Bilingualism: Language and Cognition 22 (1), 2019, 47–57 © Cambridge University Press 2017 doi:10.1017/S1366728917000530

# Language-dependent knowledge acquisition: investigating bilingual arithmetic learning*

CHRISTIAN G. K. HAHN
*Institute of Psychology, University of Göttingen, Germany*
*Faculty of Education, University of Leipzig, Germany*
HENRIK SAALBACH
*Faculty of Education, University of Leipzig, Germany*
ROLAND H. GRABNER
*Institute of Psychology, University of Graz, Austria*

(Received: January 18, 2017; final revision received: August 24, 2017; accepted: August 30, 2017; first published online 5 October 2017)

*Previous studies revealed language-switching costs (LSC) in bilingual learning settings, consisting of performance decreases when problems are solved in a language different from that of instruction. Strong costs have been found for arithmetic fact knowledge. The aim of the present study was to investigate whether LSC in arithmetic also emerge in an auditory learning task and in pure fact learning. Furthermore, we tested whether LSC are influenced by the direction of language-switching. Thirty-three university students learned arithmetic facts of three different operations (i.e., multiplication, subtraction, artificial facts) over a period of four days. The training was either in German or English. On day five, participants solved problems in both languages. Results revealed LSC in response latencies for all three types of problems, independent of the direction of language-switching. These findings suggest that LSC are modality-unspecific and occur independent of the type of arithmetic fact knowledge.*

Keywords: bilingual learning, language-switching costs, arithmetic learning, fact knowledge

## Introduction

Bilingual learning receives increasing attention both by public and research. One well-known example within the educational field is Content and Language Integrated Learning (CLIL), which represents a dual-focused instructional approach to simultaneously teach content while improving language skills in a foreign language (Eurydice, 2006; Lasagabaster & Sierra, 2009). For example, mathematics or geography are taught in English to German native speakers who have learned English as a second language. Despite the great success of these programs to foster language learning (e.g., Zaunbauer, Bonerad & Möller, 2005; Zaunbauer & Möller, 2009), it is an open question whether and to what extent the acquired knowledge is represented in a language-dependent or language-independent way. This question is not only of theoretical but also of practical relevance. Language-dependent knowledge representations may cause cognitive costs if the language of instruction differs from the language of knowledge retrieval and application. For instance, a student who acquires mathematical knowledge in a foreign language may not be able to use this knowledge in his native language as effectively as when he had learned it in his mother tongue. The costs commonly consist of longer solution times and higher error rates. So far, the so-called language-switching costs (LSC) have been reported for retrieving arithmetic (Spelke & Tsivkin, 2001; Grabner, Saalbach & Eckstein, 2012; Saalbach, Eckstein, Andri, Hobi & Grabner, 2013), and other numerical and non-numerical fact knowledge (Marian & Fausey, 2006), as well as recalling autobiographic information (Marian & Neisser, 2000). The present paper aims to further investigate the extent, correlates and mechanisms of LSC in the domain of arithmetic.

### *Language and knowledge representation in arithmetic*

Language affects how people process information and knowledge is stored in memory (e.g., Gentner & Goldin-Meadow, 2003; Gumperz & Levinson, 1996; Malt & Wolff, 2010; Wolff & Holmes, 2011 for review). As a consequence, cognitive differences between speakers of different languages can be detected across a wide range of domains (e.g., Boroditsky, Fuhrman & McCormick, 2011; Fausey & Boroditsky, 2011; Saalbach & Imai, 2007). For mathematics, Miller, Smith, Zhu and Zhang (1995) found that the structure of the numerical system affects how quickly children develop basic counting and arithmetic abilities. For instance, compared to Chinese children, U.S. children had more problems understanding the base-10 structure, committing more counting errors (e.g., counting "twenty-eight, twenty-nine, twenty-ten, twenty-eleven"; see also Fuson & Kwon, 1992; Park, 1999

* I am grateful to Dr. Stephan Vogel, Frieder Schillinger, and Maria Schneider for their support during the process of conducting this research.

Address for correspondence:
Christian G. K. Hahn, Room 216, Faculty of Education, University of Leipzig, Marschnerstr. 31, 04109 Leipzig
*christian.hahn@uni-leipzig.de*

https://doi.org/10.1017/S1366728917000530 Published online by Cambridge University Press

for an overview). In addition, the phonological structure of number words affects performance. For instance, cross-language performance differences have been reported between Mandarin and English (Chen, Cowell, Varley & Wang, 2009) and between English and Welsh speaking language groups (Ellis & Hennelly, 1980). In the study by Chen and colleagues (2009), thirty native Mandarin Chinese and thirty native English speakers were tested on verbal and visuo-spatial working memory span (e.g., forward and backward digit span task). Results revealed significantly higher scores in the Mandarin Chinese speaking group for verbal working memory span than in the English-speaking group. The advantage of Mandarin was associated with the shorter articulation time for digits in spoken Mandarin Chinese. In arithmetic, the association between language and numerical cognition has been found predominantly for exact calculation (exact solution of an arithmetic problem) rather than approximate calculation (Dehaene & Cohen, 1997; Spelke & Tsivkin, 2001; Lemer, Dehaene, Spelke & Cohen, 2003). These findings are in line with neuroimaging studies, showing that the retrieval of (exact) arithmetic facts is in close connection to brain circuits associated with language processing and storage of verbal information (e.g., Lee, 2000; Dehaene, Molko, Cohen & Wilson, 2004; Domahs & Delazer, 2005; Venkatraman, Siong, Chee & Ansari, 2006; but see Benn, Zheng, Wilkinson, Siegal & Varley, 2012; Klessinger, Szczerbinski & Varley, 2012).

In bilinguals, arithmetic knowledge seems to be strongly related to the language of acquisition, which is typically the mother tongue. For instance, Frenck-Mestre and Vaid (1993) required bilingual participants to perform simple addition problems (e.g., 2 + 5) as well as simple multiplication problems (e.g., 7 x 3). Performance was slower and less accurate when calculating in their second language (L2) than in their first language (L1). Similarly, German–French bilingual adolescents showed better performance when arithmetic tasks were presented in L1 (German) compared to L2 (French), even though later, in secondary education, mathematics had been taught in French. The effect was greater for complex addition problems (e.g., 56 + 32) compared to more simple addition problems (e.g., 4 + 2; van Rinsveld, Brunner, Landerl, Schiltz & Ugen, 2015). Taken together, research in the field of bilingual mathematics learning suggest that language is relevant for task performance. What is the implication for bilingual learning settings when language of encoding and language of retrieval differ?

