



OPEN ACCESS

EDITED BY

Lies Sercu, KU Leuven, Belgium

REVIEWED BY

Folkert Kuiken, University of Amsterdam,
NetherlandsChristiane Dalton-Puffer, University of Vienna,
Austria

*CORRESPONDENCE

Thomas Caira

✉ thomas.caira@vub.be

RECEIVED 31 October 2024

ACCEPTED 12 December 2024

PUBLISHED 24 January 2025

CITATION

Caira T, Declerck M and Struys E (2025)
Language-mixing in Content and Language
Integrated Learning: benefit or burden? An
auditory recall perspective.
Front. Educ. 9:1520791.
10.3389/feduc.2024.1520791

COPYRIGHT

© 2025 Caira, Declerck and Struys. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Language-mixing in Content and Language Integrated Learning: benefit or burden? An auditory recall perspective

Thomas Caira*, Mathieu Declerck and Esli Struys

Brussels Centre for Language Studies (BCLS), Vrije Universiteit Brussel (VUB), Brussels, Belgium

Recent pedagogical trends have seen a revival in language-mixing in CLIL contexts, thereby challenging the traditional 'one subject-one language' approach. From a cognitive standpoint, recent research indicates that the disadvantages of language mixing may not be as significant as previously thought. This is further supported by studies showing no negative effects of language-mixing on immediate memory recall in CLIL pupils. Using an Old/New recognition task design, participants in the present study had to differentiate between previously defined concepts and new ones in three different language contexts (i.e., single-language L1; single-language L2 and a mixed context). We also accounted for delayed recall with a second test phase 36 hours after the first one. Response times and accurate recall scores were retained for further analysis. Our findings reveal a nuanced picture: we found that mixed-language input negatively impacts immediate and delayed recall of information compared to L1 input. However, mixed-language input also brings about better recall of information compared to L2 input. It seems that language-mixing thus partly mitigates the disadvantage in auditory recall that occurs in a single-language L2 context. Overall, these results suggest a need to reconsider the effects of language-mixing on memory and, consequently, nuance its role in CLIL practices.

KEYWORDS

language-mixing, CLIL, language control, cognition, translanguaging, code-switching, mixing costs

1 Introduction

The presence of multiple languages within the same context (hereinafter referred to as language-mixing), is an integral and highly relatable phenomenon in multilinguals' daily lives. Language-mixing situations occur, for instance, when multilinguals freely use multiple languages in a single conversation. Despite the prevalence of language-mixing in daily life, many actors in charge of (multilingual) education (including policymakers, school boards and teachers) tend to adhere to a strict 'one subject, one language'-rule in the classroom (Antón et al., 2016; Wei, 2018). As a result, the influence of language-mixing on learning outcomes in the context of education remains largely unclear. In the present study, we aim to remedy this gap in the current literature by conducting an Old/New experiment with children enrolled in a French-Dutch multilingual program in Belgium. During that experiment, the pupils were asked to recall if certain items had been defined during a previous input phase or not. This experimental design allowed us to gain insight in pupils' ability to accurately recall information gathered in single-language and mixed-language contexts. Indeed, within this experiment, a distinction was made between input contexts where definitions were given in only one language, and another context where definitions were provided in a mix of two languages.

1.1 Language-mixing in multilingual education

Despite its prevalence in multilinguals' daily experiences, language-mixing has traditionally not been viewed positively in education, with most (multilingual) programs adhering to a strict separation of the main language of instruction and the target language of certain courses (San Isidro and Lasagabaster, 2019). One of the underlying reasons behind this linguistic segregation was the assumption that the presence of language-mixing in pupils' oral productions (as described by Poplack, 1980) represented a symptom of insufficient language proficiency (Reyes, 2004). This postulate and its subsequent 'one subject, one language'-rule already prevailed during the inception of the very first multilingual programs (Lambert et al., 1973).

Similarly to those multilingual programs, the Content and Language Integrated Learning (CLIL) approach is described as "a generic term to describe all types of provision in which a second language (a foreign, regional or minority language and/or another official state language) is used to teach certain subjects in the curriculum other than languages lessons themselves" (European Commission, 2006, p. 8). As such, it does not feature explicit language instruction, but rather functions with an additional language to the main school language, as the medium of instruction (i.e., the CLIL language) within specific subjects to acquire content knowledge (e.g., subject-specific vocabulary or discourse). From its onset, recommendations regarding CLIL education hinted at the fact that CLIL should not refuse the use of both the first language and an additional language in the learning context (Marsh, 2013), thus aiming to break this 'one subject, one language'-rule used in previous forms of multilingual education. Moreover, it was noted that CLIL has the potential to set itself apart from traditional immersion models by adopting a more flexible and balanced approach to incorporating the students' first language in its lessons (Lin, 2015). Therefore, CLIL should not strictly separate the first and the CLIL language, but rather embrace a multilingual pedagogy featuring fluid language practices (Garcia and Flores, 2012) in order to achieve its pedagogical objectives. Following these recommendations, subsequent studies have further theorized and observed the use of language-mixing, either described as code-switching (Poplack, 1980) or translanguaging (Williams, 1994), in the literature. While the distinction between these terms falls outside of the present study's scope, most of the studies pertaining to these forms of language-mixing endorse its inclusion in multilingual education practices (e.g., Nikula and Moore, 2019; Cenoz and Gorter, 2022).

Yet, even though recent theoretical discussions surrounding multilingual education have highlighted the importance of the first language in learning additional languages (e.g., Nikula and Moore, 2019; Cenoz and Gorter, 2022), ideologies that emphasize using only one language per subject still largely prevail in practice and policy, as well as in evaluating learning outcomes (Wei, 2018). One of the possible reasons underlying this reluctance to the implementation of language-mixing in the classroom, might be a lack of evidence for its effect on learning outcomes. Indeed, systematic reviews (Prilutskaya, 2021; Lu et al., 2023) report that the foci of the existing literature mainly pertained to the attitudes of students and teachers toward language-mixing (e.g., Zhou and

Mann, 2021; Serra and Feijoo, 2022; Sahan et al., 2022), as well as the forms it adopts and the function it serves (e.g., Lialikhova, 2019; San Isidro and Lasagabaster, 2019; Papaja and Wysocka-Narewska, 2020). The importance of their findings is certainly not to be dismissed, as they show that language-mixing increases student attention and participation in the classroom (Tai and Wei, 2021), especially in the case of learners with low CLIL language proficiency (Domalewska, 2017). That being said, the aforementioned reviews also report a notable gap in quantitative research examining these practices (Lu et al., 2023, p. 15). This could, in part at least, be due to the fact that classroom contexts vary a lot and have numerous factors at play simultaneously, which makes it difficult to assess practices in a quantitative manner. Nevertheless, the lack of (quantitative) assessment of language-mixing practices leaves us with very limited knowledge on their potential impact on learning. Our goal in the present study is to isolate and observe language-mixing specifically through the use of an experimental design. This allows us to test its effects in a more controlled context, compared to, for instance, classroom observations. Before expanding further on the design, let us first turn to the impact of language-mixing on cognitive functions in order to unravel the underlying mechanisms at play.

