

Analysing Mathematical Word Problem Solving with Secondary Education CLIL Students: A Pilot Study

Análisis de la resolución de problemas con estudiantes AICLE de educación secundaria: un estudio piloto

Análise da resolução de problemas com alunos CLIL do ensino médio: um estudo piloto

Víctor Pavón

Universidad de Córdoba, España.

victor.pavon@uco.es

<https://orcid.org/0000-0003-3063-1746>

Rafael Cabezuelo

Instituto de Enseñanza Secundaria López-Neyra de Córdoba, España.

rafael.cabezuelo@ieslopezneyra.com

<https://orcid.org/0000-0003-4518-7617>

Received: 05/02/2019

Sent to peer review: 04/03/2019

Accepted by peers: 09/04/2019

Approved: 23/04/2019

DOI: 10.5294/laclil.2019.12.1.2

To reference this article (APA) / Para citar este artículo (APA) / Para citar este artigo (APA)

Pavón, V. & Cabezuelo, R. (2019). Analysing mathematical word problem solving with secondary education CLIL students: A pilot study. *Latin American Journal Of Content & Language Integrated Learning*, 12(1), 18-45. DOI: <https://doi.org/10.5294/laclil.2019.12.1.2>

ABSTRACT. The purpose of this study is to investigate to what extent the use of L2 in math tests influences bilingual education learners' process of word problem solving in a mandatory secondary education school with Content and Language Integrated Learning (CLIL). The reading comprehension level of the students was analysed using a standards-based assessment and the questions used in Programme for International Student Assessment (PISA) tests. The word problems were selected according to the students' level of reading-comprehension and mathematical competence. Learners also had to answer a questionnaire, which was used to analyse if contextual factors were affecting mathematical performance in L2. To this end, the questionnaire included some questions related to the bilingual history of the students and their perception about solving word problems in English. Data were analysed through one-way or two-way ANOVA tests to find out which factors were relevant. Results show that solving word problems is not only affected by the use of L2, but that it also depends on the mathematical difficulty, irrespective of the students' level of language proficiency. The findings, hence, imply that interaction between linguistic difficulty and mathematical complexity is at the centre of the issues affecting word problem solving.

Keywords (Source: Unesco Thesaurus): CLIL; mathematics; secondary education; word problem solving; contextual factors; reading comprehension; student evaluation.

RESUMEN. El objetivo de este estudio es investigar hasta qué punto el uso de la L2 en las pruebas de matemáticas influye en el proceso de resolución de problemas por el alumnado. Esta investigación ha tenido lugar en un instituto de educación secundaria obligatoria que sigue la metodología del Aprendizaje Integrado de Contenidos y Lenguas Extranjeras (AICLE). Se analizó el nivel de comprensión lectora del alumnado por medio de una evaluación basada en estándares, utilizando las preguntas de los cuestionarios de PISA. Los problemas se seleccionaron de acuerdo con el nivel real de comprensión lectora y de competencia matemática de los estudiantes. Además, se administró un cuestionario para analizar si algunos factores contextuales estaban afectando al desempeño matemático en la L2, en el cual se incluyeron algunas cuestiones relacionadas con la historia bilingüe del alumnado y de su percepción con relación a la resolución de problemas en inglés. Los datos se analizaron mediante pruebas ANOVA (uno o dos factores) para analizar qué factores eran relevantes. Los resultados muestran que la resolución de problemas no solo se ve afectada por el uso de la L2 sino que también depende de la dificultad matemática, independientemente del nivel de dominio del idioma de los alumnos. Los hallazgos sugieren, por tanto, que la interacción entre la dificultad lingüística y la complejidad matemática es uno de los principales aspectos que afectan a la resolución de problemas.

Palabras clave (Fuente: tesauro de la Unesco): CLIL; AICLE; matemáticas; enseñanza secundaria; resolución de problemas; factores de contexto; comprensión lectora; evaluación del estudiante.

RESUMO. O objetivo deste estudo é investigar até que ponto o uso de L2 nos testes de matemática influencia o processo de resolução de problemas dos alunos. Esta pesquisa foi realizada em um instituto de ensino médio obrigatório que segue a metodologia CLIL (Content and Language Integrated Learning, em inglês). Analisou-se o nível de compreensão leitora dos alunos por meio de uma avaliação baseada em padrões, utilizando as questões dos questionários do PISA. Os problemas foram selecionados de acordo com o nível real de compreensão de leitura e de competência matemática dos alunos. Além disso, foi aplicado um questionário para analisar se alguns fatores contextuais estavam afetando o desempenho matemático em L2, o qual incluiu algumas questões relacionadas à história bilingue dos alunos e sua percepção em relação à resolução de problemas em inglês. Os dados foram analisados por meio de testes ANOVA (um ou dois fatores) para analisar quais fatores eram relevantes. Os resultados mostram que a resolução de problemas não é afetada apenas pelo uso de L2, mas também depende da dificuldade matemática, independentemente do nível de proficiência linguística dos alunos. Os resultados sugerem, portanto, que a interação entre dificuldade linguística e complexidade matemática é um dos principais aspectos que afetam a resolução de problemas.

Palavras-chave (fonte: tesauro da Unesco): CLIL; matemática; ensino médio; resolução de problemas; fatores de contexto; compreensão de leitura; avaliação do aluno.

Introduction

The interrelation between language and content matter in mathematics has a major significance: “Language-dependent knowledge representations are particularly evident in the domain of mathematics learning” (Grabner, Saalbach, & Eckstein 2012, p. 147). It occurs so because, on one hand, the language of mathematics is particularly problematic (Morgan, 2007) due to its unique register (Halliday, 1978) and, on the other hand, because mathematical processing demands a great amount of procedural knowledge essentially linked with language (Hiebert & Lefevert, 1987).

Particularly regarding the topic of this study, Swetz (2012) points out that word problems are frequently used to teach mathematics because they have been used as the primary means of instruction for thousands of years. Also, according to Bernardo (2002), word problems have always been an important part of mathematics education, in which the linguistic component is a fundamental part because problems are embedded within a text and difficulties are related to terminology: “formal mathematical language is characterized by lack of redundancy and refers to the standard use of terminology (mathematical register)” (Novotná, Háj-Mousová, & Hofmannová, 2005, p. 3). Providing an adequate assessment of mathematical performance in an immersion context of bilingual education is the key, as it is likely to give a misleading impression of the students’ academic abilities. Comprehension of content through L2 has been investigated with primary education learners in bilingual settings, including CLIL, particularly with respect to its significance when assessing mathematics skills (Abedi & Lord, 2001; Cummins, Kintsch, Reuser, and Weimer, 1988; Jiménez-Jiménez, 2015; Kempert, Saalbach, & Hardy, 2001; Moschkovich, 2005, 2007, 2015; Ouazizi, 2016; Smit & Van Eerde 2011). Besides, investigating the effect of language proficiency in the programmes where mathematics is conducted in English (as the L2) at other levels of education is increasingly frequent (Adanur, Yagiz, & Isik, 2004; Dalrymple, KaraGiannakis, & Papadopoulos, 2012; Grabner, Saalbach, & Eckstein, 2012; May & Ong, 2011; Spelke & Tsivkin, 2001; Surmont, Struys, Ven Den Noort, & Van Der Kraen, 2016; Tavares, 2015; Van Rinsveld, Brummer, Landerl, & Schiltz 2016).