### *Bilingual arithmetic learning and language-switching costs*

According to the ENCODING-SPECIFICITY HYPOTHESIS the effectiveness of retrieving facts from memory is in close relation to the context in which information had been encoded (e.g., Barber, Rajaram & Aron, 2010; Tulving & Thomson, 1973). With respect to bilingual learning, this would suggest that the retrieval and application of knowledge is most effective in the language of encoding. When a person needs to solve a task in a language that is different from the language of encoding (or instruction, respectively), cognitive costs may emerge. Such LSC have been reported in previous research (Spelke & Tsivkin, 2001, Grabner et al., 2012; Saalbach et al., 2013). Spelke and Tsivkin (2001), for example, had Russian–English bilinguals undergo two training sessions consisting of different set of problems including exact calculations (e.g., "What is the sum of fifty-four and forty-eight?"), and approximation tasks (e.g., "Estimate the approximate cube root of twenty-nine!"). The testing situation included two kind of verification tasks in which participants had to decide which one was the exact answer (exact number task), or which one is closest to the exact number (approximation number task). LSC were specific to the exact number tasks as opposed to the approximation tasks as well as to a third task, including non-numerical information. The authors concluded that exact arithmetic is more strongly language-dependent than approximate arithmetic. Saalbach and colleagues (2013) investigated to what extent LSC in arithmetic are moderated by the arithmetic operation and whether they generalize to untrained problems. Thirty-nine bilingual high school students underwent a three-day training of fourteen multiplication and fourteen subtraction facts either in German (L1) or in French (L2). During training and test, problems were displayed in number-words (e.g., "twelve times seven"). In the test session, participants were presented with the trained as well as untrained problems in both languages. Results revealed that participants had longer response latencies as well as lower accuracy rates for both multiplication and subtraction problems when language-switching was required. To notice, LSC for the trained problems did not depend on the arithmetic operation. This was unexpected, since it is commonly argued that multiplication problems rely more strongly on a verbal coding than subtraction problems, which are associated with mental manipulation of magnitude (e.g., Dehaene et al., 2004; Ischebeck, Zamarian, Siedentopf, Koppelstätter, Benke, Felber & Delazer, 2006). Thus, by manipulating the language, stronger LSC for multiplication problems had been expected. Interestingly, LSC also emerged in the untrained problems, suggesting that the impact of the language of instruction may not only affect fact retrieval but also the recall of other kinds of knowledge such as procedural knowledge. In addition, LSC were stronger when participants switched from their dominant language (L1, German) to the non-dominant language (L2, French) than vice versa (see also Marian and Fausey, 2006, for similar findings). The mechanisms underlying LSC in

https://doi.org/10.1017/S1366728917000530 Published online by Cambridge University Press

arithmetic were investigated by Grabner et al. (2012). They used functional magnetic resonance imaging (fMRI) to scrutinize which neuro-cognitive processes might be associated with LSC. During a four-day training, twenty-nine participants learned ten subtraction and ten multiplication facts presented in number-words either in German or Italian. Throughout the test session, participants had to solve trained and untrained problems in both languages. In line with Saalbach and colleagues (2013), LCS were found both for trained and untrained problems in response latencies and accuracy rates as well as for multiplication and subtraction problems. Moreover, results revealed an association between LSC and activation in areas related to magnitude processing, implying that LSC may be due to additional numerical processing rather than to mere language translation. As for the behavioral results, the association between LSC and neural correlates was independent of the arithmetic operations.

In sum, previous studies on language-dependency in arithmetic learning consistently reveal LSC in response latencies and accuracy rates. In addition, LSC appear to be independent of the arithmetic operation, arguing for a similar cognitive cause regarding rote learned information (i.e., fact knowledge). Furthermore, LSC were found for different language combinations, highlighting the important role that a mismatch of the language of instruction and language of application can have on performance. Findings also suggest a directional effect in that LSC are higher when switching is required from L1 (first language) to L2 (second language).

Even though previous research has provided first important insights into the language-dependency of knowledge representation in arithmetic, some methodological limitations and open questions need also to be taken into consideration. First, in all three studies on arithmetic, stimuli were presented in written form (e.g., "three times twelve"), which is hardly used in educational practice and thus represents a substantial limitation of ecological validity. Second, verification tasks were used, which do not resemble authentic arithmetic problem solving and may even produce undesired effects. Indeed, the solutions to problems could be guessed instead of calculated by applying certain strategies (e.g., eliminating obviously wrong answers). Moreover, verification tasks produce an interference effect in which response latencies increase and solution rates decrease the closer the numerical distance is between the correct answer and the distractor (Ashcraft & Battaglia 1978; Ashcraft & Stazyk, 1981). Third, the cognitive mechanism underlying LSC in processing arithmetic problems is still unclear. Based on neuroimaging data as stated previously, Grabner and colleagues (2012) concluded that LSC may be the result of additional quantity processing (such as calculation) rather than mere translation into the testing language after fact retrieval in the language of training. However, earlier findings by Marian and Fausey (2006) revealed that LSC also apply to the retrieval of non-arithmetic knowledge, showing that mere quantity processing is unlikely to account for the appearance of LSC alone. Finally, the potential interaction of second language proficiency and LSC requires exploration. Marian and Fausey (2006) argued that participants rely more on the higher-proficiency language during the encoding phase, therefore finding higher LSC when switching from the dominant to the non-dominant language than vice versa. However, language proficiency was assessed by means of self-reports to categorize participants as dominant or non-dominant speakers but not with an objective measure of language proficiency. As other research revealed, language-proficiency is critical for cognitive performance across different domains of academic learning including mathematics (e.g., Kempert, Saalbach & Hardy, 2011). Thus, it is important to take language proficiency into account when studying LSC within arithmetic learning.