1.2 Language-mixing from a cognitive standpoint

From a cognitive standpoint, the presence of multiple languages within the same context could lead to increased interference between languages in terms of both comprehension and production (Christoffels et al., 2007; Green and Abutalebi, 2013). The processes our brains use to mitigate this cross-language interference (Costa et al., 2000; Hoshino and Kroll, 2008; Meade et al., 2017) and allow for managing multiple languages and switching between them is referred to as language control (Declerck and Koch, 2023). The potential cognitive costs that arise from these additional control processes are known as mixing costs (e.g., de Bruin et al., 2020; Declerck et al., 2021; Ma et al., 2016). In other words, mixing costs reflect worse performance in (the language repetition trials of) mixed-language contexts relative to single language contexts (across L1 and L2 trials), due to the cognitive cost of having to maintain multiple languages at the same time (e.g., Declerck, 2020). These mixing costs are typically reported in studies examining language production (most often by means of a picture- or digit-naming task), although the reported results are largely dependent on the nature of the instructions. In fact, when participants have to name the images in a language which is indicated, for instance, by a flag of that language's country (i.e., a cued picture-naming task), mixing costs appear to be a very robust finding. On the contrary, when participants are able to freely choose the language in which they will name the picture (i.e., a voluntary picture-naming task), they usually experience a mixing benefit (e.g., de Bruin et al., 2020; Grunden et al., 2020). In other words, the participants' performances on the task increase in the mixed-language context relative to the single-language contexts in the L1 or the L2 when the choice of language that is being used is voluntary. This confirms the hypothesis that the cognitive demands involved in language control are context-dependent. Following this reasoning, the adaptive control hypothesis (Green and Abutalebi, 2013), distinguishes three

interactional contexts with different implementations of language control:

- A single-language context “in which one language is used in one environment and the other in a second distinct environment” (p. 513). When applied to a classroom context, this reflects the traditional, ‘one subject, one language’ pedagogy.
- A dual language context “in which both languages are used but typically with different speakers” (p. 513). This could be the case in a CLIL co-teaching environment, wherein one teacher speaks the first language and the other teacher speaks the CLIL language.
- A dense code-switching context “in which speakers routinely interleave their languages in the course of a single utterance and adapt words from one of their languages in the context of the other” (p. 513). This context would then echo the fluid language practices of the language-mixing approach to multilingual education.

According to this hypothesis, it is actually the latter context in which control processes are the least required. This is mainly attributed to the assumption that languages function in a co-operative relationship in a dense code-switching context, instead of in competition with one another as is the case in the other two contexts (Green and Abutalebi, 2013). Because of this lesser need for language control when languages function in co-operation, we expect less mixing costs or even mixing benefits in dense switching contexts compared to the other two contexts in which languages compete with one another. This further underscores the value of a language-mixing approach from a cognitive standpoint.

The existing body of studies investigating mixing costs and benefits in comprehension is scarcer and reported findings are less conclusive. One study examined the influence of language-mixing on comprehension (Declerck et al., 2019) through the use of a number categorization tasks (i.e., indicating whether a number is even/odd or larger/smaller than 5 and 6). The participants were presented with series of number words either in one language (i.e., single-language blocks) or both languages (mixed-language blocks). They reported no mixing costs in the participants’ performance (reaction times and error rates), except when including language pairs with a high cognate rate (i.e., words that share both phonological form and meaning across a language pair, as is the case for ‘animal’ in English (/ˈæniməl/) and French (/aniˈmal/); Proctor and Mo, 2009), namely French and Spanish. However, the absence of mixing costs or benefits in other language combinations were not related to the number task specifically (which could have been the case according to Declerck et al., 2012), as no mixing costs were found in a subsequent experiment that relied on an animacy task instead of the parity and magnitude tasks.

Although most of the language mixing literature has focused on production and, to a lesser extent, on comprehension, a couple of studies also investigated the potential cognitive costs of language-mixing on recall of information in the context of education. In Antón et al. (2015), Spanish-Basque speaking participants had to learn features of unfamiliar objects in both a single-language L1 and a mixed-language context. Although no significant differences were observed between both contexts, all the participants in their study were perfectly balanced and simultaneous bilinguals, which is not representative of most bilingual individuals or communities (Antón

et al., 2016). To remedy this limitation, they relied on unbalanced (Russian-English) bilingual pupils who followed CLIL at school in a follow-up study (Antón et al., 2016). Despite their unbalanced proficiency in both languages, the exposure of these participants to both languages of instruction was equal. In this study, the participants were first exposed to auditory input comprised of definitions for everyday concepts in either single-language Russian (i.e., their L1) or a mix of both Russian and English (in a counterbalanced order). After having heard all 14 definitions in one context, they were shown 28 images. Half of the images represented the objects which had been defined during the input phase, the other half represented new items. Participants had to accurately recall if the object at hand had been defined during the input phase or not. Afterwards, they would be exposed to 14 definitions and 28 images in the other context and repeated the same recall test. Again, no differences were found in the pupils’ performance of the task in terms of accurate (immediate) recall and response times between both language contexts. This led them to conclude that similar outcomes in recall can be obtained when languages are mixed compared to when they are kept separated. However, given the specific type of CLIL in which these participants were enrolled with an equal exposure rate to both languages, the question remained unanswered, if these findings could be generalized to settings with dissimilar rates of exposure to both instruction languages in a CLIL curriculum.

