

# The Effects of Mental Abacus Expertise on Working Memory, Mental Representations and Calculation Strategies Used for Two-Digit Hindu-Arabic Numbers

Steson Lo<sup>1</sup>, Sally Andrews<sup>1</sup>

[1] *School of Psychology, University of Sydney, Sydney, NSW, Australia.*

---

Journal of Numerical Cognition, 2022, Vol. 8(1), 89–122, <https://doi.org/10.5964/jnc.8073>

Received: 2020-07-09 • Accepted: 2021-10-28 • Published (VoR): 2022-03-31

Handling Editor: Mirjam Ebersbach, Universität Kassel, Kassel, Germany

Corresponding Author: Steson Lo, School of Psychology, University of Sydney, NSW 2006, Australia. E-mail: [stesonlo@gmail.com](mailto:stesonlo@gmail.com)

---

## Abstract

In Asia, some children are taught a calculation technique known as the ‘mental abacus’. Previous research indicated that mental abacus experts can perform extraordinary feats of mental arithmetic, but it disagrees as to whether the technique improves working memory. The present study extended and clarified these findings by contrasting performance from several numerical and working memory tasks across three groups of participants: Japanese mental abacus experts, abacus-naïve Australian undergraduates, and abacus-naïve Japanese undergraduates. It also investigated whether the mental representations and strategies used to process two-digit numbers differed across the three groups. First, the results showed that the Japanese mental abacus experts only performed better when the numerical and working memory tasks involved arithmetic problems, suggesting domain-specific transfer rather than domain-general improvements to numerical processing or working memory. Second, the results suggest that the Japanese mental abacus experts were less reliant on decomposed magnitude representations, and used a processing strategy that is less sensitive to the perceptual overlap between numbers. Finally, performance was less discrepant between the Australian and Japanese abacus-naïve undergraduates than either group with the Japanese mental abacus experts, indicating that mental abacus training, rather than socio-cultural differences, was responsible for the observed group differences.

## Keywords

mental abacus, working memory, magnitude judgement, number bisection, mental addition, unit-decade compatibility, part-whole congruency

Hindu-Arabic numbers are ubiquitous in both professional settings (e.g., accountancy, epidemiology) and everyday contexts (e.g., calendars, monetary transactions). Children acquire this symbolic number system during their formative years (Gelman & Gallistel, 1978; Sarnecka, Goldman, & Slusser, 2015). However, learning is gradual (Piaget, 1952; Siegler, Thompson, & Opfer, 2009), and characterised by substantial variability over individuals (e.g., Dowker, 2005) and countries (e.g., Miura, Okamoto, Kim, Steere, & Fayol, 1993). In the Programme for International Student Assessment (PISA; OECD, 2016, 2019) and the Third International Mathematics and Science Study (TIMSS; Thomson, Wernert, O’Grady, & Rodrigues, 2016), East Asian countries like Singapore, China, and Japan have consistently occupied a top 10 rank. Some children from those countries can add, subtract, multiply and divide sequences of very large numbers with extraordinary speed and accuracy (Frank & Barner, 2012), because they were taught (Stigler, 1984), and acquired expertise (Hatano, Miyake, & Binks, 1977) in a calculation technique known as the ‘*mental abacus*’ (Kojima, 1954). Yet, the effect of mental abacus expertise on cognition in other areas has sparked continuing debate, because some researchers have reported transfer to other domains (e.g., academic performance: Stigler, Chalip, & Miller, 1986; Wang,



2020; inductive reasoning: Jia, Zhang, Yao, Chen, & Liang, 2021; visual attention: Cheng, Ma, Hu, & Zhou, 2021) while others have found that its effects are limited to arithmetic tasks (e.g., Amaiwa & Hatano, 1989; Barner et al., 2016).

Motivated by this debate, the present research compared Japanese mental abacus experts with groups of numerically competent, abacus-naïve adults on performance in both domain-general measures of working memory and domain-specific outcomes from three numerical tasks. The stimuli for those numerical tasks were designed to assess similarities and differences between the mental representation of two-digit Hindu-Arabic numbers and the calculation strategies used by the mental abacus experts and non-experts. Below, we explain why such differences in working memory and numerical processing were expected, following a brief outline of mental abacus training in Japan – the variable which distinguished the experts from all non-experts in the present research.

Mental abacus expertise usually begins with exposure to a physical abacus – a device containing beads attached to rods on a rectilinear frame (e.g., Kojima, 1954; Frank & Barner, 2012). A contemporary version of this device known as the Japanese *soroban* (Ifrac, 2001; Menninger, 2013) is often used by novices to represent numbers and perform arithmetic. Through extensive, teacher-guided practice, students learn to visualise the movement of beads on a mental version of the physical abacus (e.g., Hanakawa, Honda, Okada, Fukuyama, & Shibasaki, 2003). At advanced levels, experts prefer visualising the movement of beads when performing arithmetic over manipulations with the physical device.

Physical abacus exposure has remained a compulsory part of Japanese early maths education since 1926 (Shwalb, Sugie, & Yang, 2004). The curriculum currently defined by the Japanese Ministry of Education, Culture, Sports, Science and Technology specifies that students in grades three and four should “become aware of how numbers are represented on the abacus...and use it to calculate simple addition and subtraction” (Takahashi, Watanabe, & Yoshida, 2008, p. 7). Although the Japanese curriculum only stipulates rudimentary understanding of physical abacus calculations, many children attend extracurricular programs which offer mental abacus instruction. There, children normally spend one to two hours after school up to four times a week performing various abacus related activities under the supervision of an accredited abacus teacher. In 2003, approximately 7 million (Shwalb et al., 2004) of the 13.5 million students enrolled in primary and secondary schools (Statistics Bureau of Japan, 2015) studied in one of approximately 20,000 private abacus clubs (known as *juku*; Shwalb et al., 2004) distributed throughout Japan (Bellos, 2010). Approximately 6% of these students (Shwalb et al., 2004) also participated in accreditation exams administered by the League for Soroban Education in Japan for which they are ranked by mental abacus expertise (Hatta, Hirose, Ikeda, & Fukuhara, 1989). Experts also compete in various regional and national competitions for trophies, prizes, experience, or prestige.

## Domain-General Effects of Mental Abacus Expertise: Working Memory

Previous research suggested that mental abacus experts and non-experts use different subsystems of working memory (Baddeley, 1986, 2003) to perform mental arithmetic (e.g., Hatano et al., 1977; Frank & Barner, 2012). The mental abacus experts were argued to store interim sums by imagining how each would look on a physical abacus (e.g., Stigler, 1984). Supporting the view that visuo-spatial working memory is required to sustain these mental images, Hatano and Osawa (1983) found that digit span for the experts was significantly reduced when a visuo-spatial task was sandwiched between encoding and recall, relative to tasks involving simple factual questions (e.g., “who is the prime minister of Japan?”) or backward naming of three-syllable words. Because the non-experts showed the opposite pattern of more interference from the aural-verbal than visuo-spatial interference tasks, Hatano and Osawa concluded that digit memory was visuo-spatially mediated for the mental abacus experts, but verbally mediated for the non-experts. This difference in visuo-spatial working memory use between groups is consistent with research showing that mental abacus experts responded more quickly than non-experts in a word-to-picture matching task which involved mental image construction (Hatta & Miyazaki, 1989), and correctly recalled the position of more objects in a simple spatial span task (Lee, Lu, & Ko, 2007). However, Barner et al.’s (2016) extended investigation of mental abacus training found no evidence that it enhanced visuo-spatial working memory, because their longitudinal growth models revealed no difference in performance on a spatial matching task between learners and control participants over a period of three years.