The context where this investigation takes place — the bilingual education programme in the Spanish Region of Andalusia — leads to an interweaving language with content as proposed in CLIL. Although previous research indicates that there could be an impact on mathematical performance when word problems are presented in L2, this particular area has not yet been studied in secondary education in the Andalusian context. After more than a decade of bilingualism in the region, it is important to know whether or not the proficiency level in the L2 in mathematics influences assessment of content matter. Specifically, the main objective is to weigh the importance of L2 in word problem solving and to find the other variables at work in the students' performance in mathematics. During the next sections, the theoretical foundations for this study will be described, namely the interactions between language and content (in L1 and L2) and interaction of L2, word problem solving processes, impact of reading comprehension proficiency level, and the study of socio economic factors and their relation to academic achievement. Analysis of the data will help obtain several useful organizational and pedagogical conclusions with respect to assessment of students' performance in mathematics.

Literature Review

Interaction between language and content in word problems

One of the most difficult assignments that a bilingual education teacher has to face is the balance between the linguistic and the cognitive demands of students' tasks: "the goal of professional development for content area teachers should be to frame their planning in terms of the question: What is the language my students need to succeed in this task?" (Hansen-Thomas, Langman, & Farias, 2018, p. 211). As Cummins (2000) states, language and content will be acquired most successfully when students are challenged cognitively and provided with the contextual and linguistic scaffolding required for successful task completion. Hence, linguistic needs should be enough to make advancement

possible, but not too high to avoid development of the content. Lyster (2007) provides another relevant contribution to this issue, noting that students constantly shift their attention between form and meaning, and that there is a need to construct a solid and balanced link between them by stimulating students' awareness of learning both language and content.

Mathematics, as every other academic discipline, has a complex and specialized language that is different from everyday conversation (Sigley & Wilkinson 2015). Therefore, proficiency in language may affect mathematical achievement on instructional and testing levels (Kempert, Saalbach, & Hardy 2011). The CLIL Matrix (Coyle, Hood, & Marsh, 2010) explains the relationship between language and content, as observed in Figure 1.

Figure 1. The CLIL matrix

Cognitive demands	High	2	3
	Low	1	4
		Low	High
		Linguistic demands	

Source: Coyle et al. (2010).

Ideally, the focus should be directed to quadrant 2 because the language is not going to impede learning. Nevertheless, moving periodically to quadrant 3 will lead to a progression in language learning without affecting the cognitive challenge of the learner. Quadrant 4 aims at fostering the linguistic potential of the students, whilst the specific work included in quadrant 1 could help build initial confidence in learners (Coyle et al., 2010). Word problems could be included in quadrant 3,

and this should be a goal to achieve in the linguistic performance of our learners. It may happen when the linguistic demands are too high for a particular group of students. Adaptations can be made in order to assess their mathematical skills, hence moving to quadrant 2. Otherwise, if the task is in quadrant 2, there is risk that the linguistic proficiency of the students is not properly developed.

According to Abedi (2002), the linguistic complexity of test items unrelated to the content being assessed may at least be partly responsible for the performance gap between non-native speakers and native speakers in an immersion context. Problematic features appear at three levels: syntactic (involving complex sentences, multiple subordinate clauses, nested constructions, long noun phrases, and passive voice), lexical (concerning unfamiliar words, unfamiliar phrases, and unfamiliar connotations of words with multiple meanings), and background (focusing on knowledge in word problems, sentence and paragraph level, and word phrases) (Moschkovich, 2015).

The link between language and mathematics is especially evident in the case of word problems. It is known that children's performance on arithmetic word problems is a reliable predictor of their subsequent mathematical competence (Kempert, Saalbach, & Hardy 2011). In fact, a highly technical, precise and densely structured language is required in the mathematics register (Sigley & Wilkinson, 2015), where solving arithmetic problems is a cognitive task that relies on language processing (Van Rinsveld, Brunner, Landerl, Schiltz, & Ugen, 2015). Thus, the learning of mathematics is more strongly related to language processes than previously assumed (Kempert, Saalbach, & Hardy, 2011; Surmont et al., 2016). According to Bialystok (2001), as language and mathematics share common critical features, such as abstract mental representation, conventional notations, and interpretive function, mathematics is a domain where cognitive effects on bilinguals are likely to occur.

Language proficiency affects mathematical achievement at the levels of instruction and testing (Kempert, Saalbach, & Hardy 2011). Hence, both oral and written language are central to mathematical teaching and learning, which rely on the discourse structures of description, sequencing, procedural iteration, and justifications (Sigley & Wilkinson, 2015). Furthermore, several studies state that the field of

mathematics reasoning is consistent with threshold hypothesis (Cummins, 2000). Even when students may not be in full command of the mathematics register until they understand math (Sigley & Wilkinson 2015), presenting mathematic problems with simplified linguistic instructions (Abedi & Lord 2001) seemed to help them overcome their linguistic complexity (Van Rinsveld et al. 2016). Following Bialystok (2001), a generous interpretation of studies is that, if language proficiency is at least adequate for understanding the problem, bilingualism has no effect on mathematical problem solving.

Language proficiency and mathematical complexity

The impact of a high linguistic proficiency on mathematics has been generally reported as positive: “bilingual pupils have an advantage in mathematics when they are highly competent in both languages, compared to their monolingual peers” (Surmont et al., 2016, p. 322). From a different perspective, Abedi and Lord (2001) also show that there is a real interaction between language and mathematics achievement. Their research included a math test with the items that are simplified in terms of language, and with modified vocabulary and linguistic structures not related to mathematics, which are compared to the original ones. They claim that this interplay has to be considered in mathematics assessment research and practice, also that language adjustments must be considered carefully. To this end, Lorenzo (2008) establishes three different categories that teachers in bilingual education contexts use: simplify the text (as it could make the test almost meaningless), elaborate it (as an attempt to reduce complexity), or re-discursify (making the outcome a student-centred text with coherent remaining ideas and enough complexity to challenge the student learning processes).

Whereas there are specific grammatical patterns to the mathematics register that include dense noun phrases, subordinators, nominalizations, logical connectors and verbs employed in arguments, justifications and constructions of mathematical ideas (Sigley & Wilkinson, 2015), it is suggested that simplifying the linguistic structure of word problems presented in L2 increases the mathematical performance of learners in L2 (Abedi & Lord, 2001). Other comprehension problems at surface level (i.e., an unknown verb or noun) may result in an extra

load of memory resources that are therefore not available for arithmetic calculations (Kempert, Saalbach, & Hardy, 2011).