The main aim of the present study was to further our knowledge about the language-dependency of arithmetic knowledge and the nature of the LSC. In particular, we first investigated whether previous findings in German–English bilinguals can be replicated by using the ecologically more valid auditory stimuli (research question 1). Second, we further examined the mechanisms underlying LSC by comparing the extent of LSC after learning artificial vs. real arithmetic facts (research question 2). Specifically, in addition to multiplication and subtraction problems, we included artificial problems requiring pure fact learning (i.e., ## box # = ##). Arithmetic problems, even if extensively trained, may not only be solved by fact retrieval but also by other (e.g., magnitude-related) processes. These processes have even been discussed as a major cause of LSC in arithmetic (Grabner et al., 2012). In the artificial problems, however, such alternative strategies can be precluded. Third, we investigated whether the extent of LSC depends on the direction of language switching (from L1 to L2 or v.v.; research question 3). Finally, we explored whether and to what extent an indicator for L2 proficiency modulates the size of the LSC (research question 4).

We hypothesized that problems for all three tasks involving auditory material are solved more slowly and less accurately when the language of instruction differs from the language of application (i.e., when language-switching is required). (Hypothesis 1). Furthermore, we predicted LSC to appear for all three tasks (i.e., multiplication, subtraction and artificial problems) since all problems are likely to represent fact knowledge after a training period of four days, independently of their individual type (Hypothesis 2). In line with previous research (Marian & Fausey, 2006; Saalbach et al., 2013)

https://doi.org/10.1017/S1366728917000530 Published online by Cambridge University Press

we expected to find stronger LSC when knowledge, which has been encoded in the dominant language, is retrieved in the non-dominant language as compared to a situation when knowledge, acquired in the non-dominant knowledge, needs to be retrieved in the dominant language (Hypothesis 3).

## Methods

### *Participants*

Thirty-three university students at the University of Göttingen, Germany, underwent the training and test procedure. One participant had to be excluded due to missing the last training session. The final sample consisted of 32 participants, with 20 being female and 12 being male. Half of the participants received the training in German (dominant language, L1), the other half in English (non-dominant language, L2). All participants were native German-speaking and had at least seven years of formal English education. The LexTALE test for vocabulary knowledge showed a mean score of 73% (SD = 12), which is supposed to indicate language proficiency at an upper intermediate level (B2; Lemhöfer & Broersma, 2012).

### *Material*

Three different arithmetic tasks were included as experimental stimuli, comprising a) six multiplication problems: two-digit times one-digit numbers with two-digit solutions (00 x 0 = 00); b) six subtraction problems: two-digit minus two-digit numbers, including only carry-over calculations with two-digit solutions (00 - 00 = 00); and c) six artificial problems: two-digit and one-digit numbers connected via an arbitrary symbol (box) with two-digit solutions (00 box 0 = 00). The multiplication and subtraction problems resembled those used in previous arithmetic training studies (Delazer, Domahs, Bartha, Brenneis, Lochy, Trieb & Benke, 2003; Ischebeck et al., 2006; Grabner & De Smedt, 2012; Saalbach et al., 2013). The first operand ranged from 14 to 17, and the second operand ranged from 3 to 7, keeping two-digit outcomes, and excluding solutions divisible by 5 or 10. Subtraction problems matched the multiplication problems regarding their difficulty (Ischebeck et al., 2006) to be comparable to each other. Artificial problems had the same structure as the multiplication problems – while having different operands – to account for best comparability. Auditory stimuli were created with the professional audio software Voice Reader Studio 15, a widely used text-to-speech program (Linguatec, 2015). The training and test program was created using *E-Prime 2.0 Professional stimulus presentation software* (Schneider, Eschmann & Zuccolotto, 2002). The online version of the LexTALE was chosen to indicate general English vocabulary knowledge. It has been developed to account for the increasing need in experimental studies to assess language vocabulary knowledge within a short time scale and been validated by Lemhöfer and Broersma (2012). The test consists of 60 items (40 words, 20 nonwords). Nonwords are orthographically legal and pronounceable, but represent nonsense strings. Participants have to indicate whether the word is an existing English word or not. Assessing language proficiency in its full detail would go beyond the scope of our present study. Nevertheless, this vocabulary test has not only been shown to be a better predictor than commonly used self-ratings for vocabulary knowledge (Lemhöfer & Broersma, 2012), but also shows a substantial correlation with more widely used vocabulary tests (e.g., r = .80 in Mochida & Harrington, 2006) as well as with the Quick Placement Test (QPT) that is frequently used to predict language proficiency (Quick Placement Test, 2001).

### *Training procedure*

Throughout the training period, participants were trained and tested in a room of the local Institute for Psychology. Training and testing was held in group sessions with up to four participants. Participants used headphones and were seated separately in cabins to eliminate distracting factors. Participants were randomly assigned to either the German training group or the English training group. They were instructed to learn 18 problems (i.e., 6 multiplications, 6 subtractions, and 6 artificial problems) over a period of four consecutive days. Instructions were given in written form in an extra session on the day of the first training session. Instructions were also repeated before each session to guarantee a problem-free run. Each training session consisted of three task-blocks (i.e., multiplication, subtraction, and artificial; see Figure 2a) including the six problems with six repetitions each. Thus, 108 problems were solved in each training session (6 problems x 6 repetitions x 3 tasks). At the end of the four training days, each problem had been repeated 24 times respectively. This number of repetitions has previously been shown to produce strong training effects (Ischebeck et al., 2006; Grabner et al., 2012; Saalbach et al., 2013). A one-minute break was included after each task-block. Overall, one training session lasted about 20 to 25 minutes. The order of the six different problems within each task block was randomized. No problem was repeated two times in a row. Trials started with a white fixation point for 1000 milliseconds on a black screen, followed by the auditory presentation of the problem while the screen remained black without the fixation point (see Figure 1). Participants were asked to press the ENTER key when having the correct answer in the instructed language ready to speak out loud. The maximum

https://doi.org/10.1017/S1366728917000530 Published online by Cambridge University Press

