Stimulus Equivalence and Increase of Correct Responses in Addition and Subtraction Problems

Marcelo Henrique Oliveira Henklain
João dos Santos Carmo
Universidade Federal de São Carlos, São Carlos-SP, Brazil

Abstract: In order to propose interventions that increase correct responses, investigations search for properties of arithmetic problems that hinder its solution. This study assessed whether the formation of two sets of addition and subtraction equivalence classes reaches that goal. It also investigated the effect of training in balance-type problem solving and the teaching of algorithms. Eight second to fifth-grade elementary school students served as participants. After the formation of classes, all participants showed an average growth of 25% at post-test 1. Later, at the end of the training in balance-type problem solving, there was an average growth of 25% in post-test 2. In Generalization Test 1, all participants showed correct responses above 75%. After the teaching of algorithms, an average growth of 8% was detected in Post-test 3, and 100% correct answers in the reaplication of the test generalization. The procedures employed produced gains in solving arithmetic problems behavior.

Keywords: stimulus equivalence, problem solving, mathematics

A very common difficulty faced by many math teachers is the teaching of addition and subtraction problem solving, especially those with unknowns in the initial position, as well as written problems with semantic structures involving static relations (Bryant, 2011; Carpenter & Moser, 1983; Carpenter, Moser, & Bebout, 1988; Fayol, 1992; Geary, 1994; Haydu, Pullin, légas, & Costa, 2010; Neef, Nelles, Iwata, & Page, 2003; Nesher, Greeno, & Riley, 1982; Sá & Fossa, 2008; Verschaffel & De Corte, 1997).

Carpenter et al. (1988) investigated the effect of two learning histories on the response of depicting written additive problems with different semantic structures and different positions of the unknown value (a and b) through
operations with numerals (number-sentences). Twenty-two second graders and four third graders from an elementary school served as participants, and were distributed into two groups. Group 1 was exposed to all possible positions of the unknown, while Group 2 was exposed only to the c position. Then, they assessed the behaviors of (a) writing number-sentences to depict written problems with different semantic structures - transformation (positive and negative), combination, comparison and equaling -, and of (b) problems solving. Participants in Group 1 had more correct responses in depicting and problem solving tasks with different semantic structures and different unknown positions than did Group 2. These results suggest the existence of a correlation between the learning history and the development of the capacity to translate word-problems into number operation problems; additionally, the degree of accuracy for that translation seems to be related with the correct solution of the problem. These data indicate the relevance of the formation of an equivalence class between the presentation of different types of problems and the problem-solving behavior.

Aiming to evaluate the difficulties for each type of semantic structure and if there was a reduction of those difficulties along the elementary school years, Magina, Santana, Carzola and Campos (2010) sought to identify the strategies used to solve additive problems by first graders from an elementary school. The participants were 1021 students from second through fifth grade that were exposed to additive problems with combination, transformation or comparison structures. The data showed that, the higher the complexity of the problem, the lower the percentage of correct answers. The more complex problems were those with comparison structure, as well as those without semantic congruency between the keywords of the enunciation of the problem and the corresponding operation to solve it. The absence of keywords seemed to hamper the choice of the correct operation. That means that the students tend to identify the operation based on the type of words and not the comprehension of the problem.

Associated with the semantic structure, one of the variables that generate more difficulties in solving additive problems is the position of the unknown value (Fayol, 1992). Problems with the unknown in the c position are usually the easiest ones, while problems with the unknown in the a position are the hardest ones. One possible explanation for that difficulty seems to lie in the fact that the child cannot use a solution strategy that is guided by the actions described in the enunciation of the problems (Bryant, 2011; Haydu et al., 2010). A classic research on the unknown position is the work of Rosenthal and Resnick (1974), who investigated the effect of the order of presentation of the problem data, the position of the unknown, and the verbs “win” and “lose” on the problem solving performance. Two experiments using the same method were conducted with 63 fourth grade students. The problems were verbal and differed in three dimensions: (a) order of mention of the chronological events, that is, the situations were mentioned in chronological order (beginning, changing, end) and in reverse order (end, changing, beginning); (b) the position of the unknown, which could be in the final or in the initial situation; (c) the verb associated to the changing situation. The results showed that the problems in the reverse order generated more errors than the problems in the direct order, and problems with the unknown in the a position produced more errors and higher latencies than problems with c position. No significant effect of latency for the chronological order was detected. Rosenthal and Resnick (1974) concluded that the position of the unknown was the variable that most affected the participants’ performance.

