

Insights into the reversal error from a study with South African and Spanish prospective primary teachers



Authors:

Calos Soneira¹ 
 Sarah Bansilal² 
 Reginald Govender² 

Affiliations:

¹Department of Pedagogy and Didactics, Faculty of Education Sciences, University of A Coruña, A Coruña, Spain

²Mathematics and Computer Science Education, School of Education, University of KwaZulu-Natal, Durban, South Africa

Corresponding author:

Reginald Govender,
 govenderr4@ukzn.ac.za

Dates:

Received: 01 Apr. 2021
 Accepted: 24 Aug. 2021
 Published: 25 Nov. 2021

How to cite this article:

Soneira, C., Bansilal, S., & Govender, R. (2021). Insights into the reversal error from a study with South African and Spanish prospective primary teachers. *Pythagoras*, 42(1), a613. <https://doi.org/10.4102/pythagoras.v42i1.613>

Copyright:

© 2021. The Authors.
 Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Read online:



Scan this QR code with your smart phone or mobile device to read online.

This study, using a quantitative approach, examined Spanish and South African pre-service teachers' responses to translating word problems based on direct proportionality into equations. The participants were 79 South African and 211 Spanish prospective primary school teachers who were in their second year of a Bachelor of Education degree. The study's general objective was to compare the students' proficiency in expressing direct proportionality word problems as equations, with a particular focus on the extent of the reversal error among the students' responses. Furthermore, the study sought to test the explanatory power of word order matching and the static comparison as causes of the reversal error in the two contexts. The study found that South African students had a higher proportion of correct responses across all the items. While nearly all the errors made by Spanish students were reversals, the South African group barely committed reversal errors. However, a subgroup of the South African students made errors consisting of equations that do not make sense in the situation, suggesting that they had poor foundational knowledge of the multiplicative comparison relation and did not understand the functioning of the algebraic language. The study also found that the word order matching strategy has some explanatory power for the reversal error in both contexts. However, the static comparison strategy offers explanatory power only in the Spanish context, suggesting that there may be a difference in curriculum and instructional approaches in the middle and secondary years of schooling, which is when equations are taught.

Keywords: algebra; reversal error; word problems; word order matching; static comparison; pre-service mathematics teachers.

Introduction

Over the past decades there has been much research about the learning of early algebra and algebra researchers have documented many misconceptions of students (e.g. Bush & Karp, 2013; Kieran, 2007; Knuth, Alibali, Hattikudur, McNeil, & Stephens, 2008; Swan, 2001). One error that has received much attention in algebraic research is the so-called *reversal error* which is exhibited by students at all levels who, when asked to express a directly proportional relationship between two quantities in terms of an equation, interchange the two variables (Clement, 1982). The error gained prominence after the famous *student-professor* problem posed by Clement (1982) and Clement, Lochhead and Monk (1981). In this problem, students were asked to write an equation using S for students and P for professors to represent the statement 'There are six times as many students as professors in this university'. Clement et al. (1981) found that 68% of the incorrect responses were $6S = P$. This error is referred to as a reversal error because of the reversal in the position of the two variables compared to their position in the correct equation $6P = S$.

The reversal error has been identified at different levels of study, among high school students in their initial stage of formal algebra learning (MacGregor & Stacey, 1993), and also among university students (Clement, 1982; Cohen & Kanim, 2005; Fisher, Borchert, & Bassok, 2011; González-Calero, Arnau, & Laserna-Berenguer, 2015; Sung-Hee, Phang, Soo, Kenney, & Uhan, 2014). It is therefore important that research is conducted in this area so as to suggest measures that can be taken in the early stages of algebra learning so as to prevent the error, as well as in the later stages in order to remedy it (Barbieri, Miller-Cotto, & Booth, 2019). For teaching interventions to be effective, these need to be based on the knowledge of the specific cognitive process and contextual circumstances that trigger the misconception. In addition, the phenomenon is complex because it involves the difficulties linked to the transition from arithmetic to algebra (MacGregor & Stacey, 1993; Malisani & Spagnolo, 2009), as well as the general difficulties related to the change of representation from the natural to a formal language (González-Calero, Berciano, & Arnau, 2020).

The present study looks into the cognitive processes that underly the reversal error, by exploring the extent of the reversal error and the explanatory power of the theoretical models suggested by previous research. The participants are prospective primary school teachers from two countries (South Africa and Spain) which represent two different educative contexts in terms of educational opportunities and resources. We take a comparative approach with the two groups (South African and Spanish prospective primary school teachers). This is accomplished by means of two versions of a questionnaire in which specific changes were made to adapt each one to the idiomatic features but keeping the same core structure in both countries. By considering these two different contexts and using these specifically designed questionnaires, this study looks at the incidence of the reversal error across the two samples. Therefore, as each sample is taken from a different context, the study can identify whether the reversal error is equally likely to occur in the two groups, which would suggest that the difficulty is embedded within the learning of algebra in general. In addition, if the strategies used by students differ across the groups, then this would suggest that there may be differences in curriculum approaches in the early learning of algebra, which could be explored in further studies. Hence, we can gain insights into how the educative context may affect some sources that seem to prompt the reversal error, while there may be other sources that seem to be inherent to the learning of algebra.

Literature review

Much research in mathematics education has focused on systematic misconceptions that commonly occur with students from different countries and students who speak other languages. As part of the learning process of any concept, students develop misconceptions as they grapple with the concept at hand. A misconception is often revealed when a teacher is alerted by a consistent error made by a student in response to a particular situation. Instructors can be alerted by a consistent error made by students, which points to a misconception. Barmby (2009) believes that misconceptions are inevitable, they occur naturally and the errors that they lead to are largely predictable.

Some students have misconceptions about the meaning of the equal sign, which seem to persist up to tertiary level studies (Booth, 1988; Fyfe, Matthews, & Amsel, 2020). Many students view the equal sign as a symbol separating two expressions, while others see it as an operational signal which indicates that the equation must be solved (Kieran, 1981; Matthews, Rittle-Johnson, McEldoon, & Taylor, 2012). Because students do so much computing of the answer, they often use it as a 'write the answer' sign, instead of seeing the sign as indicating equivalence between two expressions. There have also been many studies that have looked at students' understanding of the variable. Some studies have reported the common misconception of taking a variable as a label standing for an object instead of standing for an unknown quantity (Fisher et al., 2011;