1.3 Present study and research questions

Although previous studies by Antón et al. (2015, 2016) form a reference point for investigating language-mixing quantitatively in education, they also present some opportunities to delve further into this topic, which is exactly what we will do in the current study. The first opportunity pertains to the exposure rate to the CLIL language of the participants. Although their follow-up study included unbalanced bilinguals (Antón et al., 2016), these pupils still had perfectly balanced exposure to their first language and the CLIL language at school. This is not representative of all CLIL contexts and leaves us unknowing whether these results could be replicated with pupils in a school context that provides unbalanced (and thus less) exposure to the CLIL language. Secondly, their single-language context focused on the pupils’ first language, which does not make for a clear comparison between mixed-language classes and CLIL courses in which single-language teaching takes place in the CLIL language only. Thirdly, their Old/New-task design only accounted for immediate recall, with no indications of long(er)-term effects of language-mixing on memory recall. Given the fact that evaluations in the classroom do not (only) take place immediately after learning, delayed recall abilities are also of great importance. In order to tackle these additional questions, we firstly recruited participants with unbalanced exposure to their first and CLIL language at school. Indeed, our participants have 70% exposure to French (i.e., their first language) and 30% to Dutch (their CLIL language) throughout their weekly classes. Secondly, in addition to the single-language condition in the first language and the mixed-language condition, we added a single-language condition in the CLIL language (L2). Finally, we expanded on the initial design of Antón and colleagues with a delayed recall test, which took place 36 h (on average) after the initial immediate recall test.

With these design modifications in mind, our main investigation of the effect of language-mixing in education on learning can be broken down into the following research questions:

- 1 Can the previously reported absence of mixing costs in memory recall (cf. Antón et al., 2015, 2016), by comparing performance in a mixed-language context to a single-language context in their first language, be replicated with CLIL students who have unbalanced exposure to both instruction languages at school?
- 2 How do recall abilities in a mixed-language context compare to a single-language context in the CLIL language (L2)?
- 3 What are the consolidation effects of mixed-language input, relative to single language input, over time for recall of information?

Given the fact that our participants have less exposure to the CLIL language (L2) in their program compared to the pupils included in previous studies (30% vs. 50%), we assume that they would have had fewer opportunities to practice recalling input in that language. As a result, we expect this lower school exposure to the second language to have a negative impact on performance in the mixed-language condition of our recall experiment, which taps into a skill that is specifically practiced within a school setting. Therefore, if L2 school exposure indeed turns out to be the crucial factor here, we expect better performance in the L1 condition compared to the mixed condition (RQ 1) and, similarly, better recall in the mixed condition compared to the L2 condition as well (RQ2). As far as the time effect is concerned (RQ 3), a general decrease in recall performance is to be expected over time across conditions; a decrease which we expect to be even steeper in the two conditions containing L2 input, if school exposure is not only the relevant factor for understanding mixing costs and benefits, but also for consolidation effects. However, due to the lack of previous findings in delayed recall, this hypothesis is largely explorative.

2 Materials and methods

2.1 Participants and research context

All participants were recruited from two different schools in Wallonia (i.e., the French-speaking part of Belgium). In this region, CLIL courses may be taught in Dutch, English or German and do not exceed 40% of the total curriculum in secondary education (Mettewie and Van Mensel, 2023). The remaining courses are provided in their first language (i.e., French). Additionally, there are no official formal requirements for pupils to comply to in order to enroll (European Commission, 2006). Both schools we recruited our participants from were French-speaking and offer a Dutch-speaking CLIL program. The courses taught in Dutch in the selected schools represent 31% (i.e., 10 h) of the participants' weekly classes.

All participants were native French speakers in their first or second year of secondary education (17 first-year pupils, mean age = 13.17 years old) and had no prior experience with CLIL. This information has been collected through an adapted version of the Leap-Q Questionnaire (aimed at identifying an individual's language experience and proficiency, see Marian et al., 2007 for further details)

which was sent to the parents. The questionnaire also included questions about the level of education of the participants' mothers as a proxy for their socioeconomic status (Harding et al., 2015), as well as potential language and/or learning impairments. Moreover, the participants completed a Lextale task in both French (Brybaert, 2013) and Dutch (Lemhöfer and Broersma, 2012) prior to partaking in the actual experiment in order to account for their proficiency in both their first and the CLIL language. Means and standard deviations of those background variables can be found in Table 1. Participants and their parents were fully informed about the nature of the study, and written consent was obtained from both the participants and their parents.

2.2 Materials and procedure

In an effort to replicate Antón et al. (2016)'s design as accurately as possible, our experiment featured an Old/New task design too, which was built in PsychoPy (Version 2023.2.3) and subsequently administered using a computer. In order to increase the ecological validity of the experiment, participants completed the task in their usual classroom during one of the CLIL courses. The experiment started with written instructions. The instructions were also explained orally to the participants. The task itself included two phases: an input phase, followed by a test phase. During the input phase, participants heard 14 definitions of concrete nouns (e.g., a car) retrieved from a high frequency list in French (Liste de fréquence lexicale, 2023) (average frequency = 188.97 occurrences per 1 million words, according to Gimenes and New, 2016). Every definition is comprised of two characteristics expressed in 3–8 words each (e.g.: "it is used for transportation" and "it has four wheels and an engine" for the word 'car'). In creating the definitions, we avoided the use of homophones, homographs and cognates as much as possible. Each definition was followed by a blank screen of 500 ms before the next trail was presented. It bears noting that the participants were only exposed to auditory input during this phase. All definitions were recorded by the same bilingual speaker with a native command of both French and Dutch. After having heard the 14 definitions, the test phase started. In this second phase, participants were shown 28 unique image pairs. Each pair represented one concept (e.g., two images of different cars), and these pairs were shown in a random sequence. Images were retrieved from copyright-free databases and selected according to their accurate representation of the target nouns. For each image pair, they had to identify whether the item at hand had been defined during the input phase (i.e., an 'old' item) or not (i.e., a 'new' item). This was done by means of a keyboard response: 'y' for old items and 'n' for new

TABLE 1 Mean and standard deviations for the background variables of the recruited participants (SES represents a score out of 7). Unfortunately some questionnaire data got lost for the age and SES variables.

	n	M	SD
Gender	29	18 female	
Age	21	13,17	0,54
SES	20	5,25	0,85
French Lextale	29	75,12 %	8,86 %
Dutch Lextale	29	53,41 %	5,64 %

items (on an azerty keyboard). The new items were also frequent nouns which we retrieved from the same list as the ones selected for the input phase (115.29 occurrences per 1 million words, according to Gimenes and New, 2016). This test phase (including the same image pairs, albeit in a different random order), but not the input phase, was repeated the next day in order to account for delayed recall.