Given the difficulties, a cognitive modelling process is required to solve mathematics word problems, in which students identify and extract relevant pieces of information and, at the same time, suppress any misleading or irrelevant linguistic or numerical information (Kempert, Saalbach, & Hardy, 2011) that is embedded in the problem context. According to a prominent model of arithmetic word problem solving, the whole process takes place within three steps: forming a situational model by structuring its relevant features, extracting a mathematical problem model (translation from linguistic code to mathematical relations), and, finally, do calculations, interpret, and validate results (Kempert, Saalbach, & Hardy 2011).

As shown above, variations in the linguistic aspect of word problems affect problem representation and problem-solving performance. A focus on the problem structure of word problems when the text is presented leads to significant improvements in problem-solving accuracy (Bernardo, 2002). The author found that Filipino-English bilingual students solved word problems when they were written in their first language (Filipino) because they understood the text. He also finds that rewording the problem texts resulted in smaller gains in accuracy with the problems worded in English compared with the problems in Filipino. In other words, bilingual education students who have to solve word problems in a foreign language may have more difficulties understanding them. Kempert, Saalbach, and Hardy (2011) have also found that proficiency in the language of testing had the strongest influence on the students' mathematical achievement in both L1 and L2. At the same time, bilingual education students also partly compensated their language deficits by displaying an enhanced ability in attentional control.

According to Abedi and Lord (2001), the following features can be revised in order to adapt the linguistic register of mathematical word problems to the linguistic level of students:

- » Familiarity or frequency of non-math vocabulary: change of unfamiliar or infrequent words
- » Voice of verb phrases: change passive to active verb forms
- » Length of nominals: shorten long nominals
- » Conditional clauses: separate conditional sentences or change the

- order of conditional and main clause
- » Relative clauses: remove or recast
- » Question phrases: change complex question phrases to simple question words
- » Abstract or impersonal presentations: make information more concrete

Some changes could involve more than one feature.

However, as Lorenzo (2008) points out, linguistic adaptations must be taken with care, as undesired outcomes may appear in simplification or elaboration processes.

Kempert, Saalbach, and Hardy (2011) state that proficiency in the instructional language and arithmetic skills can strongly predict students' ability to solve mathematical word problems. In order to classify students by linguistic proficiency, reading comprehension could be used as a suitable instrument (Schleicher, Zimmer, Evans, & Clements, 2009). For this purpose, a standards-based assessment process becomes a fundamental tool to observe if differences among students can be set while a questionnaire can be passed to determine participants' language histories and to obtain an accurate picture of their degree of bilingualism (Jiménez-Jiménez, 2015).

Method

This study can be classified as *ex post facto* because the independent variables — e.g., the language in which word problems are presented or the reading comprehension according to the Common European Framework of Reference for Languages (CEFR) level — are not manipulated. A part of the study is descriptive with the use of the context questionnaire, which includes information that must also be considered as an empirical variable. Research question 1 is answered by studying the effect of one independent variable so they can be considered as prospective with one independent, simple variable. Research question 2 analyses the effect of two independent variables over the dependent variable, so it can be considered as prospective with more than one independent, factorial variable.

Table 1 presents how the main variables have been used, further including the statistical test used to answer each research question, i.e., one-way or two-way ANOVA.

Table 1. Variables, design methodology, and statistical tests

R.Q.	Variables		Design methodology	Statistical test
	Independent	Dependent		
1	Language	Score	Simple	One-way ANOVA
2	Language and mathematical difficulty	Score	Factorial	Two-way ANOVA
	Highly difficult language	Score	Simple	One-way ANOVA
	Highly difficult language	Score	Simple	One-way ANOVA
Notes. R.Q.: Research question. The design methodology is empiric with quantitative methodology, ex post facto with a prospective research.				

Source: Own elaboration.

Objectives

The main purpose of this study is to investigate if using the foreign language influences the word problem solving process when assessing the mathematical content. Following the aims of this paper, two research questions have been elaborated as follows:

1. Does the L2 level of students influence mathematical word problem solving?
2. How does mathematical complexity interact with language in the word problem solving process?

Context and participants

The study was conducted at a high school in the city of Córdoba (Andalusia, Spain) where bilingual education follows a CLIL approach. Two groups of bilingual education learners, with a total of 53 students consisting of 29 girls and 26 boys in the fourth year of secondary education (4th grade of compulsory secondary education or 10th grade) participated in the study. All of them were native speakers of Spanish who were

learning English as their second language and living in monolingual families (Spanish-speaking parents). The majority of students enrolled in this bilingual education programme when they started secondary education (1st grade of compulsory secondary education or 7th grade). Originally, the two groups were composed of 27 and 28 students, respectively, for a total of 55. Eventually, two students were not included in the investigation because they had failed several subjects, including Mathematics and English as a Foreign Language. The mean age of the students on the day the PISA test was conducted was 15.55 years old.

As for the general context concerning the bilingual education programme in Andalusia in secondary education, high schools with bilingual education must have at least two non-language subjects (which must cover a minimum of 30% of the schedule per week — 9 out of 30 hours —) in a foreign language. At least 50% of the content in those subjects must be instructed in the L2. When assessing content subjects, the linguistic level of learners in L2 is only considered to improve their qualifications (*Orden 28 de junio*, 2011). Bilingual education teachers must have a B2, C1, or C2 level of proficiency according to the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001), although C1 will be required in the future (Junta de Andalucía, 2016). Schools with bilingual education have a linguistic assistant for 12 (full-time) or 6 (part-time) hours per week. Both bilingual education teachers and language assistants collaborate to promote oral conversation with students (*Orden 28 de junio*, 2011).

Data collection procedure

The research was carried out during the academic year of 2016–2017. A first data-collecting calendar was established with the criterion of making the administration of each type of paper within the same day possible. Another criterion was the lapse of time, trying to avoid differences in their English or mathematics proficiency level or in their perception of using English as a medium to learn mathematics. The students, parents and managing staff were informed about the study prior to the data collection process, and students signed their consent to participate. The students were also informed that they could voluntarily quit the study, and that they did not receive any rewards (financial or of any other type) for participating.

The three instruments used to collect the data were a reading comprehension level test, a questionnaire, and two versions of the same word problem test.

Reading comprehension test

Participants took a standards-based assessment B1 test from the book *Cambridge English - Preliminary: Reading Parts 1 to 5* (for a total of 35 questions) from *Reading and Writing Sample Set 6* (Cambridge, 2017a). Each correct answer was worth 1 mark, with a maximum of 35 possible marks (Cambridge, 2017b). The digitalized information consisted of the student's exclusive numeric identifier his or her group, his or her name, the score, and the date of the test. Three more fields were automatically calculated: CEFR Level (A2, B1 or B2) based on the score, a detailed CEFR Level (A2, B1.1 — minimum score of 25 —, B1.2 — minimum score of 28 — or B2) also based on the score and the score out of 10. A brief explanation of the text and the answer sheet was provided during the first 10 minutes of class. Students had 50 minutes to complete the test. The students' scores were compared to other marks obtained during the academic year in order to gain plausible scores.