Another classic work on the position of the unknown is Hiebert (1982), who examined the effect of the unknown position on the behavior of representing, through concrete materials, the elements of a problem and the strategies to solve additive problems presented verbally. Forty-seven second grade children were exposed to 36 additive problems with a semantic structure of combination. Some cubes were available so that the participants could manipulate them in order to solve the problems. The results showed that the position of the unknown was critical to the way children modeled a math problem, which, in turn, is related to the probability of success in solving the problem because: (a) there were more correct strategies and responses in problems frequently modeled with cubes; and (b) in most cases, if a problem were initially modeled, the probability of a correct answer was higher.

Later studies suggest that, besides the semantic structure and the position of the unknown, the type of presentation of the problem can affect the child’s performance. Capovilla, César, Capovilla and Haydu (1997) demonstrated that presenting additive problems in a virtual scale format can help children learn math concepts and skills, like the notions of balance, unknown value, and additive problems’ solutions. In an experiment that assessed the children’s responding to scale-type additive problems, the authors verified that the participants showed the best performance in a position problems and the worst performances in c position problems, which is in contrast with Hiebert (1982). The authors suggest that this discrepancy indicates that Hiebert’s data reflect the difficulties with written problems instead of with a specific position of the unknown.

In the same sense, Haydu et al. (2001) analyzed the effect of three forms of presenting additive problems on the problem solving performance. The forms were written, numeral operations, and scale problems, with the unknown in the three possible positions (a, b, and c), and the unknown values varying from zero to four. The participants were 86 third grade students with ages ranging from 7 to 11 years. Results showed that, in written and numeral problems, the best performances were in the c position while, in scale problems, the best performances were in b position. The worst overall performances were in written problems, with the unknown in a and b position, and performances in numerical and scale problems were similar. They concluded that the form of presenting a problem is a variable that affects performance.

The research by Iégas and Haydu (2002) followed those findings and investigated if training problem solving
in a scale type could improve performances on numeric and word problems (with semantic structure of transformation), regardless of whether the unknown was in the a, b or c position, or whether the problem was addition or subtraction. The participants were 48 second grade elementary school students between six and eight years old. The results indicated that, in addition problems, participants had greater difficulty in word problems with a and b positions. In subtraction problems, the lowest percentage of correct responses was found for problems with position a, regardless of the presentation form. The teaching procedure improved the participants’ performances, although a lower percentage of correct responses remained when the unknown value was in the a and b position, which suggests the training was not sufficient to make the various problems with unknown positions equal in terms of difficulty. Nevertheless, the data indicate that the graphic representation of addition problems can improve the problem solving performance.

A study conducted by Haydu, Costa and Pullin (2006) went a step further than Légas and Haydu (2002), offering another intervention possibility in the same kind of difficulty. They assessed whether the formation of an equivalence class, involving word, numeric and scale problems, would suffice to improve the problem solving performance. The participants were seven first grade students who were not exposed to teaching to solve arithmetic problems in school. The results showed that: (a) six out of seven participants presented an increase higher than 20% in the mean percentages of correct responses between pre and post-tests; (b) in word problems, the percentage of correct responses was lower than in numeric and scale problems; (c) all participants reached the correct response criterion for emergent relations; (d) six out of seven participants presented equivalence relations between different forms of problem presentation; (e) three participants presented a small reduction in the percentage of correct response in numeric problems with the unknown in position c.