The same procedure is applied to all three of the language contexts of the experiment. The first is the single-language L1 context for which both features of every definition were recorded in French (e.g., “*on s'en sert pour se déplacer*” and “*a quatre roues et un moteur*” for the word ‘voiture’ (car)). For the Mixed-Language (ML) context, one feature of every definition remained in French, whereas the other was translated to Dutch (e.g., “*on s'en sert pour se déplacer*” and “*heeft vier wielen en een motor*”). Thirdly, the newly added single-language (L2) context consisted of two features entirely in Dutch (e.g., “*wordt gebruikt om zich te verplaatsen*” and “*heeft vier wielen en een motor*”). We created a distinct set of 14 orally recorded definitions and 28 image pairs for every condition. This thus resulted in a total of 42 defined concepts and 84 image pairs (see Appendix 3). Each set of stimuli was assigned to a context in a counter-balanced order. Similarly, all participants have been exposed to all three language contexts (i.e., L1, L2 and mixed) in a single day in a counter-balanced order. A break of 60 s was built in between each context. The experiment (including oral instructions) lasted around an hour on the first day and 30 min on the second day.

2.3 Data analysis

The participants' response times and accurate recall abilities (in terms of percentage of hits and false alarms) were collected. Furthermore, the response times' means and standard deviations were calculated for correct trials on old and new items separately for every language context at both time points in order to exclude outliers in the reaction time analysis. Subsequently, any trial exceeding the lower (500 ms) or upper ($M + 2 * SD$) limit has been removed for every language context at each timepoint (4.30% of the data). In terms of accurate recall, we used binary coding with correct responses coded as ‘1’ and incorrect responses as ‘0’. Based on these values, d' was calculated based on the Z -score of the proportions of hits (i.e., correct responses to old items) minus the Z -score of the proportion of false alarms (i.e., incorrect responses to new items). This value used for accurate recall presents the advantage of combining both the hit and false alarms rate in a single composite measure. Additionally, using this measure allows for a better comparison with previous studies, since it has also been used by Antón et al. (2016).

In order to measure the effects of language-mixing on recall abilities (both immediately and over time), a linear mixed model analysis was performed in R (version 4.3.3) using the lme4 package (Bates et al., 2015). The participants' accurate recall values were computed as the dependent variable, while Time (coded as “immediate” vs. “delayed”; “immediate” being the reference) and language context (coded as “ML” vs. “L1” vs. “L2”; “ML” being the reference) were added as fixed effects. The model was fully randomized where possible by including random intercepts for the variance between participants on the one hand and items on the other hand. In the case of the response times, the same set of fixed and random effects were used in a linear mixed-effect analysis for correct responses to old items only. By considering correct responses only, we eliminate the

variability introduced by incorrect guesses, which could have skewed the results. Moreover, this approach allowed us to more precisely measure the cognitive processes involved in recognizing previously encountered items, providing a clearer understanding of performance in memory recall. The descriptive statistics of the variables as well as the results from the analyses can be found below.

3 Results

The descriptive statistics for both accurate recall scores (d') and response times can be found in Tables 2, 3, respectively. A subsequent linear mixed-effects model analysis was conducted to examine the effects of language context (i.e., ML vs. L1 and ML vs. L2) and Time (immediate vs. delayed recall) (see Table 4). Additionally, a similar analysis was run with the response times of correct trials for Old Items as the dependent variable (see Table 5).

3.1 Performance in language-mixing vs. L1

The linear mixed-effects model confirmed that there is a significant main effect of the L1 context, indicating better performance in terms of accurate recall abilities when being exposed to French single-language input compared to mixed-language input (d' difference of 1.40, see Table 4). Additionally, we found a significant two-way interaction between the L1 context and time. The observed mixing benefit thus seemed to decrease as time progresses (minus 0.33 in d' at Time 2, see Table 4).

The analysis of the response times to correct Old Items trials also revealed a positive main effect of the L1 language context, thus also implying a mixing cost in terms of response times (167.41 ms

TABLE 2 Percentages of hits, false alarms and discriminability values for the three conditions in terms of both immediate and delayed recall.

		Hits	False alarms	d'
ML	T1	72.66% (18.69)	21.39% (20.82)	2.12 (1.61)
	T2	70.42% (18.83)	30.17% (20.81)	1.44 (1.58)
L1	T1	85.12% (14.11)	14.11% (18.81)	3.50 (2.06)
	T2	83.29% (13.62)	23.45% (19.94)	2.51 (1.98)
L2	T1	59.05% (19.86)	32.91% (19.76)	0.75 (1.92)
	T2	54.5% (20.42)	36.29% (20.33)	0.55 (1.87)

TABLE 3 Mean reaction times (in milliseconds) to both old and new items for the tree in terms of immediate and delayed recall (standard deviations are provided between brackets).

		Old items	New items
ML	T1	1708 (649)	2,116 (1189)
	T2	1,236 (653)	1,666 (1196)
L1	T1	1,544 (645)	2088 (1178)
	T2	1,171 (642)	1,503 (1182)
L2	T1	1,561 (647)	2,103 (1185)
	T2	1,365 (642)	1,506 (1181)

slower on average, see Table 5). In this case, however, we did not find a significant interaction between response times and time for the difference between the L1 language context and the mixed context.

3.2 Performance in language-mixing vs. L2

Accurate recall scores are the lowest in the context with single-language L2 input (see Table 2). In the subsequent analysis, we found a significant main effect of the L2 language context, which points toward a mixing benefit in this case (increase by 1.37 d' score in the mixed-language context; see Table 4). This shows that, although the participants seem to perform worse in our recall task after being exposed to mixed-language input compared to single-language L1 input, the opposite is true when comparing performance in the mixed-language context and the single-language L2 context. We also found a significant interaction between the L2 language context and time. This means that, at time 2, the mixing benefit which we observed at time 1, has been significantly reduced (by 0.48 in d' , see Table 4).

