CrossMark

An assessment of the sources of the reversal error through classic and new variables

Carlos Soneira[1] (iD) · José Antonio González-Calero[2] ·
David Arnau[3]

Abstract We present two empirical studies with 241 and 211 pre-service teachers that evaluate the explanatory power of word order matching and static comparison as models for the reversal error. We used tasks consisting of generating an algebraic equation representing a comparison given in a verbal statement. We introduce the types of magnitude involved in the statement as variables of analysis, something that was not previously tackled in previous works. Our results show that there are no statistical differences in the production of reversal errors depending on the information included in the name used to designate the variable, and that there are statistical differences depending on the syntactic configuration as well as the type of magnitude involved in the statement. The interpretation of these results indicates that both word order matching and static comparison have some potential as explanatory models for the reversal error, and that neither one of them, alone, is enough to completely explain the phenomenon.

Keywords Word problem solving · Algebra · Reversal error · Magnitude

1 Introduction and background

1.1 The reversal error

Several studies have addressed the difficulties students encounter when facing situations that require the use of algebraic language. Some of these difficulties, such as the polysemy

---

✉ Carlos Soneira
    carlos.soneira@udc.gal

[1] Departament of Pedagogy and Didactics, Educational Sciences Faculty, University of A Coruña,
     Campus de Elviña, 15071 A Coruña, Spain

[2] Department of Mathematics, School of Education of Albacete (Edificio Simón Abril), University of
     Castilla-La Mancha, Plaza de la Universidad, 3, 02071 Albacete, Spain

[3] Department of Didactics of Mathematics, University of Valencia, 46022 Valencia, Spain

of the letter x (Filloy, Rojano, & Solares, 2010) or the incorrect extrapolation of rules as a consequence of the overgeneralization of arithmetic properties (Matz, 1982), are typical of students getting started in the use of algebra. Other difficulties, such as the interpretation of the equals sign as a sign of comparison (Kieran, 1981), may persist even at university levels. Among this group of lingering errors, the reversal error occupies a relevant place. This error was widely described in the works of Clement (1982) and Clement, Lochhead, and Monk (1981). Specifically, these works show the results of a study with engineering students in which, among other tasks, the following was proposed: "Write an equation using the variables S and P to represent the following statement: 'There are six times as many students as professors at this university.' Use S for the number of students and P for the number of professors "(Clement et al., 1981, p. 288). Among the incorrect equations proposed, $P = 6 \cdot S$ stood out. Since P and S appear inverted with respect to the correct solution, the authors coined this as the reversal error.

1.2 Sources of the reversal error

In Clement et al. (1981), two possible causes, not necessarily exclusive, of the reversal error were identified. The first one, called *word order matching*, consists of a "literal, direct mapping of the words of English into the symbols of algebra" (p. 288). That is, the student performs a translation from natural language into algebraic language keeping the same order of the symbols in the equation as the key words in the statement, regardless of the meaning of the expressions. For example, in the previous problem, $6 \cdot S$ represents "There are six times as many students", the sign = represents "as", and P "professors," which would explain the answer $6 \cdot S = P$ (Clement, 1982).

In the second model, known as static comparison, the equation would represent the statement "one professor corresponds to 6 students" or "one professor for every six students," obtained from a mental representation of the situation.

> Apparently the expression '6S' is used to indicate the larger group and 'P' to indicate the smaller group. The letter S is not understood as a variable that represents the number of students but rather is treated like a label or unit attached to the number 6. The equals sign expresses a comparison or association, not a precise equivalence. (Clement et al., 1981, p. 288)

In this case, the student comprehends the correct ratio relationship but fails when translating it into algebraic language. Both models involve different cognitive processes. Under the static comparison model, subjects correctly identify the network of arithmetic relationships between quantities expressed by the text of the statement, that is, they build a correct problem model according to Kintsch (1998). However, in the word order matching model, the subject does not necessarily identify a problem model since he/she may build the equation taking into consideration exclusively syntactic considerations on the surface of the text.

1.3 Approaches in the study of the reversal error

Several authors studied the explanatory potential of word order matching and static comparison as models for the reversal error. One strategy used in the literature consists of the systematic modification of some task variables whose eventual impact on the incidence of the reversal error could be interpretable in terms of the explanatory power of the models.

For example, a comparison of the incidence of the reversal error between statements whose word-for-word translation leads solvers to a reversed equation, and those that do not, enables the explanatory effect of word order matching to be assessed. Similarly, the use of a multiplication or division when constructing the equation (Fisher, Borchert, & Bassok, 2011; Landy, Brookes, & Smount, 2014), the presence of contextual cues in the statement (Cohen & Kanim, 2005; González-Calero, Arnau, & Laserna-Belenguer, 2015; Wollman, 1983), or the multiplicative or additive nature of comparisons (González-Calero et al., 2015; MacGregor & Stacey, 1993) are some of the variables employed. However, dissimilar results make it impossible to achieve a consensus concerning the explanatory power of both models.