Henklain and Carmo (2013) followed the same direction. The main difference in this study (when compared to Haydu et al., 2006) is that the authors attempted to assess whether the formation of an equivalence class involving four different types of presentation of addition problems (word problems with structure of transformation, numeral, set and scale problems) would improve the performance in addition and subtraction. The word problems had semantic structures of transformation, comparison and combination. Participants were eight second through eighth grade elementary school students, who presented difficulties in the pre-test of addition and subtraction word problems, with the unknown in b and c position. The results of post-test 1 showed an increase in the percentage of correct responses in all types of problems, with a mean difference of 16% between pre and post-test, which came closer to the 20% reported by Haydu et al. (2006). Five participants had difficulties with scale problems. Henklain and Carmo suggested this happened because the problems were accompanied by a written enunciation describing the problem-situation represented by the scale. Next, given the applied nature of the conducted study, it was assessed whether a procedure for explicit teaching of an algorithm to solve additive problems with the unknown in positions a and b could increase the percentages of correct responses even further. They conducted a session for teaching the algorithm of addition, followed by post-test 2, and a session for teaching the algorithm of subtraction, followed by post-test 3 and a generalization test. Four participants showed an increase in the percentage of correct responses in post-test 2, and six in post-test 3. Although positive, the results indicated a need for changes in the procedure. The participants reached 100% of correct responses on the generalization Test.

The present study followed the same sense as the earlier studies. The main objective was to assess whether the formation of two sets of equivalence classes could improve the problem solving performance. As this is an applied investigation, we adopted two other procedures to ensure the performance improvement. Both are reported in this article. Therefore, we also aimed to investigate the effects of training the solution of scale problems, as well as the teaching of algorithms for the same performance.

Method

Participants

Eight elementary school children, with ages ranging from seven to twelve years, seven females (FG, NV, AQ, MD, CB, LP e SP) and one male (GS). Two participants attended the fifth grade (LP and SP), four attended the third grade (AQ, GS, MD, and CB), and two the second grade (FG and NV). All participants were students at a municipal school in the city of São Carlos, São Paulo, Brazil. According to their parents, none of them used medication or had any permanent sensory, motor or intellectual limitation. The presence of difficulties and learning disabilities was not evaluated in the selected sample. The participants were experimentally naive with respect to research on teaching and learning mathematics, but they were part of the reading league (an education program for children with reading difficulties) and already had experience with matching to sample (MTS) tasks. In addition, they were not naive either in relation to solving one and two-digit addition and subtraction problems with the unknown in position c because they had already been exposed to that content in school. The criterion to participate was to be able to read simple sentences and present a success rate lower than 70% in the pretest.

Data collection was carried out in a room of the Library of the Federal University of São Carlos, where the extension project Reading League functions.

Instruments

The software ProgMTS (Marcicano, Carmo, & Prado, 2011) was used. Participants were exposed to three different types of presentation forms: scale problems (these stimuli will
be designated by the letter A); number problems (designated by the letter B), word problems with one of two possible semantic structures: transformation and comparison (both designated by the letter C). Besides the different types of problem presentation, participants were also exposed to three unknown positions: a, b and c. The word problems contained a question, and scale problems were never accompanied by written instructions.

For pre-test and post-test sessions, four tests were built, one for each problem presentation form and one additional test to assess whether the participants had already formed an equivalence class between the different problems' presentation form. The fourth test was not used to calculate the performance of the participants on the pre-test and post-test. Therefore, it was not adopted as a criterion to include participants in the research. Each test consisted of 18 problems (nine addition and nine subtraction), except for the word problem test and the test to evaluate the existence of an equivalence class. There were three problems for each position of the unknown.

The word problem test consisted of 36 problems, 18 of addition and 18 of subtraction. Of these 18, nine displayed a transformation semantic structure and nine a comparison structure. Of these nine, there were three problems for each position of the unknown. These two structures have been used only in the pre-test and post-test. In the training and test sessions, participants were exposed only to the transformation semantic structure. All problems had values and results between one and nine. The position of the comparison stimuli and the order of presentation of problems between sessions varied.

**Procedure**

**Data collection.** The phases of this research were: (a) initial instruction, (b) pre-test, (c) preparatory training, (d) teaching and testing of conditional discriminations, (e) post-test 1, (f) problem solving in the form of scales, (g) post-test 2, (h) generalization test 1, (i) teaching of addition and subtraction algorithms, (j) post-test 3 and (k) generalization test 2.

In the pre-test and post-test sessions, four tests were applied, one with word problems, another one with scale problems, one with number problems and the assessment of existence of an equivalence class, totaling 108 questions.