When analyzing the response times, we still found a main effect of the L2 context in terms of response times in immediate recall (slower responses by 130.84 ms on average, see Table 5). At the first test time, participants thus take longer to correctly discriminate between old and new items after having been exposed to mixed-language input than single-language input in their L2. Yet, we also found an interaction between the L2 context and time, indicating longer response times in that context over time. In other words, when considering the delayed test time, this mixing cost disappears

TABLE 4 Estimates, (standard errors) and p -values for the different predictors of d' scores.

Predictors	Accurate recall (d')		
	Estimates	CI	p
(Intercept)	2.12	1.74–2.51	<0.001
L1 Language context	1.40	1.28–1.51	<0.001
L2 Language context	–1.37	–1.49 – –1.25	<0.001
Time 2	–0.68	–0.80 – –0.57	<0.001
L2 Language Context \times Time 2	–0.33	–0.50 – –0.17	<0.001
L3 Language Context \times Time 2	0.48	0.31–0.64	<0.001

All p -values lower than 0.05 are edited in bold in the table.

TABLE 5 Estimates, standard errors and p -values for the different predictors of response times to correct old items trials.

Predictors	Response times (ms)		
	Estimates	CI	p
(Intercept)	1725.29	1609.65–1840.93	<0.001
L1 language context	–167.41	–263.00 – –71.82	0.001
L2 Language Context	–130.84	–237.55 – –24.13	0.016
Time 2	–458.63	–558.11 – –359.14	<0.001
L1 Language Context \times Time 2	81.98	–52.97 – 216.93	0.234
L2 Language Context \times Time 2	240.48	88.44–392.52	0.002

All p -values lower than 0.05 are edited in bold in the table.

and even turns into a mixing benefit in favor of the mixed-language context.

3.3 The influence of L2 proficiency: a follow-up analysis

In a subsequent analysis, including the Lextale scores for Dutch as an additional factor (see Appendix 1), we found a significant two-way interaction between the L1 context and the participants' performance on the Dutch LexTale task, pointing toward a moderating effect of L2 proficiency on the difference in recall between language contexts. This interaction indicates that better knowledge of Dutch seemed to reduce the mixing cost that was found when comparing L1 input to mixed-language input (a decrease by 0.25 d' units, see Appendix 1). Similarly, a significant two-way interaction was found between accurate recall in the L2 context and Dutch Lextale scores. This indicates a decrease of the mixing cost associated with L2 input compared to mixed-language input as L2 proficiency increases (a decrease by 0.13 d' units, see Appendix 1).

We also conducted a follow-up analysis to investigate the potential influence of Dutch proficiency scores on response times. Yet, contrary to the follow-up analysis of the participants' accuracy scores, no significant interactions were observed with Dutch Lextale scores (see Appendix 2).

4 Discussion and conclusion

The general goal of the present study was to investigate the impact of language-mixing on recall abilities in the case of first and second-grade pupils, who were French native speakers, enrolled in a Dutch-speaking CLIL program. Our research objectives were threefold: (1) replicate previous designs aimed at comparing single-language L1 input compared to mixed-language input, (2) comparing language-mixing to a single-language L2 condition and (3) explore the effects of the type of input on recall abilities over time. Overall, we found (1) a mixing cost in both recall and response times in the comparison between the mixed-language condition and the single-language L1 condition, (2) a mixing benefit in terms of accurate recall, but a mixing cost in response times when compared to L2 single-language input and (3) an overall decrease in both accurate recall and response times in performance over time.

4.1 The influence of language-mixing compared to L1

We did observe significant differences in recall accuracy and response times depending on whether the pupils had been exposed to single-language L1 or mixed-language input. One of the possible explanations for the presence of this mixing cost, is our participants' rate of school exposure to the CLIL language (L2). Indeed, we tested pupils with 30% of exposure to the CLIL language (i.e., Dutch) through their weekly classes. Our participants also had a relatively short experience time in CLIL at the time of the experiment, as they had no prior experience with it from primary school. Moreover, in our context of research, Dutch is not mandatory as the second language

for all primary school pupils. According to the *weaker link hypothesis* (Gollan et al., 2008), associations between words and concepts are highly dependent on prior exposure and experience. Following this hypothesis, less L2 exposure would lead to weaker associations between the words in the L2 input and their conceptual counterparts and subsequently hinder the participants' ability to accurately recall them during the test phase. This partly explains the mixing cost (compared to L1) and benefit (compared to L2 input) we observed.

Our results clearly differ from those reported in the aforementioned studies on which we based our design (namely Antón et al., 2016), as they reported no mixing costs. This is most probably due to a couple of key differences between the groups of participants that were tested. Firstly, Anton et al. tested participants with a perfect 50–50% balance between L1 and L2 exposure at school, which was not the case for our participant group. Secondly, our participants were younger than those included in the study of Antón and colleagues (13.17 years old compared to 14.38 years old) which means they have had less experience with CLIL throughout their educational journey. In line with the weaker link hypothesis described above, higher L2 exposure and experience of participants featured in previous studies may have led to them having stronger word-concept associations and therefore better recall performances of L2 input. Drawing on those educational and linguistic disparities between both sets of participants, it can also be assumed that our participants show lower L2 proficiency compared to those of Antón et al. (2016). This lower L2 proficiency could partly explain some of the mechanisms underlying the mixing costs and benefits in recall. Indeed, previous studies on recall of auditory input also emphasize that their findings “should be limited to the case in which learners have sufficient L2 ability to understand the lexical items of the listening passages” (Sakai, 2009, p. 368). Other studies also underscore the importance of vocabulary size and general L2 proficiency when it comes to accurate processing of auditory input (e.g., Wang and Treffers-Daller, 2017). Moreover, our analyses clearly show the influence of L2 proficiency as the observed mixing benefits (vs. L2) and costs (vs. L1) both decrease as L2 proficiency increases.