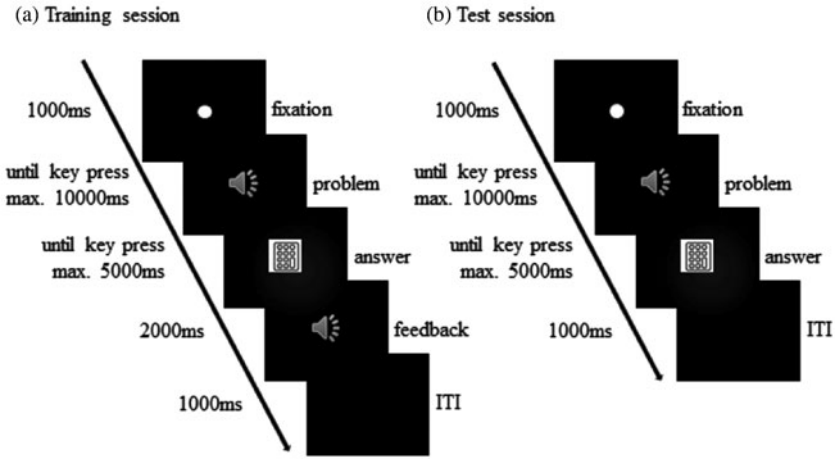



Figure 1. Schematic display of the trial time course during (a) the training and (b) the test session.

(a) Training session

| | Training 1 | Training 2 | Training 3 | Training 4 |
|---|---|---|---|---|
| Block 1 | mul | sub | art | art |
| Block 2 | art | mul | mul | sub |
| Block 3 | sub | art | sub | mul |

(b) Test session

| | Block 1 | Block 2 | Block 3 | Block 4 | Block 5 | Block 6 |
|---|---|---|---|---|---|---|
| Order A | art (ns) | art (s) | sub (ns) | sub (s) | mul (ns) | mul (s) |
| Order B | art (s) | art (ns) | sub (s) | sub (ns) | mul (s) | mul (ns) |

Figure 2. Schematic display of the block design for a) the training session and b) the test session. Within the test session, the two different orders are represented (ns: no-switching condition; s: switching condition).

time-frame to answer a problem was ten seconds from the start of the auditory stimuli. Next, the correct answer had to be typed in via a key-pad, confirming it with a key-press within five seconds. The typed numbers were visible on the screen in order to correct the answer if needed. Corrective feedback was given by a green or red display, followed by the auditory presentation of the correct answer, irrespectively of whether the given answer was correct or not. Thereby, we attempted to strengthen the training process and the connection with the training language. Further, the auditory feedback was essential in order for the artificial problems to be learned in the first place. Before the next trial started, an inter-trial-interval (ITI) of one second appeared. Reaction time (RT) and accuracy rate (ACC) were recorded.

***Test procedure***

On day five, participants underwent the test session, taking about 45 minutes. Compared to the training session, each block was presented in both, the trained language (no language-switching required) and the untrained language (language-switching required). The test consisted of six task blocks. In order to avoid additional executive-control processes and unlike in previous studies, participants did not switch between operations or languages within blocks. Each block consisted of 36 problems, which were either presented in the language of training or in the untrained language. Test order was counter-balanced, so that half of the sample started the test session with a task-block (e.g., multiplication) in the no language-switching condition (ns for no-switching) followed by a block of the same task

https://doi.org/10.1017/S1366728917000530 Published online by Cambridge University Press

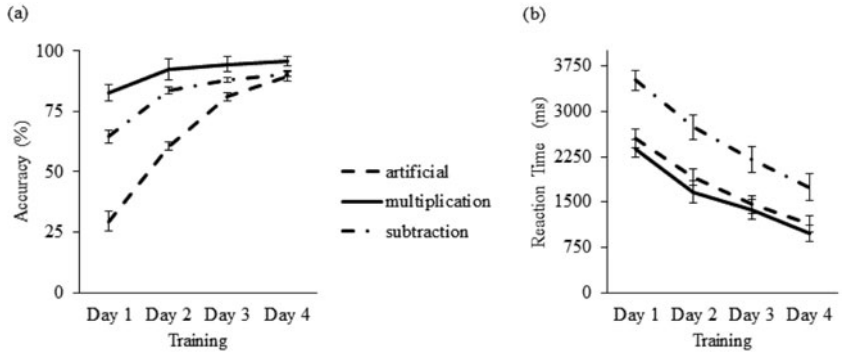



Figure 3. Training data for a) accuracy rates and b) reaction times. Separate lines represent the three different tasks.

(e.g., multiplication) in the language-switching condition (s for switching; order A) and vice versa (order B, see Figure 2b). Neither corrective feedback nor auditory presentation of the correct results was provided. RT and ACC were recorded.

### *Data analysis*

For statistical analyses, IBM SPSS Statistics 20 was used. To analyze the data, mixed design ANOVAs for RT and ACC were computed. To analyze the training data, the ANOVA contained the two within-subject factors Arithmetic Task (artificial vs. multiplication vs. subtraction) and Training Day (day 1 vs. day 2 vs. day 3 vs. day 4), and the between-subject factor Language of Training (German vs. English). The ANOVA for the testing data contained the two within-subject factors Arithmetic Task (artificial vs. multiplication vs. subtraction) and Language Switching (switching vs. no switching), as well as the between-subject factors Language of Training (German vs. English). Test Order (i.e., switching followed by no-switching vs. no-switching followed by switching) was included as covariate. RT data was only analyzed for correct trials. For effect sizes, Cohen's $d$ and partial eta-squared ($\eta_p^2$) were computed. In case of violation of the assumption of sphericity (Mauchly's test), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

## Results

### *Training data*

Training data for ACC and RT are displayed in Figure 3. For ACC, analyses revealed a main effect for the factor Training Day**:** $F(2.10, 62.84) = 142.96$, $p < .001$, $\eta_p^2 = .83$. Post-hoc t-tests showed a persistent effect for each consecutive day (all $p_s < .001$), indicating that ACC increased significantly in the course of the training. Moreover, a main effect was found for Arithmetic Task, $F(1.40, 41.81) = 64.33$, $p < .001$, $\eta_p^2 = .68$, indicating that ACC were not equal for all three arithmetic tasks. Post-hoc t-tests showed that – averaged over all four days – artificial problems were solved less accurately than multiplication problems (65% vs. 91%; $t(31) = -10.962$, $p < .001$, $d = 1.94$) and subtraction problems (65% vs. 82%; $t(31) = -5.831$, $p < .001$, $d = 1.03$), while subtraction problems were solved less accurately than multiplication problems (82% vs. 91%; $t(31) = -6.601$, $p < .001$, $d = 1.17$). Finally, an interaction effect was observed for Arithmetic Task and Training Day**,** ($F(3.85, 115.50) = 44.03$, $p < .001$, $\eta_p^2 = .59$, indicating that the magnitude for the daily training effects was different depending on the arithmetic task.