Questionnaire

A paper-and-pencil questionnaire was elaborated in English based on different context questionnaires, such as PISA (OECD, 2016; Willms & Tramonte, 2015) and Andalusian Diagnostic Assessment Tests (*Pruebas de Evaluación de Diagnóstico*) (AGAEVE, 2017). It was structured in three parts. The first one (questions 1 to 4) provided information about the English language learning record of the participants; the second part (questions 5 to 8) was about the mathematical-bilingual education record and personal perceptions of content learning process through a second language; finally, the third part (questions 9 to 13) provided information about the familiar structure, home possessions (number of books at home) and parents' expectations, education and occupation. A brief explanation of the questionnaire was given during the first 10 minutes of the class. Students were told about the importance of the questions, that gathering individual perceptions and experiences was relevant, that there were no right or wrong answers, and that the utmost sincerity was needed. They had 50 minutes to complete it.

PISA word problem test

The mathematical word problem test was made using selected PISA-released mathematics items (INEE, 2013; OECD, 2006, 2013). Questions were available in both English and Spanish, and they were also balanced in terms of general difficulty. This level of difficulty was previously evaluated based on the percentage of right answers for the word problems in the OECD and Spain (INEE, 2013). Three word problems were pre-selected from each content area: arithmetic and algebra, geometry and functions, and graphics. To fit the extension of the test to the time available, one word problem from each content area was discarded, for a total of 6 word problems and 11 questions. To avoid the possible bias caused by the difficulty of the language or the order in which the word problems were presented to the students, questions were arranged in two booklets (paper A and paper B), both with six word problems, three in Spanish and three in English. The two booklets had the same word problems, but in different languages (i.e., if one problem is in Spanish in paper A, then the same problem is in English in paper B). The order of the word problems within each booklet was also randomized. Each question was marked according to the PISA test scoring with a maximum of 1 point each. Only two questions from word problem 4 (Ice-Cream Shop) had codes for intermediate scoring of 0.5 points, while the rest of the questions could only be marked as correct (1 point) or incorrect (0 points). The resulting information was digitalized in a spreadsheet containing the student's exclusive numeric identifier, his or her group, his or her name, and the type of paper (A or B) solved. A first sheet was used to introduce the code for each word problem number and question number, and this information was automatically converted into score values (0, 0.5, 1 or "-") in another sheet. One column had the score of the questions solved in Spanish and another had the score of the questions solved in English, adding the total out of the 11 maximum score in a third column. As paper A had 6 questions in Spanish and 5 in English and paper B was the other way around, scores were automatically transformed into a scale from 0 to 10 rounded to two decimal places to make all of them comparable. A brief explanation of the PISA word problems test and the sequence of each booklet were given within the first 10 minutes of the class. Then, they had 50 minutes to complete

it. All word problems in English had a glossary adapted to the B1 reading comprehension level of most of the students. The most important thing was to evaluate mathematical performance. Questions on the vocabulary were allowed, but not about any mathematical content or processes although a very few were made.

Results

In this section of the paper, we will try to address the research questions using the statistical techniques detailed in the previous section. All data were analysed using the IBM® SPSS® Statistics 24 software.

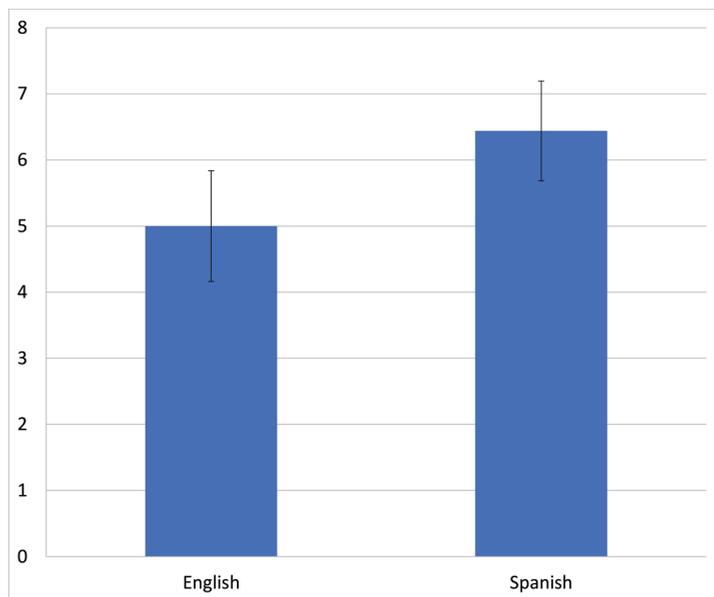
Research question 1: Does the L2 level of students influence solving mathematical word problems?

To answer the first research question, a one-way ANOVA was performed with language as the independent variable and the total score per language as the dependent variable. Each student has, therefore, two scores—one for the word problems written in English and the other one for the word problems written in Spanish. As shown in Figure 2, descriptive data indicate that the mean for the word problems written in English ($M = 5.000$; $SD = 3.039$) is lower than the mean for the word problems in Spanish ($M = 6.440$; $SD = 2.737$).

The ANOVA test was conducted in order to assess whether or not the difference was significant. Homogeneity of variances can be assumed with a p -value = 0.234 for the Levene test. The test revealed that the difference of the scores obtained in solving the word problems in English or Spanish was statistically significant ($F = 6.569 > F_{1,105,0.05} = 3.932$). The p -value = 0.012 confirmed that the scores in English were significantly lower than the scores in Spanish at 0.05 level of significance.

The importance of this finding is that the language affects the assessment of mathematical proficiency. This has to be taken into consideration when evaluating word problem solving processes in bilingual education contexts.

Figure 2. Mean and 95% confidence interval of the scores in PISA test papers in English and Spanish



Source: Own elaboration.

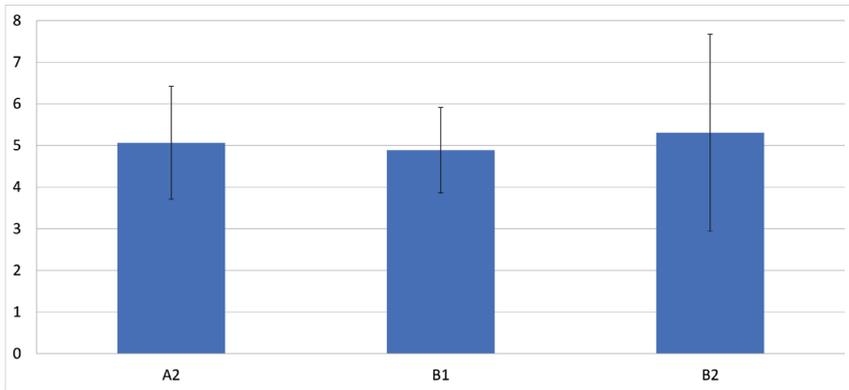
Two one-way ANOVA tests were also completed with the CEFR reading comprehension level as the independent variable, and the total score of the word problems written in English as the dependent value. As seen in Figures 3 and 4, descriptive data shows that the mean for the word problems written in English was similar for each CEFR level (Table 2).