The present work focuses on studying word order matching and static comparison as explanatory models of the reversal error by means of three variables: notation, syntactic structure of verbal statements, and type of magnitude. Next, we present the reasons that lead us to consider these variables as useful for this purpose.

1.4 The influence of notation

The static comparison model assumes that letters are used as labels to designate objects, rather than to indicate the cardinal of a set. Related to this, Rosnick (1981) designed an experiment with undergraduate students that aimed at determining the students' interpretation of letters in the *Students and Professors* problem. Participants were given the statement: "At this university, there are six times as many students as professors. This fact is represented by the equation $S = 6P$ [...] In this equation, what does the letter P stand for?" (p. 419). The answer was to be selected from the following items: (1) Professors, (2) Teacher, (3) Number of professors, (4) None of the above, (5) More than one of the above (if so, indicate which ones), and (6) Do not know. More than 40% of the participants were not able to identify the letter P with the "number of professors." As a conclusion, Rosnick (1981) hypothesized that most of students who believe that P represents "professors" will also think that the correct equation is 6 students = 1 professor ($6S = P$). As a didactic implication, Rosnick (1981) suggested teachers should always write "P = number of professors" instead of "P = professors". The underlying idea behind this recommendation is that, if static comparison were the main source of the reversal error, this error should occur more frequently when the word professors is provided, than the proposition number of professors. Following this line of thought, Fisher (1988) carried out an experiment in which the letters Ns and Np were used, since "Ns is more likely to be read as number of students and Np as number of professors than the corresponding notations S and P" (p. 260). However, the participants who employed the notation Ns-Np made a greater number of reversal errors. As an explanation, Fisher (1988) suggested that this fact could be due to the greater complexity of the notation Ns-Np. Given the aforementioned facts, new studies seem to be necessary in order to test Rosnick's (1981) hypotheses on the influence of the different types of names given to designate quantities.

1.5 The influence of sentence structure

Due to the flexibility of natural language, different wording of a statement can express the same network of relationships between quantities, leading, consequently, to the same problem model. Several studies have found that the syntactic structure of statements affects the understanding of problems and the rate of errors. For example, Lewis and Mayer (1987)

addressed this issue with arithmetic problems, finding that the rate of reversal errors varied according to the ordering of the quantities in the statements.

The analysis of the influence of the syntactic structure of the statement can also be extended to those that describe multiplicative (or additive) algebraic comparisons. Related to this, we distinguish between *statements with syntactic obstruction* and *statements without syntactic obstruction*. We employ the term *statement with syntactic obstruction* to refer to those statements where a reversal error would take place when a word-for-word translation is applied. An example of this would be the version of the *Students and professors* problem: "There are six times as many students as professors at this university". By contrast, the alternative wording "The number of students is six times greater than the number of professors at this university" would be classified as a *statement without syntactic obstruction* because a process of word order matching would lead to a correct equation. A comparison of incidences of the reversal error between statements with and without syntactic obstruction would be useful to study the explanatory power of word order matching. This would be the underlying argument in Cohen and Kanim (2005) and MacGregor and Stacey (1993). In fact, if word order matching were not a relevant cause of the reversal error, there should be no difference in the incidence of the reversal error based on the syntactic obstruction of the statement. The results obtained to date are not conclusive. MacGregor and Stacey (1993) found no evidence of a relationship between the order of quantities in the equations constructed by the participants and the order of appearance of the quantities in the statement. However, Cohen and Kanim (2005) obtained a greater number of reversal errors for statements with syntactic obstruction.

1.6 The influence of the type of magnitude

As we mentioned above, the literature has considered several variables to analyze the explanatory power of the two models of the reversal error. However, we believe that there is at least one relevant variable that has not been systematically considered so far. In particular, most of the works aimed at studying the reversal error only employ items with discrete magnitudes. Although some studies have considered items with continuous magnitudes (e.g., Landy et al., 2014; MacGregor & Stacey, 1993; Rosnick & Clement, 1980; Sung-Hee et al., 2014; Wollman, 1983), the eventual effect of this variable has not been analyzed.

However, from our point of view, the analysis of the type of magnitude may be of interest when studying the explanatory power of static comparison, since this model entails the establishment of a correspondence between two sets of different sizes that reflect a comprehension of the relative size between sets. In discrete situations, solvers can easily construct a mental image associated with a problem model, e.g., a classroom with a professor and six students. This mental image may result in a correspondence between sets that leads to establishing the ratio 1 professor per 6 students, that in turn can lead to $1P = 6 S$ due to the application of the static comparison approach (Clement, 1982). When the magnitudes are continuous, the construction of these situation models could be hindered, because the mental image requires the representation of a single element per set (e.g., quantity of water/lemon juice).

Nevertheless, the subject could discretize the situation. This process is illustrated by the following example: "In a lemonade drink there is three times as much water as lemon juice." The subject could model this situation as if the magnitudes were discrete, interpreting them as "parts of lemonade" or as "the cardinal of glasses of water or lemon juice used to make the lemonade." This discretization process allows the solver to reinterpret the statement as: "In a lemonade drink there are three parts of water for each part of lemon juice", or "In this

lemonade drink there are three times as many glasses of water as glasses of lemon juice." This process of measurement transforms the original problem into a new situation with discrete magnitudes. After this variation of the original problem, the subject could apply a static comparison in a strictly analogous way that in the *Students and Professors* problem. In both situations, the solver can now build a mental representation of the situation in which two sets of different cardinal are compared.

Nonetheless, the phenomenon of discretization can be conditioned by the students' own way of thinking, ability, or even be favored by certain continuous magnitudes. Related to this, we consider certain extensive continuous magnitudes (e.g., volume or weight) to be more prone to discretization, since they can be more easily represented by images. In fact, we observe this phenomenon of discretization with precisely these magnitudes in the case study reported by Wollman (1983).

The possibility of discretization leads us to extend the analysis to intensive magnitudes derived from continuous magnitudes (e.g., acceleration or density). In this case, the quantity of magnitude does not depend on the amount of the substance it is applied to, but refers to a quality of this substance defined as the ratio between two different magnitudes. This prevents an immediate visual representation of a unit of measurement corresponding to intensive quantities. Indeed, according to Schwartz (1988), since intensive quantities are based on a relationship, their representation is difficult. For these reasons, discretization phenomena with intensive magnitudes may be considered less likely. Consequently, situations involving these magnitudes would be less likely to trigger static comparison processes. In short, if the static comparison were a relevant cause of the reversal error, such an error should appear with a higher incidence when the quantities compared are extensive discrete, compared to when they are continuous intensive. At the same time, since the discretization of extensive continuous magnitudes may or may not occur, if the reversal error were strongly linked to static comparison, the incidence of this error should be lower with these types of magnitude than with discrete magnitudes and higher than with intensive magnitudes derived from continuous ones.

2 The present work

In this article, we offer results from two empirical studies whose general aim is to determine the explanatory power of word order matching and static comparison as models for the reversal error. In other words, we intend to assess to what extent each model can explain the origin of this error. To this end, we follow the aforementioned strategy consisting of analyzing whether the incidence of the reversal error is affected by variations of some task variables.

Specifically, we approach this main objective by examining three research questions through two complementary experiments. In experiment 1, we study whether there are differences in the incidence of reversal error depending on the verbal expressions (e.g., professors versus number of professors) employed to represent the quantities involved in the equation (RQ1). Thus, we wonder if the type of information conveyed by these expressions, which hereinafter we call names, influences the commission of the reversal error. Specifically, RQ1 makes it possible to evaluate the static comparison model since, according to Rosnick (1981) and Fisher (1988), differences in the production of reversal errors due to the use of different notations can only be explained by this model. In experiment 1, with the aim of reducing the ambiguity that the use of letters may cause regarding the students' interpretation of variables, we introduce the novel idea of having participants denominate the variables by

means of names in natural language (e.g., professors and number of professors), instead of letters (e.g., "*P*" and "*Np*").

experiment 2 aimed at analyzing the effect on the reversal error of two factors: the presence of syntactic obstruction in the statement (RQ2) and the type of magnitude of the quantities used in the comparison (RQ3).

In relation to RQ2, although previous studies have already considered the influence of sentence structure to assess the explanatory power of word order matching (e.g., Cohen & Kanim, 2005; MacGregor & Stacey, 1993), in experiment 2, this issue is studied more in deep, given the fact that a uniform syntactic structure is employed in all the items. This novel idea avoids the interference of different students' interpretations of variables due to the use of different structures.

Lastly, RQ3 takes into consideration the influence that different magnitudes (extensive discrete, e.g., number of professors; extensive continuous, e.g., volume; or intensive continuous, e.g., density) may have. This research question, inspired by Schwartz (1988) and a transcription from Wollman (1983), enables a specific assessment of the static comparison model and entails a novel contribution concerning the study of reversal error.

With the aim of properly answering the research questions, different instruments need to be used, since the semantic structure of items used to answer RQ1 is not appropriate in the study of RQ2 and RQ3. Thus, an experimental design with only one sample of participants to deal with all the research questions would imply that each participant would have to complete a large number of tasks. This fact may lead to the participants eventually feeling tired during the experiment, or to a lack of engagement with the tasks. Accordingly, we opted to structure the study in two experiments.

3 General method

Within a quantitative approach, we conducted two experiments for which we constructed two questionnaires with verbal statements expressing comparative relationships. The tasks consisted of producing the corresponding algebraic equations. These tests were administered by using a computer application designed ad hoc, which is based on the software used in González-Calero et al. (2015). The computer application offers the statement and a list of buttons with the names expressed in natural language that can be used in the task. The user may introduce mathematical expressions by using the buttons for each quantity, along with buttons with the equals sign and the arithmetic operations. While constructing the equation, the user may always modify or delete it. When the expression is accepted by the user, it is stored in a database and a new task is loaded. In order to avoid possible bias due to the training effect, or the order in which the names were offered, the application loaded the tasks and ordered the labels randomly. The answers were codified as correct, reversal error, and another type of error (in the case of "Students and professors" an example of each one would be STUDENTS = PROFESSORS·6, 6·STUDENTS = PROFESSORS, and PROFESSORS·STUDENTS = 6, respectively).