RT-analyses revealed a main effect for Training Day, $F(1.89, 56.61) = 113.85$, $p < .001$, $\eta_p^2 = .79$. Post-hoc t-tests showed that the effect was persistent for each consecutive day (all $p_s < .001$), indicating that response time decreased significantly in the course of the training. We further found a main effect for Arithmetic Task, $F(2, 90) = 33.63$, $p < .001$, $\eta_p^2 = .53$. Post-hoc t-tests showed no significant difference in RT between artificial and multiplication problems (1601 ms vs. 1762 ms, $t(31) = -1.239$, $p = .225$, $d = 0.22$), but between artificial problems and subtraction problems (1601 ms vs. 2548 ms; $t(31) = -6.858$, $p < .001$, $d = 1.21$), and between multiplication problems and subtraction problems (1762 vs. 2548 ms; $t(31) = -6.615$, $p < .001$, $d = 1.17$). Finally, we found an interaction effect between Arithmetic Task and Language of Training, $F(2, 60) = 3.99$, $p = .024$, $\eta_p^2 = .12$. Post-hoc analyses revealed that subtraction problems account for this interaction since they were solved more slowly in the English training group than in the German training group (2928 ms vs. 2167 ms; $t(30) = -2.12$, $p = .042$, $d = 0.37$).

### *Test data*

The ACC- and RT-results are shown in Table 1.

#### *Hypothesis 1: Problems are solved more slowly and less accurately when the language of instruction differs from the language of application.*

ACC and RT data for language-switching are depicted in Figure 4. For ACC, no effect for Language Switching was

https://doi.org/10.1017/S1366728917000530 Published online by Cambridge University Press

Table 1. *Mean reaction time in milliseconds (upper rows) and accuracy in percentage correct (lower rows) as a function of arithmetic task, and type of switching condition. Standard errors are enclosed in parentheses.*

| | Artificial | Multiplication | Subtraction |
|---|---|---|---|
| **Reaction time in milliseconds** | | | |
| No language switching | 889 (117) | 973 (154) | 1715 (286) |
| Language switching | 1156 (141) | 1240 (162) | 1879 (249) |
| **Accuracy in percentage correct** | | | |
| No language switching | 93.5 (1.9) | 97.5 (0.9) | 92.3 (1.8) |
| Language switching | 92.8 (2.4) | 97.2 (0.9) | 90.6 (2.4) |

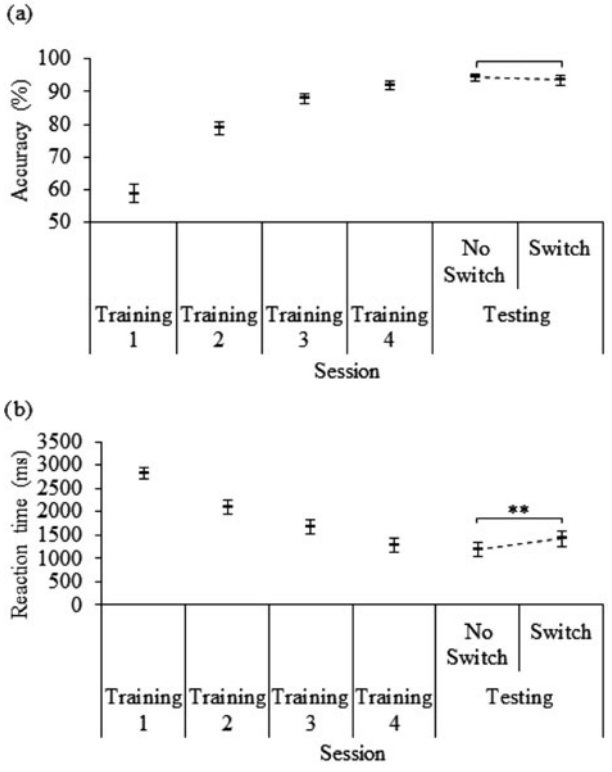



Figure 4. Illustration of accuracy and reaction time from training to testing. Error bars presenting the standard error. $^{*}p < .05$. $^{**}p < .01$.

found ($F(1, 29) = 1.36, p = .25, \eta_p^2 = .05$), demonstrating that the answers to language-switching problems (93.5%) were as accurate as the answers to problems where no language-switching was required (94.4%). For RT, however, we found the predicted main effect for Language Switching ($F(1, 29) = 7.26$, $p = .012$, $\eta_p^2 = .20$), indicating that problems requiring language-switching were solved more slowly than problems requiring no language-switching (1425 ms vs. 1192 ms).

Unexpectedly, a significant interaction effect between LANGUAGE SWITCHING and the covariate Test Order emerged ($F(1, 29) = 17.97$, $p < .001$, $\eta_p^2 = .38$). To break down this interaction, additional post hoc t-tests were performed, showing that LSC only appeared in test order B, when – for each operation separately (see Figure 2b) – participants were tested first in the language-switching condition followed by the no language-switching condition (1492 ms vs. 1044 ms; $t(16) = 5.94$, $p < .001$, $d = 1.44$). No significant LSC emerged for the other sequence of testing, i.e., when testing started with the presentation of problems in the no language-switching condition followed by the language-switching condition (1360 ms vs. 1350 ms; $t(16) = 0.14$, $p = .89$, $d = 0.04$). The comparisons between test order A and order B in the switching condition (1350 ms vs. 1492 ms; $t(30) = -0.428, p = .67, d = 0.03$) and in the no-switching condition (1360 ms vs. 1044 ms; $t(30) = 0.951$, $p = .35$, $d = 0.17$) did not reveal significant effects.