Table 2. Mean and standard deviation in English scores per CEFR level (simple and detailed)

CEFR Level	M	SD	Detailed CEFR Level	M	SD
A2	5.066	1.092	A2	5.066	1.092
B1	4.889	3.030	B1.1	4.686	3.097
B2	5.306	3.724	B1.2	5.092	3.037
			B2	5.306	3.724

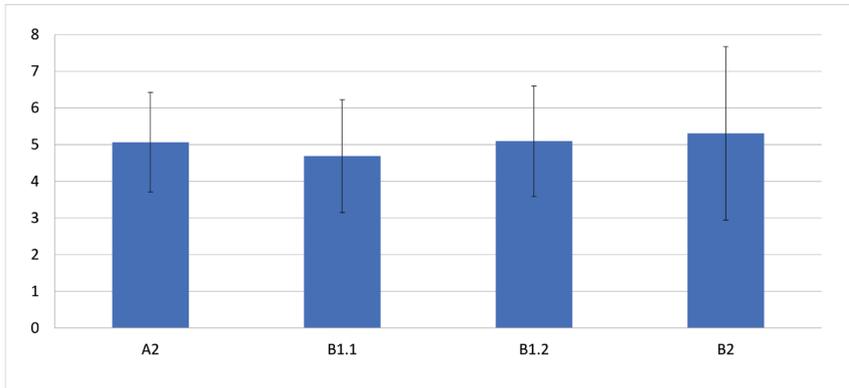
Source: Own elaboration.

Figure 3. Mean and 95% confidence interval of English scores per CEFR level



Source: Own elaboration.

Figure 4. Mean and 95% confidence interval of English scores per detailed CEFR level



Source: Own elaboration.

Two ANOVA tests—one for the simple and another for the detailed CEFR level—were made to establish whether or not the differences were significant. Homogeneity of variances cannot be assumed ($p\text{-value}_{\text{simple}} = 0.003$; $p\text{-value}_{\text{detailed}} = 0.005$ for the Levene tests), implying that other analysis must be done to confirm the results. The ANOVA tests indicated that the difference of the scores in English obtained in solving the word problems per CEFR level was not statistically significant ($F_{\text{simple}} = 0.083$; $F_{\text{detailed}} = 0.105$; both less than $F_{3,52,0.05} = 2.783$). Moreover, $p\text{-value}_{\text{simple}} = 0.921$ and $p\text{-value}_{\text{detailed}} = 0.957$ confirmed that the scores in English did not differ significantly between CEFR levels (simple or de-

tailed) at 0.05 significance level. The robust tests of equality of means (Brown-Forsythe) confirmed that the scores were not significantly different with a p -value > 0.898 .

Multiple comparisons were made by a *post hoc* test (Tamhane) to evaluate the effect of each pair of CEFR levels, resulting in no significant differences between any such pair, being simple with a minimum p -value = 0.980, or detailed with a minimum p -value = 0.998.

Although in the first research question it was posited that there might be differences in solving word problems in English compared to Spanish, it seems that the reading comprehension level does not affect the final score obtained when solving word problems in English. However, as for the finding that the reading comprehension level does not affect the score of the word problems written in English, students think otherwise. Other variables that learners perceive as influencing the word problem solving are the non-mathematical vocabulary and the greater time needed to find out the solution. To avoid the extra weight that these factors could have, a possible recommendation would be that students have both enough time to complete the test and vocabulary help, e.g., in the form of a glossary. However, as glossaries deal with technical terms and one of the objectives in CLIL is to express content material cohesively and in a meaningful way, it can be suggested that embedding word problem solving as another component of regular lessons could compensate the possible limitation of only using glossaries.

Research question 2: How does mathematical complexity interact with language in the word problem solving process?

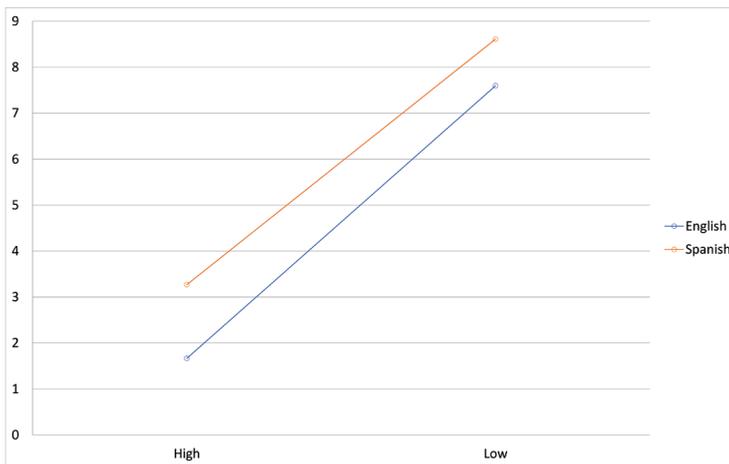
Interaction

It must be noted that the discussion on mathematical complexity should include the cognitive language processes that allow them to learn what was previously unknown. For this reason, emphasising the interaction between content and language by means of maximising communication becomes a truly decisive factor. The first analysis performed was a two-way ANOVA test with language and math difficulty as factors and the score obtained in each case as the dependent value. Questions were classified into high, medium and low mathematical difficulty according to the percentage of students with the maximum

score and the average score. Then, the word problems to be included were selected. The idea was to classify all the questions for each word problem according to cognitive difficulty, and that the maximum number of word problems could be arranged per language at a particular level (see Table 2). Word problem 6 was discarded because it had one low and one medium difficulty question, whereas word problem 1 was not selected because the rest of the low difficulty word problems (2 and 3) were written in a different language. Hence, the final selection included word problems 2 and 3 for low mathematical difficulty and word problems 4 and 5 for high mathematical difficulty.

Descriptive data (see Figure 5) coincide in the fact that the score was higher in Spanish than in English, although the difference between the two languages seems to be greater for word problems with the highest mathematical difficulty.

Figure 5. Mean score for word problems according to language and mathematical difficulty



Source: Own elaboration.