González-Calero et al., 2015; Soneira, González-Calero, & Arnau, 2013; Stacey & MacGregor, 1999). Other misconceptions that have been identified include students thinking that the same letter appearing at different points in a number sentence could not represent the same number (Fillooy, Rojano, & Puig, 2008), that letters can only stand for whole numbers, among many others. Although rooted in arithmetic thinking, all these errors are subsequently transferred to their algebraic thinking, causing the corresponding misconceptions about the notion of equation (Fillooy et al., 2008; Humberstone & Reeve 2018). Moreover, students tend to face difficulties when representing mathematical relationships by means of algebraic language (Fillooy et al., 2008; Kieran, 2007; Koedinger & Nathan, 2004; Stacey & MacGregor, 1999). Specifically, when translating word problems into equations, they tend to keep the same order of the main pieces of information as displayed in natural language although it may lead to errors (Duval, 2017). Martin and Bassok (2005) pointed out that students often rely on translation cues to move directly from the words to writing the equation, without actually thinking about whether the solution makes sense in the setting or whether the relational statement was appropriately represented by the algebraic statement. The use of these cues can lead to correct solutions but it is often the case that students' reliance on these translation cues leads to systematic errors such as the reversal error (Landy & Goldstone, 2007; Lewis & Mayer, 1987). Another explanative model of the reversal error is that students understand the mathematical relationship, but hold misconceptions about the equal sign (e.g. Clement, 1982; Cohen & Kanim, 2005; Fisher et al., 2011; González-Calero et al., 2015). Whatever the explanation, it is clear that the reversal error seems to be rooted in common algebra learners' misconceptions.

Many studies have investigated the influence of factors such as variable symbol choice, sentence structure, semantic cues, context familiarity, among others, on students' tendency to commit the reversal error (e.g. Cohen & Kanim, 2005; Clement, 1982; Martin & Bassok, 2005; Fisher et al., 2011). In terms of variable choice, some researchers (Crowley, Thomas, & Tall, 1994; Fisher, 1988; Sims-Knight & Kaput, 1983) consider that the use of specific letters as variables may influence students when making up the equation to represent a relationship. For example, the use of the first letter of a word as the variable, such as P for the number of professors and S for the number of students, may be more likely to prompt students to make the reversal error as compared to the use of the more generic x and y symbols to represent the quantities (Crowley et al., 1994; Sims-Knight & Kaput, 1983). It was argued that the use of these first letters as symbols may prompt students to use the symbols as labels rather than as variables (Crowley et al., 1994; Fisher, 1988). However, research by Cohen and Kanim (2005) as well as Soneira et al. (2018) did not find any differences in the rate of the reversal error in problems that took the variable as the first letter of the quantity and those that took the variable as a more general symbol.

Conceptual framework: Word order matching and static comparison models

Basically, studies that assess the explanatory power of different students' strategies that lead them to commit a reversal error have consisted of modifying some variable of the research instrument and then observing whether the change prompted different results in the students' responses. In this study we investigate the explanatory power of the word order matching and static comparison strategies as well as their contextual dependence or independence.

Word order matching strategy

The word order matching strategy involves a 'literal, direct mapping of the words of English into the symbols of algebra' (Clement et al., 1981, p. 288). For example, in the students and professors problem, if the reversed equation $6S = P$ is posed, $6S$ would represent the clause 'There are six times as many students', the sign $=$ would stand for the word 'as', and P for 'professors'. Students who rely on this word order strategy assumes that the order of the key words in the problem statement can be mapped directly into the order of symbols that appear in the equation. However, sometimes because of the flexibility of natural language the same multiplicative relationship can be expressed by using a syntactic structure such that the application of the word order matching leads to a reversal error, or by using a different syntactic structure such that the application of this incorrect strategy will however lead to a correct equation. We differentiate between these two versions by referring to statements 'with syntactic obstruction' as opposed to statements 'without syntactic obstruction' (Soneira et al., 2018, p. 4.). Statements with syntactic obstruction are those where the use of the word order matching would result in a reversal error (Soneira et al., 2018, p. 4). An example of such a statement is 'At this company, there are six times as many workers as managers'. Assuming that the variable W represents the number of workers and M the number of managers, a word order matching translation would result in the equation $6W = M$, which is a reversal error. In contrast, the same relationship can be expressed using a statement without syntactic obstruction (Soneira et al., 2018, p. 4) as 'In this company, the number of workers is six times the number of managers'. A word order matching translation of this latter statement will result in the equation $W = 6M$, which is the correct equation to represent the multiplicative comparison conveyed in both statements (with and without syntactic obstruction).

Students who opt for the reversal error strategy because they have been misled by the translation cues have some algorithmic understanding of equations; however, their algebraic skills may not have included a consideration of the modelling component (Martin & Bassok, 2005). Therefore, these reversal errors may be considered as evidence of some algorithmic comprehension but rudimentary knowledge of the multiplicative comparison relation (Martin & Bassok, 2005).

Static comparison strategy

Research has also identified that some students may display the reversal error even in statements without syntactic obstruction (Cohen & Kanim, 2005; MacGregor & Stacey, 1993; Martin & Bassok, 2005; Soneira et al., 2018). It can be inferred in this case that these students did not rely on the word order matching strategy because if they did so consistently, then they would not have obtained the reversal error. Hence, in this case the cause of the error must be a different one, the static comparison being the most widely considered in the literature.

With the static comparison strategy, the student conceives the equation as representing a correspondence or ratio between groups instead of an algebraic equivalence between quantities. In the previous student and professor example, using the equation $6S = P$, the equal sign would mean that 'six students correspond to one professor'. As asserted by Clement et al. (1981):

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. (p. 288)

Regarding the assessment of the explanative power of the static comparison model, most studies (e.g. Clement, 1982; Cohen & Kanim, 2005; Fisher et al., 2011; Martin & Bossok, 2005) considered only items with quantities that had discrete magnitudes. Furthermore, those that use items with continuous variables, namely Landy, Brookes and Smount (2014), MacGregor and Stacey (1993), Sung-Hee et al. (2014) and Wollman (1983), do not take into account this continuous character of the magnitude in their analysis. The only exception is the work by Soneira et al. (2018), who went further to explore whether the type of magnitude used in setting up the direct proportional relationship between two quantities influenced the rate of the reversal errors in expressing the relationship. Those researchers (Soneira et al., 2018) considered items based on quantities that were discrete, extensive continuous and intensive continuous in magnitude and we use the same distinction in this study. The intention was to observe more closely the explanatory power of the static comparison model, which in effect builds a correspondence between two sets, for example the set of workers and the set of managers. Since the sets are of different sizes, establishing the correspondence would require an understanding of the relative size between the sets. For quantities that are discrete, the student could easily set up a mental picture by taking one element of the set of managers and considering the corresponding elements of the second, for example taking a company with one manager and seeing that there will be six workers. Hence, the student will infer that $6W = 1M$ within the static comparison model, where the sign $=$ indicates correspondence instead of equality between the quantities represented in both sides. When magnitudes are continuous, it is harder to consider one element of a set in