Evidence about the relationship between mental abacus expertise and central executive processes is similarly equivocal. Using a 4-back task to assess executive function, Dong et al. (2016) found that novices who received 20

consecutive days of abacus instruction performed better than control participants. Also using a training design, Wang et al. (2015) observed greater task switching ability among mental abacus users. However, mental abacus users did not significantly differ from naïve participants when the central executive was assessed using a go/no-go task by Hatta and Miyazaki (1989) or Wang et al. (2015). Null effects were also found using tasks which involved mental updating (Hanakawa et al., 2003), mental rotation (Barner et al., 2016), and backward recall of digits (Lee et al., 2007) or letters (Hatano & Osawa, 1983) to assess the central executive.

These discrepant findings about the impact of mental abacus training on domain-general working memory processes might reflect two methodological limitations. First, studies comparing experts to control participants have often used small, elite samples – for example, Hanakawa et al. (2003) tested individuals who attained the highest mental abacus rank (10<sup>th</sup> *dan*), but only recruited 6 participants. Second, training studies which provide the other major source of evidence often use short to medium durations of mental abacus instruction that might not yield sufficient expertise (see also Barner et al., 2016). For example, Dong et al.'s (2016) experimental group included 18 participants, but they only received 90 minutes of instruction per day for 20 consecutive days – substantially less than any mental abacus expert. The present study sought to overcome these limitations in skill or statistical power by comparing large samples of mental abacus experts and naïve participants (at least 75 students per group) using a standardised working memory battery (Lewandowsky, Oberauer, Yang, & Ecker, 2010).

## Domain-Specific Effects of Mental Abacus Expertise: Mental Representations and Strategies for Two-Digit Numbers

Rather than strengthening domain-general working memory processes, differences between the mental abacus experts and non-experts might arise from variability in domain-specific processes, such as their mental representations and strategies for two-digit numbers. Conjectures about the former have either assumed a single magnitude representation based on the entire sequence of printed digits (i.e., the *holistic* theory: Dehaene, Dupoux, & Mehler, 1990; Dehaene, 1992) or separate magnitude representations based on the position of each Hindu-Arabic digit (i.e., the *decomposed* theory: Poltrock & Schwartz, 1984; Nuerk, Weger, & Willmes, 2001).

Previous research on populations of adults without abacus exposure suggests that both types of mental representation were constructed. Support for this view derives from the *magnitude judgement task* in which participants select the larger of two simultaneously presented numbers. Consistent with the decomposed theory, Nuerk et al. (2001) found that performance was faster and more accurate for unit-decade *compatible* stimuli (e.g., 42\_57) in which the larger number contains the bigger digit in both the decade and unit positions (viz.  $4 < 5$  and  $2 < 7$ ) than for unit-decade *incompatible* stimuli (e.g., 47\_62) in which the larger decade digit is in one number and the larger unit digit is in the other (viz.  $4 < 6$  but  $7 > 2$ ). These sets were equated by Nuerk et al. (2001) on absolute and logarithmic distance between each number pair (viz. 15 in the example above) to ensure that differences between the compatible and incompatible stimuli could not be due to holistic magnitude. Nevertheless, their data showed evidence of the overall distance and problem size effects (e.g., Moyer & Landauer, 1967) that index reliance on holistic mental representations, because performance was slower and less accurate for problems with numbers that were closer together (the overall distance effect) or larger (the problem size effect). Nuerk et al.'s (2001) findings therefore implied that both types of mental representation were used.

Whether mental representations differ between the mental abacus experts and non-experts has not, to our knowledge, been directly investigated. From the perspective of skill development, Tzelgov, Ganor-Stern, Kallai, and Pinhas (2015) hypothesised that, with enough practice, multi-digit numbers might eventually be represented holistically in long-term memory such that, for expert arithmeticians, “what appears to us as an ordinary series of digits assumes a singular meaning” (Dehaene, 2011, p. 132). This view is exemplified by the lightning calculator G. P. Bidder: “The number 763 is represented symbolically by three figures 7-6-3; but 763 is only one quantity, one number, one idea, and it presents itself to my mind just as the word ‘hippopotamus’ presents the idea of one animal” (Smith, 1983, pp. 55-56). Mental abacus experts typically spend several hours per week operating with multi-digit numbers as part of their extracurricular training (e.g., Stigler, 1984). This practice might have stimulated development of holistic magnitude representations. Consequently, the mental abacus experts should be insensitive to unit-decade compatibility and show larger overall distance and problem size effects.

Inspired by the unit-decade compatibility, overall distance and problem size effects, the present research tested for differences in the mental representations of two-digit numbers between the mental abacus experts and non-experts using a magnitude judgement task. Because the mental representations used for simple magnitude judgements might differ from those used to solve arithmetic problems, novel variants of the unit-decade compatibility effect, detailed below, were constructed for a *number bisection* and *mental addition* task. These cross-task comparisons allowed us to evaluate whether the mental representation differences between the experts and non-experts are specific to arithmetic tasks.

Another domain-specific process examined in the present research was the extent to which different strategies were used by the experts and non-experts for the mental addition task. Numeracy curricula in Australia, like those in the US (e.g., Fuson & Briars, 1990), encourage *decomposition algorithms* for multi-digit problems (also known as ‘1010’, ‘Separation’, ‘Tens & Units’ or ‘Full Decomposition’; Blöte, Klein, & Beishuizen, 2000; Nys & Content, 2010), because children are taught strategies which separately combine the unit and decade digits of the problem (e.g.,  $67 + 31 = 60 + 30$  and  $7 + 1$ ) before the interim sums are finally integrated (viz.  $90 + 08$ ). Other instructional methods, including the mental abacus (Stigler, 1984), encourage *sequential algorithms* for multi-digit problems (also known as ‘N10’, ‘Aggregation’, ‘Split and Jump’ or ‘Partial Decomposition’; Klein, Beishuizen, & Treffers, 1998; Lemaire & Callies, 2009). These are more consistent with holistic magnitude representations because their execution requires starting with the whole augend, before adding on values corresponding to the decade and unit digits of the addend (e.g.,  $67 + 31 = 67 + 30$  followed by  $97 + 01$ ). A subset of the items in the mental addition task, detailed below, were designed to evaluate whether the mental abacus experts and non-experts differed in their reliance on decomposition or sequential algorithms when solving arithmetic problems.

The final domain-specific process examined in the present research focussed on whether sensitivity to perceptual or semantic overlap between numbers for the mental addition task differed between the mental abacus experts and non-experts. As detailed below, this task required participants to decide whether two equations yielded the same total. Perceptual overlap referred to the number of shared digits between the two equations (e.g.,  $67+31$ \_61+37 shares all digits, but  $67+31$ \_56+42 only shares one), while semantic overlap referred to the magnitude of difference between the equations’ totals (e.g.,  $67+31$ \_85+12 differs by 1, but  $67+31$ \_85+42 differs by 29). If the mental abacus experts can rapidly compute a holistic representation of the sum for each equation, their mental addition performance should be insensitive to perceptual overlap but show a larger semantic overlap effect than the non-experts.