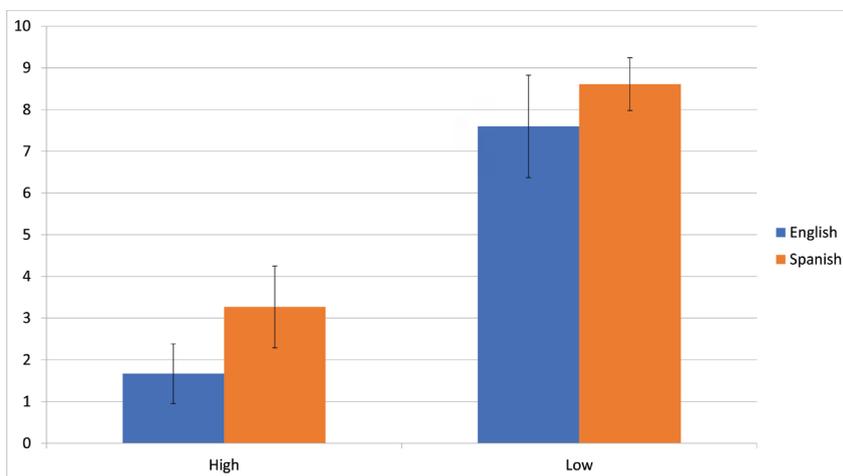
The two-way ANOVA test evaluated the importance of the two factors in the score differences. The ANOVA test showed that there were statistically significant differences in the scores regarding the mathematical complexity ($F = 162.178 > F_{1,105,0.05} = 3.932$), which was confirmed by the p -value = $8.3524 \cdot 10^{-23}$. As expected from the findings in the answer to research question 1, there were also statistically significant differences in the scores regarding the language ($F = 8.746 > F_{1,105,0.05} = 3.932$),

which was confirmed by the p -value = 0.004. The test also indicates that interaction of the two factors, difficulty and language, was not statistically significant ($F = 0.441 > F_{1,105,0.05} = 3.932$), which was confirmed by the p -value = 0.508. Analysis of the partial eta squared showed that the difference was mostly explained by the mathematical difficulty, as language was responsible for 7.9% of the differences and interaction for 0.4%.

Separate analysis

Given the fact that the mathematical difficulty was the main factor affecting the differences, a separate one-way ANOVA was performed for high and low difficulty levels. Language was the independent variable and the score was the dependent variable. Again, descriptive data (see Figure 6) shows that the difference was greater for word problems with high mathematical difficulty than for problems with low difficulty, with a mean value of 1.667 (SD = 1.801) for word problems written in English and a mean of 3.269 (SD = 2.426) for word problems written in Spanish. In contrast, the difference was diminished for word problems with low mathematical difficulty, with a mean value of 7.596 (SD = 3.040) for the word problems written in English and a mean of 8.611 (SD = 1.601) for the word problems written in Spanish.

Figure 6. Mean and 95% confidence interval of the scores per difficulty level



Source: Own elaboration.

The ANOVA test for high difficulty word problems shows that homogeneity of variances could be assumed with a p -value = 0.127 for the Levene test. ANOVA informed that the difference of the scores obtained in solving high-difficulty word problems in English or Spanish was statistically significant ($F = 7.494 > F_{1,52,0.05} = 4.027$). P -value = 0.009 confirmed that the scores in English were significantly lower than those in Spanish at 0.05 significance level.

On the other hand, the ANOVA test for low-difficulty word problems indicated that homogeneity of variances could not be assumed with a p -value = 0.012 for the Levene test. ANOVA pointed out that the difference of the scores obtained in solving low-difficulty word problems in English or Spanish was not statistically significant ($F = 2.338 < F_{1,52,0.05} = 4.027$). P -value = 0.132 confirmed that the scores in English and Spanish did not differ significantly at 0.05 significance level.

Although there was no significant interaction of language and mathematical difficulty, the separate analysis revealed that, when the mathematical difficulty of the word problems was low, there was no significant difference in the scores. However, when the difficulty was high, the difference was significant. This must be taken into consideration when assessing word problem solving processes in bilingual education contexts.

Although the CEFR reading comprehension level did not seem to affect the word problem solving processes, this could be due to the fact that only five students were classified as being in level A2. According to this finding, the mathematical register of the word problems may not be necessarily adapted to the linguistic level of the students.

Discussion

Several main ideas can be drawn from the analysis of the data. Two of them have clear pedagogical implications in the assessment of mathematical performance regarding word problem solving processes. The first one is that language affects evaluation, as the word problems written in English have significantly lower scores than those in Spanish. This is consistent with the students' perception that solving word problems in English is more difficult than doing the same in Spanish.

It also concurs with other findings that language proficiency affects mathematical achievement in testing (Bernardo, 1999; Bernardo, 2002; Kempert, Saalbach, & Hardy, 2011), but not with other studies that declare that, if language proficiency is adequate for understanding the problem, bilingualism has no effect on mathematical problem-solving (Bialystok, 2001; Surmont et al., 2016). The reason for this discrepancy is clear in the researched context, in which a significant number of students exhibit a non-ideal command of the English language, and thus the influence of the language on word solving problems is higher than in contexts where the language is comprehensible. The second one is that this difference appears in word problems with high mathematical difficulty, but not when the difficulty is low. It can be concluded that one possibility to help the assessment of mathematical proficiency using word problem solving processes in bilingual education contexts is to choose word problems with low mathematical difficulty to be written in the L2 (quadrant 4 in the CLIL matrix), and to complement them with the other word problems written in the students' L1. Use of word problems in both languages in the same paper could be recommended at the beginning of bilingual education, expecting responses in the learners' L1 as well until students are cognitively not discouraged through language and complexity (Lyster, 2007).

According to the interaction between language and word solving problems, students also reported that word problem solving processes require more time in L2 than in L1, and the unknown non-mathematical vocabulary is problematic. Studies state that simplified linguistic instructions help students overcome the linguistic complexity of the word problems (Abedi & Lord, 2001; Van Rinsveld et al., 2016), and known vocabulary unloads memory resources that are needed when comprehension difficulties occur at surface level, which can lead to the unavailability for carrying out arithmetic calculations (Kempert, Saalbach, & Hardy, 2011). This can be partially solved by including glossaries or word and expression banks in tests where word problems in the L2 appear, something that Van Rinsveld et al. (2016) refer to as the benefits of the "contextual linguistic factors," which may help less proficient students successfully perform in word solving problems: "language context helped bilingual participants to solve additions faster in their less dominant language" (p. 78). Besides, providing students

with more time to do the test can help them read more carefully and, consequently, understand word problems more easily, keeping the focus on the solution process and not on the time they have to perform the task. Familiarizing students with mind mapping for visualization of the idea in word problems in mathematics can easily bring solutions for understanding them.

Finally, no contextual factors—such as family income, parental education, and parental occupation—have been found (as described in Caro & Cortés, 2012; Ensminger, Forrest, Riley, Kang, Green, Starfield, & Ryan, 2000; Hattie, 2009; Rutkowski & Rutkowski, 2013; Sirin, 2005; White, 1982; Willms & Tramonte, 2015) to influence students' performance when solving the word problems written in English. This could be due to the fact that the group is quite homogeneous. Another reason that can explain the moderate influence of contextual factors could be the decrease of their influence as the students grow up. Thus, in line with the findings (Jäppinen, 2005), home environment or the socio-economic statuses of parents are clearly influential during the first stages of education, but decline along time when they are covered and homogenised by other variables (Lancaster, 2018).