4.2 The influence of language-mixing compared to L2

When expanding on previous studies' designs (cf. Antón et al., 2015, 2016), we report a mixing benefit in terms of recall abilities in the mixed-language condition compared to the single-language L2 condition. Similar to the L1 mixing cost discussed in the previous section, this finding seems to indicate a recall disadvantage related to low L2 exposure and proficiency, which is partly compensated for by the presence of the L1 in mixed-language input. Since every object definition was comprised of two distinct characteristics, they can rely on the L1 characteristic to help them decode the L2 characteristic and the conceptual information of the mixed-language input. This scaffolding technique is, of course, not possible in the single-language L2 context.

In terms of response latencies, the mixing cost we observed at the first time point in comparison to the L2 context is not completely in line with previous studies (Declerck et al., 2019) and may point toward a greater use of cognitive resources when decoding a mix of

two languages within the definition of a single object. In a series of three experiments, Declerck and colleagues only found mixing costs in the case of language pairs with a high cognate rate (French and Spanish). Our findings provide an account of mixing costs even in languages with a lower cognate rate (French and Dutch). This possibly indicates the presence of (proactive) language control when decoding mixed-language input, thereby requiring additional cognitive resources, which results in longer response latencies. It bears reiterating that these longer response times do not lead to less accurate recall when compared to single-language L2 input. Moreover, they are reversed in delayed recall. This finding thus highlights the need for both speed and comprehension measures in order to gain a more exhaustive understanding of the impact of language-mixing on (auditory) recall.

4.3 The influence of language-mixing over time

Overall, accurate recall decreases over time across all conditions, which is in line with our expectations given the nature of the design. More interestingly, both the observed mixing benefit and mixing cost follow a similar downward pattern as time progresses, although this finding deserves some nuancing. With regards to the L2 context, this could be due to the fact that accuracy scores were already very low at the immediate test time, thus pointing toward a floor effect in this context. When compared to the L1 context, the mixing cost indeed decreases, but the mixed context still presents the steepest decrease between both test times. Our presented findings are therefore certainly not exhaustive as far as the effect of language-mixing over time is concerned.

Response times, on the other hand, were also significantly faster at the delayed test time. Higher processing speed at the second test phase could be explained by the fact that participants were already familiar with the visual stimuli at this point, which could have led to priming effects that allow for faster responses (e.g., Misra et al., 2012). However, the faster response latencies do not omit that we report less accurate recall scores at the delayed test time. Indeed, it would seem that participants forgot some of the information provided during the first day of the experiment. This significant decrease in accurate recall of information over time reinforces the well-known need for spaced repetition of learning materials in education and for memory retention in general (e.g., Weinstein et al., 2018; Abbas et al., 2023).

4.4 Implications for research and practice

Regarding research implications, our study underscores the importance of considering the specificities of the participants' linguistic profiles and educational contexts. As we report in the present research, taking into account different levels of L2 proficiency and exposure at school brings about different outcomes in comparison to previous studies (Antón et al., 2015, 2016), even when replicating their design. Moreover, our design allowed to shed light on unraveled evolutions in the findings over time, especially in the case of response times. Our study also underscores the importance of replicating studies in general, especially in cases where the existing body of

literature is scarce. Finally, there is a need for additional studies aiming at bridging the gap between laboratory settings and classroom practices, by increasing the ecological validity of their experimental paradigms.

When transposed to classroom practices, these findings may raise concerns regarding the implementation of language-mixing in CLIL. Although the mixing cost we found compared to L1 input should certainly not be disregarded, the pedagogical goals of a CLIL program encompass both content and language outcomes (Martens et al., 2023). As previous studies have shown, CLIL pupils often outperform their peers in terms of L2 proficiency (see systematic reviews by Graham et al., 2018; Goris et al., 2019), even in CLIL programs with rather low L2 exposure (Bulté et al., 2021). We expect that, as the pupils' L2 proficiency grows over time, the mixing cost compared to learning in their first language will subside, especially if L1 and L2 exposure at school is kept more or less in balance (as was the case in Antón et al., 2016). Yet, more research needs to be done to confirm this hypothesis.

In the meantime, our study demonstrates the utility of language-mixing for pupils who find themselves at the onset of their CLIL trajectory. For this profile of learners in particular, language-mixing even mitigates the learning costs (in terms of memory recall at least), which would otherwise occur in a single-language L2 environment. While we only tested their recall of conceptual information, our findings could also be applied to recall of subject specific discourse and other contents. Incorporating language-mixing as a classroom practice then, appears to represent a great scaffold for those learners who have not yet attained the L2 proficiency level which is required to avoid those L2 learning costs.

4.5 Limitations and future research

Despite the relevant implications of our findings, the current study also presents some limitations which we would like to address. The first one pertains to our research design. Although the experimental task we used allowed us to single out the impact of language-mixing, this controlled computer task entails a more artificial setting for students compared to a real classroom experience. The same critique could be expressed with regards to the stimuli we created. The fact that these were definitions of everyday objects which did not pertain to an actual school subject might have reinforced the artificial nature of our design. Moreover, our stimuli only included inter-sentential mixing, leaving us with no findings concerning other language-mixing types (i.e., intra-sentential and tag mixing, Poplack, 1980). Additionally, because we had an identical number of participants, this allowed for a more clear-cut comparison with the study we aimed to replicate (cf. Antón et al., 2016). Yet, our relatively low sample size of 29 participants is another limitation that should be addressed as it may raise questions with regards to the generalizability of the results.

In order to further our understanding the influence of the type of input on learning outcomes, we would advise future research to investigate different types of language-mixing and/or other student profiles in terms of L2 proficiency and exposure. A significant increase in sample size is also recommended for even stronger arguments in favor of the generalizability of the findings. Even more

importantly, perhaps, studies in the future should investigate the influence of language-mixing on other learning processes. In the present study, we explored auditory recall abilities, which is situated at a low level of learning compared to other processes like understanding and applying knowledge (Anderson, 2010). Ultimately, while our study sheds light on the nuanced effects of language-mixing on recall abilities, future research must delve deeper into diverse language-mixing types and broader learning processes, thereby paving the way for more evidence-based and effective educational strategies. That being said, our study most importantly reveals that language-mixing in CLIL provides significant scaffolding benefits for learners with lower L2 proficiency thereby enhancing their learning outcomes. In light of our findings, we would certainly join “the plea to end the language mixing-taboo” (Antón et al., 2016) initiated by previous studies, although we also underscore the nuances of the implementation of language-mixing depending on the specific context in terms of L2 exposure and pupil characteristics with regard to their proficiency in the CLIL language.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Human Sciences Ethics Committee, Vrije Universiteit Brussel (ECHW_490). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