## Overview of Experiments

Motivated by the ongoing debate about the possible effects of mental abacus expertise, the present research compared young mental abacus experts with numerically competent, but abacus-naïve, adults on performance in domain-general measures of working memory and domain-specific tests of numerical processing. The mental abacus experts (hereafter *Abacus Experts*) were all Japanese and recruited from private abacus clubs in Tokyo. Their performance was compared against samples of Australian university students without mental abacus training (hereafter *Australian students*), for whom the numerical and working memory tasks were originally developed and tested. As detailed below, the young Abacus Experts produced equivalent or better task performance than the older and more educationally mature Australian students, consistent with an effect of mental abacus expertise. However, the Abacus Experts and Australian students also differed in socio-cultural background. Consequently, a third sample comprising Japanese university students without mental abacus training (hereafter *Japanese students*) was tested. Comparisons which differentiate the Abacus Experts from both the Australian and Japanese students would therefore suggest an effect of mental abacus expertise. Alternatively, comparisons which discriminate the Australian students from both the Japanese students and Abacus Experts would implicate socio-cultural factors.

The specific hypotheses for each experiment were as follows. Experiment 1 tested whether mental abacus training enhanced domain-general working memory by specifically augmenting visuo-spatial and central executive processes. The remaining experiments tested whether domain-specific mental representations and strategies for two-digit numbers differed between the mental abacus experts and non-experts. Such differences were inferred from performance in the magnitude judgement, number bisection, and mental addition tasks in Experiments 2, 3, and 4, respectively. The Abacus

Experts were expected to rely more on holistic than decomposed mental representations, show greater preference for a sequential than decomposition algorithm, and demonstrate increased sensitivity to semantic relative to perceptual overlap than both the Australian and Japanese students. Direct comparisons between the Australian and Japanese students tested whether this pattern was affected by socio-cultural factors.

## General Method

### Participants

#### Australian Students

Different samples of undergraduates from the University of Sydney, who each provided informed consent, were recruited for each numerical task: Ninety students (42 male; 10 left-handers; mean age = 19.1 years) participated in the magnitude judgement task, 102 students (39 male; 13 left-handers; mean age = 19.1 years) in the number bisection task, and 255 students (73 male<sup>1</sup>; 35 left-handers; mean age = 21.6 years) in the mental addition task. Participants in the number bisection task completed the entire working memory battery, but those in the mental addition task only completed the operation span and spatial short-term memory tasks due to constraints on available testing time. Students from the magnitude judgement and number bisection tasks participated in exchange for course credit, while students from the mental addition task participated during a class exercise on cognitive psychology. Based on the prevalence of instruction in Australia, it would be unusual for any to have received mental abacus training.<sup>2</sup>

#### Abacus Experts

Altogether, 79 experts participated after providing informed consent. Parental consent was also obtained for all minors from Soroban USA, a *juku* in Japan which specialises in after-school abacus instruction. Four participants were members of the Abacus Club at Waseda University, two were recruited from the broader student population at Waseda University, and the rest were students at Soroban USA.<sup>3</sup> All students from Soroban USA were Japanese, but only those formally ranked on mental abacus ability, or had received more than two years of training, were selected for participation.

The 73 experts recruited from the intermediate or advanced classes of Soroban USA (38 male; 67 right-handers; mean age = 11.33 years) averaged 5.05 years of extracurricular abacus experience ( $SD = 2.10$  years). Many (87.7%) also hold ranks from official accreditation exams: 15 are 10<sup>th</sup> *dan* ‘grand masters’, 30 with *dan* rankings from 1 to 9 are considered ‘experts’, and 19 with a *kyu* rank are ‘intermediate’ users (Hatta & Miyazaki, 1989; Hishitani, 1990). The remaining 9 participants did not report a formal rank, but 6 were awarded positions in at least one regional or national competition for abacus users. Participants recruited from Soroban USA were therefore highly skilled users of the mental abacus. The remaining 6 experts (2 male; 5 right-handers) recruited from Waseda University were also formally ranked Japanese individuals (range: 2<sup>nd</sup> to 10<sup>th</sup> *dan*), but they were older (mean age = 20.8 years) and had spent more time practising the mental abacus (average duration = 9.17 years,  $SD = 3.54$  years) than participants from Soroban USA. All Abacus Experts received a ¥1,000 book voucher for participation.

#### Japanese Students

Ninety-nine Japanese undergraduates participated after providing informed consent in exchange for a ¥1,000 book voucher: 29 from Rikkyo University (6 male; 27 right-handers; mean age = 20.10 years), and 70 from Waseda University (37 male; 66 right-handers; mean age = 20.73 years). None reported having previously received any extracurricular mental abacus lessons.

---

1) Demographic information was missing for 4 participants.

2) None of the Australian students were explicitly quizzed about mental abacus exposure, because they were tested before the first author initiated this research with the Abacus Experts. No training was instead inferred from a lack of public awareness or opportunity to engage with mental abacus lessons outside Asia 10-15 years ago when the Australian students were children.

3) Data collected from 10 other mental abacus users were excluded because three were soroban teachers, and seven reported less than a year of instruction at Soroban USA.

## Measures of Working Memory

The memory updating, operation span and spatial short-term memory tasks were selected from the working memory battery developed by Lewandowsky et al. (2010). The memory updating and operation span tasks measure central executive working memory, while the spatial memory task assesses visuo-spatial working memory (e.g., Ecker, Lewandowsky, Oberauer, & Chee, 2010). Details about the stimuli and scoring for each task can be found in Lewandowsky et al. (2010).

The two Japanese samples completed variants of the three working memory tasks. First, all instructions were translated from English to Japanese. Second, each letter in the operation span task was replaced with one of 46 hiragana characters used in the Japanese writing system, because formal exposure to the English alphabet only commences in Grade 4 for Japanese children (approximately 9 years of age). Participants responded using a standard US keyboard (see Figure 1) with labels reflecting the order presented in charts shown during early education (Mabuchi, 1993). None of the hiragana sequences spelled proper Japanese words, and different keys were labelled correct and incorrect for the interleaved maths verification problems used to assess processing in the operation span task. Following Lewandowsky et al. (2010), performance in the *operation span task* was defined by recall of the stored sequence of letters, while performance in the secondary *operation span processing task* was defined by maths verification accuracy.

**Figure 1**

*The Keyboard Used by the Abacus Experts and Japanese Students for the Operation Span Task*



*Note.* Hiragana labels were attached to each key using coloured stickers.

## Procedure

The DMDX display software (Forster & Forster, 2003) was used to control stimulus presentation and record reaction time (RT) and errors for the three numerical tasks. Numbers were presented in black monospace font (Consolas) on a white background with each digit spanning a height of 12 mm.

Stimulus presentation and response detection for all three working memory tasks was controlled using MATLAB 8.1.0.604 (R2013a) and Psychtoolbox Version 3.0.9 with code from Lewandowsky et al. (2010). For the Australian students, the memory updating, operation span and spatial memory tasks were delivered in that order following the number bisection or mental addition tasks. For the Abacus Experts, tasks were subdivided into sessions based on equipment and participant availability. The Japanese students experienced the following sequence of tasks: magnitude judgement, mental addition, memory updating, operation span, spatial memory, and number bisection.