One of the implications of this study has to do with the levels of assessment and to the pedagogical dimension and curricular aspects which associated to the assessment procedure. Any evaluation test contains questions with different degrees of complexity, but according to the findings in research questions 1 and 2, the judgement of the mathematical skills of bilingual education learners should be based on tests where the word problems presented in L2 are chosen among those with low mathematical difficulty. The complex word problems written in L1 can also be used to completely estimate mathematical competence level. If a math teacher wants to translate questions from a test in L1 into L2, all these factors must be considered in order to diminish any possible mismatch of the mathematical performance due to linguistic difficulties: a) choosing the easiest word problems to be translated into L2; b) eliminating some questions to provide extra time; and c) including a glossary (also suggested by Dalrymple, Karagiannakis, & Papadopoulos, 2012, as well as by Tavares, 2015). However, it can be claimed that some of these suggestions may be unfair for peers in monolingual education (a) or that they could not actually support the

use of non-mathematical vocabulary (c). In any case, the data analysed suggest, in line with the findings of Ouazizi (2016) and Surmont et al. (2016), that the content matter and the language benefit reciprocally as they help activate underlying brain mechanisms and foster metalinguistic awareness: “increased metalinguistic awareness could lead to a better understanding of and insight into the abstract language of math” (Surmont et al., 2016, p. 329).

Conclusion and Limitations

The data was analysed through ANOVA tests mostly used with a minimum of three groups. Perhaps a t-test of independent samples would have been a better option to answer research question 1, since this instrument is commonly used for comparing two groups.

It should be noted that the research took place in only one high school in the city centre of an average city in Andalusia, a fact that definitely affects the external validity and, thus, makes it difficult to extrapolate the results to other contexts. Nevertheless, participants are students who could be representatives of high schools with similar contextual and socio-economic terms. In any case, further research is needed in order to guarantee that the conclusions are valid in general, including schools with different situations (e.g., urban, suburban or rural) and contexts. Another factor that could affect the external validity is the homogeneity of the group in terms of CEFR reading comprehension level.

Hence, further studies with students having different reading comprehension levels should be carried out to confirm if solving word problems in English is truly affected by their reading performance, but also to verify if the reading performance combined with the mathematical difficulty can be used for the adequate selection of the word problems, particularly in tests in which the main objective is to assess the mathematical performance. Finally, homogeneity also appears in the contextual factors. For this reason, a profound investigation regarding context and mathematical performance in English as an L2 is required.

References

- Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment*, 8(3), 231–257. DOI: https://doi.org/10.1207/S15326977EA0803_02
- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14(3), 219–234. DOI: https://doi.org/10.1207/S15324818AME1403_2
- Adanur, Y., Yagiz, O., & Işık, A. (2004). Mathematics and language. *Journal of the Korea Society of Mathematical Education Series D: Research in Mathematical Education*, 8(1), 31–37. Retrieved from <https://www.semanticscholar.org/paper/Mathematics-and-Language-Adanur-Ya%C4%9F%C4%B1z/358c34c870078bd4a42568fdec-29838d442e37c5>
- AGAEVE. (2017). *Evaluación ESCALA: cuestionario de contexto. Familia*. Retrieved from <http://www.juntadeandalucia.es/educacion/agaeve/profesorado-primaria-escala-contexto.html>
- Bernardo, A. B. (1999). Overcoming obstacles to understanding and solving word problems in mathematics. *Educational Psychology*, 19(2), 149–163. DOI: <https://doi.org/10.1080/0144341990190203>
- Bernardo, A. B. (2002). Language and mathematical problem solving among bilinguals. *Journal of Psychology: Interdisciplinary and Applied*, 136(3), 283–297. DOI: <https://doi.org/10.1080/00223980209604156>
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge, UK: Cambridge University Press.
- Cambridge. (2017a). *Cambridge English: Preliminary (PET) exam format*. Retrieved from <http://www.cambridgeenglish.org/exams-and-tests/preliminary/exam-format>
- Cambridge. (2017b). *The Cambridge English scale*. Retrieved from <http://www.cambridgeenglish.org/exams-and-tests/cambridge-english-scale>
- Caro, D. H., & Cortés, D. (2012). Measuring family socioeconomic status: An illustration using data from PIRLS 2006. *IERI Monograph Series. Issues and Methodologies in Large-Scale Assessments*, 5, 9–33. Retrieved from http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_05_Chapter_1.pdf

- Council of Europe. (2001). *Common European framework of reference for languages*. Cambridge, UK: Cambridge University Press.
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge, UK: Cambridge University Press.
- Cummins, J. (2000). *Bilingual education and bilingualism: Language, power and pedagogy: Bilingual children in the crossfire (1)*. Bristol, UK: Multilingual Matters.
- Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive psychology*, 20(4), 405–438. DOI: [https://doi.org/10.1016/0010-0285\(88\)90011-4](https://doi.org/10.1016/0010-0285(88)90011-4)
- Dalrymple-Smith, A., Karagiannakis, V., & Papadopoulos, K. (2012). *Does international education risk the development of students' mother tongues? An investigation into language use in mathematics classrooms* (Unpublished project-based research). Utrecht University, Utrecht, The Netherlands.
- Ensminger, M. E., Forrest, C. B., Riley, A. W., Kang, M., Green, B. F., Starfield, B., & Ryan, S. A. (2000). The validity of measures of socioeconomic status of adolescents. *Journal of Adolescent Research*, 15(3), 392–419. DOI: <https://doi.org/10.1177/0743558400153005>
- Grabner, R. H., Saalbach, H., & Eckstein, D. (2012). Language-switching costs in bilingual mathematical learning. *Mind, Brain, and Education*, 6(3), 147–155. DOI: <https://doi.org/10.1111/j.1751-228X.2012.01150.x>
- Halliday, M. A. K. (1978). Sociolinguistic aspects of mathematical education. In M. A. K. Halliday (Ed.), *Language as social semiotic: The social interpretation of language and meaning* (pp. 194–204). London, UK: Edward Arnold.
- Hansen-Thomas, H., Langman, J., & Farias, T. (2018). The role of language objectives: Strengthening math and science teachers' language awareness with emergent bilinguals in secondary classrooms. *LA-CLIL*, 11(2), 193–214. DOI: <https://doi.org/10.5294/laclil.2018.11.2.2>
- Hattie, J. A. (2009). *Visible learning: A synthesis of 800 + meta-analyses on achievement*. Abingdon, UK: Routledge.
- Hiebert, J., & LeFever, P. (1987). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge in mathematics* (pp. 1–27). Hillsdale, NJ: Erlbaum.