Before the pre-test, all participants watched a video that provided the initial instruction. This video showed one problem for each form of presentation, followed by a recording that indicated that the participant should find which number should be put in the place of the question mark to solve it. Therefore, during the video, the participant was asked to solve some problems with the same standard adopted for the problems in the pre-test and post-test. In all stages of the research, the test coordinator explained to the participants that they could solve the problems in the way they considered most convenient, like with paper and pencil (setting the calculation or by making risks to represent the values of the problem), with the use of fingers, etc. In this stage, no feedback was given as to right or wrong answers. At the end of the pre-test application of word, scale and number problems, before evaluating the presence of a class of equivalence between different forms of problem presentation, a video was presented to explain how the participants should undertake this task. It was explained that they should identify which comparison stimulus corresponded to the model stimulus. In addition to this instruction, after the end of the pre-test, a preparatory training for scale problem solving was offered, in which the participants were explicitly taught to balance the scales (they were instructed on how to solve and then solved some problems alone) and received feedback for trial and error. All videos and feedback were provided using the software ProgMTS.

To produce the formation of an equivalence class, conditional discriminations between three types of problem (A, B, C) were taught, one for each unknown position (A1, A2, A3, B1, B2, B3, C1, C2, C3); there were three tasks to teach (with feedback) and three probes (without feedback) for each problem. The criterion for advancing was 100% in the training sessions and 70% in the test. There were two sets of conditional discriminations: one for addition problems and another for subtraction problems. First, the relations of addition were taught and tested, followed by the subtraction. Two relations were taught for each set of classes: scale problem and numeral problem (AB); balance problem and the word problem (CA). Then, we tested the emergence of four new relationships through the tests of symmetry (BA and CA), transitivity (BC) and equivalence (CB). At the end of the training, post-test 1 was applied.

In the stage of solving scale problems, the criterion adopted to move forward was 100% of correct answers. Thus, it would be possible to determine if the scale model influenced the participants’ performance in other types of problems. At the end of this training, post-test 2 was applied.

In the algorithm teaching phase, only one session was held for joint teaching of addition and subtraction algorithms. The criterion for advancing was 100% of correct responses. The algorithms were taught to the participants through a verbal intervention that followed the same sequence, independently of the participant: identification of the type of operation, the location of the unknown and the identification of which operation should be carried out and with what values. The test coordinator asked questions to the participants, asking them to identify these problem properties. If the participants identified something incorrectly, the experimenter would correct them.

After the algorithm execution, the experimenter asked the participant to put the value obtained in the place of the question mark and calculate the algorithm to check if the result was correct. For example, if the account was “5 - ? = 2” and the participant got the value “3” after performing the algorithm, he should then replace the question mark with the value found, “3”, and verify if the account was correct: “five minus three equals two?” If the answer was yes, the value found was the correct response (and could replace the question mark); otherwise, the algorithm had to be reinitiated. After we finished teaching the algorithms, post-test 3 was applied.
Data analysis. To analyze the performance of the participants, the percentage of correct answers was calculated before and after the teaching procedures. Based on the average performance of the eight participants, we used two statistical tests, Wilcoxon and Friedman, with a significance level of 5%.

Ethical Considerations

This study was approved by the Ethics Committee on Human Research at the Federal University of São Carlos (no. 013/2012).

Results

Formation of Equivalence Classes

All participants reached percentages of accuracy above 88% in all probes and tests for symmetry, transitivity and equivalence, which indicates the formation of two sets of equivalence classes. It should be kept in mind that we found high percentages of correct responses in a test that assessed whether participants had already formed these two sets of classes even before the conditional discrimination sessions. AQ showed 78% accuracy in addition training classes and 61% in subtraction classes, SP 56% and 61%, CB 56% and 72%, GS 72% and 89%, NV 72% and 56%, MD presented 94% in both sets of classes; LP reached 89% and 100%; FG showed 94% at both. In the training class on addition problems, in the “scale problem – number problem” session (AB), NV and SP repeated the session twice to reach the criterion, AQ and FB repeated it once. Only GS had to repeat the symmetry test of this relationship. In teaching “scale problem – word problem” (AC), LP, SP and MD repeated the session thrice, CB twice and FB one. Again, only GS repeated the symmetry test of this relationship. In the training class on subtraction problems, when teaching the “scale problem – number problem” (AB), LP and GS repeated the session. In teaching the relation “scale problem – word problem” (AC), MD repeated the session twice and NV and SP once.