response to a corresponding element from the second set. For example, it is harder to consider a quantity of lemon juice or water as an element of a set. However, it may be possible to consider glasses of water or lemon juice, thus discretising that continuous quantity so as to make the mental model. In these cases where the continuous magnitudes are easier to discretise, such as water or juice, we follow Soneira et al. (2018) and use the term extensive continuous to describe them. However, there are other continuous magnitude situations such as when the acceleration of a car is being compared to that of a motorbike, or when the density of two quantities are being compared, when the magnitude is harder to discretise. In such instances, we refer to these quantities as having an intensive continuous magnitude. The key point regarding intensive magnitudes is that because they are harder to discretise, it would be also harder to establish a correspondence between sets that underlies the static comparison strategy. Therefore, if the static comparison has, indeed, explanative power as a model for the reversal error, a student most likely will apply it when dealing with discrete magnitudes, which in turn will lead to a higher rate of reversal errors. As this strategy is less likely to be applied when dealing with extensive continuous but discretisable magnitudes, the rate of reversal errors will be lower with this type of item. Finally, it is highly improbable that the static comparison could be applied in dealing with intensive continuous magnitudes, and therefore the rate of reversal errors would be the lowest. On the other hand, if the static comparison did not have explanative power as a model for the reversal error, then there would be no differences in the rate of reversal errors depending on the type of magnitude. This is why in this study we also investigate whether the type of magnitude, whether it is discrete, extensive continuous or intensive continuous, influences the tendency to commit the reversal error.

Note that each model involves different cognitive processes. Indeed, within the static comparison the student understands the correct ratio between quantities but falls short when trying to translate the relationship into an equation. In contrast, word order matching can be applied without comprehending the multiplicative relationship between quantities, but by incorrectly trying to replicate the verbal syntax in algebraic language.

Aim and objectives

The purpose of this study was to consider the explanatory power of certain theoretical models as causes of the reversal error using items set within two languages: Spanish and English. This study was guided by the following research objectives with respect to the samples of Spanish and South African prospective primary school teachers:

- To examine the proficiency of the students from the two contexts in being able to express direct proportionality word problems into equations.
- To test the explanatory power of word order matching and the static comparison as causes of the reversal error within the two groups.

Research design and methodology

In order to assess the explicative power and the context's (in)dependence of the word order matching and static comparison as models for the reversal errors, we followed a quantitative approach with two samples. First, within a repeated measures design in each sample, we conduct hypothesis testing so as to assess in each context the student's proficiency when translating verbal proportional statements into equations and the explicative power of each model for the reversal error. Afterwards, as both experiments were analogous, we follow Ellis (2010) and compare the effect sizes obtained in the corresponding tests in each context in order to get insight into the context's dependence on each model's explicative power. Indeed, although the sample sizes are different, the effect sizes allow us to compare the corresponding results since the sample size is taken into account in the formulae for the effect size. In fact, the goal of comparative research is 'to search for similarity and variation' across different contexts or categories (Given, 2008, p. 100) and in this study we look at the cognitive processes underlying the reversal error. By taking a comparative approach to the two different groups, we hope to gain deeper insights into what seems to prompt the reversal error.

Participants

The participants were made up of two groups of prospective primary school teachers from universities in Spain and South Africa based on convenience sampling since the institutions were accessible to the researchers. However, the sampling criteria were taken as the educational backgrounds, admission requirements, field of study and future job qualification which were equivalent in both countries. Both groups of students were in their second year of a Bachelor of Education (B.Ed.) degree with a primary school specialisation. The sample from South Africa consisted of 79 students most of whom (95%) spoke English as a second language. Two-thirds of the group had attended non-fee-paying schools showing that they come from very poor socio-economic backgrounds. Researchers have found that students from non-fee-paying schools in South Africa generally perform less well in mathematics national and international exams when compared to their peers from more affluent schools (Reddy et al., 2012; SAIRR, 2016).

The sample from Spain consisted of 211 students at a public university in their second year of study towards a B.Ed. degree with a primary school specialisation. Regarding their socio-economic background, they mostly belong to the low-middle and middle class and attended non-fee-paying schools. The analysis of the international assessment tests conducted in Spain also conclude that students belonging to high socio-economic backgrounds obtain better academic results than their peers from poorer backgrounds (MEFP, 2019, 2020). However, in Spain the difference in the academic achievement between the higher and lower achievers has received much media attention in the Organisation for Economic Co-operation and Development countries (MEFP, 2019).

Data collection instruments

Concerning the instrument, we adapted the questionnaire designed and already applied in quantitative research about the reversal error by Soneira et al. (2018). The questionnaire consists of 12 verbal statements expressing proportionality relationships in everyday situations where six items were expressed using syntactic obstruction and six corresponding items without syntactic obstruction. Furthermore, the statements involved comparisons of quantities, two of which were discrete, two extensive continuous and two intensive continuous.

In the Spanish university, we used the original questionnaire of Soneira et al. (2018), which was written in Spanish. In order to be applied in the South Africa university, we translated the items into English and ensured that all the contextual situations described in the items were familiar to students in both countries. Sometimes the wording and word order needed to be changed to ensure the word problem made sense and was relatable. However, the translation into English was done in a manner that preserved whether the item was expressed with or without syntactic obstruction. Table 1 presents some examples of Spanish statements, their direct translation to English – which could be grammatically incorrect, or at least awkward – and their modified versions, which were used in the English test.

Data collection took place across two countries, first at the university in Spain and thereafter at the university in South Africa. At both sites, the university learning management system was used as a medium to administer the 12 word problem statements in random order that formed the data collection instrument. Each item was presented with a list of operators and numbers and with instructions to click and drag into position to create the equation (Figure 1). While the user poses the equation, the program allows them to modify their attempt and once the student accepts it the equation is saved in a database and a new task is loaded.

Coding of variables

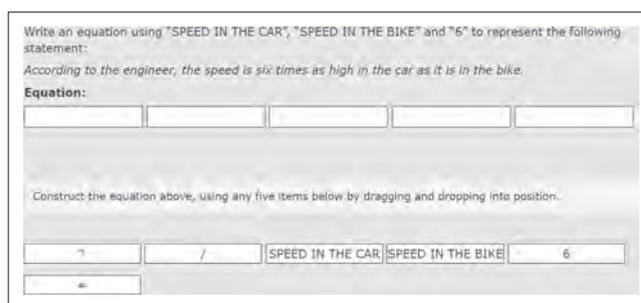
When coding the students' responses, we defined three variables: 'Cor' (correct), 'Rev' (reversal error) and 'Other' (other error) to code the response of each student to each problem. The variable Cor takes the value 1 if the response is correct and 0 otherwise. The variable Rev equals 1 if the answer corresponds to a reversal error and 0 otherwise. The variable Other takes the value 1 whenever the answer is incorrect, but the error is not a reversal, otherwise it equals 0. In the case of the students and professors problem, the responses $Students = 6 \times Professors$ and $Students \div Professors = 6$ are examples of correct answers ($Cor = 1$), $6 \times Students = Professors$ and $Students = Professors \div 6$ are examples of reversal errors ($Rev = 1$), while $Students \times Professors = 6$ and $Students + 6 = Professors$ are errors different from reversals ($Other = 1$). Thus, the variables Cor and Other account for the student's proficiency and difficulties with each type of statement, and the variable Rev allows for assessment of the explicative power of each model for the reversal error.