***Hypothesis 2: LSC emerge in all three task operations***

As expected there was no interaction between Arithmetic Task and Language Switching (for RT: $F(1.64, 47.65) = .48$, $p = .59$, $\eta_p^2 = .02$, for ACC: $F(1.31, 37.91) = .09$, $p = .83$, $\eta_p^2 < .01$), indicating that LSC appear in all three included operations (i.e., multiplication, subtraction and artificial problems).

***Hypothesis 3: More LSC when switching to the non-dominant language as compared to switching to the dominant language***

Inconsistent with our prediction, no interaction was found between Language of Training and Language Switching (for RT: $F(1,28) = 0.32$, $p = .86$, for ACC: $F(1,28) = 0.24$, $p = .63$) indicating that LSC do not significantly differ across training languages.

***Explorative analysis: Is L2 vocabulary knowledge related to LSC in general?***

Individual scores for vocabulary knowledge were correlated with the respective LSC for each arithmetic task separately and for the overall LSC (i.e., including all

https://doi.org/10.1017/S1366728917000530 Published online by Cambridge University Press

three arithmetic tasks). There was no correlation between vocabulary knowledge and overall LSC ($r(32) = -.08$, $p = .68$) nor between vocabulary knowledge and operation-specific LSC (for artificial problems: $r(32) = -.15$, $p = .42$; for multiplication problems: $r(32) = .20$, $p = .28$; for subtraction problems: $r(32) = -.14$, $p = .46$).

## Discussion

The main aim of the present study was to further investigate language-switching costs (LSC) in the domain of arithmetic. Therefore, thirty-two university students learned eighteen problems of three different arithmetic operations in German (L1) or English (L2) over four consecutive training days and were tested in both languages on the fifth day. We found significant LSC for RT but not for ACC. Results further revealed LSC for RT in all three task (i.e., multiplication, subtraction and artificial problems). However, LSC due to learning in the dominant language and retrieval in the non-dominant language did not differ from LSC due to learning in the non-dominant language and retrieval in the dominant language. Finally, there was no significant relation between vocabulary knowledge of L2 and LSC.

The present design provides an important extension of prior research. While previous studies on LSC in arithmetic learning used visual stimuli in the form of written number words (Spelke & Tsivkin, 2001, Grabner et al., 2012; Saalbach et al., 2013), the present study was the first to show that LSC appear when arithmetic problems are learned and tested auditorily. LSC using auditory stimuli is an important finding, since numerical information is commonly presented either auditorily or as digits rather than as words during instruction. Further, it was the first study to show LSC in a block-wise language switching design, compared to random switching of language and task within blocks (e.g., Grabner et al., 2012; Saalbach et al., 2013; see also Meuter & Allport, 1999; and Campbell, 2005, for studies on cued language switching). Especially, if we are interested in implications for bilingual educational programs, a closer look at testing formats is necessary.

The first hypothesis, expecting problems to be solved more slowly and less accurately when the language of instruction differs from the language of application, was partly confirmed. In contrast to previous studies, LSC were limited to RT. The absence of LSC for ACC might be explained by adaptations made in the present study design, which led to a ceiling effect in ACC (ranging between 90% and 98%). The preceding studies used verification tasks, which required participants to choose among two or more answers. In the present study, a production task was administered in which participants had to type in their answers after they indicated the completion of problem solving by keypress. In addition, due to having only one language and one specific arithmetic operation within each block during testing, participants did not have to switch the language or operation type from trial to trial (but block-wise), which was required in the previous studies. This lower level of cognitive load within each block may have facilitated problem-solving, resulting in comparably high solution rates for all three tasks, even in the language-switching condition.

Interestingly, another methodological change during the testing phase led to unexpected results regarding LSC. LSC were only found in test order B, when participants started with a block in the language-switching condition followed by a block in the no-switching condition (see figure 2b). No LSC emerged in the reversed order A. It could be speculated that these results are due to a differential overlay of language-switching and practice effects. Overall, we found that RT for earlier trials within each block were significantly longer than for later trials, indicating a typical practice effect over the test session (post-hoc analysis showed a training effect within each block (all $p_s < .001$, all $d_s > .98$). In test order A, the practice effect may have counteracted the LSC resulting in similar RT in the switching-blocks (blocks 2, 4, and 6) compared to the respective no-switching blocks (blocks 1, 3 and 5). In test order B, however, the practice effect may have even amplified LSC as already the first blocks of each operation required language switching. Thus, despite of the clear advantages of the block design in examining LSC (e.g., preventing item-wise switching) it may partly have resulted in a confounding of practice and language-switching effects. This post-hoc finding and the following interpretation on practice effects remains still vague. It may give us a first insight on possible interventions to lower the likelihood of LSC within a short period of time. We may then ask the question whether only one or two short training session in the untrained language can prevent LSC appearing. Future studies may directly compare different designs to shed more light on the question of the robustness of LSC and possible interventions.

Regarding our second hypothesis, predicting LSC to appear for all three tasks, we found LSC for RT not only for typical arithmetic problems (i.e., multiplication and subtraction problems; replicating findings from Grabner et al., 2012, and Saalbach et al., 2013), but also for atypical arithmetic problems (i.e., artificial problems). Notably, LSC for artificial problems did not differ from LSC for multiplication or subtraction problems. This finding suggests that LSC cannot be solely explained by additional magnitude processing as suggested by the fMRI findings in previous research (Grabner et al., 2012). To identify underlying mechanisms, studies on LSC might benefit from the use of strategy reports after each trial. It is well understood that individuals use different strategies when performing arithmetic problems (e.g., LeFevre, Sadesky

https://doi.org/10.1017/S1366728917000530 Published online by Cambridge University Press

& Bisanz, 1996; Campbell & Xue, 2001). Overlearned problems are commonly retrieved from memory as facts, while new or large problems are indicated as solveable by the use of procedural strategies. Different strategies have also been found to be accompanied by specific neural correlates in fMRI as well as EEG (e.g., Dehaene, Piazza, Pinel & Cohen, 2003; Jost, Beinhoff, Hennighausen & Rösler, 2004; Núñez-Peña, Cortiñas & Escera, 2006; De Smedt, Grabner & Studer, 2009; Grabner & De Smedt, 2012). Thus, future research should employ strategy reports to further study the cognitive mechanisms underlying LSC. Such reports could also indicate what length of training is sufficient for problems to be rote-learned.