TC: Conceptualization, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing. MD: Conceptualization, Funding acquisition, Supervision, Writing – review & editing. ES: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This publication was made possible by the funding of research project SRP88 by the Vrije Universiteit Brussel (VUB).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations,

References

- Abbas, A. M., Hamid, T., Iwendi, C., Morrissey, F., and Garg, A. (2023). "Improving learning effectiveness by leveraging spaced repetition (SR)" in Big data and cloud computing. eds. N. Venkataraman, L. Wang, X. Fernando and A. F. Zobaa (Singapore: Springer Nature) 1021, 145–160.
- Anderson, A. (2010). 1. Varieties, taxonomies, and definitions. In A. Anderson, M. Bergunder, A. F. Droogers and LaanC. van der (Eds.), *Studying global pentecostalism: theories and methods* (pp. 13–29). Berkeley, California, United States of America: University of California Press
- Antón, E., Thierry, G., and Duñabeitia, J. A. (2015). Mixing languages during learning? Testing the one subject—one language rule. *PLoS One* 10:e0130069. doi: 10.1371/journal.pone.0130069
- Antón, E., Thierry, G., Gaborov, A., Anasagasti, J., and Duñabeitia, J. A. (2016). Testing bilingual educational methods: a Plea to end the language-mixing taboo. *Lang. Learn.* 66, 29–50. doi: 10.1111/lang.12173
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Brybaert, M. (2013). LexTALE_FR: a fast, free, and efficient test to measure language proficiency in French. *Psychol. Belgica* 53, 23–37. doi: 10.5334/pb-53-1-23
- Bulté, B., Surmont, J., and Martens, L. (2021). The impact of CLIL on the L2 French and L1 Dutch proficiency of Flemish secondary school pupils. *Int. J. Biling. Educ. Biling.* 25, 3151–3170. doi: 10.1080/13670050.2021.2018400
- Cenoz, J., and Gorter, D. (2022). Pedagogical Translanguaging and its application to language classes. *RELC J.* 53, 342–354. doi: 10.1177/00336882221082751
- Christoffels, I. K., Firk, C., and Schiller, N. O. (2007). Bilingual language control: an event-related brain potential study. *Brain Res.* 1147, 192–208. doi: 10.1016/j.brainres.2007.01.137
- Costa, A., Caramazza, A., and Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 1283–1296. doi: 10.1037/0278-7393.26.5.1283
- de Bruin, A., Samuel, A. G., and Duñabeitia, J. A. (2020). Examining bilingual language switching across the lifespan in cued and voluntary switching contexts. *J. Exp. Psychol. Hum. Percept. Perform.* 46, 759–788. doi: 10.1037/xhp0000746
- Declerck, M. (2020). What about proactive language control? *Psychon. Bull. Rev.* 27, 24–35. doi: 10.3758/s13423-019-01654-1
- Declerck, M., Grainger, J., and Hartsuiker, R. (2021). Proactive language control during bilingual sentence production. *Int. J. Biling.* 25, 1813–1824. doi: 10.1177/13670069211047803
- Declerck, M., and Koch, I. (2023). The concept of inhibition in bilingual control. *Psychol. Rev.* 130, 953–976. doi: 10.1037/rev0000367
- Declerck, M., Koch, I., Duñabeitia, J. A., Grainger, J., and Stephan, D. (2019). What absent switch costs and mixing costs during bilingual language comprehension can tell us about language control. *Journal of experimental psychology. J. Exp. Psychol. Hum. Percept. Perform.* 45, 771–789. doi: 10.1037/xhp0000627
- Declerck, M., Koch, I., and Philipp, A. M. (2012). Digits vs. pictures: the influence of stimulus type on language switching. *Biling. Lang. Cogn.* 15, 896–904. doi: 10.1017/S1366728912000193
- Domalewska, D. (2017). Discourse analysis of teacher talk: code switching in content and language integrated learning (CLIL) classrooms in Thailand. *Asian J. Educ. E-Learning* 5, 36–43. doi: 10.24203/ajeel.v5i2.3962
- European Commission: Directorate-General for Education, Youth, Sport and Culture. (2006). *Content and language integrated learning (CLIL) at school in Europe*.
- García, O., and Flores, N. (2012). Multilingual pedagogies. In: Martin-Jones, M., Blackledge, A. and Creese, A. (eds.) *The Routledge handbook of multilingualism*, (e-book). Routledge. 232–246.
- Gimenes, M., and New, B. (2016). Worldlex: twitter and blog word frequencies for 66 languages. *Behav. Res. Methods* 48, 963–972. doi: 10.3758/s13428-015-0621-0
- Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001
- Goris, J., Denessen, E., and Verhoeven, L. (2019). Effects of content and language integrated learning in Europe: a systematic review of longitudinal experimental studies. *Eur. Educ. Res. J.* 18, 675–698. doi: 10.1177/1474904119872426
- Graham, K., Choi, Y., Davoodi, A., Razmeh, S., and Dixon, L. (2018). Language and content outcomes of CLIL and EMI: a systematic review. *Latin Am. J. Content Lang. Integrated Learn.* 11, 19–38. doi: 10.5294/laclil.2018.11.1.2
- Green, D. W., and Abutalebi, J. (2013). Language control in bilinguals: the adaptive control hypothesis. *J. Cogn. Psychol.* 25, 515–530. doi: 10.1080/20445911.2013.796377
- Grunden, N., Piazza, G., García-Sánchez, C., and Calabria, M. (2020). Voluntary language switching in the context of bilingual aphasia. *Behav. Sci.* 10:9. doi: 10.3390/bs10090141
- Harding, J. F., Morris, P. A., and Hughes, D. (2015). The Relationship Between Maternal Education and Children's Academic Outcomes: A Theoretical Framework. *J. Marriage Fam.* 77, 60–76. doi: 10.1111/jomf.12156
- Hoshino, N., and Kroll, J. F. (2008). Cognate effects in picture naming: does cross-language activation survive a change of script? *Cognition* 106, 501–511. doi: 10.1016/j.cognition.2007.02.001
- Lambert, W. E., Tucker, G. R., and d'Anglejan, A. (1973). Cognitive and attitudinal consequences of bilingual schooling. *J. Educ. Psychol.* 65, 141–159. doi: 10.1037/h0034983
- Lemhöfer, K., and Broersma, M. (2012). Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behav. Res. Methods* 44, 325–343. doi: 10.3758/s13428-011-0146-0
- Lialikhova, D. (2019). "We can do it together!" – but can they? How Norwegian ninth graders co-constructed content and language knowledge through peer interaction in CLIL. *Linguist. Educ.* 54:100764. doi: 10.1016/j.linged.2019.100764
- Lin, A. M. Y. (2015). Conceptualising the potential role of L1 in CLIL. *Lang. Cult. Curric.* 28, 74–89. doi: 10.1080/07908318.2014.1000926
- Liste de fréquence lexicale. (2023). Éduscol | Ministère de l'Éducation Nationale | Direction générale de l'enseignement scolaire. Available at: <https://eduscol.education.fr/186/liste-de-frequence-lexicale> (Accessed October 22, 2024)
- Lu, C., Gu, M. M., and Lee, J. C.-K. (2023). A systematic review of research on translanguaging in EMI and CLIL classrooms. *Int. J. Multiling.*, 1–21. doi: 10.1080/14790718.2023.2256775
- Ma, F., Li, S., and Guo, T. (2016). Reactive and proactive control in bilingual word production: an investigation of influential factors. *J. Mem. Lang.* 86, 35–59. doi: 10.1016/j.jml.2015.08.004
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- Marsh, D. (2013). *Content an language integrated learning (CLIL). A development trajectory.* [Doctoral dissertation, Universidad de Córdoba]. Dialnet. Available at: <http://helvia.uco.es/xmlui/handle/10396/8689>
- Martens, L., Mettwie, L., and Elen, J. (2023). Looking for the i in CLIL: a literature review on the implementation of dual focus in both subject and language classrooms. *Nordic J. Lang. Teach. Learn.* 11, 255–277. doi: 10.46364/njtl.v11i3.1155
- Meade, G., Midgley, K. J., Sevcikova Sehyr, Z., Holcomb, P. J., and Emmorey, K. (2017). Implicit co-activation of American sign language in deaf readers: an ERP study. *Brain Lang.* 170, 50–61. doi: 10.1016/j.bandl.2017.03.004
- Mettwie, L., and Van Mensel, L. (2023). Understanding foreign language education and bilingual education in Belgium: a (surreal) piece of cake. *Int. J. Biling. Educ. Biling.* 26, 639–657. doi: 10.1080/13670050.2020.1768211