- INEE.** (2013). *Estímulos PISA de Matemáticas liberados: aplicación como recurso didáctico en la ESO*. Retrieved from http://recursostic.educacion.es/inee/pisa/matematicas/_private/pisamatematicas2013.pdf
- Jäppinen, A-K.** (2005). Thinking and content learning of Mathematics and Science as cognitional development in content and language integrated learning (CLIL): Teaching through a foreign language in Finland. *Language and Education*, 19(2), 148–169. Retrieved from https://www.unifg.it/sites/default/files/allegatiparagrafo/20-01-2014/jap-pinen_thinking_and_content_learning_in_mathematics_and_science.pdf
- Jiménez-Jiménez, A. F.** (2015). Private speech during problem-solving activities in bilingual speakers. *International Journal of Bilingualism*, 19(3), 259–281. DOI: <https://doi.org/10.1177/1367006913509902>
- Junta de Andalucía.** (2016). *Plan estratégico de desarrollo de las lenguas en Andalucía. Horizonte 2020*. Retrieved from https://www.juntadeandalucia.es/export/drupaljda/plan_estrategico.pdf
- Kempert, S., Saalbach, H., & Hardy, I.** (2011). Cognitive benefits and costs of bilingualism in elementary school students: The case of mathematical word problems. *Journal of Educational Psychology*, 103(3), 547–561. DOI: <http://dx.doi.org/10.1037/a0023619>
- Lancaster, N.** (2018). Extramural exposure and language attainment: The examination of input-related variables in CLIL programmes. *Porta Linguarum*, 29, 91–114. Retrieved from <http://digibug.ugr.es/handle/10481/54024>
- Lorenzo, F.** (2008). Instructional discourse in bilingual settings. An empirical study of linguistic adjustments in content and language-integrated learning. *Language Learning Journal*, 36(1), 21–33. DOI: <https://doi.org/10.1080/09571730801988470>
- Lyster, R.** (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam, The Netherlands: John Benjamins.
- May, T., & Ong, S.** (2011). Teaching Mathematics and Science in English in Malaysian classrooms: The impact of teacher beliefs on classroom practices and student learning. *Journal of English for Academic Purposes*, 10, 2–18. DOI: <https://doi.org/10.1016/j.jeap.2010.11.001>
- Morgan, C.** (2007). Who is not multilingual now? *Educational Studies in Mathematics*, 64, 239–242. DOI: <http://doi.org/10.1007/s10649-006-9064-y>

- Moschkovich, J. N. (2005). Using two languages when learning mathematics. *Educational Studies in Mathematics*, 64, 121–144. DOI: <https://doi.org/10.1007/s10649-005-9005-1>
- Moschkovich, J. N. (2007). Beyond words to mathematical content: Assessing English learners in the mathematics classroom. *Assessing Mathematical Proficiency*, 53, 345–352. Retrieved from https://www.researchgate.net/publication/237281918_Beyond_Words_to_Mathematical_Content_Assessing_English_Learners_in_the_Mathematics_Classroom
- Moschkovich, J. N. (2015). Academic literacy in mathematics for English learners. *The Journal of Mathematical Behavior*, 40, 43–62. DOI: <https://doi.org/10.1016/j.jmathb.2015.01.005>
- Novotná, J., Hadj-Moussová, Z., & Hofmannová, M. (2005). *Teacher training for CLIL – Competences of a CLIL teacher*. Retrieved from: <http://people.fjfi.cvut.cz/novotant/jarmila.novotna/Novotna-et-alsemt01.pdf>
- OECD. (2006). *Education at a Glance*. Paris, France: OECD Publishing.
- OECD. (2013). *Education at a Glance*, Paris, France: OECD Publishing.
- OECD. (2014). *PISA 2012 technical report*. Paris, France: OECD Publishing.
- OECD. (2016). *PISA 2015 assessment and analytical framework: Science, reading, mathematics and financial literacy*. Paris, France: OECD publishing.
- Orden de 28 de junio de 2011, by which state bilingual schools in the Autonomous Region of Andalusia are regulated.
- Ouazizi, K. (2016). The effects of CLIL education on the subject matter (Mathematics) and the target language (English), *LACLIL*, 9(1), 110–137. DOI: <https://doi.org/10.5294/laclil.2016.9.1.5>
- Rutkowski, D., & Rutkowski, L. (2013). Measuring socioeconomic background in PISA: One size might not fit all. *Research in Comparative and International Education*, 8(3), 259–278. DOI: <https://doi.org/10.2304/rcie.2013.8.3.259>
- Schleicher, A., Zimmer, K., Evans, J., & Clements, N. (2009). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris, France: OECD Publishing (NJ1).
- Sigley, R., & Wilkinson, L. C. (2015). Ariel's cycles of problem solving: An adolescent acquires the mathematics register. *Journal of Mathematical Behavior*, 40, 75–87. DOI: <https://doi.org/10.1016/j.jmathb.2015.03.001>
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453. DOI: <https://doi.org/10.3102/00346543075003417>

- Smit, J., & Van Eerde, A. A. (2011). A teacher's learning process in dual design research: Learning to scaffold language in a multilingual mathematics classroom. *Mathematics Education*, 43, 889–900. DOI: <https://doi.org/10.1007/s11858-011-0350-5>
- Spelke, E. S., & Tsivkin, S. (2001). Language and number: A bilingual training study. *Cognition*, 78, 45–88. DOI: [https://doi.org/10.1016/S0010-0277\(00\)00108-6](https://doi.org/10.1016/S0010-0277(00)00108-6)
- Surmont, J., Struys, E., Ven Den Noort, M., & Van Der Craen, P. (2016). The effects of CLIL in mathematical content learning: A longitudinal study. *Studies in Second Language Learning and Teaching*, 6(2), 319–337. DOI: <http://doi.org/10.14746/ssl.2016.6.2.7>
- Swetz, F. J. (2012). *Mathematical expeditions: Exploring word problems across the ages*. Baltimore, MD: John Hopkins University Press.
- Tavares, N. J. (2015). How strategic use of L1 in an L2-medium mathematics classroom facilitates L2 interaction and comprehension, *International Journal of Bilingual Education and Bilingualism*, 18(3), 319–335. DOI: <https://doi.org/10.1080/13670050.2014.988115>
- Van Rinsveld, A., Brunner, M., Landerl, K., Schiltz, C., & Ugen, S. (2015). The relation between language and arithmetic in bilinguals: Insights from different stages of language acquisition. *Frontiers in Psychology*, 6(265), 1–15. DOI: <https://doi.org/10.3389/fpsyg.2015.00265>
- Van Rinsveld, A., Schiltz, C., Brunner, M., Landerl, K., & Ugen, S. (2016). Solving arithmetic problems in first and second language: Does the language context matter? *Learning and Instruction*, 42, 72–82. DOI: <https://doi.org/10.1016/j.learninstruc.2016.01.003>
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91(3), 461–481. DOI: <http://dx.doi.org/10.1037/0033-2909.91.3.461>
- Willms, J. D., & Tramonte, L. (2015). *Towards the development of contextual questionnaires for the PISA for development study*. Retrieved from <https://www.oecd-ilibrary.org/docserver/5js1kv8crsjf-en.pdf?expires=1563951052&id=id&accname=guest&checksum=404F909C23A-D91A5DF544DB9E2783E7C>