According to our third hypothesis, we expected more LSC for the German training group than for the English training groups, as participants of the former group have to switch from their dominant language (i.e., German, L1) to their non-dominant language (i.e., English, L2). Results revealed that LSC did not depend on whether the training was carried out in the dominant language or the non-dominant language. This finding contrasts with one of the previous studies, showing more LSC when switching from the dominant to the non-dominant than vice versa (Saalbach et al., 2013), but is in line with the study by Grabner and colleagues (2012) on Italian–German bilinguals. Contradicting results concerning the directional effect may be attributed to the specific language combination used in the previous studies. So far, training studies on LSC used different language combinations (i.e., German and Italian; German and French; German and English). Importantly, the order of the Arabic digit notation for two-digit number words differs cross-linguistically (Campbell and Xue, 2001). German uses a unit-ten order (e.g., "24" = four-and-twenty), whereas Italian, French, and English (for numbers higher than twenty) use a ten-unit order (e.g., "24" = twenty-four). This difference in word structure has been shown to influence arithmetic performance (e.g., Ellis & Hennelly, 1980; Göbel, Moeller, Pixner, Kaufmann & Nuerk, 2014; van Rinsveld et al., 2015). French, however, adds a second interference by making use of a base-20 structure for numbers between 70 and 99, while the other languages have a clear base-10 structure. This additional interference might have led to a directional effect of LSC in Saalbach et al. (2013). Finally, in an exploratory analysis, no relationship between L2-vocabulary knowledge scores and LSC was found. The validity of these findings is limited in two ways: First, the LexTALE represent only an indication of language-proficiency and is not equal to the concept of language proficiency. Second, the present sample represents a rather homogeneous group with regard to L2 vocabulary knowledge. Thus, future research on LSC needs to assess language proficiency in a more comprehensive way within a group of bilingual speakers being also heterogeneous with respect to their L2 proficiency.

The present study provides both theoretical and practical implications. With regard to the former, our findings give further insights into the interplay of language and arithmetic knowledge acquisition. So far, different arithmetic operations were considered to rely differently on language-based processing. For example, previous research suggests that multiplication problems rely more strongly on a verbal coding than subtraction problems (e.g., Dehaene et al., 2004; Ischebeck et al., 2006; see introduction). The present study, however, does not reveal differences between these two operations with regard to LSC using auditory stimuli (see also Saalbach et al., 2013, using visual stimuli). Thus, we find no indication that auditorily presented multiplication problems rely more strongly on verbal coding than subtraction problems. Furthermore, finding no difference between LSC in the two arithmetic operations and LSC in the artificial task, requiring pure fact retrieval, suggests that arithmetic problems are stored as factual knowledge after an extended time of rote-learning. However, this assumption requires further and more direct examination, for example, by means of strategy reports or specific neuroscientific approaches.

Findings of the present study also provide implications for CLIL settings. A lot of content learned in school represents factual knowledge (e.g., rote-learning the multiplication table, remembering capital cities, historical dates, etc.). Given our findings that rote-learned information is applied more efficiently in the language of instruction, LSC may also occur in school settings when language of application differs from language of instruction. This effect may be particular relevant for learners performing a task in limited time, such as classroom exams or other assessments. However, we need to be cautious in drawing inferences from laboratory studies to real-life classrooms. Teaching at school does not normally contain such massive rote learning as in the paradigm of the present study. Furthermore, the content used in this study and in most previous studies on language switching costs is limited to the effects on factual knowledge, representing only a part of what is learned in school. Thus, future research needs to examine the effects of language switching across learning and testing on the acquisition of conceptual as well as procedural knowledge within more complex kinds of task. In other words, highly controlled experimental studies on LSC should be complemented by research in more authentic settings. One way would be the scientifically based evaluation of implemented CLIL programs. Although a large evaluation of a specific CLIL program is being carried out (the *Europe School Berlin* program; Möller, Hohenstein, Fleckenstein, Köller & Baumert, 2017), this does not include an examination of possible LSC yet.

https://doi.org/10.1017/S1366728917000530 Published online by Cambridge University Press

To conclude, the present study revealed that cognitive costs arise when the language of instruction is different from the language of knowledge retrieval and application in the domain of arithmetic. This finding adds new evidence that language affects the way knowledge is stored in memory. To widen the extent to which these assumptions can be generalized, future research on cognitive costs through switching languages across instruction and retrieval needs to target other kinds of knowledge and more complex task settings. Then, it may be possible and justified to draw important implications for the design of effective CLIL programs.

## References


Ashcraft, M. H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory, 4(5),* 527.

Ashcraft, M. H., & Stazyk, E. H. (1981). Menatal addition: A test of three verification models. *Memory & Cognition, 9(2),* 185–196.

Barber, S. J., Rajaram, S., & Aron, A. (2010). When two is too many: Collaborative encoding impairs memory. *Memory & Cognition, 38(3),* 255–264.

Benn, Y., Zheng, Y., Wilkinson, I. D., Siegal, M., & Varley, R. (2012). Language in calculation: A core mechanism? *Neuropsychologia, 50(1),* 1–10.

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition, 118(1),* 123–129.

Campbell, J. I., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General, 130(2),* 299–315.

Campbell, J. I. (2005). Asymmetrical language switching costs in Chinese–English bilinguals' number naming and simple arithmetic. *Bilingualism: Language and Cognition, 8(1),* 85–91.

Chen, Z. Y., Cowell, P. E., Varley, R., & Wang, Y. C. (2009). A cross-language study of verbal and visuospatial working memory span. *Journal of Clinical and Experimental Neuropsychology, 31(4),* 385–391.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex, 33(2),* 219–250.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20(3-6),* 487–506.

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current opinion in neurobiology, 14(2),* 218–224.

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., & Benke, T. (2003). Learning complex arithmetic—an fMRI study. *Cognitive Brain Research, 18(1),* 76–88.