TABLE 1: Some examples of word problem statements used in Spain and South Africa.

Magnitude type:	Language	Item
Syntactic obstruction		
Intensive continuous		
Yes	Spanish	<i>Según el ingeniero, es nueve veces mayor la aceleración de la motocicleta que la de la bicicleta.</i>
	Direct translation to English	According to the engineer, [it] is nine times higher the acceleration of the motorbike than that of the bicycle.
	Modified and used	According to the engineer, the speed is nine times as high in the car as it is in the bike.
No	Spanish	<i>Según el ingeniero, la aceleración del ciclomotor es cuatro veces mayor que la del monopatín.</i>
	Direct translation to English	According to the engineer, the acceleration of the scooter is four times greater than that of the skateboard.
	Modified and used	According to the engineer, the speed of the scooter is eight times as great as that of the tricycle.
Discrete		
Yes	Spanish	<i>En esta universidad, es seis veces mayor el colectivo de estudiantes que el de profesores.</i>
	Direct translation to English	At this university, [it] is six times greater the community of students than that of professors.
	Modified and used	At this university, there are six times as many students as professors.
No	Spanish	<i>En este hospital, el colectivo de pacientes es cinco veces mayor que el de médicos.</i>
	Direct translation to English	At this hospital, the community of patients is five times greater than that of doctors.
	Modified	At this hospital, the community of patients is eight times as large as that of doctors.
Extensive continuous		
Yes	Spanish	<i>En esta limonada, es cuatro veces mayor el volumen de agua que el de limón.</i>
	Direct translation to English	In this lemonade drink, [it] is four times greater the volume of water than that of lemon juice.
	Modified and used	In this drink, there is four times as much the volume of water as that of lime juice.
No	Spanish	<i>En este cóctel, el volumen de tónica es cinco veces mayor que el de ginebra.</i>
	Direct translation to English	In this cocktail, the volume of tonic is five times greater than that of gin.
	Modified and used	In this cocktail, the volume of tonic is five times as much as that of gin.

Source: Adapted from Soneira, C., González-Calero, J.A., & Arnau, D. (2018). An assessment of the sources of the reversal error through classic and new variables. *Educational Studies in Mathematics*, 99(1), 43–56. <https://doi.org/10.1007/s10649-018-9828-1>

Note: The translation was done by the authors, two of them being native English speakers.



Source: Direct screenshot from learning management system.

FIGURE 1: Example of a word problem appearing on the learning management system.

Statistical tests

Regarding the statistical tests we use, note that some choices depend on the fulfilment of certain assumptions, which in turn relies on the sample sizes and the results of other statistical tests. For the sake of clarity and to avoid repetition, in this section we describe the general procedure and in the results section we provide the statistical values that were used to check if the assumptions were fulfilled.

In order to assess the explanative power of the word order matching, our assumption is that students who use the word order matching strategy are likely to respond correctly to items that are expressed without syntactic obstruction and for items expressed with syntactic obstruction, these students' guiding strategy would prompt them to produce a reversal error. We take the factor 'syntactic obstruction' (our instrument is made up of six items with syntactic obstruction and six without syntactic obstruction), which has two levels: items expressed with syntactic obstruction and those that are not. Given the sample sizes (79 and 211), we opted to use the paired t-test which is robust for sample sizes greater than 30 (Pallant, 2010).

In order to assess the explanative power of the static comparison, as explained in the conceptual framework section, our assumption is that this strategy is more or less likely to be applied depending on the type of magnitude, which will be reflected in the rate of reversal errors. Hence, we take the factor 'type of magnitude'. Specifically, this factor has three levels (intensive continuous, extensive continuous, extensive discrete) which are ordered from lowest to highest according to how likely it is that the statement triggers the static comparison strategy. Thus, we use a one-way within-subject factorial design. The sphericity assumption was checked by means of the Mauchly's test. If the sample distributions meet the sphericity condition, the analysis of variance test is used, otherwise we follow the protocol suggested by Pardo and San Martin (2010, p. 305). The latter involves applying first the multivariate approximations, specifically the Pillai's Trace, the Wilks's Lambda, and the Hotelling's Trace. If these show statistical differences, then the factorial analysis is complete and we test pairwise comparisons by means of the paired t-test. If the multivariate approximations do not show significant differences, then we go on to calculate the statistic F with the modified degrees of freedom by the Greenhouse-Geisser and the Huynh-Feldt tests. If the latter show statistical differences, we test pairwise comparisons by means of the paired t-test.

In all tests, we take a p -value of 0.05 as the decision criterion. We calculate the Cohen's delta to estimate the effect sizes of each paired t-test, and in order to provide more confidence in the results we report the corresponding p -value and the 95% confidence interval (CI) in all cases. As we conduct multiple hypothesis testing with each sample, in order to prevent type I errors we use the Holm-Bonferroni correction (all the reported p -values are those calculated after applying the correction).

Ethical considerations

The ethical committees granted ethical clearance at the universities for this study. The researchers complied with all prescribed ethical measures, such as getting informed consent by the participants and keeping their anonymity. Moreover, all participants were volunteers.

Results

We first present the descriptive statistics as a general overview of the results, before going into more specific detail about the results of the statistical tests.

Descriptive statistics

Regarding the factor of syntactic obstruction, for each student and each of the variables Cor, Rev and Other, we computed the mean in statements with syntactic obstruction, and the mean in statements without syntactic obstruction (Table 2).

The results for the variable Cor in Table 2 point out that the group of South African students outperforms the Spanish one regardless of the syntactic obstruction. However, the results for the variables Rev and Other suggest that the pattern of errors in the South African context is quite different from the Spanish. In particular, regardless of the syntactic obstruction, nearly all the errors made by Spanish students are reversals, while South African students barely commit reversal errors but errors consisting of equations that do not make sense in the situation stated in the problem statement, such as additive or inverse proportionality relationships.

Concerning the type of magnitude, we computed three scores for each student and variable. Each one was the mean of the scores for the items involving discrete magnitudes, extensive continuous magnitudes and intensive continuous magnitudes (Table 3).