As we were focused on the reversal error, participants' answers were scored as 1 in case of reversal error and 0 otherwise. For each research question, each student was assigned a score that represents the mean of the number of reversal errors committed in those items related to the corresponding research question. Thus, the ranges of the reversal error scores on both experiments are standardized to [0,1] regardless of the number of items considered in each question.

Concerning the statistical analysis, it should be noted that when assumptions for parametric test were violated, bootstrap methods were employed instead of traditional non-parametric tests. This decision relies on the fact that bootstrap methods commonly offer more accurate inferences and statistical power in situations of unequal variances and non-normality (Wilcox, 2005). In addition, results were reported with confidence intervals along with p values in order to give an estimate of effect size and confidence in results (Maxwell & Delaney, 2004). A significance level of 0.05 was adopted.

4 Experiment 1

Experiment 1 is a between-group study aimed at analyzing the influence of the type of name for quantities on the incidence of the reversal error (RQ1). A group of students had to build the equations using names for quantities that referred only to the objects, hereinafter object names (e.g., professors). The other group had to use names that made reference to quantities of magnitude, hereinafter quantity name (e.g., number of professors). As the null hypothesis, H01_E1, we assumed that there were no statistically significant differences in the incidence of the reversal error depending on the type of name. According to the considerations of Rosnick (1981), the application of a static comparison should be less likely when using quantity rather than object names. Consequently, if static comparison were the single cause (or at least the most important cause) of the reversal error, the results should prompt us to reject H01_E1.

4.1 Participants

Participants were 241 pre-service teachers in the second year of a Bachelor degree in education at a Spanish university. They belonged to seven classes taking a subject aimed at providing them with a sufficient level of mathematics to qualify them to teach in primary education. Algebra was included as a topic in this course. In addition, one of the objectives of the compulsory mathematics subjects the participants studied during their secondary education was to qualify them to solve these kinds of tasks.

4.2 Instrument

Studies concerning the reversal error have been carried out mainly with tasks written in English. Hence, the predominant way of expressing multiplicative comparisons, see for example Clements et al. (1981), is "N times as many X as Y" (e.g., "there are six times as many students as professors"). The most literal translation into Spanish would be "hay N veces tantos X como Y" (e.g., "hay seis veces tantos estudiantes como profesores"). However, in Spanish, the most usual way to express multiplicative comparisons between discrete magnitudes employs the propositions "N veces más que" or "N veces menos que" (Castro, 1995). In other languages such as Arabic, Czech, or Hebrew, something similar happens (Nesher, Hershkovitz, & Novotna, 2003).

We employed a questionnaire with eight comparison statements. Items were constructed taking into account three dichotomous variables: type of comparison (additive-multiplicative), direction of comparison (increasing, decreasing), and inclusion, or not, of contextual information allowing the solvers to anticipate which of the compared quantities is greater. In the experiment, the effect of these variables was not analyzed but employed to generate a greater number of items with different characteristics for the purpose of avoiding any result being

linked to a specific kind of statement. An analysis of these variables can be found in González-Calero et al. (2015). In this questionnaire, we follow the criterion of expressing comparisons in the most usual way in Spanish. Consequently, the syntagmas N veces más que (with a direct translation in English, "N times more than") or "N veces menos que" ("N times less than") were used for the multiplicative comparisons, while for the additive ones, we used "más que" ("more than") and "menos que" ("less/ fewer than"). By way of example, the following statements are two word-for-word translation into English of one multiplicative and one additive item from the questionnaire: *There are five times fewer doctors than patients in this hospital* (*Hay cinco veces menos médicos que pacientes en este hospital*) and *There are six men more than women in a cinema* (*Hay seis hombres más que mujeres en un cine*).

4.3 Procedure

The experiment took place in the computer classrooms of the schools that the students attended. Each participant was randomly assigned a computer equipped with the program, which provided each subject items with either object or quantity names. In the end, 121 and 120 participants worked with object names and quantity names, respectively. At the beginning of the experiment, one of the authors showed the students how the computer application worked by setting up an equation to solve the item: "Write an equation using "Z", "Y," and "3" to represent the following statement: "Z is equal to 3 plus Y". Then, the application provided the items from the questionnaire following the general procedure described earlier.

4.4 Results

The means of reversal errors were 0.28 ($SD = 0.24$) in the group that used object names and 0.27 ($SD = 0.27$) in the group that employed quantity names. The Shapiro-Wilk tests for both groups ($W = 0.91$, $p < .001$ and $W = 0.87$, $p < .001$), in conjunction with Q-Q plots, indicated that both groups were significantly non-normal. Consequently, we used a bootstrap t-test (Wilcox, 2005), whose results did not allow us to reject the null hypothesis H01_E1 ($Yt = -1.13$, (-0.14, 0.04), $p = .2570$). Although at a descriptive level, a slightly higher incidence of reversal error is observed in the object name group, we cannot conclude that there were significant differences in the incidence of reversal errors depending on the type of name provided to construct the equation. Hence, the results of experiment 1 do not allow us to conclude that *static comparison* is a fundamental source of the reversal error or, at least, we cannot assert that the use of object names triggers different cognitive processes compared to quantity names.

5 Experiment 2

Experiment 2 is designed to study the effect of syntactical obstruction and the type of magnitude on the incidence of reversal errors. In order to design an instrument with a reasonable number of items, we restrict ourselves to multiplicative comparisons with contextual clues and increasing direction. We test two null hypotheses: there are no differences in the incidence of the reversal error between statements with and without syntactic obstruction (H01_E2), and there are no differences in the incidence of the reversal error depending on the type of magnitude (extensive discrete, extensive continuous or intensive continuous) (H02_E2). Moreover, we carry out post hoc comparisons in order to answer RQ3. As argued

in previous sections, if word order matching were not a relevant source of reversal errors, the results should prompt us to retain H01_E2. Moreover, if static comparison were not a relevant cause, the results should indicate that we should retain H02_E2.

5.1 Participants

The sample consists of 211 undergraduate students in the second year of a Bachelor degree in education at two Spanish universities. The group of students in experiment 2 was different from the group in experiment 1, although they were taking the same subject and their educational backgrounds were completely analogous.

5.2 Instrument

In experiment 1, we followed the criterion of using the most usual syntactic structure in Spanish when comparing discrete magnitudes. This implies that all items used in experiment 1 had syntactic obstruction. Hence, we designed a new instrument because the study of RQ2 and RQ3 requires us to vary the syntactic organization of the statements as well as the type of magnitude, resorting to forms of writing slightly less frequently used. Specifically, in experiment 2, we put together a questionnaire with 12 items with a syntactic configuration that allows us, regardless of the type of magnitude, to write items either with or without syntactic obstruction (Table 1). Evidently, these forms are also grammatically correct and used in daily life, although less frequently than those in experiment 1.

To avoid confounding effects from other variables, we employed a uniform syntactic and semantic structure regardless of the presence of syntactic obstructions or the type of magnitude. Moreover, in the case of discrete magnitudes, we used a collective name because otherwise, the comparison cannot be correctly formed in Spanish when writing items without syntactic obstruction. Specifically, we use the structure:

[Name of the magnitude/collective name] + [of] + [object name].

Moreover, the results of experiment 1 allow us to reduce the number of items used in experiment 2, since no differences in the incidence of reversal error were observed depending on the names given to the variables when posing the equation. Nevertheless, we adopted a conservative position and offered students names related to magnitude (e.g., "volume of water") instead of names associated with objects (e.g., "water") to construct the equations.

5.3 Procedure

A within-subject design was carried out. Regarding the gathering of data, the procedure was analogous to that of experiment 1. Additionally, in order to avoid type I errors, we applied a Holm-Bonferroni correction, and only the corrected p values are reported.

5.4 Results

In order to analyze the role of syntactic obstruction, we computed two scores for each student, one calculating the mean of reversal errors in statements with syntactic obstruction ($M = 0.39$, $SD = 0.02$) and the other the items without syntactic obstruction ($M = 0.34$, $SD = 0.03$).

Table 1 Examples of items used in experiment 2 in Spanish and in a word-for-word translation into English

MT	SO	Item
Intensive continuos	Yes	Según el ingeniero, es nueve veces mayor la aceleración de la motocicleta que la de la bicicleta.
		According to the engineer, [it][a] is nine times higher the acceleration_of the motorbike than that of the bicycle.
	No	Según el ingeniero, la aceleración del ciclomotor es cuatro veces mayor que la del monopatín.
		According to the engineer, the acceleration of the scooter is four times greater than that of the skateboard.
Extensive discrete	Yes	En esta universidad, es seis veces mayor el colectivo de estudiantes que el de profesores.
		At this university, [it][a] is six times greater the community of students than that of professors.
	No	En este hospital, el colectivo de pacientes es cinco veces mayor que el de médicos.
		At this hospital, the community of patients is five times greater than that of doctors.
Discretizable extensive continue	Yes	En esta limonada, es cuatro veces mayor el volumen de agua que el de limón.
		In this lemonade drink, [it][a] is four times greater the volume of water that that of lemon juice.
	No	En este cóctel, el volumen de tónica es cinco veces mayor que el de ginebra.
		In this cocktail, the volume of tonic is five times greater than that of gin.

The grammatical freedom in relation to the order of words is higher in Spanish than in English. Thus, in this case, in Spanish, the word *it* does not need to appear and the clause "the acceleration of the motorbike" can act as the subject without being placed before the verb. Hence, in Spanish, both the first and second statements in Table 1 are grammatically correct and differ exclusively in the presence of syntactic obstruction. We are aware that the word-for-word translation is not grammatically correct in English. A less literal translation with the same ordering of syntagmas would be "According to the engineer, there is a nine times higher acceleration of the motorbike than that of the bicycle." The same grammatical considerations may be applied to the rest of items with syntactic obstruction

MT magnitude type, *SO* syntactic obstruction

Using Shapiro-Wilk's tests, we rejected normality assumption in the difference variable, so we used a bootstrapped robust method for comparing dependent measures (Wilcox, 2005). The test showed statistical differences when comparing the number of reversal errors between statements with, and without, syntactic obstruction ($Yt = 0.10$ (0.06, 0.14), $p < .001$, $r = .20$), so we rejected the null hypothesis H01_E2.

Secondly, in order to test H02_E2, we assigned three scores to each student, computed respectively as the mean of reversal errors made with items containing extensive discrete magnitudes ($M = 0.41$, $SD = 0.02$), extensive continuous ones ($M = 0.36$, $SD = 0.03$), and intensive ones ($M = 0.33$, $SD = 0.03$). Shapiro-Wilk and Mauchly's tests showed that the conditions of normality and homocestacity required for a one-way repeated measure ANOVA were not met. Hence, we conducted a robust bootstrapped ANOVA (Wilcox, 2005) that showed statistical differences in the incidence of reversal errors depending on the magnitude type ($F_t = 3.67$, $p = .0300$). Consequently, we rejected the null hypothesis H02_E2 which stated that statistical differences should not exist depending on the type of magnitude involved in the statement. Post hoc comparisons were used to delve into the effect of each type of magnitude. Again, as the Shapiro-Wilk tests ($p < .001$ in all comparisons) showed that the assumption of normal distributions was not met, we opted for a robust analysis in all the pairwise comparisons. Results showed that there were statistical differences in all possible comparisons between magnitude types. Specifically, we got the values ($Yt = 0.15$, (0.09, 0.21),

$p < .001$, $r = .22$) between the discrete and intensive ones, ($Yt = 0.09$, (0.03, 0.15), $p = .0177$, $r = .13$) between discrete and extensive continuous ones, and ($Yt = 0.06$, (0.01, 0.11), $p = .0300$, $r = .11$) between intensive and extensive continuous magnitudes.

In summary, the results of experiment 2 indicate that, when translating a multiplicative comparison from natural to algebraic language, the syntactic organization of the verbal statement and the type of magnitude involved are relevant factors in the occurrence of reversal errors. Indeed, the differences in the incidence of reversal error depending on the syntactic obstruction can be explained from the word order matching procedure, and those depending on the type of magnitude, from the static comparison model.

6 General discussion

In this section. we present the conclusions from both experiments in a more detailed and comprehensive way. In addition, we offer plausible explanations for the results and indicate some teaching implications.

With regard to the first research question, in experiment 1, the analysis does not indicate the existence of statistically significant differences in the incidence of reversal error depending on whether the names used to designate the variables refer to the objects (e.g., "Professors") or the quantities (e.g., "Number of professors"). These results permit us to reject Rosnick's hypothesis (1981). This finding is in line with previous research such as Fisher (1988), where the hypothesis was not statistically tested. Moreover, it must be noted that in these previous studies, only the *Students and professors* problem is used and that, regardless of the type of name provided, the students were explicitly told that the variables must be interpreted as quantities.

Following Rosnick's (1981) argument, the current results have at least two possible interpretations. On one hand, they could indicate that static comparison is not a relevant source of the reversal error. So, either the use of object names does not mean that the solvers are more likely to consider "1 professor for each 6 students," or this modeling of the statement does not prompt solvers to be more likely to pose the equation $P = 6S$.

On the other hand, it is possible that, when faced with the term number of professors, certain subjects mentally process the term by associating it with objects instead of quantities. The reason could be an erroneous conceptualization of the meaning of variables and equations, consolidated over time. This possibility is a study limitation which is difficult to overcome, because it would require knowledge of the mental representation carried out by the subject. Related to this, it would be also problematic to ask the student directly without conditioning her/his answer. In addition, the question would allude, at least implicitly, to notions such as mental representation or information processing. This would imply difficulties for the subject's comprehension of the question, and in turn, it makes it difficult to interpret the answer.

Regarding experiment 2, note first that the instrument used (Table 1) was specifically designed in such a way that all its items, regardless of syntactic obstruction or the type of magnitude, have the same syntactic-semantic structure and analogous context information. This makes it more plausible that the observed differences are due to the variables under study, as they are the only elements that change between statements, unlike those of Cohen and Kanim (2005). Hence, resorting to collective names in the case of items with discrete magnitudes, avoiding the word number, and the homogeneous syntactic structure across all items makes it possible to minimize confounding variables. Thus, this instrument itself constitutes a relevant contribution of this work, which may be used in future research.

The results of experiment 2 give affirmative answers to the two remaining research questions concerning the influence of syntactic obstruction (RQ2) and the type of magnitude (RQ3), respectively. Regarding RQ2, we found that when constructing the equation, students committed a significantly greater number of reversal errors in statements with syntactic obstruction. At a descriptive level, our results agree with those of Cohen and Kanim (2005) obtained with a questionnaire with just discrete magnitudes and, what is more relevant, where items without syntactic obstruction always included the word number in the statement, while items with syntactic obstruction did not. MacGregor and Stacey (1993) found that the reversal error also occurred in items without syntactic obstruction. Thus, our results are to some extent aligned with those of MacGregor and Stacey (1993), since we also report reversal errors for these types of items.

