the experimental group performed lower in solving mentally the 10 packs of 2 pens problem (see the post hoc observation at the end of the section "Results"). But the content of the ACE program can explain this finding (see the sub-section "A mutual influence of two subdomains"), which is a locally negative one for the ACE program.

Conclusion

The present study assessed the impact on second-grade students of implementing the ACE (Arithmetic Comprehension in Elementary school) program, a complete 1-year program for teaching arithmetic, which was first applied to first-grade students. Our objective was to address a major question in didactics: Is it possible to teach arithmetic in such a way that students do not misinterpret the notion of mathematical equality? The answer was positive, as the experimental group performed much better than the control group on the section of the post-test that assessed, at least indirectly, whether students were subject to this misinterpretation. The experimental teaching program had a larger effect in the equality writing subdomain ($d = 0.485$: see Table 2) than in the other subdomains. The effect of the ACE program in the equality writing subdomain was not obtained at the expense of learning in the other subdomains. Encouragingly, the effect of the ACE program was durable and cumulative. At the very least, the present results open the way for using the ACE approach to teach the mathematical notion of equality, first in second grade and then at higher school grades.

References

Alibali, M. W., Phillips, K. M. O., & Fischer, A. D. (2009). Learning new problem-solving strategies leads to changes in problem representation. *Cognitive Development, 24*(4), 89–101. https://doi.org/10.1080/15248372.2010.516417.

Bailey, D. H., Nguyen, T., Jenkin, J. M., Domina, T., Clements, D. H., & Sarama, J. S. (2016). Fadeout in an early mathematics intervention: constraining content or preexisting differences? *Developmental Psychology, 52*(9), 1457–1469. https://doi.org/10.1037/dev0000188.

Barmby, P., Harries, T., Higgins, S., & Suggate, J. (2009). The array representation and primary children's understanding and reasoning in multiplication. *Educational Studies in Mathematics, 70*(3), 217–241. https://doi.org/10.1007/s10649-008-9145-1.

Baroody, A. J. (1999). Children's relational knowledge of addition and subtraction. *Cognition and Instruction, 17*(2), 137–175. https://doi.org/10.1207/S1532690XCI170201.

Baroody, A. J., & Ginsburg, H. P. (1983). The effects of instruction on children's understanding of the "equals" sign. *The Elementary School Journal, 84*(2), 199–212.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01.

Brissiaud, R. (1994). Teaching and development: solving "missing addend" problems using subtraction. *European Journal of Psychology of Education, 9*(4), 343–365. https://doi.org/10.1007/BF03172907.

Byrd, C. E., McNeil, N. M., Chesney, D. L., & Matthews, P. G. (2015). A specific misconception of the equal sign acts as a barrier to children's learning of early algebra. *Learning and Individual Differences, 38*, 61–67. https://doi.org/10.1016/j.lindif.2015.01.001.

Byrd-Hornburg, C., Rieber, M. L., & McNeil, N. M. (2017). An integrative data analysis of gender differences in children's understanding of mathematical equivalence. *Journal of Experimental Child Psychology, 163*, 140–150. https://doi.org/10.1016/j.jecp.2017.06.002.

Chesney, D. L., McNeil, N. M., Matthews, P. G., Byrd, C. E., Petersen, L. A., Wheeler, M. C., Fyfe, E. R., & Dunwiddie, A. E. (2014). Organization matters: mental organization of addition knowledge relates to understanding math equivalence in symbolic form. *Cognitive Development, 30*, 30–46. https://doi.org/10.1016/j.cogdev.2014.01.001.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Taylor & Francis.

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

DeCaro, M. S. (2016). Inducing mental set constrains procedural flexibility and conceptual understanding in mathematics. *Memory & Cognition, 44*(7), 1138–1148. https://doi.org/10.3758/s13421-016-0614-y.

Dehaene, S. (2011). *The number sense* (2nd ed.). New York: Oxford University Press.

Dussuc, M. P., Charnay, R., & Madier, D. (2009). *Cap maths cycle 2, CE1: Nouveaux programmes*. Paris: Hatier.

Fischbein, E. (1989). Tacit models and mathematical reasoning. *For the Learning of Mathematics, 9*(2), 9–14.

Fischer, J. P., & Tazouti, Y. (2012). Unraveling the mystery of mirror writing in typically developing children. *Journal of Educational Psychology, 104*(1), 193–205. https://doi.org/10.1037/a0025735.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2–18. https://doi.org/10.1037/a0024338.

Fuchs, L. S., Fuchs, D., Hamlet, C. L., Powell, S. R., Capizzi, A. M., & Seethaler, P. M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *Journal of Learning Disabilities, 39*(5), 467–475. https://doi.org/10.1177/00222194060390050701.