## Model Specification for the Numerical Tasks

RT and accuracy from the magnitude judgement, number bisection and mental addition tasks were separately analysed using generalised linear mixed-effect models (GLMMs) based on version 1.1-21 from the lme4 package (Bates, Mächler,

Bolker, & Walker, 2015) in the R program for statistical computing (R Core Team, 2015). All analyses treated participants and stimuli as crossed random factors, and statistically controlled for the effects of trial-level predictors such as RT of the previous response and spatial position of the larger number (e.g., Andrews & Lo, 2012, 2013). These extraneous trial-level effects on performance in the magnitude judgement, number bisection, and mental addition tasks are summarised in Appendix A. By accounting for structural dependencies in the data and controlling for systematic relations with participant and item covariates (e.g., overall distance between numbers for each problem), GLMMs offer greater power to detect treatment effects on the mean for each condition than either analysis of variance or regression (e.g., Baayen, Davidson, & Bates, 2008; Bolker et al., 2009). In line with previous research, GLMMs of RT used an Inverse Gaussian distribution as a model of skewed waiting time with an identity link function (e.g., Lo & Andrews, 2015), while GLMMs of accuracy used a binomial distribution with a logit link function (e.g., Jaeger, 2008). Models estimated random slopes for all effects of interest as recommended by Barr, Levy, Scheepers, and Tily (2013), and all continuous predictors were standardised prior to analysis. Normalised sum contrasts (Venables & Ripley, 2002) were specified to separately compare the Abacus Experts with the Australian and Japanese students in one set of GLMMs, while another set separately compared the Australian students with the Abacus Experts and Japanese students. The degrees of freedom are currently unknown for the *t*-statistic in GLMMs (e.g., Kliegl, Masson, & Richter, 2010), but the asymptotic *p*-values reported below should be accurate for datasets with many hundreds or thousands of observations like ours (Baayen et al., 2008). The significance criterion for the group interaction contrasts was adjusted to  $.05/3 = .017$  using the Bonferroni procedure to correct for multiple comparisons.

## Experiment 1: Working Memory Tasks

The goal of Experiment 1 was to evaluate whether central executive and visuo-spatial working memory differed according to group using a larger sample of mental abacus experts than previous research (cf. Hanakawa et al., 2003; Dong et al., 2016). We also explored whether the Abacus Experts' performance depended more on visuo-spatial working memory by testing for group differences in the strength of relationship between the working memory measures.

### Method

#### Participants

Altogether, 102 Australian students, 79 Abacus Experts, and 99 Japanese students participated in all three working memory tasks. Another 255 Australian students participated in the operation span and spatial memory tasks, but not the memory updating task.

### Results

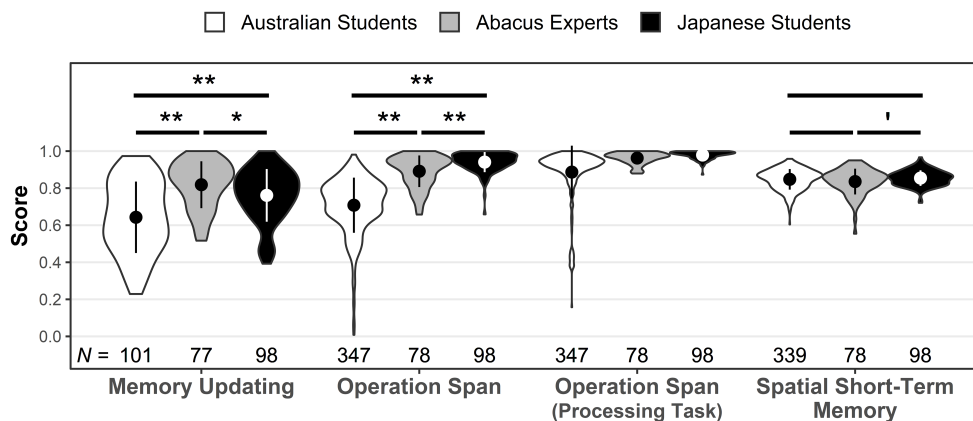
Nine Australian and one Japanese student scoring more than 4 *SD* below the mean on the operation span processing task were excluded from the analyses. Following Lewandowsky et al. (2010), one Abacus Expert and another Australian student was eliminated because they scored more than 3 *SD* below the mean on more than two tasks. Average performance for each task is presented in Figure 2 based on the remaining 347 Australian students, 78 Abacus Experts, and 98 Japanese students. Data were missing for eight Australian students on the spatial memory task, and one Abacus Expert on the memory updating task.

Within each task, performance was examined using independent samples *t*-tests. For the memory updating task, scores were significantly higher for the Abacus Experts than either the Australian,  $t(176) = -6.94$ ,  $p < .001$ , Cohen's  $d = -1.05$ , or Japanese students,  $t(173) = -2.78$ ,  $p = .01$ , Cohen's  $d = -0.42$ , and significantly lower for the Australian than Japanese students,  $t(197) = 4.92$ ,  $p < .001$ , Cohen's  $d = 0.70$ . For the operation span task, the Japanese students performed significantly better than the Abacus Experts,  $t(174) = 4.66$ ,  $p < .001$ , Cohen's  $d = 0.71$ , or the Australian students,  $t(443) = 15.18$ ,  $p < .001$ , Cohen's  $d = -1.74$ , while the Abacus Experts significantly outperformed the Australian students,  $t(423) = -10.49$ ,  $p < .001$ , Cohen's  $d = -1.31$ . For the spatial memory task, recall did not significantly differ between the Australian and Japanese students,  $t(435) = 1.09$ ,  $p = .28$ , Cohen's  $d = 0.12$ , the Australian students and Abacus Experts,  $t(415) = 1.64$ ,

$p = .10$ , Cohen's  $d = 0.21$ , or the Japanese students and Abacus Experts,  $t(174) = 2.20$ ,  $p = .03$ , Cohen's  $d = 0.33$ , using the Bonferroni-adjusted criterion of  $.017$ .<sup>4</sup>

Figure 2

Score for Each Working Memory Task According to Group



Note. Performance in the operation span task was defined by the proportion of letters recalled, while performance in the operation span (processing task) was defined by maths verification accuracy. Error bars represent 1 SD above and below each sample mean.  $^*p < .05$ .  $^*p < .017$ .  $^{**}p < .001$ .

The correlations between tasks for each group are presented in Table 1. All were significant except for the relationship between spatial memory and operation span among the Japanese students. Comparisons using Fishers' z-tests revealed significantly weaker correlations for the Japanese students than either the Australian students,  $|z| > 3.06$ , all  $p < .01$ , or Abacus Experts,  $|z| > 2.91$ , all  $p < .01$ . The correlation between memory updating and spatial memory was not significantly different between the Australian and Japanese students,  $z = -1.22$ ,  $p = .22$ , nor did any of the correlations between tasks differ significantly between the Australian students and Abacus Experts,  $|z| < 1.79$ , all  $p > .07$ .

Table 1

Correlations Between the Working Memory Tasks for Each Participant Group

Task	Memory Updating	Operation Span
<b>Australian Students</b>		
Operation Span	0.60**	
Spatial Short-Term Memory	0.42**	0.36**
<b>Abacus Experts</b>		
Operation Span	0.59**	
Spatial Short-Term Memory	0.62**	0.46**
<b>Japanese Students</b>		
Operation Span	0.22*	
Spatial Short-Term Memory	0.27*	0.02

$^*p < .05$ .  $^{**}p < .001$ .