Thus, the results depending on the type of magnitude (Table 3) have an analogous interpretation to those depending on the syntactic obstruction. That is, South African students outperform the Spaniards in all types of magnitude, and the pattern of errors was different between contexts. In the

TABLE 2: Differences depending on the syntactic obstruction. Descriptive statistics.

Response: Syntactic obstruction	South Africa (N = 79)		Spain (N = 211)	
	Mean	Standard deviation	Mean	Standard deviation
Cor				
Yes	0.6962	0.27575	0.5940	0.36404
No	0.7489	0.32344	0.6517	0.41136
Rev				
Yes	0.1013	0.14719	0.3926	0.35789
No	0.0422	0.09038	0.3397	0.40462
Other				
Yes	0.2025	0.25413	0.0134	0.06865
No	0.2089	0.32846	0.0087	0.06549

Cor, Correct; Rev, Reversal error; Other, Other error.

Spanish case the reversal error is, by far, the predominant, while it was scarce among the South African students, whose errors consisted mainly of equations that did not make sense.

Taken together, the descriptive results (Table 2 and Table 3) suggest that among the South African students there is a subgroup who do not have rudimentary knowledge of the multiplicative comparison relation and do not understand the functioning of the algebraic language, especially in terms of expressing word problems as equations.

The word order matching strategy

We report next (Table 4) the results of the hypothesis testing regarding the differences depending on the syntactic obstruction.

We obtain (Table 4) that in the South African context, there were statistically significant differences just for the variable Rev, with a weak-to-moderate effect size, while in the Spanish context we found statistically significant differences, with a weak effect size, for the variables Cor and Rev (Table 3). Hence, with respect to the second research objective, the word order

matching has some explanatory power as a source for the reversal error in both contexts; otherwise we would not find statistical differences for the variable Rev depending on the syntactic obstruction. The preceding tests also provide further insight into differences between the groups, thus addressing the first objective. Indeed, in the Spanish case there are differences for both the variable Cor and the variable Rev, and in the South African context only for Rev. This is consistent with the fact that the word order matching is a model only for the reversal error and nearly all the errors in the Spanish case are of this type (Table 2); thus, the effect on the variables Cor and Rev must be similar in this context.

With respect to the South African context, the fact that there are not significant differences depending on the syntactic obstruction neither for the variable Cor nor the variable Other would mean that, among those students who have problems with the algebraic modelling process that we identified above, the syntactic obstruction is not the main source of the error. Otherwise, there would be differences depending on the syntactic obstruction. Furthermore, these students did not follow, or their command of the algebraic language is so poor that they were not able to follow, the word order matching strategy.

The static comparison strategy

We report next the results of the hypothesis testing about the differences depending on the type of magnitude. For the sake of clarity, this is done for the variables Rev, Cor and Other in each context, separately. Concerning the variable Rev, in the South African context we cannot assume sphericity ($W = 0.952, p = 0.040$) and neither the multivariate methods nor the methods with modified degrees of freedom showed statistical differences. In the Spanish context, we cannot assume sphericity ($W = 0.952, p = 0.006$), but we found statistical differences by the multivariate methods. In particular, we obtained ($F = 12.266, p < 0.001$). Therefore, we conducted all the pairwise comparisons, obtaining significant differences between all possible pairs (Table 5).

TABLE 3: Descriptive statistics by country and magnitude type.

Response: Magnitude type	South Africa (N = 79)		Spain (N = 211)	
	Mean	Standard deviation	Mean	Standard deviation
Cor				
Intensive continuous	0.6677	0.31704	0.6600	0.41641
Extensive continuous	0.7278	0.33539	0.6268	0.41717
Discrete	0.7722	0.32073	0.5818	0.35792
Rev				
Intensive continuous	0.0601	0.13406	0.3258	0.40846
Extensive continuous	0.0728	0.14518	0.3637	0.41487
Discrete	0.0823	0.15873	0.4088	0.35225
Other				
Intensive continuous	0.2722	0.31315	0.0142	0.07582
Extensive continuous	0.1994	0.32369	0.0095	0.06835
Discrete	0.1456	0.31165	0.0095	0.06385

Cor, Correct; Rev, Reversal error; Other, Other error.

TABLE 4: Contrasts for differences depending on the syntactic obstruction.

Variable	South Africa (N = 79)					Spain (N = 211)				
	t(78)	p	95%CI		d	t(210)	p	95%CI		d
			LL	UL				LL	UL	
Cor	-1.76	0.082	-0.112	0.007	-	-4.165	0.001**	-0.085	-0.030	0.287
Rev	3.37	0.001**	0.024	0.094	0.379	3.774	0.001**	0.026	0.081	0.260
Other	-0.281	0.780	-0.051	0.039	-	1.226	0.221	-0.003	0.012	-

CI, confidence interval; LL, lower limit; UL, upper limit.

d = Cohen's delta.

**p < 0.01

TABLE 5: Differences depending on the type of magnitude for the variable Rev in the Spanish context.

Difference variables for Rev	Mean	Standard deviation	t(210)	p	95%CI		Cohen's d
					LL	UL	
Extensive continuous – Intensive continuous	0.03791	0.21069	2.614	0.030*	0.00932	0.06651	0.17993
Extensive discrete – Intensive continuous	0.08294	0.24566	4.904	0.001**	0.04960	0.11628	0.33762
Extensive discrete – Extensive continuous	0.04502	0.25422	-2.573	0.022*	0.01052	0.07952	0.17709

CI, confidence interval; LL, lower limit; UL, upper limit.

*p < 0.05; **p < 0.01

Therefore, in respect of the explanatory power of the static comparison as a model for the reversal error, firstly, our results point out that when taking together statements with and without syntactic obstruction, the rate of reversal errors among South African students is very low regardless of the type of magnitude (Table 3). Moreover, there are no differences in the variable Rev depending on the type of magnitude but the effect of the syntactic obstruction was weak to moderate, stronger than in the Spanish sample. These two results considered together would mean that, although South African students barely commit reversal errors, whenever this happens it is mainly due to the application of the word order matching, and not the static comparison strategy. This is also different from the Spanish context, in which the statistical differences depending on the type of magnitude mean that the static comparison has some explanatory power as a source of the reversal error (Table 5).

Regarding the variable Cor, in the South African group, by Mauchly's test, sphericity can be assumed ($W = 0.998$, $p = 0.941$). Thus, we apply the analysis of variance test for repeated measures, by means of which we found statistically significant differences ($F = 4.426$, $p = 0.013$). We then conducted pairwise comparisons and found statistically significant differences between intensive continuous and the extensive discrete items ($t = -2.907$, $p = 0.005$, $CI = [-0.17594, -0.03592]$, $d = 0.327$), with a weak effect size. In the Spanish context, we cannot assume sphericity ($W = 0.923$, $p < 0.001$). Multivariate methods showed statistical differences, so we conducted pairwise comparisons, obtaining significant differences between all possible pairs of types of magnitudes (Table 6).