Fuchs, L. S., Powell, S. R., Cirino, P. T., Schumacher, R. F., Marrin, S., Hamlett, C. L., Fuchs, D., Compton, D. L., & Changas, P. C. (2014). Does calculation or word-problem instruction provide a stronger route to prealgebraic knowledge? *Journal of Educational Psychology, 106*(4), 990–1006. https://doi.org/10.1037/a0036793.

Ginsburg, H. (1977). *Children's arithmetic*. New York: Van Nostrand.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage.

Hattikudur, S., & Alibali, M. W. (2010). Learning about the equal sign: does comparing with inequality symbols help? *Journal of Experimental Child Psychology, 107*(1), 15–30. https://doi.org/10.1016/j.jecp.2010.03.004.

Jara-Ettinger, J., Piantadosi, S., Spelke, E. S., Levy, R., & Gibson, E. (2017). Mastery of the logic of natural numbers is not the result of mastery of counting: evidence from late counters. *Developmental Science, 20*, e12459. https://doi.org/10.1111/desc.12459.

Joffredo-Le Brun, S., Morellato, M., Sensevy, G., & Quilio, S. (2018). Cooperative engineering as a joint action. *European Educational Research Journal, 17*(1), 187–208. https://doi.org/10.1177/1474904117690006.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics, 12*(3), 317–326.

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

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage Publications.

McNeil, N. M. (2007). U-shaped development in math: 7-year-olds outperform 9-year-olds on equivalence problems. *Developmental Psychology, 43*(3), 687–695. https://doi.org/10.1037/0012-1649.43.3.687.

McNeil, N. M., Rittle-Johnson, B., Hattikudur, S., & Petersen, L. A. (2010). Continuity in representation between children and adults: arithmetic knowledge hinders undergraduates' algebraic problem solving. *Journal of Cognition and Development, 11*(4), 437–457. https://doi.org/10.1080/15248372.2010.516421.

McNeil, N. M., Fyfe, E. R., Petersen, L. A., Dunwiddie, A. E., & Brletic-Shipley, H. (2011). Benefits of practicing 4 = 2 + 2: nontraditional problem formats facilitate children's understanding of mathematical equivalence. *Child Development, 82*(5), 1620–1633. https://doi.org/10.1111/j.1467-8624.2011.01622.x.

McNeil, N. M., Fyfe, E. R., & Dunwiddie, A. E. (2015). Arithmetic practice can be modified to promote understanding of mathematical equivalence. *Journal of Educational Psychology, 107*(2), 423–436. https://doi.org/10.1037/a0037687.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2017). *nlme: Linear and Nonlinear Mixed Effects Models* (R package version 3.1–131). Retrieved from https://cran.r-project.org/web/packages/nlme/nlme.pdf.

Powell, S. R. (2012). Equations and the equal sign in elementary mathematics textbooks. *The Elementary School Journal, 112*(4), 627–648.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Verschaffel, L., Bryant, P., & Torbeyns, J. (2012). Introduction. *Educational Studies in Mathematics, 79*(3), 327–334. https://doi.org/10.1007/s10649-012-9381-2.

Zieffler, A. S., Harring, J. R., & Long, J. D. (2011). *Comparing groups: randomization and bootstrap methods using R*. Hoboken: Wiley.

Jean-Paul Fischer. Lorraine Laboratory of Psychology and Neuroscience (2LPN), University of Lorraine, 54015 Nancy, France. E-mail: jean-paul.fischer@univ-lorraine.fr

Current themes of research:

Interactive development of procedural and declarative knowledge in the field of number and arithmetic. Mirror writing in typically developing children.

Most relevant publications in the field of Psychology of Education:

Fischer, J. P., (2017, on line). Studies on the written characters orientation and its influence on digit reversal by children. *Educational Psychology.* https://doi.org/10.1080/01443410.2017.1359239

Fischer J. P., 2017. Does finger sense really have a delayed effect on children's addition performance? *Cognitive Processing, 18(1),* 105-106. https://doi.org/10.1007/s10339-016-0782-5

Fischer, J. P. (2015). La distinction procédural/déclaratif appliquée aux premiers apprentissages numériques. In D. Butlen, I. Bloch, M. Bosch, C. Chambris, G. Cirade, S. Clivaz, S. Gobert, C. Hache, M. Hersant & C. Mangiante-Orsola (Eds), *Rôles et places de la didactique et des didacticiens des mathématiques dans la société et le système éducatif : vol. I* (pp. 97-122). Grenoble : La Pensée Sauvage.