4) The proportion of male and female participants was approximately equal for the Abacus Experts (51.3%) and Japanese students (42.9%), but there was roughly twice as many female participants among the Australian students (on average, 66.3%). Sex differences in working memory performance have previously been reported (e.g., Harness, Jacot, Scherf, White, & Warnick, 2008). However, the pattern of group differences for the male and female participants within each working memory task was the same as for the combined sample, and none of the comparisons involving group interacted significantly with sex (all  $p > .05$ ). The working memory results for the full sample therefore appear to hold for both male and female participants.



## Discussion

The benefits of mental abacus expertise were confined to the memory updating task which requires mental calculation with Hindu-Arabic numbers. Although single-digit equations were interleaved in the operation span task (e.g.,  $4 + 2 = 7$ , true or false?), verification could be achieved by checking the retrieved solution from long-term memory (e.g., Ashcraft & Stazyk, 1981). These task-specific factors complicate inferences about the effect of mental abacus expertise on central executive working memory. Crucially, the Abacus Experts did not perform better than either abacus-naïve group in the spatial memory task. The correlations between measures were also significantly weaker for the Japanese students, rather than stronger for the Abacus Experts than both non-expert groups. One interpretation for these results is that mental abacus expertise does *not* augment visuo-spatial working memory nor increase the extent to which this resource is used for other tasks (e.g., Barner et al., 2016). Alternatively, the lack of group differences might reflect the ceiling effect observed in the operation span and spatial memory tasks for both Japanese samples. Distinguishing between these accounts may require the operation span and spatial memory task difficulty to be adjusted in future experiments.

## Experiment 2: Magnitude Judgement Task

The overarching goal of the remaining experiments was to explore whether mental abacus experts and non-experts differed in both mental representations and strategies used for two-digit numbers. The magnitude judgement task of Experiment 2 tested whether the groups differed in reliance on holistic or decomposed magnitude representations for two-digit Hindu-Arabic numbers using the unit-decade compatibility, overall distance, and problem size effects introduced earlier.

## Method

### Participants

Altogether, 90 Australian students, 79 Abacus Experts, and 99 Japanese students from the previous experiment participated.

### Design and Stimuli

The critical between-decade stimuli in the magnitude judgement task comprised 240 pairs of two-digit numbers constructed by Nuerk et al. (2001). Half of these pairs were unit-decade compatible (e.g., 42\_57); the rest were unit-decade incompatible (e.g., 47\_62). Unit and decade distance between each pair of numbers was orthogonally manipulated over each of the unit-decade compatible and incompatible conditions (small: 1-3; large: 4-8). As described in Nuerk et al. (2001), overall distance between each pair of numbers, problem size and unit distance were all matched absolutely and logarithmically between the two conditions,  $|t(238)| < 1.07$ ,  $p > .29$ . Decade distance was significantly larger for unit-decade incompatible items,  $|t(238)| > 2.34$ ,  $p < .02$ , because each pair of numbers crossed an additional decade boundary to match overall distance with the unit-decade compatible items (Nuerk, Weger, & Willmes, 2004). Therefore, following Ganor-Stern, Pinhas, and Tzelgov (2009, Experiment 2), 60 filler trials with numbers containing the same decade digit (e.g., 32\_36) were intermixed to prevent adoption of task-specific strategies based on the decade digits of each pair.<sup>5</sup>

### Procedure

Each trial in the magnitude judgement task comprised an initial display of two fixation markers for 300 ms on the left and right sides of the screen horizontally separated by 10 cm of empty space that was immediately replaced by two numbers for 2000 ms or until response. Number pairs were subdivided into three lists with the larger number appearing

---

<sup>5</sup> The Australian students and Abacus Experts completed twice as many trials as the Japanese students, once with the larger number on the left and once with the larger number on the right, across three blocks varying in how each number was presented (cf. Ganor-Stern et al., 2009). This latter manipulation was not administered to the Japanese students, so only the results for simultaneously presented number pairs are reported here.

equally often on the left and right sides of the screen. Participants completed 18 practice trials followed by the 300 experimental items presented in an individually randomised order. Instructions at the beginning of the task specified that participants should press a key with their left or right hand matching whichever side of the screen contained the larger number.

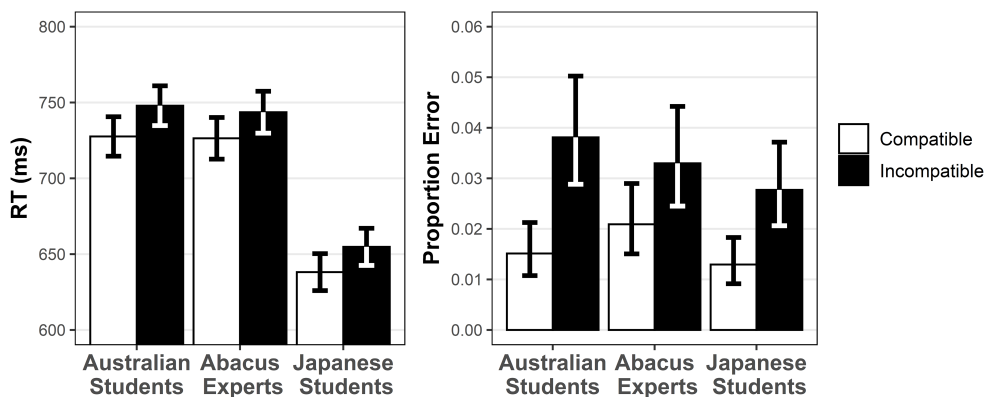
## Results

Data was missing for one Abacus Expert due to a computer recording error, and one Japanese student due to experimenter error. Two Australian students and one Abacus Expert with unusually high error rates exceeding 15% were also excluded. RTs faster than 100 ms or exceeding 2000 ms were excluded from the analyses for the remaining 88 Australian students, 77 Abacus Experts and 98 Japanese students (0.47%, 0.32%, and 0.01% of each group's total, respectively).

Figure 3 summarises the mean RTs and error proportions for each group according to unit-decade compatibility. Averaged over group, a unit-decade compatibility effect was observed, because incompatible trials yielded significantly longer RT,  $t = 5.72$ ,  $p < .001$ , RT estimate = 18 ms, and lower accuracy,  $z = 5.95$ ,  $p < .001$ , log-odds estimate = 0.73, than compatible trials. The GLMMs also revealed a significantly weaker unit-decade compatibility effect on both RT,  $t = 3.87$ ,  $p < .001$ , RT estimate = 16 ms, and accuracy,  $z = 2.59$ ,  $p = .01$ , log-odds estimate = 0.52, for problems with a smaller distance between the unit digits. They also showed that performance was affected by overall distance and problem size, because problems with a larger difference between numbers,  $t = -24.49$ ,  $p < .001$ , RT estimate = -41 ms;  $z = -15.66$ ,  $p < .001$ , log-odds estimate = -0.98, or problems with larger numbers,  $t = -3.76$ ,  $p < .001$ , RT estimate = -6 ms;  $z = 1.78$ ,  $p = .08$ , log-odds estimate = 0.09, yielded significantly faster RT or higher accuracy.