De Smedt, B., Grabner, R. H., & Studer, B. (2009). Oscillatory EEG correlates of arithmetic strategy use in addition and subtraction. *Experimental brain research, 195(4),* 635–642.

Domahs, F., & Delazer, M. (2005). Some assumptions and facts about arithmetic facts. *Psychology Science, 47(1),* 96–111.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology, 71(1),* 43–51.

Eurydice. (2006). *Content and language integrated learning (CLIL) at school in Europe*. Brussels, Belgium: Eurydice European Unit.

Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic bulletin & review, 18(1),* 150–157.

Frenck-Mestre, C., & Vaid, J. (1993). Activation of number facts in bilinguals. *Memory and Cognition 21*, 809–818.

Fuson, K. C., & Kwon, Y. (1992). Learning addition and subtraction: Effects of number words and other cultural tools. In Bideaud, J., Meljac, C., & Fischer, J. P. (Eds.). (2013). *Pathways to number: Children's developing numerical abilities*. Psychology Press.

Gentner, D., & Goldin-Meadow, S. (2003). Whither whorf. *Language in mind: Advances in the study of language and cognition*, 3–14.

Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. C. (2014). Language affects symbolic arithmetic in children: the case of number word inversion. *Journal of experimental child psychology, 119*, 17–25.

Grabner, R. H., & De Smedt, B. (2012). Oscillatory EEG correlates of arithmetic strategies: a training study. *Frontiers in psychology, 3*.

Grabner, R. H., Saalbach, H., & Eckstein, D. (2012). Language-Switching Costs in Bilingual Mathematics Learning. *Mind, Brain, and Education, 6(3),* 147–155.

Gumperz, J. J., & Levinson, S. C. (1996). Introduction to part I. In Gumperz, J. J. (1996). *Rethinking Linguistic Relativity*, pp. 21–36. UK: Cambridge University Press.

Ischebeck, A., Zamarian, L., Siedentopf, C., Koppelstätter, F., Benke, T., Felber, S., & Delazer, M. (2006). How specifically do we learn? Imaging the learning of multiplication and subtraction. *Neuroimage, 30(4),* 1365–1375.

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.

Jost, K., Beinhoff, U., Hennighausen, E., & Rösler, F. (2004). Facts, rules, and strategies in single-digit multiplication: evidence from event-related brain potentials. *Cognitive Brain Research, 20(2),* 183–193.

Klessinger, N., Szczerbinski, M., & Varley, R. (2012). The role of number words: the phonological length effect in multidigit addition. *Memory & cognition, 40(8),* 1289–1302.

Lasagabaster, D., & Sierra, J. M. (2009). Immersion and CLIL in English: more differences than similarities. *ELT journal*, *64(4)*, 367–375.

Lee, K. M. (2000). Cortical areas differentially involved in multiplication and subtraction: a functional magnetic resonance imaging study and correlation with a case of selective acalculia. *Annals of neurology, 48(4),* 657–661.


https://doi.org/10.1017/S1366728917000530 Published online by Cambridge University Press

LeFevre, J. A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(1),* 216.

Lemer, C., Dehaene, S., Spelke, E., & Cohen, L. (2003). Approximate quantities and exact number words: Dissociable systems. *Neuropsychologia, 41(14),* 1942–1958.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44(2),* 325–343.

Linguatec (2015). Retrieved from http://www.linguatec.de/en/text-to-speech/voice-reader-studio-15/.

Malt, B., & Wolff, P. (Eds.). (2010). *Words and the mind: How words capture human experience*. New York: Oxford University Press.

Marian, V., & Fausey, C. M. (2006). Language-dependent memory in bilingual learning. *Applied Cognitive Psychology, 20(8),* 1025–1047.

Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General, 129(3),* 361.

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of memory and language, 40(1),* 25–40.

Miller, K. F., Smith, C. M., Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science, 6(1),* 56–60.

Mochida, K., & Harrington, M. (2006). The yes/no test as a measure of receptive vocabulary knowledge. *Language Testing, 2*, 73–98.

Möller, J., Hohenstein, F., Fleckenstein, J., Köller, O., & Baumert, J. (Eds.). (2017). Erfolgreich integrieren-die Staatliche Europa-Schule Berlin. Waxmann Verlag.

Núñez-Peña, M. I., Cortiñas, M., & Escera, C. (2006). Problem size effect and processing strategies in mental arithmetic. *Neuroreport, 17(4),* 357–360.

Park, M. (1999). Linguistic influence on numerical development. *The Mathematics Educator, 10(1)*.

Quick Placement Test. (2001). Oxford: Oxford University Press.

Saalbach, H., & Imai, M. (2007). Scope of linguistic influence: Does a classifier system alter object concepts? *Journal of Experimental Psychology: General, 136(3),* 485.

Saalbach, H., Eckstein, D., Andri, N., Hobi, R., & Grabner, R. H. (2013). When language of instruction and language of application differ: Cognitive costs of bilingual mathematics learning. *Learning and Instruction, 26*, 36–44.

Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). E-Prime v1. 1. *Pittsburgh, PA: Psychology Software Tools Inc.*

Spelke, E. S., & Tsivkin, S. (2001). Language and number: a bilingual training study. *Cognition, 78(1)*.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review, 80(5),* 352.

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.*

Venkatraman, V., Siong, S. C., Chee, M. W., & Ansari, D. (2006). Effect of language switching on arithmetic: A bilingual fMRI study. *Journal of Cognitive Neuroscience, 18(1),* 64–74.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. Wiley Interdisciplinary Reviews: *Cognitive Science, 2(3),* 253–265.

Zaunbauer, A. C. M., Bonerad, E. M., & Möller, J. (2005). Muttersprachliches Leseverständnis immersiv unterrichteter Kinder 1 Dieser Beitrag wurde von DH Rost akzeptiert. *Zeitschrift für Pädagogische Psychologie, 19(4),* 263–265.

Zaunbauer, A. C. M., & Möller, J. (2009). Schulleistungsentwicklung immersiv unterrichteter Grundschüler in den ersten zwei Schuljahren. *Psychologie in Erziehung und Unterricht, (1),* 30–45.

https://doi.org/10.1017/S1366728917000530 Published online by Cambridge University Press