Participants’ Performance in Pre-Test and Post-Tests

Table 1 displays the participants’ percentage of correct response on each of the independent variables manipulated. In the scale problem training, SP and CB reached the criteria expected in eight sessions, NV and MD in six, GS and LP in five, AQ and FB in four. All participants reached the criteria for the algorithm teaching phase.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Tests</th>
<th>Operation</th>
<th>Form</th>
<th>Structure</th>
<th>Unknown</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Ad</td>
<td>Sub</td>
<td>NP</td>
<td>SP</td>
<td>WP</td>
</tr>
<tr>
<td>FG 2nd year</td>
<td>Pre</td>
<td>61</td>
<td>56</td>
<td>100</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Post1</td>
<td>92</td>
<td>75</td>
<td>83</td>
<td>67</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Post2</td>
<td>94</td>
<td>97</td>
<td>94</td>
<td>100</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Post3</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>NV 2nd year</td>
<td>Pre</td>
<td>22</td>
<td>25</td>
<td>22</td>
<td>11</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Post1</td>
<td>61</td>
<td>67</td>
<td>78</td>
<td>89</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Post2</td>
<td>94</td>
<td>89</td>
<td>94</td>
<td>83</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Post3</td>
<td>97</td>
<td>92</td>
<td>89</td>
<td>100</td>
<td>94</td>
</tr>
<tr>
<td>AQ 3rd year</td>
<td>Pre</td>
<td>56</td>
<td>28</td>
<td>67</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Post1</td>
<td>78</td>
<td>75</td>
<td>94</td>
<td>100</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Post2</td>
<td>92</td>
<td>89</td>
<td>78</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Post3</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>GS 3rd year</td>
<td>Pre</td>
<td>44</td>
<td>50</td>
<td>44</td>
<td>44</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Post1</td>
<td>56</td>
<td>47</td>
<td>78</td>
<td>56</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Post2</td>
<td>83</td>
<td>83</td>
<td>94</td>
<td>72</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Post3</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>CB 3rd year</td>
<td>Pre</td>
<td>44</td>
<td>58</td>
<td>83</td>
<td>39</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Post1</td>
<td>56</td>
<td>53</td>
<td>78</td>
<td>67</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Post2</td>
<td>78</td>
<td>89</td>
<td>100</td>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Post3</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>97</td>
</tr>
<tr>
<td>MD 3rd year</td>
<td>Pre</td>
<td>42</td>
<td>44</td>
<td>78</td>
<td>22</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Post1</td>
<td>78</td>
<td>83</td>
<td>83</td>
<td>94</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Post2</td>
<td>92</td>
<td>94</td>
<td>89</td>
<td>100</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Post3</td>
<td>100</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>97</td>
</tr>
</tbody>
</table>
We can see in Table 1 that all participants showed an increase in the percentage of correct responses between pre-test and post-test 3. FG, AQ and GS achieved 100% accuracy in the last post-test, CB, MD and LP 99% and NV and SP 94%. Noteworthy is the fact that participants with different educational levels and with different percentages of correct responses in the pre-test, reached a similar pattern of performance at the end of the study, in which virtually no more errors occurred, even in the face of problems that produced many errors early in the research.

AQ, SP and LP had greater difficulty with subtraction problems. FG started the research with greater difficulties in these problems and ended the study solving addition and subtraction problems correctly. NV, however, used to solve more subtraction problems correctly at baseline and changed this pattern at the end of the experiment. The remaining participants, MD, CB and GS, showed higher percentages of correct responses for subtraction problems than addition problems but, at the end of the research, the percentages of success achieved were similar in both types of problem.

With regard to the presentation form, it was found that word problems generated smaller percentages of correct answers. In problems with comparison structures, the percentages were lower than in problems with transformation structures. The highest percentages of correct problems were observed in the number problems. Problems with the unknown in position a also resulted in lower percentages of accuracy when compared to the problems with the unknown positions at b and c.

At the end of this study, these difficulties were overcome. This could be inferred from the fact that, when analyzing the percentage of correct responses in each variable, the lowest percentage was 88%, obtained by SP in trouble with the unknown in a. In all other types of problems, the percentages of correct responses were higher.