**Figure 3**

Mean RT and Error for Each Group and Level of Unit-Decade Compatibility in the Magnitude Judgement Task



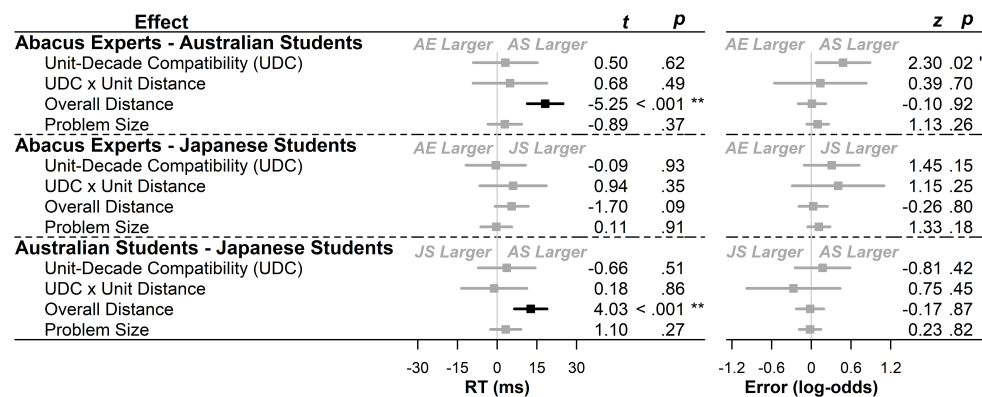
Note. Error bars represent 95% confidence intervals estimated from GLMMs based purely on fixed effect uncertainty.

Averaged over stimuli, the GLMMs revealed no significant difference in accuracy, but significantly faster RT for the Japanese students than either the Abacus Experts,  $t = -11.07$ ,  $p < .001$ , RT estimate = -89 ms;  $z = -2.11$ ,  $p = .04$ , log-odds estimate = -0.33, or Australian students,  $t = -11.85$ ,  $p < .001$ , RT estimate = 91 ms;  $z = 1.57$ ,  $p = .12$ , log-odds estimate = -0.24. Neither RT,  $t = 0.33$ ,  $p = .74$ , RT estimate = 3 ms, nor accuracy,  $z = -0.58$ ,  $p = .56$ , log-odds estimate = -0.09, differed significantly between the Abacus Experts and Australian students.

The group comparisons for each effect are presented in Figure 4, which summarises the descriptive estimates and test statistics from the GLMMs in the graphical and tabular components, respectively. Most of these group comparisons were not significant. The only exception was a significantly larger overall distance effect on RT for the Australian students than either the Abacus Experts or Japanese students. The unit-decade compatibility effect on accuracy was smaller for the Abacus Experts than Australian students, but this comparison is not significant using the Bonferroni-adjusted criterion of .017.

Figure 4

The GLMM Group Comparisons for Each Effect in the Magnitude Judgement Task



Note. Error bars represent 95% confidence intervals for each comparison. AS = Australian students; AE = Abacus Experts; JS = Japanese students.  $p < .05$ .  $*p < .017$ .  $**p < .001$ .

## Discussion

Data from the magnitude judgement task yielded three major findings. First, the unit-decade compatibility effect was successfully replicated, implicating the use of decomposed magnitude representations in all three samples. Second, performance was comparable between all three groups: the Japanese students' faster RT might be due to those participants having completed fewer trials, given the significant positive association between task length and RTs reported in Appendix A. Finally, the unit-decade compatibility effect on accuracy was marginally smaller for the Abacus Experts than Australian students. Although this result suggests greater reliance on holistic magnitude representations by the Abacus Experts, a larger overall distance effect was not observed as expected from Tzelgov et al.'s (2015) hypothesis. Instead, the overall distance effect was significantly larger for the Australian students than either Japanese sample. Limited and somewhat contradictory evidence was therefore obtained from the magnitude judgement task about the effect of mental abacus expertise on mental representations of two-digit numbers.

## Experiment 3: Number Bisection Task

Experiment 3 investigated whether clearer evidence of group differences would emerge in a number bisection task that required mental calculation to achieve accurate performance. Participants were shown three two-digit numbers and instructed to judge whether the medial number was the midpoint of the two outer numbers. Decomposed mental representations were tested using the outer numbers to define the unit-decade compatibility effect. Decomposition was also tested using a novel variant – the part-whole congruency effect – based on response consistency between the entire stimulus and subset of unit digits. For example, the stimulus 45\_67\_89 is part-whole *congruent*, because decisions based on the entire stimulus or unit digits both yield the same response (i.e., 67 correctly bisects the whole numbers 45 and 89, and 7 also correctly bisects the unit digits 5 and 9). Conversely, the stimulus 42\_58\_74 is part-whole *incongruent*, because 58 correctly bisects 42 and 74, but the unit digit 8 is not the midpoint of 2 and 4. Converging evidence for differences in reliance on holistic mental representations was again assessed using the overall distance and problem size effects. Two indices of overall distance were defined for this task: the difference between each pair of outer numbers (hereafter *Outer Numbers*), and the difference between each medial number and true arithmetic midpoint (hereafter *Medial to Midpoint*). If the Abacus Experts relied more on holistic magnitude representations, they should be less sensitive to unit-decade compatibility and part-whole congruency, and more sensitive to overall distance and problem size than either the Australian or Japanese students.

## Method

### Participants

Together, 102 Australian students, 79 Abacus Experts, and 99 Japanese students from Experiment 1 participated.

### Design and Stimuli

Sixty-four pairs of two-digit numbers were selected as the outer numbers for the number bisection task.<sup>6</sup> Half were unit-decade *compatible* (e.g., 42\_XX\_74), and the rest were unit-decade *incompatible* (e.g., 45\_XX\_71). Overall distance, problem size and unit distance between each pair of outer numbers were matched both absolutely and logarithmically between these two conditions (all  $p > .20$ ; see [Appendix B](#) for the means). Medial numbers were then chosen to correctly or incorrectly bisect the interval defined by each outer number pair (see [Appendix B](#) for example stimuli). For the correctly bisected problems, stimuli were part-whole *congruent* (e.g., 45\_67\_89) if the medial number's unit digit matched the midpoint of the outer number's unit digits (viz. 7 correctly bisects 5 and 9, and 67 correctly bisects 45 and 89). Stimuli were otherwise part-whole *incongruent* (e.g., 42\_58\_74) if the medial number's unit digit deviated from the midpoint of the unit digits from both outer numbers (viz. 8 incorrectly bisects 2 and 4, but 58 correctly bisects 42 and 74). For the incorrectly bisected problems, this criterion was inverted, such that stimuli were part-whole *incongruent* (e.g., 42\_53\_74) if the medial number's unit digit coincided with the midpoint of the outer numbers' unit digits (viz. 3 correctly bisects 2 and 4, but 53 incorrectly bisects 42 and 74). Stimuli were otherwise part-whole *congruent* (e.g., 45\_76\_89) if the medial number's unit digit deviated from the midpoint of the unit digits from both outer numbers (viz. 6 incorrectly bisects 5 and 9, and 76 incorrectly bisects 45 and 89).

The incorrectly bisected problems were further subdivided according to the discrepancy between the medial number and true midpoint. Half of the medial numbers were *close* to the arithmetic middle (distance of 1-3 for the congruent trials; distance of 5 for the incongruent trials; both containing the same decade digit as the midpoint), and half were *far* from the arithmetic middle (distance of 4-8 for the congruent trials; distance of 5 or 15 for the incongruent trials; both containing different decade digits from the midpoint). Approximately half of these medial numbers fell above or below the arithmetic midpoint. Heeding [Nuerk, Geppert, van Herten, and Willmes' \(2002\)](#) suggestions, all pairs of outer numbers were bisectable (i.e., parity of the outer numbers was homogenous), all four digits comprising each pair of outer numbers were unique, none of the outer numbers were decade numbers (e.g., 70), and none of the numbers in each problem fell onto the same row or column in the standard 12 x 12 multiplication tables.