Fischer, J. P. (2015). Commentary: Development of magnitude processing in children with developmental dyscalculia: space, time, and number. *Frontiers in Psychology (Developmental Psychology), 6*:804. https://doi.org/10.3389/fpsyg.2015.00804

Fischer, J. P., & Tazouti, Y. (2012). Unraveling the mystery of mirror writing in typically developing children. *Journal of Educational Psychology, 104(1),* 193-205. https://doi.org/10.1037/a0025735

Emmanuel Sander. Developmental psychology and educational sciences, IDEA Lab, University of Geneva. Email: emmanuel.sander@unige.ch

Current themes of research:

Arithmetical word problem-solving. Analogical reasoning.

Most relevant publications in the field of Psychology of Education:

Brissiaud, R., & Sander, E., (2010). Arithmetic word problem solving: a situation strategy first framework. *Developmental Science, 13(1)*, 92-107. https://doi.org/10.1111/j.1467-7687.2009.00866.x

Gamo, S., Nogry, S., & Sander, E. (2014). Apprendre à résoudre des problèmes en favorisant la construction d'une représentation alternative chez des élèves scolarisés en éducation prioritaire. *Psychologie française, 59(3),* https://doi.org/10.1016/j.psfr.2014.02.002.

Gamo, S., Sander, E., & Richard, J. F. (2010). Transfer of strategy use by semantic recoding in arithmetic problem solving. *Learning and Instruction, 20*(5), 400-410. https://doi.org/10.1016/j.learninstruc.2009.04.001

Hofstadter, D., & Sander, E. (2013). *Surfaces and essences: analogy as the fuel and fire of thinking.* Philadelphia: Basic Books.

Gérard Sensevy. Educational Sciences, CREAD Laboratory, University Bretagne Occidentale. E-mail: gerard.sensevy@espe-bretagne.fr

Current themes of research:

Didactics, teaching and learning, and their relationships.

Most relevant publications in the field of Psychology of Education:

Sensevy, G., Forest, D., Quilio, S. & Morales, G. (2013). Cooperative engineering as a specific design-based research. *ZDM - The International Journal on Mathematics Education, 45(7),* 1031-1043. https://doi.org/10.1007/s11858-013-0532-4

Sensevy, G., Mercier, A., Schubauer-Leoni, M.L., Ligozat, F. & Perrot, G. (2005). An attempt to model the teacher's action in mathematics. *Educational Studies in Mathematics, 59(1),* 153-181. https://doi.org/10.1007/s10649-005-5887-1

Sensevy, G., Quilio, S., & Mercier, A. (2015). Arithmetic and comprehension at elementary school. In X. Sun, B. Kaur, & J. Novotná (Eds.), *The twenty-third ICMI study: Primary mathematics study on whole numbers* (pp. 472-479). Macau (China): University of Macau.

Bruno Vilette. Developmental Psychology, PSITEC Laboratory, University Lille. E-mail: bruno.vilette@univ-lille3.f

Current themes of research:

Numerical cognition and development.

Most relevant publications in the field of Psychology of Education:

De Clercq-Quaegebeur, M., Rassel, A., Casalis, S., Vilette, B., Lemaitre, M.-P. & Vallée, L. (2017). Arithmetic abilities in children with developmental dyslexia : Performances on French ZAREKI-R Test. *Journal of Learning Disabilities.* https://doi.org/10.1177/0022219417690355

Vilette, B. (2016). Effets d'entraînements basés sur l'estimation numérique auprès d'enfants avec une dyscalculie ou des troubles du calcul. *Développements, 20-21,* 57-77.

Jean-François Richard. Paragraphe Laboratory, University Paris 8. E-mail: jf.richard.paris8@gmail.com

Current themes of research:

Modelling problem-solving behavior. Role of problem-solving activity in learning mathematics. Analysis of errors as a tool for cognitive diagnosis.

Most relevant publications in the field of Psychology of Education:

Richard, J. F., & Sander, E. (2000). Activités d'interprétation et de recherche de solution dans la résolution de problème. In J. N. Foulin & C. Ponce (Eds) *Lire, écrire, compter, apprendre : les apports de la psychologie des apprentissages*. Bordeaux : CRDP d'Aquitaine.

Gamo, S., Sander, E., & Richard, J. F. (2010). Transfer of strategy use by semantic recoding in arithmetic problem solving. *Learning and Instruction*, 20(5), 400-410. https://doi.org/10.1016/j.learninstruc.2009.04.001

Mengue-Metoule, E., Richard, J. F., & Sander, E. (2015). Classification des problèmes additifs : statut intermédiaire de la transformation relativement au complément et à la comparaison. *L'année psychologique*, 115(4), 497-531.