Therefore, in relation to the second objective, in the light of Table 5 and Table 6, we observe that in the Spanish context the results for Cor and Rev are consistent. Moreover, given the fact that in this context nearly all the errors are reversals (Table 3), this corroborates the explanatory power of the static comparison as a source of the reversal error in the Spanish context.

Note also that in the South African sample, reversal errors were barely made, and the highest rate of correct answers was obtained for discrete magnitudes and the lowest for intensive magnitudes. With Spanish students, the opposite happened (Table 3). Regarding this, the results about the variable Other complement those about Cor concerning the first objective. Indeed, in the Spanish group, the means of Other are so low for all magnitude types (Table 2) that it makes no sense to test whether there are differences between them. In the South African context, we cannot assume sphericity for the variable Other ($W = 0.914$, $p = 0.032$), so we begin by applying multivariate methods. As these methods showed statistical differences, we conducted pairwise comparisons and we got statistical differences for two pairs, with a moderate effect size in the difference between intensive continuous and extensive discrete magnitudes (Table 7). Notice also that in the pair in which we cannot reject the null hypothesis, even after having applied the Holm-Bonferroni correction, the p -value is less than 0.1.

Thus, although most errors in the South African context are not reversals, the type of magnitude still has an effect on the incidence of the errors. In addition, the performance by the South African students regarding these other kinds of errors is better with the extensive discrete magnitudes and worse with the intensive continuous ones. Comparisons of quantities with extensive discrete magnitudes are less abstract, while those for intensive continuous ones are the most abstract, with the extensive continuous in the intermediate position. This means that for the South African students of our sample, the more abstract the magnitudes involved were, the more difficult they found it to express the word problem into an equation. This did not happen in the Spanish sample, which contributes to the answer to the first research objective.

Discussion and concluding remarks

Overall, our results point out that the South African pre-service teachers in our sample outperform the ones in

TABLE 6: Differences depending on the type of magnitude for the variable Cor in the Spanish context.

Difference variables for Cor	Mean	Standard deviation	$t(210)$	p	95% CI		Cohen's d
					LL	UL	
Intensive continuous – Extensive continuous	-0.03318	0.19691	-2.447	0.015*	-0.05990	-0.00645	0.16850
Intensive continuous – Extensive discrete	-0.07820	0.24113	-4.711	0.001	-0.11092	-0.04548	0.32431
Extensive continuous – Extensive discrete	-0.04502	0.25069	-2.609	0.04*	-0.07904	-0.01100	0.17958

CI, confidence interval; LL, lower limit; UL, upper limit.

* $p < 0.01$

TABLE 7: Differences depending on the type of magnitude for the variable Other in the South African context.

Difference variables for Other	Mean	Standard deviation	$t(210)$	p	95% CI		Cohen's d
					LL	UL	
Intensive continuous – Extensive continuous	0.07278	0.29730	2.176	0.066	0.00619	0.13938	0.2448
Intensive continuous – Extensive discrete	0.12658	0.27114	4.150	0.001**	0.06585	0.18731	0.4668
Extensive continuous – Extensive discrete	0.05380	0.22882	2.090	0.040*	0.00255	0.10505	0.2351

CI, confidence interval; LL, lower limit; UL, upper limit.

* $p < 0.05$; ** $p < 0.01$

the Spanish sample when expressing proportionality word problems in terms of equations. Indeed, although the results are based on descriptive statistics (Table 2 and Table 3), it is worthwhile to note that we obtained consistent results regardless of whether we considered statements with or without syntactic obstruction and across all types of magnitude. But the difference between the groups goes beyond a mere difference in algebraic language proficiency. Indeed, the type of errors were distinct between contexts.

Firstly, our results point out that the errors when expressing proportionality word problems in terms of equations differ between the South African and Spanish pre-service teacher participants. Specifically, in the Spanish case nearly all the errors were reversals, while in the South African case there were other errors with a higher incidence. Furthermore, even restricting our attention to the reversal error, its sources are different depending on the country. In the Spanish case the word order matching and the syntactic comparison would both have some explanatory power for the reversal error, which is in line with other results about Spain (González-Calero et al., 2015; González-Calero et al., 2020; Soneira et al., 2018) and the United States (Clement, 1982; Clement et al., 1981; Cohen & Kanim, 2005; Fisher et al., 2011; Martin & Bassok, 2005; Wollman, 1983). In the South African context, just the word order matching and not the static comparison would have some explanatory power. The role of the word order matching and the visual-spatial display of the text and the equation has been highlighted in prior research by Fisher et al. (2011) and Landy et al. (2014). However, these previous studies did not address the explanatory power of the static comparison. Moreover, among the studies conducted in English, just those by Landy et al. (2014) and Sung-Hee et al. (2014) consider continuous magnitudes, but even these studies do not analyse the role of the type of magnitude.

The results of our study contribute to improve our understanding of translation of word problems involving multiplicative comparisons into equations in two ways. Firstly, we have shown that the South African students also have the tendency to keep the structure or the ordering of meaningful components of the statements in natural language when changing the register of representation, even if this leads them to make errors. Note that in this sample 90% of the students spoke isiZulu as a first language; this therefore shows that isiZulu-speaking students also display the tendency of the word order strategy. This tendency has been identified in studies with students who speak English (e.g. Cohen & Kanim, 2005; Fisher et al., 2011; Sung et al., 2014) and those who speak Spanish (González-Calero et al., 2015; González-Calero et al., 2020; Soneira et al., 2018). Our study now extends this result to students who speak isiZulu. Regarding this, although it was 90% likely that the mother tongue of a part of the South African students was isiZulu, when translating into equations from statements written in English, we observed that the word order matching had explanatory power. Even more, the effect size regarding the

syntactic obstruction was higher in the South African ($d = 0.379$) than in the Spanish context ($d = 0.260$).

Therefore, our results in the South African context point out that the tendency to apply the word order matching does not depend on the context but it is common for different contexts. On the one hand, this is in line with the theoretical ideas of Duval (2017) or Kaput (1987) about the special relevance of the natural language in the mathematical activity. On the other hand, as word order matching does not require understanding the mathematical relationship, the fact that the tendency to apply the word order matching does not depend on the context is consistent with the natural tendency to save cognitive effort (Uesaka & Manalo, 2012).

Secondly, our study points out that the static comparison may be context dependent, diminishing its explanatory power as a model for the reversal error despite being one of the classic models. Indeed, we did not find evidence of the static comparison being a tendency among South African pre-service teachers; therefore, it should be context dependent. This highlights the importance of word order matching as a source of the reversal error because it is shared by groups from different populations that differ regarding other strategies when structuring equations, specifically the static comparison.