### Procedure

The displays were identical to the magnitude judgement task except that three fixation markers were shown in the first display, an empty space of 3.5 cm separated each of the outer numbers from the medial number, and all three numbers were displayed for 10,000 ms or until response. Each stimulus was randomly sampled without replacement by the DMDX software from four lists which contained an equal number of correctly and incorrectly bisected trials. Like the magnitude judgement task, the largest number appeared equally often on the left and right sides of the screen; 18 practice trials preceded the critical items; and the order of problems was individually randomised for each participant without repetition. Before the practice trials, participants were instructed to press one key with their right hand if the medial number correctly bisected the outer numbers, and another key with their left hand if the outer numbers were incorrectly bisected by the centrally presented medial number. The critical trials were randomly separated into blocks which had a self-paced break imposed between them.

---

6) The Australian students and Abacus Experts were presented with an additional 128 pairs of outer numbers matched on stimulus characteristics with the critical items that were not presented to the Japanese students because of time constraints. These pairs were randomly distributed to yield four blocks of 48 trials for the Australian students and Abacus Experts, while the critical items were partitioned to yield two blocks of 32 trials for the Japanese students. The reduced subset of items produced very similar outcomes to the complete set, so only the results for these common stimuli are reported.

## Results

Data were missing for two Abacus Experts due to a computer recording error, and one Japanese student due to experimenter error. One Abacus Expert, two Australian, and three Japanese students who responded to fewer than 90% of problems were excluded from the analyses. Another four Abacus Experts, 25 Australian students, and 11 Japanese students with more than 30% incorrect responses were also excluded.<sup>7</sup> RTs below 200 ms or exceeding 10,000 ms were removed for the remaining 75 Australians, 72 Abacus Experts and 84 Japanese students (1.58%, 0.78% and 1.58% of each group's total, respectively).

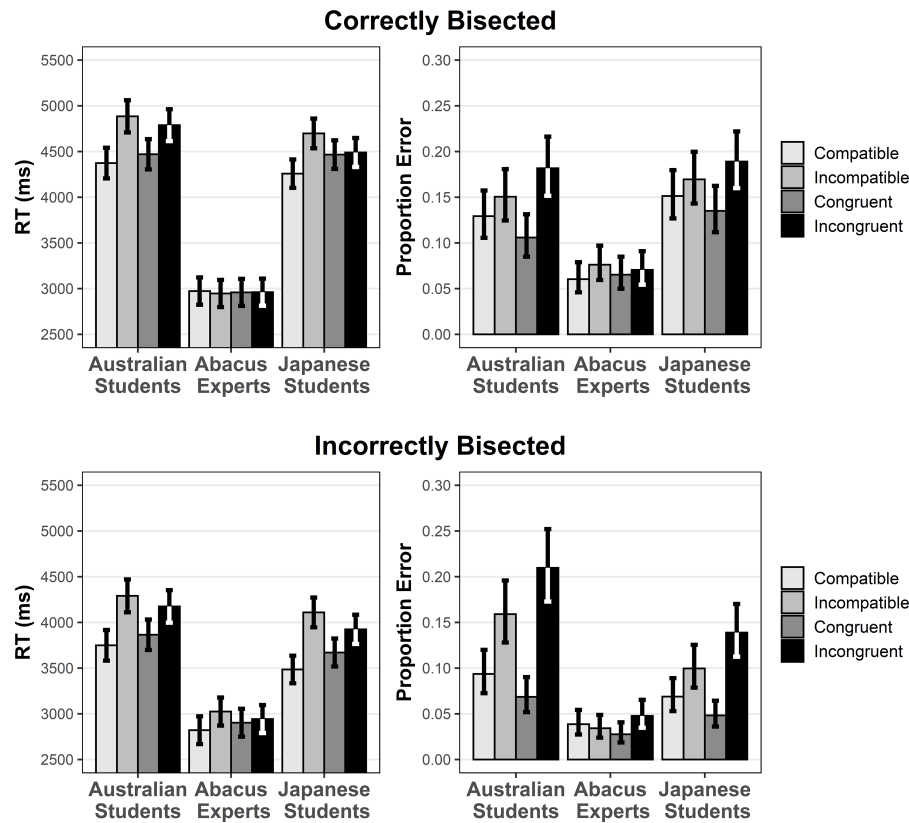
Figure 5 separately presents the mean RTs and error proportions for the correctly and incorrectly bisected problems according to group, unit-decade compatibility, and part-whole congruency. Averaged over group, performance was affected by unit-decade compatibility and part-whole congruency, because the GLMMs revealed significantly longer RT and lower accuracy for the incompatible than compatible stimuli (correctly bisected:  $t = 6.65$ ,  $p < .001$ , RT estimate = 309 ms;  $z = 2.03$ ,  $p = .04$ , log-odds estimate = 0.19; incorrectly bisected:  $t = 8.32$ ,  $p < .001$ , RT estimate = 456 ms;  $z = 2.43$ ,  $p = .02$ , log-odds estimate = 0.29), and the incongruent than congruent stimuli (correctly bisected:  $t = 2.59$ ,  $p = .01$ , RT estimate = 115 ms;  $z = 3.82$ ,  $p < .001$ , log-odds estimate = 0.37; incorrectly bisected:  $t = 3.72$ ,  $p < .001$ , RT estimate = 201 ms;  $z = 8.10$ ,  $p < .001$ , log-odds estimate = 1.00). As in the magnitude judgment task, this unit-decade compatibility effect was significantly weaker for problems with a smaller distance between the unit digits (correctly bisected:  $t = 4.13$ ,  $p < .001$ , RT estimate = 104 ms;  $z = 1.34$ ,  $p = .18$ , log-odds estimate = 0.08; incorrectly bisected:  $t = 2.86$ ,  $p = .01$ , RT estimate = 92 ms;  $z = 0.72$ ,  $p = .47$ , log-odds estimate = 0.06). Average performance also correlated with overall distance, because stimuli with a greater difference between the outer numbers (correctly bisected:  $t = 9.12$ ,  $p < .001$ , RT estimate = 217 ms;  $z = 1.46$ ,  $p = .15$ , log-odds estimate = 0.07; incorrectly bisected:  $t = 9.51$ ,  $p < .001$ , RT estimate = 285 ms;  $z = 2.55$ ,  $p = .01$ , log-odds estimate = 0.16), or had a small than large difference between the medial number and true midpoint ( $t = -6.62$ ,  $p < .001$ , RT estimate = -195 ms;  $z = -3.09$ ,  $p = .002$ , log-odds estimate = -0.29), were associated with significantly longer RT or lower accuracy. The problem size effect on average performance was less consistently observed, because stimuli with larger numbers yielded significantly longer RT for the correctly bisected ( $t = 4.68$ ,  $p < .001$ , RT estimate = 97 ms;  $z = 1.12$ ,  $p = .26$ , log-odds estimate = 0.05), but not the incorrectly bisected ( $t = 0.50$ ,  $p = .62$ , RT estimate = 13 ms;  $z = -1.55$ ,  $p = .12$ , log-odds estimate = -0.10), problems.