**Group’s Performance in the Pre-Test and Post-Tests**

Figure 1 displays the average percentage of correct answers of the eight participants on the pre-test and post-test, depending on the type of operation: addition or subtraction.

![Figure 1](image-url)
In terms of semantic structure (Figure 3), there was an unexpected proximity between the percentage of correct responses in problems with comparison and transformation structure, particularly in addition problems. In the post-test, this pattern persisted, with lower intensity in subtraction problems (with 85% accuracy in the transformation structure and 75% in the comparison). There was also an increase in the percentage of correct answers, which strengthens the argument that there was an improvement in problem solving behavior.

Figure 4 displays the average percentage of correct responses in the pre-test and post-test, depending on the problem presentation form: number, scale or writing.

In both applications of the generalization test, all participants showed a high percentage of correct answers. SP, MD and FG showed 100% accuracy for all test applications. AQ and LP showed 92% of correct responses in the first application and 100% in the second. GS achieved 83% accuracy in the first application and 100% in the second. NV reached 75% in the first application and 100% in the second. Only CB showed a reduction in the percentage of correct responses between the first and second test. In the first application, she showed 100% of correct responses, one with the unknown in position b and the other with the unknown in c; since she got the dictated subtraction problems with unknowns in b and c right and provided the same answer for the two problems that she missed, it can be assumed that these errors were due to distraction. The next day, the experimenter reapplied both of these problems, and, on this occasion, she got both right.

In the two-tailed Wilcoxon test, we found a statistically significant difference between pre-test and post-test 1 ($z = -2.521, p = .012, r = -.63$), post-tests 1 and 2 ($z = -2.524, p = .012, r = -.63$) and post-tests 2 and 3 ($z = -2.533, p = .011, r = -.63$). The Friedman test indicated a statistically
significant change in performance among the participants at the beginning and end of the experiment ($\chi^2 = 24, p = .001$).

**Discussion**

The procedures used in the present study rest on studies that demonstrated the reduction of difficulties in solving problems based on problem-solving training represented by a scale (Capovilla et al., 1997; Haydu et al., 2001; légas & Haydu, 2002) and on the formation of equivalence classes involving different forms of presenting addition problems (Haydu et al., 2006, 2010). Regarding the teaching of algorithms, in the research by Henklain and Carmo (2013), the rule was long and probably insufficiently clear. In the present investigation, based on the studies by Neef et al. (2003), we sought to simplify the rules by dividing them into small steps: identifying (a) the operation, (b) the position of the unknown, (c) the available values; and, based on the previous steps, (d) conducting the sum or the subtraction with the elements presented in the problem. The results of this research can be divided into two categories: assessment of problem-solving performance, and verification of the improvement of this performance as a function of the adopted teaching procedures.

**Participants’ Performances**

There was a lower concentration of correct responses around written subtraction problems with a semantic structure of comparison, especially when the unknown was in position a. These data sustain the investigations presented in the introduction and, once again, stress the prominence of understanding the causes of those difficulties and how to reduce them. Carpenter et al. (1988), Fayol (1992), Geary (1994), Magina et al. (2010), Nunes and Bryant (1996), Sá and Fossa (2008), and Verschaffel and De Corte (1997) suggest that children frequently solve word problems by reproducing the actions described in the problem’s enunciation. This strategy is appropriate when the problem entails transformation and the unknown lies in position c, because there is an initial value that undergoes a (positive or negative) change, and this leads to a correct final result. But this way of solving the problem does not work out when it entails static relations among two or more sets (semantic structures of comparison and combination) and, above all, when the unknown is in position b or c.

Regarding the unknown position, problems with the unknown in the initial positions are more difficult (Hiebert, 1982; Rosenthal & Resnick, 1974; Sá & Fossa, 2008). This general pattern was observed in the present study. However, when the performance of each participant is analyzed separately, one can observe a variation in this overall pattern because of the form of presenting the problem. According to Capovilla et al. (1997), Haydu et al. (2001, 2006) and légas and Haydu (2002), the position a offers more difficulties when the problem is a word-problem. However, it is not true that this data can be systematically replicated with other forms of problem presentation. Regarding the difficulty offered by each form of presentation, it is important to emphasize that the inclusion of a phase of preparatory training reduced the difficulties with scale problems, as identified in the work by Henklain and Carmo (2013).