It is also noteworthy that, in contrast to the students of the Spanish sample, the students of the South African sample committed the highest rate of errors with intensive continuous magnitudes, and the lowest with extensive discrete magnitudes. This suggests that the more abstract the magnitude, the harder they find it to mathematise the situation. Even more, when dealing with discrete magnitudes they perform clearly better than Spanish participants, but the difference fades with the items that have intensive magnitudes. This is another difference between contexts. It could be due to differences in curriculum or instructional approaches at school level, although more research about the matter is needed.

Moreover, success in comparisons of quantities with extensive continuous magnitudes, which are easy to visualise and to reduce to discrete magnitudes by means of taking units of measurement, placed between the other two types. This could mean that some students spontaneously apply a process of discretisation whenever the magnitude's unit of measurement allows a corporeal interpretation. For example, in the cocktail problem, it is possible to think in terms of five tablespoonfuls of tonic for each tablespoonful of gin. But, if we do so, the situation and the subsequent reasonings would be the same as with discrete magnitudes. Indeed, the new magnitudes would be the number of tablespoonfuls of tonic and gin. Note that this process cannot be carried out with intensive magnitudes. In this case, the quantity of substance does not correspond to the quantity of magnitude. For example, the density of the sauce can be greater than that of the water, but the quantity of water greater than that of the sauce. Thus, it would be more difficult to mentally corporealise the unit of measurement, which in turn would

make discretisation more difficult. Soneira et al. (2018) reported that the discretisation may affect the reversal error making. In the present work we see that this phenomenon may affect the solver's approach to the problem in a more general way, which suggests that it could be a solving strategy. Even more, the discretisation process seems to be common to the South African and the Spanish contexts, beyond the fact that the subsequent solving strategies differ.

Regarding curriculum and instructional implications, our results indicate that the teaching interventions to be conducted in order to remediate the reversal error should be different depending on the country. Specifically, Spanish educators should take into account both the word order matching and the static comparison strategies. On the one hand, when first introducing the algebraic language in Spanish schools, the study suggests that teachers need to highlight the meaning of the equal sign as an algebraic equivalence instead of correspondence, and the fact that letters represent quantities instead of being labels for objects. This would be so because these aspects have been proposed as the reasoning underlying the use of the static comparison strategy (Clement, 1982; Cohen & Kanim, 2005; Fisher, 1988; González-Calero et al. 2020; Soneira et al., 2018; Sung-Hee et al., 2014). Teachers in Spain need to pay attention to the tendency to keep the structure or the visual display of the sentences in natural language when expressing them as equations, as this would underlie the use of the word order matching strategy (Clement, 1982; Fisher et al., 2011; Landy et al., 2014). For their part, teachers in South Africa could focus more on preventing the word order matching.

On the other hand, the answers to problems with statements presented with syntactic obstruction point out that the cause of most errors made by the South African students was not word order matching. The errors they made, equations that did not make sense in the problem situation, suggest that their mental scheme to express multiplicative relationships by means of equations is poorly developed. Thus, although South African pre-service teachers' performance was, overall, acceptable or, at least, better than that of the Spanish ones, there was a small subgroup of them that seem to lack the rudiments of algebraic language. This suggests that when students are introduced to the concept of equation during the intermediate and senior phase schooling, it should be done in a manner that allows them to develop a conceptual understanding and to make meaning of the notion of equation, instead of proving rules for manipulation of equations.

In terms of the study's limitations, it would have been preferable to have a larger sample size, particularly for the South African context. Given that we did not find evidence of the static comparison being a tendency in the South African context and that it means a key difference with respect to others (e.g. Spain, United States), it may be convenient to increase the sample. Furthermore, it would be convenient to conduct similar experiments in the South African context,

both with pre-service teachers and at other educative levels. In addition, since we have detected that South African students barely commit reversal errors, but other types of errors, and that the source of the former seems not to be static comparison, further studies could be interesting. In particular, qualitative studies would be helpful to provide deeper insights into the present quantitative one. These would shed light on how students think when translating proportionality word problems into equations. It could also be of interest to identify whether the students' specialisation phase affects the extent to which the word order matching strategy usually arises to prevent the reversal error. Along similar lines, it could be interesting to identify whether South African students' specialisation phase is related to the extent to which they face difficulties with intensive continuous magnitudes.

Acknowledgements

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced the writing of this article.

Authors' contributions

This article was developed collaboratively by the three authors. C.S. led the conceptualisation of the article, led the experimental design, carried out the analysis of the data, and drew up the results and discussion. S.B. led in the introduction, literature review, collaborated on the analysis of the South African data, and collaborated in preparing the final version of the discussion. R.G. developed the tool, cleaned and coded the data, led the methodology, technical aspects related to standard setting, and prepared the manuscript to comply with publication standards.

Funding information

This work was supported by Spanish Government through the project PGC2018-096463-B-I00 as well as the National Research Foundation (NRF) grant number UID 118377.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Disclaimer

The views, opinions, findings and conclusions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the authors.

References

- Barbieri, C.A., Miller-Cotto, D., & Booth, J.L. (2019). Lessening the load of misconceptions: Design-based principles for algebra learning. *Journal of the Learning Sciences*, 28(3), 381–417. <https://doi.org/10.1080/10508406.2019.1573428>
- Barmby, P. (2009). *Primary mathematics: Teaching for understanding*. Maidenhead: McGraw-Hill Education.