Nevertheless, a plausible explanation for the results of the present work would be that a relevant number of subjects apply word order matching when posing the equation, which leads them to error. Thus, regarding the goals of the paper, these results indicate that word order matching can explain a considerable number of reversal errors. This interpretation concurs with the considerations given by Kaput (1987) concerning the tendency to apply natural language encoding processes to algebraic language. In addition, it is consistent with the considerations given by Duval (2006) concerning the greater difficulty of non-congruent conversions, since statements with syntactic obstruction are examples of the former.

Regarding the third research question, the results from experiment 2 show that statistical differences in the production of reversal errors exist with a high statistical significance depending on the type of magnitude involved. Also, post hoc tests show that in order to analyze the influence of the type of magnitude, it seems necessary to distinguish between at least three categories: extensive discrete magnitudes (e.g., cardinal of a discrete set), extensive continuous susceptible to be discretized (e.g., volume), and intensive continuous (e.g., acceleration). Actually, in all three possible comparisons, statistical differences exist. It is also relevant to point out that the highest rate of reversal errors occurs with discrete magnitudes, extensive continuous ones produce an intermediate rate, and the lowest rate is achieved with intensive ones. Hence, the empirical results confirm the hypothesis presented in the theoretical framework concerning the influence of the type of magnitude on the phenomenon of the reversal error.

Indeed, the above ordering can be explained as an effect of static comparison and the possibility of constructing a mental representation of the situation based on a relative size relationship between two groups. When dealing with extensive discrete magnitudes, it is always possible to construct a mental image (1 professor in front of 6 students), which can prompt, due to the application of a static comparison, the incorrect conversion sequence "1 professor for each 6 students, $1P = 6S$, $P = 6S$." In the case of extensive continuous magnitudes, at the beginning, such a representation would not be immediate. Nevertheless, as was previously shown, it is possible for the student to spontaneously discretize certain magnitudes of this type, such as the volume, and then construct a mental image of such a type (1 glass of lemon juice for each 4 glasses of water). In turn, it makes it easier to carry out the same reasoning used with discrete magnitudes. In the case of intensive continuous magnitudes, as the magnitude itself is defined as a ratio between other different magnitudes, these types of mental image would not be likely to happen due to difficulty in carrying out an immediate discretization. Then, such magnitudes would not prompt students to apply reasoning based on static comparisons. The intermediate position of extensive continuous magnitudes could be explained because their discretization may not always happen, since this depends on each subject's ability. Hence, we conclude that static comparison has some explanatory power as a

model for the reversal error, because otherwise, differences in the incidence of reversal error depending on the type of magnitude should not exist.

The above conclusion concerning static comparison agrees with that of Clement (1982), MacGregor and Stacey (1993), or Sung-Hee et al. (2014). Nevertheless, in the present work, unlike all previous studies, we reached it by explicitly studying the influence of the type of magnitude involved in the statement. Specifically, our results suggest that reversal errors can be related to the possibility of carrying out a mental representation of the situation by means of discrete sets. Therefore, taking into consideration the type of magnitude as a variable under study has proved to be one of the contributions of the present work. This provides new information that makes possible it to assess the explanatory power of static comparison, opening up a new research line in the study of reversal error.

Nevertheless, it should be acknowledged that reversal errors do not disappear in statements with intensive magnitudes, that effect sizes in comparisons between different magnitudes are not high, and that the results from experiment 1 indicate that differences in the incidence of reversal errors depending on the type of names given to the variables do not exist. On one hand, this indicates that static comparison, alone, is not enough to completely explain the phenomenon of reversal error. On the other hand, the effect size reported when studying the influence of the syntactic obstruction suggests that, although relevant, the word order matching model, alone, is not enough, to explain the phenomenon. Hence, on assessing whether the goals of this paper are reached, a global analysis of both experiments indicates that the reversal error is a complex phenomenon that cannot be explained using just one framework. At least two explanatory models coexist, word order matching and static comparison, agreeing with those identified in Clement (1982).

We now present the teaching implications of the above conclusions in the design of didactic interventions aimed at correcting the occurrence of reversal errors. Many studies show that the reversal error is a persistent phenomenon that occurs with secondary school and even undergraduate students. Most students in these educational levels have already studied, or are studying, intensive magnitudes such as those in Table 1 in science subjects, so they should be able to carry out tasks like those in experiment 2. Hence, the ordering of the rates of reversal errors, according to the type of magnitude, suggests its replication in a teaching intervention based on statements analogous to those in Table 1. The idea is inspired by the bridging strategy used by Brown and Clement (1989) for the teaching of physical magnitudes but applied here to correct reversal error production. In our case, the anchoring example would be conversion with intensive magnitudes; for the bridging one, conversion with extensive continuous, susceptible to discretization, and for the target example, with discrete magnitudes. Our results suggest that intensive magnitudes can ease the identification of the algebraic structure of the problem, and the generation of a correct equation. By analogical transfer, this would allow students to realize that the algebraic structure, and therefore the equation, is the same when the statement involves discrete magnitudes. The feasibility of this proposal relies on the fact that several studies have proved the effectiveness of such instruction designs in algebraic word problem solving (Gómez-Ferragud, Solaz-Portolés, & Sanjosé, 2013). In addition, it would be useful, especially to prevent errors due to static comparison, to approach the teaching of equation posing by means of the triadic nested model of semiosis exposed in Presmeg (2006). In our case, the first object would be the situation model and the first referent the problem statement, the next level would be the diagrammatic representation of the hypothetical active operation presented in Clement (1982), and the final level would be the equation.