or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2024.1520791/full#supplementary-material>

- Misra, M., Guo, T., Bobb, S. C., and Kroll, J. F. (2012). When bilinguals choose a single word to speak: electrophysiological evidence for inhibition of the native language. *J. Mem. Lang.* 67, 224–237. doi: 10.1016/j.jml.2012.05.001
- Nikula, T., and Moore, P. (2019). Exploring translanguaging in CLIL. *Int. J. Biling. Educ. Biling.* 22, 237–249. doi: 10.1080/13670050.2016.1254151
- Papaja, K. L., and Wysocka-Narewska, M. S. (2020). Investigating code-switching in a content and language integrated learning (CLIL) classroom. *Theory Pract. Second Lang. Acquisit.* 6, 51–63. doi: 10.31261/TAPSLA.7808
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching. *Linguistics* 18, 581–618. doi: 10.1515/ling.1980.18.7-8.581
- Prilutskaya, M. (2021). Examining pedagogical Translanguaging: a systematic review of the literature. *Languages* 6:4. doi: 10.3390/languages6040180
- Proctor, C. P., and Mo, E. (2009). The relationship between cognate awareness and English comprehension among Spanish–English bilingual fourth grade students. *TESOL Q.* 43, 126–136. doi: 10.1002/j.1545-7249.2009.tb00232.x
- Reyes, I. (2004). Functions of code switching in Schoolchildren's conversations. *Biling. Res. J.* 28, 77–98. doi: 10.1080/15235882.2004.10162613
- Sahan, K., Galloway, N., and McKinley, J. (2022). 'English-only' English medium instruction: mixed views in Thai and Vietnamese higher education. *Lang. Teach. Res.*:13621688211072632. doi: 10.1177/13621688211072632
- Sakai, H. (2009). Effect of repetition of exposure and proficiency level in L2 listening tests. *TESOL Q.* 43, 360–372. doi: 10.1002/j.1545-7249.2009.tb00179.x
- San Isidro, X., and Lasagabaster, D. (2019). Code-switching in a CLIL multilingual setting: a longitudinal qualitative study. *Int. J. Multiling.* 16, 336–356. doi: 10.1080/14790718.2018.1477781
- Serra, J., and Feijoo, S. (2022). Teacher translanguaging in CLIL primary education: Do teachers' perceptions match their real practices? (Translingüismo y aprendizaje integrado de contenidos y lengua extranjera (AICLE) en la educación primaria: percepciones y prácticas reales del profesorado). *J. Study Educ. Dev.* 45, 280–310. doi: 10.1080/02103702.2021.2009294
- Tai, K. W. H., and Wei, L. (2021). Constructing playful talk through Translanguaging in English medium instruction mathematics classrooms. *Appl. Linguis.* 42, 607–640. doi: 10.1093/applin/amaa043
- Wang, Y., and Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: the contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System* 65, 139–150. doi: 10.1016/j.system.2016.12.013
- Wei, L. (2018). Translanguaging as a practical theory of language. *Appl. Linguis.* 39, 9–30. doi: 10.1093/applin/amx039
- Weinstein, Y., Madan, C. R., and Sumeracki, M. A. (2018). Teaching the science of learning. *Cognit. Res.* 3:2. doi: 10.1186/s41235-017-0087-y
- Williams, C. (1994). An evaluation of teaching and learning methods in the context of bilingual secondary education. [Doctoral dissertation, Bangor University].
- Zhou, X. E., and Mann, S. (2021). Translanguaging in a Chinese university CLIL classroom: teacher strategies and student attitudes. *Stud. Second Lang. Learn. Teach.* 11, 265–289. doi: 10.14746/sslit.2021.11.2.5