- Booth, L.R. (1988). Children's difficulties in beginning algebra. In A.F. Coxford (Ed.), *The ideas of algebra, K-12* (pp. 20–32). Reston, VA: National Council of Teachers of Mathematics.
- Bush, S.B., & Karp, K.S. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior*, 32, 613–632. <https://doi.org/10.1016/j.jmathb.2013.07.002>
- Clement, J. (1982). Algebra word problem solutions: Thought processes underlying a common misconception. *Journal for Research in Mathematics Education*, 13(1), 16–30. <https://doi.org/10.2307/748434>
- Clement, J., Lochhead, J., & Monk, G. (1981). Translation difficulties in learning mathematics. *The American Mathematical Monthly*, 88(4), 286–290. <https://doi.org/10.1080/00029890.1981.11995253>
- Cohen, E., & Kanim, S.E. (2005). Factors influencing the algebra reversal error. *American Journal of Physics*, 73(11), 1072–1078. <https://doi.org/10.1119/1.2063048>
- Crowley, L., Thomas, M., & Tall, D. (1994). Algebra, symbols, and translation of meaning. In *Proceedings of the 18th Conference of the International Group for the Psychology of Mathematics Education* (pp. 240–247). Lisbon, 29 July–03 August.
- Duval, R. (2017). *Understanding the mathematical way of thinking – The registers of semiotic representations*. Cham: Springer.
- Ellis, P. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511761676>
- Filloy, E., Rojano T., & Puig, L. (2008). *Educational algebra: A theoretical and empirical approach*. New York, NY: Springer.
- Fisher, K.J., Borchert, K., & Bassok, M. (2011). Following the standard form: Effects of equation format on algebraic modelling. *Memory & Cognition*, 39(3), 502–515. <https://doi.org/10.3758/s13421-010-0031-6>
- Fisher, K.M. (1988). The students-and-professors problem revisited. *Journal for Research in Mathematics Education*, 19(3), 260–262. <https://doi.org/10.2307/749069>
- Fyfe, E.R., Matthews, P.G., & Amsel, E. (2020). College developmental math students' knowledge of the equal sign. *Educational Studies in Mathematics*, 104, 65–85. <https://doi.org/10.1007/s10649-020-09947-2>
- Given, L.M. (Ed.). (2008). *The Sage encyclopedia of qualitative research methods*. London: Sage.
- 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(1), 133–147. <https://doi.org/10.1007/s10649-015-9596-0>
- González-Calero, J.A., Berciano, A., & Arnau, D. (2020). The role of language on the reversal error. A study with bilingual Basque-Spanish students. *Mathematical Thinking and Learning*, 22(3), 214–232. <https://doi.org/10.1080/10986065.2020.1681100>
- Humberstone, J., & Reeve, R. (2018). The conceptual overlap between arithmetic and algebraic referential mapping. *Learning and Instruction*, 54, 138–146. Retrieved from <https://psycnet.apa.org/doi/10.1016/j.learninstruc.2017.09.001>
- 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. <https://doi.org/10.1007/BF00311062>
- Kieran, C. (2007). Learning and teaching algebra at the middle school through college levels: Building meaning for symbols and their manipulation. In F.K. Lester, Jr (Ed.), *Second handbook of research on mathematics teaching and learning*. (pp. 707–762). Greenwich: Information Age.
- Koedinger, K.R., & Nathan, M.J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences*, 13(2), 129–164. https://doi.org/10.1207/s15327809jls1302_1
- Knuth, E.J., Alibali, M.W., Hattikudur, S., McNeil, N.M., & Stephens, A.C. (2008). The importance of equal sign understanding in the middle grades. *Mathematics Teaching in the Middle School*, 13(9), 514–519. <https://doi.org/10.5951/MTMS.13.9.0514>
- Landy, D., Brookes, D., & Smount, R. (2014). Abstract numeric relations and the visual structure of algebra. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1404–1418. <https://doi.org/10.1037/a0036823>
- Landy, D., & Goldstone, R.L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 720–733. <https://doi.org/10.1037/0278-7393.33.4.720>
- Lewis, A.B., & Mayer, R.E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology*, 79(4), 363–371. <https://doi.org/10.1037/0022-0663.79.4.363>
- MacGregor, M., & Stacey, K. (1993). Cognitive models underlying students' formulation of simple linear equations. *Journal for Research in Mathematics Education*, 24(3), 217–232. <https://doi.org/10.2307/749345>
- Malisani, E., & Spagnolo, F. (2009). From arithmetical thought to algebraic thought: The role of the "variable". *Educational Studies in Mathematics*, 71, 19–41. <https://doi.org/10.1007/s10649-008-9157-x>
- Matthews, P., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign as an indicator of mathematical equality. *Journal for Research in Mathematics Education*, 43(3), 316–350. <https://doi.org/10.5951/jresmetheduc.43.3.0316>
- Martin, S.A., & Bassok, M. (2005). Effects of semantic cues on mathematical modelling: Evidence from word-problem solving and equation construction tasks. *Memory & Cognition*, 33(3), 471–478. <https://doi.org/10.3758/BF03193064>
- Ministerio de Educación y Formación Profesional (MEFP). (2019). *PISA 2018. Programa para la Evaluación Internacional de los Estudiantes. Informe español*. Retrieved from https://sede.educacion.gob.es/publivena/descarga.action?f_codigo_agc=20372
- Ministerio de Educación y Formación Profesional (MEFP). Gobierno de España. (2020). *TIMSS 2019. Estudio Internacional de Tendencias en Matemáticas y Ciencias. Informe español*. Retrieved from https://sede.educacion.gob.es/publivena/descarga.action?f_codigo_agc=21925&request_locale=en
- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. London: Routledge.
- Pardo, A., & San Martín, R. (2010). Análisis de datos en ciencias sociales y de la salud II (2nd ed.). [Data analysis in social sciences life sciences II, 2nd edition]. Madrid: Editorial Síntesis.
- Reddy, V., Prinsloo, C., Arends, F., Visser, M., Winnaar, L., Feza, N., et al. (2012). *Highlights from TIMSS 2011: The South African perspective*. Cape Town: HSRC Press. Retrieved from <http://www.hsrc.ac.za/en/research-data/ktree-doc/12417>
- Sims-Knight, J.E., & Kaput, J.J. (1983). Exploring difficulties in transforming between natural language and image based representations and abstract symbol systems of mathematics. In D. Rogers & J.A. Sloboda (Eds.), *The acquisition of symbolic skills* (pp. 561–570). Boston, MA: Springer.
- Soneira, C., González-Calero, J.A., & Arnau, D. (2018). An assessment of the sources of the reversal error through classic and new variables. *Educational Studies in Mathematics*, 99(1), 43–56. <https://doi.org/10.1007/s10649-018-9828-1>
- South African Institute of Race Relations (SAIRR). (2016). *Reasons for hope*. Johannesburg: SAIRR. Retrieved from <https://irr.org.za/reports/occasional-reports/files/race-relations-in-sa-2013-reasons-for-hope-2013-29feb16.pdf>
- 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(1), 12–22. <https://doi.org/10.1016/j.ijhcs.2013.09.010>
- Stacey, K., & MacGregor, M. (1999). Learning the algebraic method of solving problems. *The Journal of Mathematical Behavior*, 18(2), 149–167. [https://doi.org/10.1016/S0732-3123\(99\)00026-7](https://doi.org/10.1016/S0732-3123(99)00026-7)
- Swan, M. (2001). Dealing with misconceptions in mathematics. In P. Gates (Ed.), *Issues in mathematics teaching* (pp. 147–165). London: Routledge Falmer.
- Uesaka, Y., & Manalo, E. (2012). Task-related factors that influence the spontaneous use of diagrams in math word problems. *Applied Cognitive Psychology*, 26(2), 251–260. <https://doi.org/10.1002/acp.1816>
- Wollman, W. (1983). Determining the sources of error in a translation from sentence to equation. *Journal for Research in Mathematics Education*, 14(3), 169–181. <https://doi.org/10.2307/748380>