Regarding the limitations of this study, as the influence of the magnitude type had not been analyzed until now, more works concerning the topic are needed. In particular, it would be convenient to analyze distinct samples, such as secondary school students. As a possible future

line of research, qualitative studies could be conducted to explore the subjects' mental representations depending on the type of magnitudes, improving our understanding of the reversal error phenomenon.

References

Brown, D. E., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: Abstract transfer versus explanatory model construction. *Instructional Science, 18*, 237–261.

Castro, E. (1995). *Niveles de comprensión en problemas verbales de comparación multiplicativa [Levels of comprehension in multiplicative comparison word problems]*. Granada: Comares.

Clement, J. J. (1982). Algebra word problem solutions: Thought processes underlying a common misconception. *Journal for Research in Mathematics Education, 13*(1), 16–30.

Clement, J., Lochhead, J., & Monk, G. (1981). Translation difficulties in learning mathematics. *The American Mathematical Monthly, 88*(4), 286–290.

Cohen, E., & Kanim, S. E. (2005). Factors influencing the algebra reversal error. *American Journal of Physics, 73*(11), 1072–1078.

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics, 61*, 103–131.

Filloy, E., Rojano, T., & Solares, A. (2010). Problems dealing with unknown quantities and two different levels of representing unknowns. *Journal for Research in Mathematics Education, 41*(1), 52–80.

Fisher, K. J., Borchert, K., & Bassok, M. (2011). Following the standard form: Effects of equation format on algebraic modeling. *Memory & Cognition, 39*(3), 502–515.

Fisher, K. M. (1988). The students-and-professors problem revisited. *Journal for Research in Mathematics Education, 19*(3), 260–262.

Gómez-Ferragud, C. B., Solaz-Portolés, J. J., & Sanjosé, V. (2013). Analogy construction and success in mathematics and science problem-solving: A study with secondary students. *Revista de Psicodidáctica, 18*(1), 81–108.

González-Calero, J. A., Arnau, D., & Laserna-Belenguer, B. (2015). Influence of additive and multiplicative structure and direction of comparison on the reversal error. *Educational Studies in Mathematics, 89*, 133–147.

Kaput, J. (1987). Towards a theory of symbol use in mathematics. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 159–195). Hillsdale, NJ: Erlbaum.

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

Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.

Landy, D., Brookes, D., & Smount, R. (2014). Abstract numeric relations and the visual structure of algebra. *Journal of Experimental Psychology: Leaning, Memory, and Cognition, 40*(5), 1404–1418.

Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology, 79*(4), 363–371.

MacGregor, M., & Stacey, K. (1993). Cognitive models underlying students' formulation of simple linear equations. *Journal for Research in Mathematics Education, 24*(3), 217–232.

Matz, M. (1982). Towards a process model for high school algebra errors. In D. Gleeman & J. S. Brown (Eds.), *Intelligent tutoring systems* (pp. 25–50). New York: Academic Press.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Nesher, P., Hershkovitz, S., & Novotna, J. (2003). Situation model, text base and what else? Factors affecting problem solving. *Educational Studies in Mathematics, 52*(2), 151–156.

Presmeg, N. (2006). Semiotics and the "connections" standard: Significance of semiotics for teachers of mathematics. *Educational Studies in Mathematics, 61*, 163–182.

Rosnick, P. (1981). Some misconceptions concerning the concept of variable. *Mathematics Teacher, 74*, 418–420.

Rosnick, P., & Clement, J. (1980). Learning without understanding: The effect of tutoring strategies on algebra misconceptions. *Journal of Mathematical Behaviour, 3*(1), 3–27.

Schwartz, J. L. (1988). Intensive quantity and referent transforming arithmetic operations. In J. Hiebert & M. J. Behr (Eds.), *Number concepts and operations in the middle grades* (pp. 41–52). Hillsdale, NJ: Erlbaum.

Sung-Hee, K., Phang, D., An, T., Ji Soo, Y., Kenney, R., & Uhan, N. (2014). POETIC: Interactive solutions to alleviate the reversal error in student–professor type problems. *International Journal of Human-Computer Studies, 72*, 12–22.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing*. New York: Elsevier Academic Press.

Wollman, W. (1983). Determining the sources of error in a translation from sentence to equation. *Journal for Research in Mathematics Education, 14*, 169–181.