**Effectiveness of the Teaching Procedures**

All participants presented an initial performance below 60% of correct responses, and the results showed there was an improvement at each post-test in relation to the previous one. Besides, all participants completed the experiment with a significant difference between pre-test and post-test 3. There was a general trend to approach the percentage of correct responses on the different types of problems, probably as an effect of the formation of equivalence classes.

Regarding the high percentage of correct responses on the fourth probe, which assessed whether the two sets of equivalence classes had been formed, it is possible that the same data would be obtained in the experiments by Henklain and Carmo (2013) and Haydu et al. (2006), because the teaching of some of those relations is usual in school routine, such as the translation of word problems to numeric problems for example (Carpenter et al., 1988). The conditional discrimination training is probably efficient because it helps the participant to control the relevant properties of the problem, which reduces the chances of errors. Furthermore, the equivalence class is also expanded as, for example, the scale becomes related to other presentation forms. This can help children, as the strategies used to solve that kind of problems can also be applied to other types of problems that are part of the class (Haydu et al., 2010).

The training in solving scale problems helped to improve the overall problem solving performance, supporting the data from légas and Haydu (2002). There was also an improvement on participants’ performance from the teaching of algorithms, indicating that the teaching of algorithms for sum and subtraction algorithms in a single session helped participants.

The experimental design did not provide clear information as to whether the teaching procedures with scales and the teaching of algorithms were, in fact, effective because the problem solving behaviors of presented in post-tests 2 and 3 were the product of the participants’ entire experimental history along the research. What one can assert is that those procedures seem to have affected the problem solving behavior and, generally speaking, contributed to increase the percentage of correct responses in both experiments. Considering the applied nature of this research, the procedures were useful because the objective was exactly a better performance.

It was clear that consistent training on the scale functioning and the formation of an equivalence class involving addition and subtraction problems could produce even better results than the ones by Henklain and Carmo (2013), leading the participants to a balanced performance,
with a higher percentage of correct responses, regardless of the specific features of the problem, and therefore reducing the difficulties to solve additive problems.

Data from the generalization test demonstrated that participants not only learned how to solve the trained problems, but also became able to solve new problems with values ranging from 10 to 15. Nevertheless, one should remember that this ability was strongly influenced by the school experience.

A statistically significant improvement occurred at each post-test in relation to previous tests, as indicated by the results of the Wilcoxon tests. The Friedman test showed that there was a statistically significant difference between the initial and the last experiment. The average gain was 44%. Strictly speaking, there was a general trend to approach the percentage of correct responses, so that, regardless of the features of the problems, participants were able to solve them correctly.

Conclusions

The main objective of this research was to evaluate whether the formation of equivalence classes between different forms of problem presentation could produce an increase in the percentage of correct answers in solving addition and subtraction problems; it was also evaluated if the teaching of algorithms and the training to solve scale problems could further improve this performance. The results suggest that the teaching programs employed were able to achieve this goal for all participants.

The difficulties produced by the form, semantic structure and position of the unknown should draw the attention of educators because these variables are easy for teachers to manipulate. They just have to change the order of the information presentation, the words or the type of problem situation. We need to consider that a change in the problem structure can demand different behaviors. That is, the student’s error can be the result of an ineffective teaching.

It is suggested that future studies continue mapping difficulties in solving problems and building strategies and resources, such as software, to reduce these difficulties. In parallel, we need to expand our understanding about the behaviors that have to be learned in order to solve addition and subtraction problems. This requires careful analysis of these behaviors and empirical research to demonstrate the validity of these analyses. It is considered important that new studies expand the sample and employ standardized instruments to better characterize the participants and to facilitate replication of the study and data generalization. These precautions help to assess the effects of teaching procedures on children’s performance more clearly.

References


Marcelo Henrique Oliveira Henklain is a M.S. in Psychology from Universidade Federal de São Carlos.

João dos Santos Carmo is a Professor at the Universidade Federal de São Carlos and a researcher at the National Institute of Science and Technology about Behavior, Cognition and Teaching (INCT-ECCE).

Received: Jul. 15th 2012
1st Revision: Jan. 24th 2013
2nd Revision: Apr. 15th 2013
Approved: May 1st 2013

How to cite this article: