

The role of number words: the phonological length effect in multidigit addition

Nicolai Klessinger · Marcin Szczerbinski · Rosemary Varley

Published online: 30 June 2012
© Psychonomic Society, Inc. 2012

Abstract We explored the role of phonological representations of number words in exact calculation. The reaction times and accuracy of responses in multidigit addition problems were compared across three groups of participants (young healthy, older healthy, and 3 patients with severe aphasia) and two types of addition problems: phonologically long in English (containing the bisyllabic number word “seven”) and short in English (monosyllabic number words—e.g., “six”). Older healthy participants were significantly faster and more accurate in calculation than younger healthy participants. The older participants showed no evidence of a phonological length effect. However this effect was apparent in the younger adults, with longer reaction times on phonologically long problems. Furthermore, there was an association between the presence of a phonological length effect and the overall speed of response, suggesting that less proficient calculators were more reliant on phonological mediation of performance. The aphasic participants retained the ability to complete multidigit additions and were as accurate as the younger healthy group, although the response times of two of the 3 patients were slow. The aphasic participants varied with regard to the presence of a phonological length effect. Two participants showed no evidence of phonological mediation, while 1 displayed a phonological length effect. The results suggest that

language resources are not mandatory for exact addition, although they may be used to scaffold math performance in less competent calculators. Evidence of phonological mediation of performance in aphasic participants may provide insight into the integrity or otherwise of inner speech in severe aphasia.

Keywords Calculation · Number · Language · Aphasia · Phonological length · Inner speech

Introduction

The role of language in other high-order cognitive functions has been the topic of considerable debate. In the case of language and calculation, it has been suggested that mathematics, as a more recent acquisition and a culture-specific capability, borrows formal properties from the phylogenetically older and culture-universal capacity of language (Chomsky, 1988; Hurford, 1987). Both language and mathematics share the formal properties of an abstract symbol set and sensitivity to structural properties of expressions. The importance of structure in language is evident in the different role of nouns in the subject and object positions in sentences (e.g., *the lion killed the man/the man killed the lion*), while in mathematics, the noncommutative operations of subtraction and division (e.g., $7 - 2/2 - 7$) and bracketed expressions display the importance of structural principles. In addition to these general common characteristics between language and mathematics, there are also proposals that language representations constitute the underlying code in which exact calculations are performed (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Spelke & Tsivkin, 2001).

Numbers can be represented in a variety of surface formats, such as Arabic numerals and written and spoken number words. While some models propose that computations take place in an amodal semantic code (McCloskey, 1992), others suggest that calculation takes place in surface number forms,

N. Klessinger · R. Varley
Department of Human Communication Sciences,
University of Sheffield,
Sheffield, UK

M. Szczerbinski
School of Applied Psychology, University College Cork,
Cork, Ireland

R. Varley (✉)
Division of Psychology & Language Sciences,
University College London,
Chandler House, 2 Wakefield Street,
London WC1N 1PF, UK
e-mail: rosemary.varley@ucl.ac.uk

such as Arabic numerals or words. The symbol set that is used is likely to be determined by the nature of the calculation and the input/output format of the problem (Campbell, 1994). For example, single-digit addition problems (e.g., $2 + 4$) or information from rote-learned multiplication tables (4×8) might rely largely on linguistic forms (Ashcraft, 1992), while multi-digit calculations or mathematical operations, such as subtraction, that do not involve the same degree of rote learning might involve mixed representational formats, such as linguistic forms combined with visuospatial representations (Dehaene, 1992; Dehaene & Cohen, 1995). More complex calculations place demands on additional cognitive mechanisms, such as working memory and sequential planning and control. For example, multidigit arithmetic requires working memory capacity to store the intermediate products of calculation in order to support procedures such as borrowing and carrying (Adams & Hitch, 1998).

However, there are claims that language is essential for exact but not estimated calculation. Spelke and Tsivkin (2001) trained Russian–English bilinguals on exact or approximate addition problems in one of their languages. In exact calculation, there was a transfer cost if problems were trained in one language and then tested in the second language. By contrast, approximation showed no advantage for the language of training. In a functional imaging study, Dehaene et al. (1999) went on to show that approximate calculation recruited both right and left parietal zones, while exact arithmetic elicited activation near to the left-hemisphere language networks (Cohen, Dehaene, Chochon, Lehericy, & Naccache, 2000; Gruber, Indefrey, Steinmetz, & Kleinschmidt, 2001; Stanescu-Cosson et al., 2000; Venkatraman, Siong, Chee, & Ansari, 2006). Anthropological studies have reported restrictions of numerical cognition in individuals whose languages have small number vocabularies (Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004). Pica et al. explored the capacity for exact and approximate calculation in an Amazonian tribe who speak Mundurukú—a language that has number words only for 1 to 5. The Mundurukú performed in a similar manner to monolingual speakers of French in approximate calculation, but the performance of the groups differed on exact calculation. Furthermore, evidence for a link between language and calculation can be found in neuropsychological studies showing an association between aphasia and acalculia (Delazer, Girelli, Semenza, & Denes, 1999).

Other investigations, however, have not provided evidence for language mediation of exact calculation. Brysbaert, Fias, and Noel (1998) reported language mediation in Dutch or French speakers when spoken responses were required, but this effect disappeared when problems were presented in Arabic numerals and participants indicated their answer through typing digits. The results suggested a flexible processing network, where resources suitable to the specific task are recruited, rather than mandatory linguistic representation.

The results of some functional imaging studies have not revealed activation of left perisylvian language zones in exact calculation (Benn, Zheng, Siegal, Wilkinson, & Varley, 2012; Houde & Tzourio-Mazoyer, 2003; Pesenti, Thioux, Seron, & De Volder, 2000; Zago et al., 2001; Zago & Tzourio-Mazoyer, 2002). For example, Zago et al. reported activation of brain regions implicated in visuospatial working memory and imagery during complex calculation. Similarly, a study involving a comparison of indigenous Australian children who were monolingual speakers of languages with limited number vocabularies and English-speaking children revealed no differences in performance on a range of numeric tasks (Butterworth, Reeve, Reynolds, & Lloyd, 2008). Finally, a number of neuropsychological studies have reported dissociations between language and calculation, with relatively preserved mathematical performance in the presence of aphasia (Baldo & Dronkers, 2007; Basso, Burgio, & Caporali, 2000; Cappelletti, Butterworth, & Kopelman, 2001; Crutch & Warrington, 2002; Klessinger, Szczerbinski, & Varley, 2007; Rossor, Warrington, & Cipolotti, 1995; Varley, Klessinger, Romanowski, & Siegal, 2005). These findings indicate that there may be considerable functional independence between language and calculation and that linguistic representations are not the mandatory code of exact calculation.

In order to explore the nature of the representations used in exact calculation, we examined whether there was evidence of phonological mediation in mental arithmetic by manipulating the phonological length of numerals (Baddeley, 1997, 2003). The phonological component of Baddeley's working memory model consists of two subsystems: a passive phonological store capable of holding speech-based information and a rehearsal process that maintains this information. Use of a phonological code is indicated if performance is modulated by phonological properties of the processed stimuli, such as phonological similarity or length. A demand to maintain longer words or words of a similar sound structure is likely to have a detrimental effect on performance (Baddeley, Thomson, & Buchanan, 1975).

A number of studies investigating the involvement of phonological working memory in arithmetic have used a dual-task paradigm, in which the primary cognitive task of interest (e.g., multidigit addition) is combined with a secondary articulatory task. If both tasks require the resources of the phonological loop, performance on the primary and/or secondary task should deteriorate. These investigations have revealed variable involvement of phonological resources in calculation. Phonological coding is evident where problems are presented in a horizontal rather than a vertical format (Trbovich & LeFevre, 2003) or where elements of the problem are presented only briefly and have to be maintained while calculation routines are completed (Fürst & Hitch, 2000). By contrast, Noel, Desert, Aubrun, and Seron (2001)

reported a negative effect of phonological similarity on multidigit addition. However, their experiment required verbal responses from participants and maintenance of briefly presented operands. When problems are continuously visible during experimental trials, there is less evidence for phonological mediation (Seitz & Schumann-Hengsteler, 2002; Trbovich & LeFevre, 2003).

In summary, phonological codes may be involved in mental arithmetic, but their use depends on factors such as presentation format, duration of presentation, or involvement of more complex procedures, such as carrying and borrowing. These findings suggest that linguistic-phonological resources scaffold calculation when there are particular performance demands, but the variability of phonological mediation provides little support for the claim that it is mandatory in exact calculation.

With regard to word length effects, languages differ with regard to the phonological structure of number words. For example, Mandarin Chinese has number words shorter than those of English, while English number words are shorter than those of Welsh. This results in cross-language differences in performance on digit span tasks and also in mental arithmetic (Chen, Cowell, Varley, & Wang, 2009; Naveh-Benjamin & Ayres, 1986). Presentation of calculation problems in Welsh to English–Welsh bilinguals resulted in slower calculation times and more errors than did problems in English (Ellis, 1992; Ellis & Hennelly, 1980). Similarly, Chinese and Japanese speakers are faster calculators than are speakers of English (Lau & Hoosain, 1999). The causes of these differences are contentious, due to cultural and experiential confounds inherent in cross-linguistic studies. However, number words differ in length not only across languages, but also within languages. Within-language variability offers a more controlled opportunity to test the potential effect of phonological length on exact calculation. It is therefore surprising that, to our knowledge, only one study has explored this. Lemer, Leybaert, and Content (2001) examined exact addition in French speakers, manipulating the phonological length of addends. They detected a phonological length effect. However, because of presentation format (one addend at a time), it is not clear whether the effect represented the necessary involvement of phonological representations in calculation or a more peripheral involvement of the phonological loop in maintaining the addends until the whole problem was presented and the calculation completed.

In the present study, we examined the influence of number word length on multidigit addition (e.g., $15 + 16$) within a language (English). The English number system contains a single bisyllabic number (seven), which combines with other morphological markers (seventeen, seventy) to produce longer forms than do the monosyllabic numbers (e.g., eight, eighteen, eighty). Thus, calculation problems involving seven and its derivatives (seventeen, seventy) are phonologically longer than others. If phonological forms are mandatory for

calculation, phonologically long problems would result in longer calculation times than would phonologically short problems. Problems were presented in Arabic numerals, and participants responded by a keypress in order to eliminate linguistic input/output factors from performance. Problems remained on a screen until participants responded, in order to eliminate the need for short-term maintenance.

Participants were drawn from three groups: younger and older healthy adults and older adults with severe aphasia who retained calculation ability. The younger healthy controls were university students and were included to allow comparison of the results with those of other studies that typically recruit participants from this population. The older controls were roughly age- and education-matched to the aphasic participants and provided an appropriate normative comparison for the patient group. The inclusion of two groups of healthy participants also allowed some evaluation of the effects of different educational and cultural practices with regard to calculation. The older participants were educated in school systems that required more practice of mental arithmetic and a greater emphasis on rote learning of mathematical information. The calculation ability of the three aphasic participants was reported in Varley et al. (2005). Despite severe aphasia and marked syntactic impairment, each patient retained calculation ability across all four operations and also displayed insight into the structural-syntactic properties of mathematical expressions (e.g., non-commutative operations of subtraction and division and the hierarchical structure of bracket expressions). In the present investigation, we explored whether these patients showed phonological mediation of calculation. This exploration may also shed light on the integrity or otherwise of inner speech in individuals with severe aphasia. Early research on this question (Goodglass, Denes, & Calderon, 1974) suggested that inner speech impairment was present in aphasia. However, few studies have systematically explored the influence of phonological-linguistic factors such as phonological length on the performance of aphasic individuals on nonlanguage cognitive tasks.

Method

Participants

Three groups of participants were recruited to the study: 32 healthy younger adults, all of whom were university students (19 females and 13 males; age range, 18–39 years; median, 24.5), 25 healthy older adults (all male; age range, 47–75; median, 56), and 3 men with severe aphasia but preserved calculation (S.A., S.O., and P.R.; age range, 56–59). The 3 patients had previously participated in a study of calculation, and details of their calculation abilities and neurological status

can be found in Varley et al. (2005). All 3 had severely impaired syntactic ability, with comprehension of spoken and written reversible sentences at chance level. Similarly, grammatical output was markedly impaired, with output (spoken and written) restricted to single words (usually nouns or social forms; e.g., “bye”). Comprehension of words was less impaired, although all showed some degree of difficulty in understanding abstract words. Spoken naming was profoundly disrupted, with performance at or near floor level on a picture-naming test. One patient (S.A.) showed some residual capacity in written naming. Phonological digit span was tested in a recognition paradigm due to the patients’ severe difficulties in word retrieval and production. S.A. had a span of three items, S.O. five items, and P.R. four items. [Appendix 1](#) provides details on test scores for each patient. With regard to their ability to understand and use phonological number words, all 3 patients showed some residual comprehension of spoken number words between zero and twenty. Spoken number comprehension was tested by selecting the corresponding number of counters in response to a spoken word (S.A., 76 % correct; S.O., 90 % correct; P.R., 95 % correct) and spoken number production through spoken labeling of the same quantities (S.A., 80 % correct; S.O., 62 % correct; P.R., 95 % correct). They displayed variable capacity to transcode between phonological words and Arabic numerals (S.A., 71 % correct; S.O., 100 % correct; PR, 100 % correct). The healthy older adults had varied educational attainment. Twenty had a university degree, and 5 were in formal education until 15 years old. Of the aphasic calculators, S.O. was a retired university professor and had advanced premorbid competence in mathematics, while S.A. and P.R. both left formal education at age 15. Of the younger university student group, 19 had formal mathematics education up to 16 years (12 females, 7 males), and 13 (7 females, 6 males) to 18 years or above. All participants were native speakers of English and were educated in English. The healthy participants had no known history of neurological impairment, and all had normal or corrected-to-normal vision. Participants gave informed consent to participation in the study, and the study was approved by the local NHS Research Ethics Committee (NS200291449).

Materials

Three sets of multidigit addition (one digit plus two digits or two digits plus two digits) problems were created. Each set contained 30 items and was designated as short_1, short_2, or long in terms of number of syllables required to name the numerals. Problem size was matched across sets (e.g., 45 + 18 [short_1, five syllables], 48 + 15 [short_2, five syllables], 47 + 16 [long, six syllables]). Because 7 is the only single digit numeral with a two-syllable number word in English, problems in the long condition had at least one 7 in one of the addends. Two short conditions were included in order to

reduce the proportion of addition problems that included 7 in the overall set. Mean syllable length was 4.40 ($SD = 1.13$) for short_1, 4.73 ($SD = 0.98$) for short_2, and 5.83 ($SD = 1.09$) for long. ANOVA and post hoc *t*-test pairwise comparisons, with Bonferroni adjustment for multiple comparisons, revealed an overall effect of syllable length, $F(2, 87) = 14.80$, $p < .001$, $\eta^2 = .254$,¹ with differences between the short_1 and long ($p < .001$) and the short_2 and long ($p < .001$) conditions, but not between the short_1 and short_2 conditions ($p = .69$). Mean number of phonemes was 12.20 ($SD = 3.03$) for short_1, 12.93 ($SD = 2.64$) for short_2, and 14.97 ($SD = 2.70$) for long. An ANOVA revealed an overall effect of phoneme length, $F(2, 87) = 7.89$, $p = .001$, $\eta^2 = .154$, and post hoc comparisons confirmed differences between the short and long conditions (short_1 vs. long, $p = .001$; short_2 vs. long, $p = .018$) but no difference between the short_1 and short_2 conditions ($p = .94$). Stimulus lists are presented in [Appendix 2](#).

The problems were presented in Arabic numerals and in a horizontal array. Below the problem, two potential answers were displayed, and participants indicated via a left/right buttonpress which was correct. Each of the 90 addition problems was presented twice. On one occasion, the false solution differed with regard to the unit, and on the second, with regard to the decade (unit vs. decade conditions). The difference between the wrong and correct results was always 2 (e.g., 31 vs. 33) in the unit condition and always 10 (e.g., 31 vs. 21) in the decade condition. The resulting 180 trials were counterbalanced with regard to the serial position of the larger addend and the position of the correct solution.

Procedure

Stimuli were presented on a laptop computer with a 14-in. monitor. Addends were separated by a plus-operator (+) and were followed by an equal-sign (=) and a question mark (?). The two alternative answers to the problem were presented in the lower left and right corners of the screen, together with two arrow heads to indicate the corresponding answer key. Both addends stayed on the screen until a response was made. Participants in the younger adult group responded via two arrow keys on the right of the keyboard. Because of sensory–motor impairments in the patients, patients and healthy older controls responded by pressing one of two buttons of a computer mouse. The task was modeled using different fingers for keypresses (index for the left button and middle for the right); however, participants were not corrected if they chose to use a single finger to make responses. Participants were given the following instruction: “A series of addition problems like this (pointing to example) will appear on the screen. Two possible answers will be given,

¹ η^2 = partial eta squared.

one correct, one false. Your task is to calculate the solution and press the corresponding left- or right-arrow key. It is important that you do this as quickly and accurately as possible.” Instructions were modified for the aphasic participants, with linguistic information supported by gestures and diagrams.

The 180 trials were presented in two blocks of 90 problems in random order. Each problem from the problem-size triplet was presented only once in each block. For healthy participants, the order of blocks was counterbalanced, while the patients received the two blocks in the same order. Each trial started with a central fixation mark (*). After 500 ms, the addition problem was presented together with the two potential answers. No feedback was provided on response accuracy. The intertrial interval was 1,000 ms. The presentation setup was designed and programmed in Visual Basic, and response time (RT) was recorded to the nearest millisecond.

Participants completed 30 practice trials prior to the first block. None of the practice trials included values used in the experimental trials. There was a 2-min break between practice and the first experimental block and a 3-min break between the two experimental blocks.

Statistical analysis

RT data were trimmed as follows. First, means and standard deviations (*SDs*) were calculated separately for each participant on each experimental condition. All RTs that were more than 3 *SDs* above their respective mean were then removed from the analysis. This led to the elimination of 121 out of 10,800 data points (1.1 %). Trimmed data were then averaged across items or participants, depending on the level of analysis.

Group RTs and calculation accuracy were analyzed at two levels (within participants and between items) for both older and younger healthy participants. In the within-participants analysis, RTs and errors were averaged over items for each condition (i.e., separately for each participant). In the between-items analysis, RTs and errors were averaged over participants for each condition (i.e., separately for each item). ANOVAs were separately conducted within participants (F_1) or between the 180 items (F_2). With regard to analysis of

patient results, RT data were analyzed on a case-by-case basis to investigate whether any effect of phonological length could be established for individuals.

Results

Healthy participants

Response times

Mean RTs and *SEs* for both experimental factors for the two groups of healthy participants (aggregated data from within-participants analysis) are presented in Table 1.

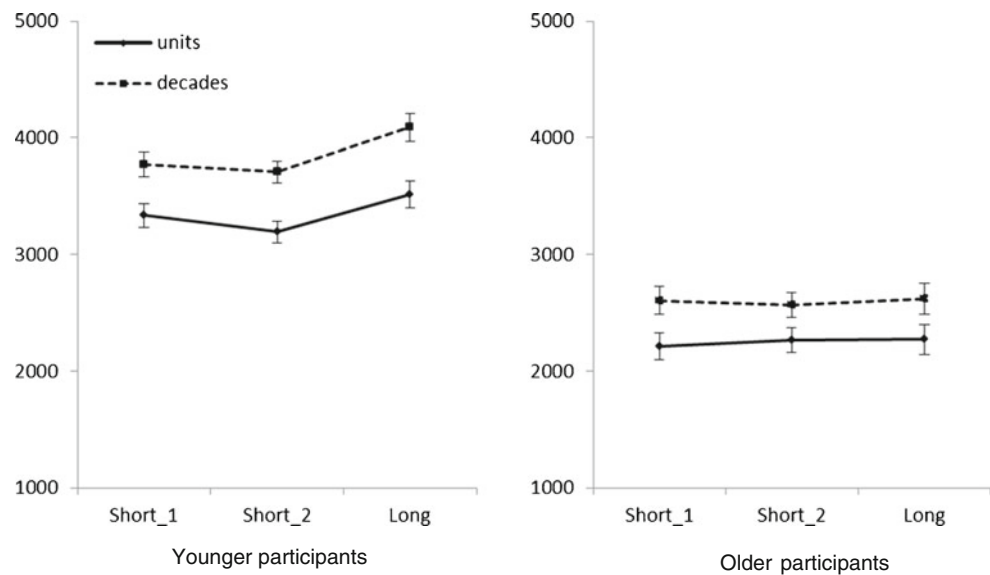
A mixed $3 \times 2 \times 2$ ANOVA was performed, with phonological length (Phono) and unit versus decade (UD) as within-participants factors and group (younger vs. older adults) as a between-participants factor. This analysis revealed a significant main effect of Phono, $F_1(2, 110) = 9.40, p < .001, \eta^2 = .146$, and UD, $F_1(1, 55) = 27.11, p < .001, \eta^2 = .330$, but also revealed a significant interaction between Phono and group, $F_1(2, 110) = 6.39, p = .002, \eta^2 = .104$, and an additional main effect of group, $F_1(1, 55) = 15.17, p < .001, \eta^2 = .216$. This indicated that the Phono effect was present only in the younger group and that older participants responded faster overall. No other effects were found, indicating that the effect of UD was present in both groups and irrespective of Phono [$UD \times group, F_1(1, 55) = 0.91, n.s., \eta^2 = .016$; $Phono \times UD, F_1(2, 110) = 0.35, n.s., \eta^2 = .006$; $Phono \times UD \times group, F_1(2, 110) = 1.08, n.s., \eta^2 = .019$]. The results are illustrated in Fig. 1. Because of the significant interaction between Phono and group, all further analyses were conducted separately for younger and older participants.

We conducted 3×2 (Phono \times UD) repeated measures ANOVAs separately within participants (F_1) and between the 180 items (F_2). For the 32 younger adults, the Phono effect was observed: RTs were longer for phonologically long than for short items [$Phono: F_1(2, 62) = 11.54, p < .001, \eta^2 = .271$; $F_2(2, 174) = 3.01, p = .052, \eta^2 = .033$], although the effect reached statistical significance only in the by-participants analysis). Additionally, RTs were longer

Table 1 Means (and standard errors) of response times (in milliseconds) by experimental conditions for younger and older healthy participants (within-participants analysis)

Group	Phonological length	Unit vs. Decade		Total
Younger adults <i>n</i> = 32	short_1	3,336 (206.2)	3,772 (215.1)	3,554 (201.1)
	short_2	3,193 (184.9)	3,707 (190.3)	3,450 (178.3)
	long	3,516 (232.7)	4,090 (239.1)	3,802 (227.8)
Older adults <i>n</i> = 25	short_1	2,210 (233.3)	2,607 (243.3)	2,409 (227.5)
	short_2	2,265 (209.2)	2,568 (215.3)	2,417 (201.8)
	long	2,271 (263.2)	2,622 (270.5)	2,446 (257.7)

Fig. 1 Mean response times (in milliseconds) across phonological length conditions for younger and older participants (within-participants analysis). Confidence bars represent standard errors



in the decade than in the unit condition [UD: $F_1(1, 31) = 15.71, p < .001, \eta^2 = .336$; $F_2(1, 174) = 17.72, p < .001, \eta^2 = .092$]. No significant interaction between the two factors was found [Phono \times UD: $F_1(2, 62) = 0.71, p = .49, \eta^2 = .022$; $F_2(2, 174) = 0.11, p = .90, \eta^2 = .001$]. Post hoc *t*-test pairwise comparisons with Bonferroni adjustment for multiple comparisons revealed that differences in Phono were significant between long and short_1 (mean difference \pm SE: 249.0 ± 70.1 ms; $p = .004$) and between long and short_2 (352.6 ± 92.4 ms; $p = .002$), but not between short_1 and short_2 (103.6 ± 60.2 ms; $p = .286$).

For the group of 25 older participants, only a main effect of the UD condition was found, $F_1(1, 24) = 15.73, p < .001, \eta^2 = .396$; $F_2(1, 174) = 20.69, p < .001, \eta^2 = .106$, indicating slower responses to problems in the decade condition. There was no main effect of Phono, $F_1(2, 48) = 0.56, p = .58, \eta^2 = .023$; $F_2(2, 174) = 0.08, p = .93, \eta^2 = .001$, and no interaction [Phono \times UD: $F_1(2, 48) = 1.41, p = .25, \eta^2 = .056$; $F_2(2, 174) = 0.11, p = .90, \eta^2 = .001$].

The stimulus lists provided in Appendix 2 reveal that there were five problems that did not require carry operations (e.g., $40 + 14$). All these problems were in the short_1 (7 % of items) and short_2 (10 % of items) lists. Data were reanalyzed with the triplets including a noncarry problem removed. The results replicated the analysis of the full lists, with shorter RTs in the older adult group, robust UD effects, and a significant effect of Phono in the young adult group, $F_1(2, 62) = 8.816, p < .001$, although the Phono effect failed to reach significance in the between-items analysis, $F_2(2, 144) = 1.896, p = .154$.

Analyses of differences linked to gender and level of mathematics education were performed for the younger adult group. A 2×2 (gender \times math education level, up to vs. higher than 16 years) ANOVA for RT revealed a main effect for gender, $F_1(1, 28) = 4.96, p = .034, \eta^2 = .150$, but no main effect for math education level, $F_1(1, 28) = 1.77, p = .195, \eta^2 =$

$.059$, and no interaction, $F_1(1, 28) = 2.18, p = .151, \eta^2 = .072$. In general, males responded faster than females. Possible interactions of gender or math education level with Phono were explored by two mixed design 2×2 (Phono \times gender and Phono \times math) ANOVAs, with a within-participants factor of Phono and a between-participants factor of gender or math education level. The results revealed no significant interactions of Phono with either gender or math education [Phono \times gender, $F(1, 30) = 0.48, p = .495, \eta^2 = .016$; Phono \times math, $F(1, 30) = 1.49, p = .232, \eta^2 = .047$].

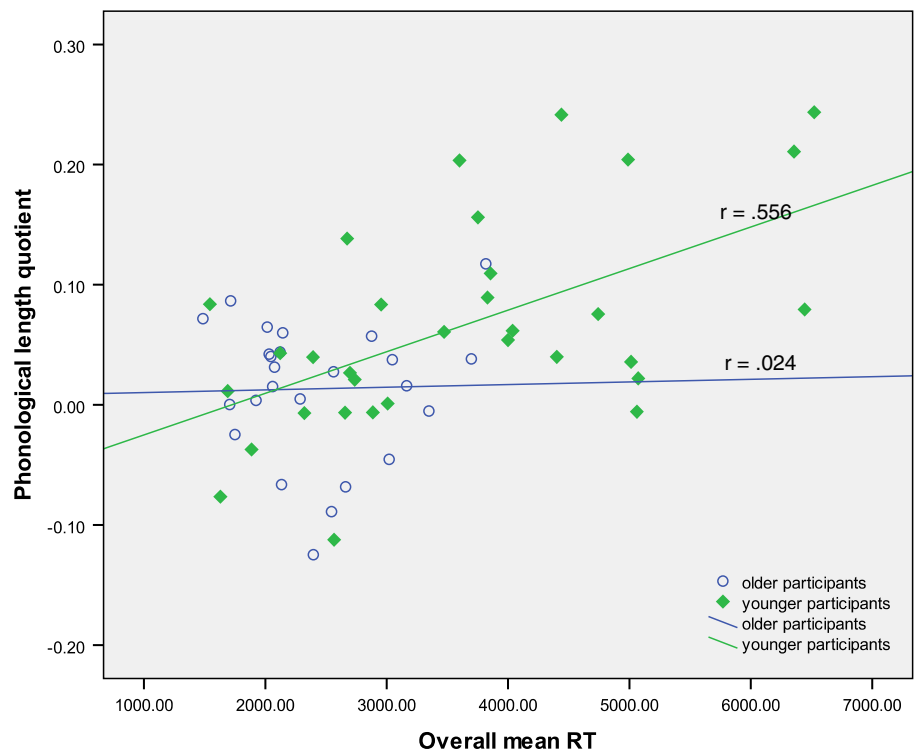
For the group of older participants, differences linked to level of mathematics education were explored with a mixed 2×2 (Phono \times math) ANOVA. It revealed no significant effect, indicating that the performance of participants with a lower math education level (up to age 15, $n = 11$) did not differ in from the ones with a higher math education level ($n = 14$) with regard to calculation speed or phonological length [Phono, $F(1, 23) = 0.95, p = .340, \eta^2 = .040$; math, $F(1, 23) = 0.0001, p = .994, \eta^2 = .000$; Phono \times math, $F(1, 23) = 1.04, p = .318, \eta^2 = .043$].

Finally, the relationship between individual calculation speed and the Phono effect was also examined. A quotient of the difference in RT for phonologically long minus short items and the overall RT [$Q_{RT} = (\text{Mean}_{RT\text{long}} - \text{Mean}_{RT\text{short}}) / \text{Mean}_{RT\text{overall}}$] was calculated (Fig. 2). For the younger participants, this quotient correlated significantly with overall calculation speed, $r = .556, p < .001$, indicating that the Phono effect increases with longer response times. For the older participants, no such association was found, $r = .024, p = .908$.

Accuracy

Accuracy was examined using the trimmed data set; that is, items with RTs of more than three standard deviations above the individual mean were removed from the analysis (see the

Fig. 2 The relationship between individual calculation speed and the magnitude of the phonological length effect



Method section). The results are presented in Table 2. Mean error rates for younger participants ranged from 1.1 % to 10.7 % ($M = 4.07\%$, $SE = 0.43\%$), and for older participants from 0 % to 7.8 % ($M = 2.20\%$, $SE = 0.40\%$). A mixed 3 (Phono) \times 2 (UD) \times 2 (group) ANOVA was performed. This analysis revealed a significant UD effect, $F_1(1, 55) = 27.65$, $p < .001$, $\eta^2 = .335$, and a group effect, $F_1(1, 55) = 9.70$, $p = .003$, $\eta^2 = .150$, but no main effect of Phono, $F_1(2, 110) = 0.45$, $p = .641$, $\eta^2 = 008$. No other effects were found. Thus, the older participants were more accurate overall, and the decade condition elicited higher rate of errors in both participant groups.

Given the significant between-group difference, separate 3 \times 2 (Phono \times UD) ANOVAs were conducted for younger and older groups, within-participants (F_1) and between the 180 items (F_2). The results confirmed the same pattern of behavior in the within-participants as well as the between-

items analyses for both groups. In the younger group, error rates were higher for problems in the decade, as compared with the unit, condition, $F_1(1, 31) = 18.6$, $p < .001$, $\eta^2 = .375$; $F_2(1, 174) = 23.79$, $p < .001$, $\eta^2 = .120$. There was no significant effect of Phono on accuracy, $F_1(2, 62) = 2.31$, $p = .11$, $\eta^2 = .069$; $F_2(2, 172) = 1.74$, $p = .18$, $\eta^2 = .020$, and no interactions between the two factors, $F_1(2, 62) = 1.97$, $p = .149$, $\eta^2 = .060$; $F_2(2, 174) = 1.44$, $p = .24$, $\eta^2 = .016$. For the older participants, error rates were also higher for problems in the decade, as compared with the unit, condition, $F_1(1, 24) = 11.2$, $p = .003$, $\eta^2 = .317$; $F_2(1, 174) = 17.10$, $p < .001$, $\eta^2 = .089$. There was no difference in error rates for phonologically long problems, as compared with short ones, $F_1(2, 48) = 0.85$, $p = .44$, $\eta^2 = .034$; $F_2(2, 172) = 0.57$, $p = .57$, $\eta^2 = .007$, and no interaction between the two factors, $F_1(2, 48) = 1.65$, $p = .20$, $\eta^2 = .064$; $F_2(2, 174) = 1.12$, $p = .33$, $\eta^2 = .013$.

A 2 \times 2 (gender \times math education level) ANOVA revealed no main effects for gender, $F(1, 28) = 1.45$, $p = .24$, $\eta^2 = .049$, or math education level, $F(1, 28) = 1.88$, $p = .18$, $\eta^2 = .063$, and no interaction, $F(1, 28) = 0.17$, $p = .68$, $\eta^2 = .006$, in the younger participant group. Similarly, there was no effect of level of math education in the older adult group (education to 15 years vs. education beyond 15 years; error rates: 2.9 % vs. 1.7 %, respectively), $t(23) = 1.50$, $p = .15$, $d = 0.603$.

Calculation speed (measured as mean RT for each participant) and calculation accuracy (error rate) were weakly related, $r = .238$, $p = .075$. Since the correlation was positive, it indicated no speed–accuracy trade-off. Calculation accuracy was not related to the size of the Phono effect in RTs, as indicated by a negligible correlation, $r = -.015$, n.s.,

Table 2 Means (and standard errors) of mean error rates (%) by experimental conditions for younger and older adults (within-participants analysis)

Group	Phonological length	Unit vs. Decade		Total
Younger adults $n = 32$	short_1	1.99 (0.4)	5.69 (0.9)	3.8 (0.5)
	short_2	1.78 (0.4)	5.28 (0.9)	3.5 (0.5)
	long	4.10 (0.8)	5.58 (0.8)	4.8 (0.6)
Older adults $n = 25$	short_1	1.50 (0.4)	2.95 (0.3)	2.2 (0.4)
	short_2	1.08 (0.3)	3.91 (0.9)	2.5 (0.6)
	long	1.22 (0.4)	2.55 (0.8)	1.9 (0.5)

between mean error rate and the Phono effect RT quotient introduced earlier $(M_{RT\ long} - M_{RT\ short})/M_{overall}$.

Aphasic calculators

Response times

RTs for the aphasic calculators (S.A., S.O., and P.R.) are presented in Table 3 and displayed in Fig. 3. Trimming of data (see the Method section) resulted in the removal of 2 data points (out of 180) for S.A. and 1 data point each for S.O. and P.R. S.A. and S.O. displayed longer calculation times than did all the healthy participants, while P.R.'s calculation speed was comparable to that of the younger healthy adults.

RT data were analyzed on an individual basis, and 3×2 (Phono \times UD) repeated measurement ANOVAs were performed. The results revealed a main effect of Phono for S.A., $F_{S.A.}(2, 27) = 11.45, p < .001, \eta^2 = .298$, who responded more slowly to phonologically long addition problems, but not for S.O. or P.R., $F_{S.O.}(2, 28) = 1.59, p = .213, \eta^2 = .054$; $F_{P.R.}(2, 28) = 0.81, p = .451, \eta^2 = .028$. Bonferroni-adjusted post hoc analysis revealed that S.A.'s calculation times were increased in the long condition, in comparison with both the short_1 and short_2 conditions ($p < .05$). There was no difference between the latter two ($p = .151$). In addition, S.A. responded quicker to items in the unit condition than to those in the decade condition, $F_{S.A.}(1, 27) = 13.06, p = .001, \eta^2 = .326$. No such effect was found in S.O. or P.R., $F_{S.O.}(1, 28) = 0.26, p = .618, \eta^2 = .009$; $F_{P.R.}(1, 28) = 1.59, p = .218, \eta^2 = .054$, nor were any interactions, $F_{S.A.}(2, 54) = 1.446, p = .245, \eta^2 = .051$; $F_{S.O.}(2, 56) = 0.53, p = .591, \eta^2 = .019$; $F_{P.R.}(2, 56) = 2.20, p = .120, \eta^2 = .073$.

Accuracy

Error rates for all 3 aphasic calculators were low. After elimination of accuracy data for RT outliers (4 out of 540 data points; see above), S.A. produced 6 errors out of 178 (3.37 %), and S.O.

and P.R. gave incorrect answers to 6 out of 179 items (3.35 %). The mean error rate of 3.36 % was higher than that of the age-matched control group (2.20 %; range, 0 %–7.8 %) but lower than that of the younger healthy group (4.07 %; range, 1.1 %–10.7 %). With regard to Phono and UD, S.A. showed a scatter of errors across all conditions, while S.O. and P.R. both produced more errors on phonologically short than on long items.

Discussion

The results provide evidence of phonological mediation of exact calculation only in the younger adults and in one of the aphasic calculators (S.A.). Consequently, there is little evidence that phonological codes are mandatory for exact calculation. Instead, the results suggest that calculation routines may be very different across participants and are influenced by educational and cohort experience (Noel, 2001; Noel & Seron, 1992). The older participants were faster and more accurate at mental arithmetic than the younger cohort, and it is likely that the educational experiences of the older cohort involved more practice of mental calculation and less reliance on external tools, such as calculators. As a result, the older adults were more expert at mental calculation and perhaps utilized either visuospatial forms or amodal semantic representations in mediating performance. The association of slower calculation with increased phonological mediation suggests that the resources of the language system are utilized by less proficient calculators. In this way, language can be seen as a resource that scaffolds and supports other high-order cognitive functions. Through encoding elements of a problem into phonological form, they can be maintained and manipulated in phonological working memory, allowing time for less automatized calculation processes to be performed.

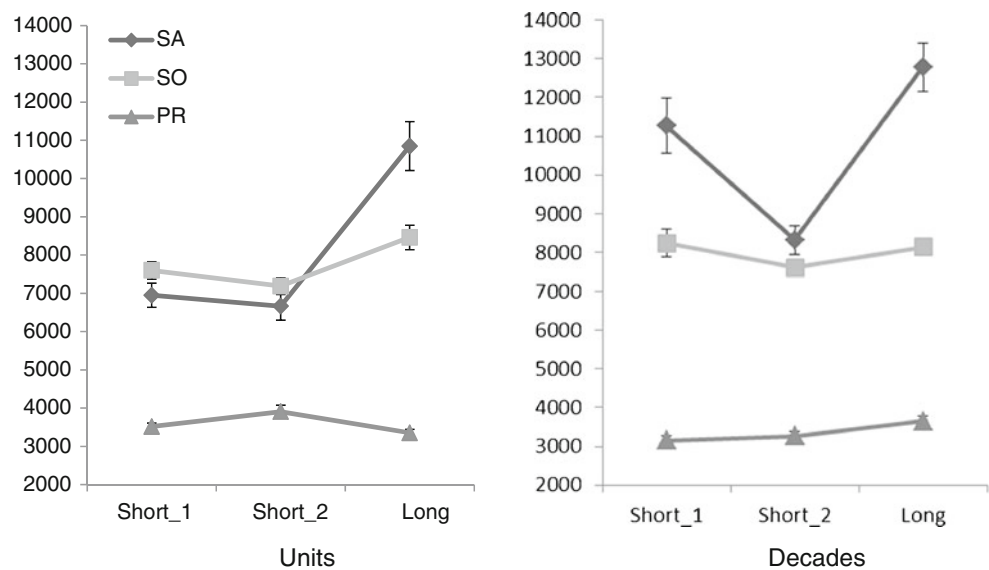
The argument that language can be employed to scaffold cognitive performances can be used to account for the findings of the bilingual training study of Spelke and Tsivkin (2001). An effective strategy in a training task is to encode information

Table 3 Means (and standard errors) of response times (in milliseconds) by experimental conditions for the aphasic participants (case-by-case analysis)

Patient	Phonological length	Unit vs. Decade		Total
S.A. <i>n</i> = 28	short_1	6,950 (631.8)	11,271 (1421.0)	9,111 (891.8)
	short_2	6,665 (737.4)	8,316 (728.1)	7,490 (574.8)
	long	10,849 (1,274.4)	12,787 (1,249.6)	11,818 (1,034.9)
S.O. <i>n</i> = 29	short_1	7,599 (453.8)	8,237 (717.2)	7,918 (414.2)
	short_2	7,189 (433.2)	7,605 (472.3)	7,397 (318.7)
	long	8,462 (637.2)	8,136 (469.7)	8,299 (437.5)
P.R. <i>n</i> = 29	short_1	3,516 (187.3)	3,154 (221.2)	3,331 (155.8)
	short_2	3,907 (333.9)	3,270 (222.7)	3,588 (197.6)
	long	3,351 (190.1)	3,645 (275.8)	3,498 (190.7)

Note. The value for *n* is the number of complete triplets across conditions.

Fig. 3 Mean response times (in milliseconds) across experimental conditions for S.A., S.O., and P.R. Confidence bars represent standard errors



into language forms to enable storage in longer term memory. If this information is probed in a language different from the one in which it was rote-encoded, there is an inevitable translation cost. However, these costs are attributable to memory storage rather than a necessary dependence on language representations in exact calculation. A similar argument can be constructed to explain the results of anthropological studies that suggest that the ability to track quantities is dependent upon the availability of exact number words (Gordon, 2004; Pica et al., 2004). For example, in a task where a participant is required to track the number of items remaining in a container when objects are being individually removed, the easy availability of a number word series simplifies processing demands. Instead of maintaining and continually updating a visual representation of the number of remaining items, a participant who has access to number words can simply count backward. However, this does not necessarily indicate that language is a mandatory component of the processing complex. Instead, it suggests that language is a powerful cognitive resource that can be employed across a range of tasks.

The results from the aphasic participants confirm the findings of Varley et al. (2005) that all 3 patients were competent calculators. Their error rates were below those of younger healthy controls, although above those of healthy age-matched controls. All 3 were slower in calculation than age-matched controls; however, P.R. showed calculation speeds comparable to those of the younger healthy group. Slowed calculation processes may partly account for the RTs of S.A. and S.O., but other factors linked to their brain injury may also contribute. For example, both S.A. and S.O. used their pre-morbidly nonpreferred hand to respond, while all the other participants (including P.R.) used their preferred hand. In addition, severe brain injuries can result in a complex of attentional and sensory-perceptual deficits that can contribute to lengthened RTs (Godefroy, Lhullier, & Rousseaux, 1994; Milner, 1986).

The finding of a Phono effect in the performance of one of the aphasic calculators (S.A.) was surprising, given the extent of his aphasic impairment. S.A. was severely impaired in spoken naming of pictures (0/60). However, and by contrast to the two other aphasic participants, he displayed considerable residual capacity in written naming of pictures (24/60). The difference between his performance in the spoken and written modalities might, to some degree, be due to a severe speech and oral apraxia that may prevent phonetic output of retrievable phonological forms. Number words are high-frequency language forms (Dehaene & Mehler, 1992) and of much higher frequency than the word forms typically tested in spoken and written object-naming tests. S.A. may, therefore, be able to access number words, and his mathematical performance may be mediated by a spoken number word lexicon. His capacity to use inner speech to support cognitive processing might be more limited where lower frequency word forms are required. There was no evidence of phonological mediation in either S.O. or P.R.

The evidence of consistent UD effects across participant groups in terms of shorter RTs and lower error rates in the unit condition provides some insight into the calculation processes adopted in the answer verification task used here. One proposal might be that participants calculated the result of the unit addition and then checked the value of the units in the response choices. If one of the units did not match the calculated result, processing could be terminated, and an answer selected. In instances where the value of both units was correct, processing had to continue to the decade value, with resulting increase in calculation times and increased possibility of an error. This account suggests right-to-left processing of the digits in the paradigm used here. An alternative account would be that unit problems are highly overlearned and solutions are retrieved rather than calculated. By contrast, solutions to the decade problems must be

calculated, either in right-to-left sequence or in parallel, which slows down processing. However, if this were the case, one might predict stronger Phono effects within the unit condition, due to dependence on verbally encoded rote knowledge.

This study was conducted in English, which contains only one bisyllabic number word. Given this single exemplar, the apparent Phono effects observed in some participants might be attributable not to the bisyllabic structure of “seven,” but to other of its characteristics. For example, seven is an odd number and represents a relatively high value in the zero-to-nine number sequence, and it is also a number word of relatively low frequency. Using the British National Corpus data (Leech, Rayson, & Wilson, 2001), we established that the number words used in the long condition indeed tended to be less frequent on average ($M = 127.82$, median = 86, $SD = 148.60$), as compared with the words used in the short_1 condition ($M = 169.27$, median = 66, $SD = 291.00$) and short_2 condition ($M = 218.64$ median = 92, $SD = 375.69$) (frequency per million is reported). While those differences were not statistically significant [$F(2, 252) = 2.141$, $p = .120$, $\eta^2 = .017$, for raw frequency data; $F(2, 252) = 2.068$, $p = .129$, $\eta^2 = .016$, for log-transformed data], it remains possible that the apparent increases in calculation time on the phonologically long problems might be an artifact of combination of the properties of “seven,” including its frequency. Investigation of the Phono effect in a language other than English could rule out this possibility. For example, Hebrew names for digits all consist of two syllables, apart from the digit 6, which is labeled by the monosyllabic form “Shesh.” If the claim of phonological mediation of calculation performance in less competent calculators is upheld, the prediction in studies of Hebrew calculators would be that six-problems would show shorter calculation times than non-six-problems. An analogous Polish study might find shorter calculation times

for two-, three-, five-, and six-problems (with monosyllabic number words), as compared with one-, four-, seven-, eight-, and nine-problems (with two-syllable number words).

Conclusions

The findings indicate that phonological number representations are not mandatorily accessed during the processing of Arabic digits in exact calculation in healthy, arithmetically skillful adults and support the view that mathematical computations can be functionally independent of language (Gelman & Butterworth, 2005). The evidence of different patterns of performance across healthy participant groups suggests considerable flexibility in the resources recruited in calculation, and this flexibility might be linked to differences in educational experience and the degree of automatization of calculation routines. The results of the aphasic calculators confirm that calculation can be retained in the face of severe language impairment. However, silent lexical mediation of performance may still be possible in some individuals, despite severe overt speech production deficits. Our findings address the issue of calculation only in the mature cognitive system, and it remains possible that number words are necessary for the learning of number representations and calculation routines.

Acknowledgements We thank S.A., S.O., and P.R. for their willing participation in this experiment and Catherine Lemer for sight of an unpublished manuscript (Lemer, Leybaert, & Content, *The effect of phonological length on multi-digit mental addition*) that assisted us in the design of this study. The software program for the calculation task was designed by Dr. Mike Coleman from University College London.

Appendix 1

Table 4 Language test scores for patients S.A., S.O., and P.R

	Chance score	S.A.	S.O.	P.R.
ADA Spoken Word Picture Matching	16.5	60/66*	61/66*	60/66*
ADA Written Word Picture Matching	16.5	62/66*	57/66*	66/66*
ADA Spoken Synonym Matching	80	123/160*	145/160*	121/160*
ADA Written Synonym Matching	80	121/160*	112/160*	139/160*
PALPA 54 Spoken Picture Naming	–	0/60	0/60	1/60
PALPA 54 Written Picture Naming	–	24/60	0/60	2/60
Comprehension of Spoken Reversible Sentences	50	49/100	32/100	38/100
Comprehension of Written Reversible Sentences	50	42/100	43/100	49/100
PALPA 13 Digit Span (Recognition)	-	3 items	5 items	4 items

*Scores significantly above chance ($p < .01$ level)

Note. Tests were taken from the ADA Auditory Comprehension Battery (Franklin, Turner, & Ellis, 1992) and Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay, Lesser, & Coltheart, 1992) or were devised for the purposes of this study

Appendix 2

Table 5 Stimulus sets. Five problems in the short_1 and short_2 conditions did not require carry operations, and these are marked with *. Problems including the numeral 12 in the short_1 condition are coded as having one syllable (“twelve”); however, they could be

realized as “one–two.” This would increase the syllable length by one syllable, but they remain comparable in total syllable length to those in the short_2 condition

Experimental stimuli

Triplet	Addition problem	No. of syllables	No. of phonemes	Assigned phonological length condition
1	7 + 13	4	10	long
	8 + 12	2	7	short_1
	6 + 14	3	9	short_2
2	7 + 14	4	10	long
	9 + 12	2	8	short_1
	8 + 13	3	7	short_2
3	7 + 15	4	11	long
	8 + 14	3	7	short_1
	6 + 16	3	11	short_2
4	7 + 16	4	12	long
	8 + 15	3	8	short_1
	5 + 18	3	8	short_2
5	7 + 18	4	9	long
	6 + 19	3	10	short_1
	9 + 16	3	10	short_2
6	17 + 13	5	13	long
	18 + 12	3	9	short_1
	16 + 14	4	12	short_2
7	17 + 14	5	13	long
	19 + 12	3	11	short_1
	16 + 15	4	13	short_2
8	17 + 15	5	14	long
	18 + 14	4	9	short_1
	19 + 13	4	11	short_2
9	17 + 16	5	15	long
	18 + 15	4	10	short_1
	19 + 14	4	11	short_2
10	17 + 18	5	12	long
	16 + 19	4	13	short_1
	14 + 21	5	14	short_2 *
11	27 + 17	7	16	long
	28 + 16	5	15	short_1
	29 + 15	5	15	short_2
12	37 + 17	7	17	long
	40 + 14	4	9	short_1*
	38 + 16	5	13	short_2
13	47 + 17	7	17	long
	46 + 18	5	12	short_1
	44 + 20	5	12	short_2*
14	57 + 17	7	18	long
	56 + 18	5	13	short_1
	52 + 22	6	13	short_2*
15	67 + 17	7	19	long

Table 5 (continued)

Experimental stimuli				
Triplet	Addition problem	No. of syllables	No. of phonemes	Assigned phonological length condition
16	68 + 16	5	15	short_1
	69 + 15	5	15	short_2
	17 + 64	6	16	long
17	18 + 63	5	13	short_1
	16 + 65	5	16	short_2
	17 + 53	6	16	long
18	16 + 54	5	14	short_1
	15 + 55	5	14	short_2
	17 + 45	6	15	long
19	16 + 46	5	15	short_1
	19 + 43	5	13	short_2
	17 + 36	6	16	long
20	16 + 38	5	13	short_1
	18 + 35	5	11	short_2
	17 + 28	6	16	long
21	19 + 26	5	15	short_1
	1629	5	16	short_2
	27 + 18	6	15	long
22	29 + 16	5	16	short_1
	26 + 19	5	15	short_2
	37 + 13	6	14	long
23	38 + 12	4	11	short_1
	36 + 14	5	13	short_2
	47 + 16	6	16	long
24	45 + 18	5	11	short_1
	48 + 15	5	12	short_2
	57 + 14	6	15	long
25	59 + 12	4	13	short_1
	58 + 13	5	12	short_2
	67 + 15	6	17	long
26	68 + 14	5	13	short_1
	69 + 13	5	14	short_2
	27 + 68	7	19	long
27	26 + 69	6	18	short_1
	29 + 66	6	19	short_2
	37 + 54	7	16	long
28	40 + 51	5	12	short_1*
	38 + 53	6	14	short_2
	47 + 43	7	16	long
29	48 + 42	6	12	short_1
	49 + 41	6	14	short_2
	57 + 35	7	17	long
30	56 + 36	6	17	short_1
	58 + 34	6	13	short_2
	67 + 24	7	19	long
	68 + 23	6	17	short_1
	65 + 26	6	18	short_2

References

- Adams, J., & Hitch, G. J. (1998). Mental arithmetic and working memory. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 153–173). Hove, U.K.: Psychology Press.
- Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*, 75–106.
- Baddeley, A. (1997). *Human memory. Theory and practice*. Hove, U.K.: Psychology Press. Revised ed.
- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839.
- Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word-length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, *14*, 575–589.
- Baldo, J. V., & Dronkers, N. F. (2007). Neural correlates of arithmetic and language comprehension: A common substrate? *Neuropsychologia*, *45*, 229–235.
- Basso, A., Burgio, F., & Caporali, A. (2000). Acalculia, aphasia and spatial disorders in left and right brain-damaged patients. *Cortex*, *36*, 265–280.
- Benn, Y., Zheng, Y., Siegal, M., Wilkinson, I. D., & Varley, R. (2012). Language in calculation: A core mechanism? *Neuropsychologia*, *50*, 1–10.
- Brysbart, M., Fias, W., & Noel, M. P. (1998). The Whorfian hypothesis and numerical cognition: Is 'twenty-four' processed in the same way as 'four-and-twenty'? *Cognition*, *66*, 51–77.
- Butterworth, B., Reeve, R., Reynolds, F., & Lloyd, D. (2008). Numerical thought with and without words: Evidence from Indigenous Australian children. *Proceedings of the National Academy of Sciences of the USA*, *105*, 13179–13184.
- Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition*, *53*, 1–44.
- Cappelletti, M., Butterworth, B., & Kopelman, M. (2001). Spared numerical abilities in a case of semantic dementia. *Neuropsychologia*, *39*, 1224–1239.
- 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*, 385–391.
- Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures*. Cambridge, MA: MIT Press.
- Cohen, L., Dehaene, S., Chochon, F., Lehéricy, S., & Naccache, L. (2000). Language and calculation within the parietal lobe: A combined cognitive, anatomical and fMRI study. *Neuropsychologia*, *38*, 1426–1440.
- Crutch, S. J., & Warrington, E. K. (2002). Preserved calculation skills in a case of semantic dementia. *Cortex*, *38*, 389–399.
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*, 1–42.
- Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*, 83–120.
- Dehaene, S., & Mehler, J. (1992). Cross-linguistic regulations in the frequency of number words. *Cognition*, *43*, 1–29.
- Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970–974.
- Delazer, M., Girelli, L., Semenza, C., & Denes, G. (1999). Numerical skills and aphasia. *Journal of the International Neuropsychological Society*, *5*, 213–221.
- Ellis, N. C. (1992). Linguistic relativity revisited: The bilingual word-length effect in working memory during counting, remembering numbers, and mental calculation. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 137–155). Amsterdam: North-Holland.
- Ellis, N. C., & Hennesly, 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*, 43–51.
- Franklin, S., Turner, J. E., & Ellis, A.W. (1992). *The ADA auditory comprehension battery*. ADA, University of York, York, U.K.
- Fürst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & Cognition*, *28*, 774–782.
- Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences*, *9*, 6–10.
- Godefroy, O., Lhullier, C., & Rousseaux, M. (1994). Reliability of reaction time measurements in brain-damaged patients. *Journal of the Neurological Sciences*, *126*, 168–171.
- Goodglass, H., Denes, G., & Calderon, M. (1974). The absence of covert verbal mediation in aphasia. *Cortex*, *10*, 264–269.
- Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*, 496–499.
- Gruber, O., Indefrey, P., Steinmetz, H., & Kleinschmidt, A. (2001). Dissociating neural correlates of cognitive components in mental calculation. *Cerebral Cortex*, *11*, 350–359.
- Houde, O., & Tzourio-Mazoyer, N. (2003). Neural foundations of logical and mathematical cognition. *Nature Reviews Neuroscience*, *4*, 507–514.
- Hurford, J. R. (1987). *Language and number*. Oxford: Blackwell.
- Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic assessments of language processing in Aphasia*. Hove, UK: Psychology Press.
- Klessinger, N., Szczerbinski, M., & Varley, R. (2007). Algebra in a man with severe aphasia. *Neuropsychologia*, *45*, 1642–1648.
- Lau, C. W., & Hoosain, R. (1999). Working memory and language difference in sound duration: A comparison of mental arithmetic in Chinese, Japanese, and English. *Psychologia*, *42*, 139–144.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken english*. Harlow: Longman.
- Lemer, C., Leybaert, J., & Content, A. (2001). *The effect of phonological length on multi-digit mental addition*. Unpublished manuscript.
- McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*, 107–157.
- Milner, A. D. (1986). Chronometric analysis in neuropsychology. *Neuropsychologia*, *24*, 115–128.
- Naveh-Benjamin, M., & Ayres, T. J. (1986). Digit span, reading rate, and linguistic relativity. *The Quarterly Journal of Experimental Psychology*, *38*, 739–751.
- Noel, M.-P. (2001). Numerical cognition. In B. Rapp (Ed.), *The handbook of cognitive neuropsychology: What deficits reveal about the human mind*. Philadelphia: Psychology Press.
- Noel, M.-P., Desert, M., Aubrun, A., & Seron, X. (2001). Involvement of short-term memory in complex mental calculation. *Memory & Cognition*, *29*, 34–42.
- Noel, M.-P., & Seron, X. (1992). Notational constraints and number processing: Reappraisal of the Gonzalez and Kolars (1982) study. *The Quarterly Journal of Experimental Psychology*, *45A*, 451–478.
- Pesenti, M., Thioux, M., Seron, X., & De Volder, A. (2000). Neuro-anatomical substrates of Arabic number processing, numerical comparison, and simple addition: A PET study. *Journal of Cognitive Neuroscience*, *12*, 461–479.
- Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*, 499–503.
- Rossor, M. N., Warrington, E. K., & Cipolotti, L. (1995). The isolation of calculation skills. *Journal of Neurology*, *242*, 78–81.

- Seitz, K., & Schumann-Hengsteler, R. (2002). Phonological loop and central executive processes in mental addition and multiplication. *Psychologische Beiträge, 44*, 275–302.
- Spelke, E. S., & Tsivkin, S. (2001). Language and number: A bilingual training study. *Cognition, 78*, 45–88.
- Stanesco-Cosson, R., Pinel, P., van de Moortele, P.-F., Le Bihan, D., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia: A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain, 123*, 2240–2255.
- Trbovich, P. L., & LeFevre, J. A. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition, 31*, 738–745.
- Varley, R. A., Klessinger, N. J. C., Romanowski, C. A. J., & Siegal, M. (2005). Agrammatic but numerate. *Proceedings of the National Academy of Sciences, 102*, 3519–3524.
- Venkatraman, V., Siong, S. C., Chee, M. W. L., & Ansari, D. (2006). Effect of language switching on arithmetic: A bilingual fMRI study. *Journal of Cognitive Neuroscience, 18*, 64–74.
- Zago, L., Pesenti, M., Mellet, E., Crivello, F., Mazoyer, B., & Tzourio-Mazoyer, N. (2001). Neural correlates of simple and complex mental calculation. *NeuroImage, 13*, 314–327.
- Zago, L., & Tzourio-Mazoyer, N. (2002). Distinguishing visuospatial working memory and complex mental calculation areas within the parietal lobes. *Neuroscience Letters, 331*, 45–49.