In contrast to the magnitude judgement task, the GLMMs revealed that for both the correctly and incorrectly bisected problems, the Abacus Experts produced significantly faster RT and fewer errors than either the Japanese (correctly bisected:  $t = 15.29$ ,  $p < .001$ , RT estimate = 1518 ms;  $z = 7.14$ ,  $p < .001$ , log-odds estimate = 0.96; incorrectly bisected:  $t = 9.17$ ,  $p < .001$ , RT estimate = 874 ms;  $z = 4.95$ ,  $p < .001$ , log-odds estimate = 0.87) or Australian students (correctly bisected:  $t = 16.21$ ,  $p < .001$ , RT estimate = 1670 ms;  $z = 5.73$ ,  $p < .001$ , log-odds estimate = 0.80; incorrectly bisected:  $t = 10.95$ ,  $p < .001$ , RT estimate = 1097 ms;  $z = 7.44$ ,  $p < .001$ , log-odds estimate = 1.31). Neither RT,  $t = -1.46$ ,  $p = .15$ , RT estimate = -151 ms, nor accuracy,  $z = 1.34$ ,  $p = .18$ , log-odds estimate = 0.16, differed significantly between the Japanese and Australian students for the correctly bisected problems. However, the incorrectly bisected problems yielded slower,  $t = -2.22$ ,  $p = .03$ , RT estimate = -223 ms, and significantly less accurate responses,  $z = -2.92$ ,  $p = .004$ , log-odds estimate = -0.44, for the Australian than Japanese students.

7) This overall accuracy criterion was more lenient than the 75% cut-off used by Nuerk et al. (2002). However, the present experiment was more difficult due to the removal of both non-bisectable and multiplicative trials.

Figure 5

Mean RT and Error for Each Group, Level of Unit-Decade Compatibility and Part-Whole Congruency in the Number Bisection Task Separately Calculated for the Correctly (Top Row) and Incorrectly (Bottom Row) Bisected Problems

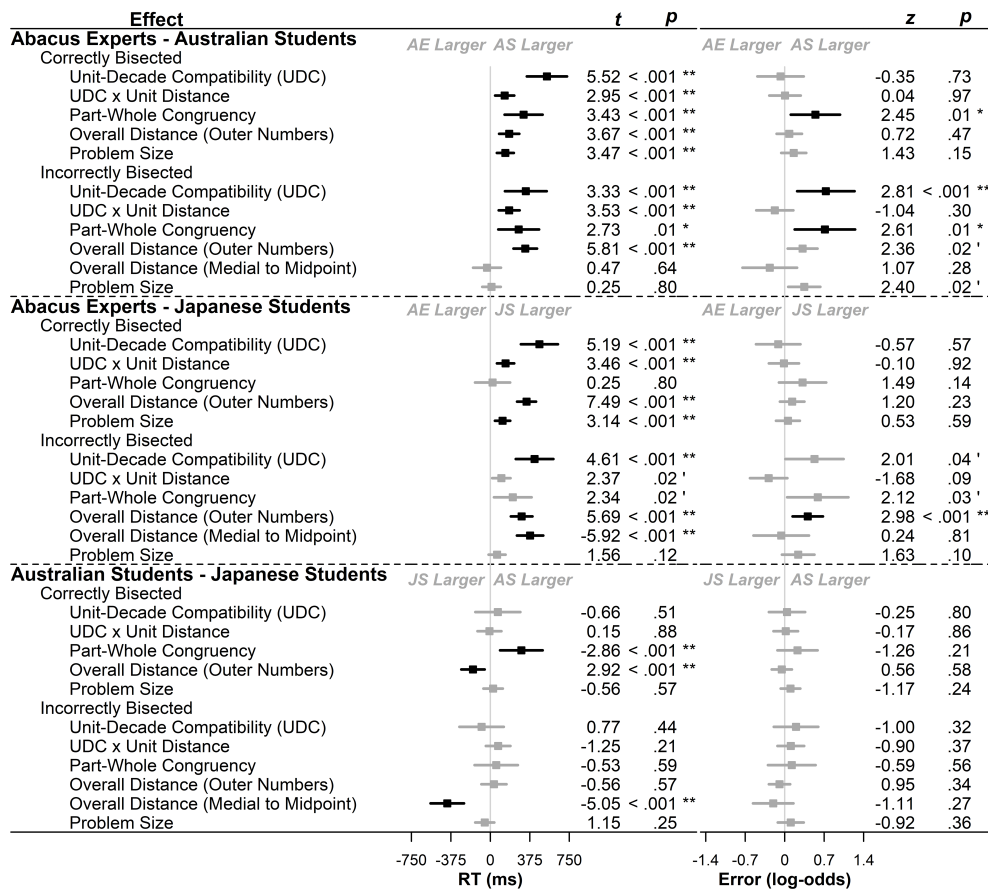


*Note.* The means for the compatible and incompatible trials were averaged over part-whole congruency. The means for the congruent and incongruent trials were averaged over unit-decade compatibility. Error bars represent 95% confidence intervals estimated from GLMMs based purely on fixed effect uncertainty.

The descriptive estimates and test statistics from the group comparisons for each effect are presented in Figure 6. Unlike the magnitude judgement task, many comparisons were significant, and they discriminated the Abacus Experts from the Australian and Japanese students by showing that most effects (i.e., unit-decade compatibility, part-whole congruency, overall distance, and problem size) were smaller among the Abacus Experts than either student group. None of these significant group differences revealed a larger effect for the Abacus Experts than either the Australian or Japanese students. Relatively few differences were observed between the Australian and Japanese students. Although the Japanese students produced a significantly larger overall distance effect on RT for the correctly and incorrectly bisected problems, their part-whole congruency effect on RT for the correctly bisected problems was significantly weaker than the Australian students'.

Figure 6

The GLMM Group Comparisons for Each Effect in the Number Bisection Task



Note. Error bars represent 95% confidence intervals for each comparison. AS = Australian students; AE = Abacus Experts; JS = Japanese students. 'p < .05. \*p < .017. \*\*p < .001.

### Discussion

Contrary to Experiment 2, the number bisection task revealed strong benefits of mental abacus expertise on overall performance: for both the correctly and incorrectly bisected problems, the Abacus Experts were faster and more accurate than either non-expert group. These gains in overall performance were accompanied by weaker reliance on decomposed magnitude representations, indexed by the Abacus Experts' smaller unit-decade compatibility and part-whole congruency effects. Mental abacus expertise, rather than socio-cultural factors, seemed to drive both the Abacus Experts' superior performance and weaker reliance on decomposed representations, because the Australian and Japanese student differences were descriptively smaller and statistically fewer. Yet, as elaborated in the General Discussion, socio-cultural factors might explain the more accurate responses and larger overall distance effect on RT for the incorrectly bisected problems by the Japanese than Australian students. Like the magnitude judgement task, this experiment provided no clear evidence of greater reliance on holistic magnitude representations by the Abacus Experts: both the overall distance and problem size effects were generally weaker than the other two groups.

### Experiment 4: Mental Addition Task

Introducing problems that required mental calculation revealed significant benefits of mental abacus expertise. Although the Abacus Experts were substantially younger and less educationally mature, they solved the bisection prob-

lems more quickly, and at least as accurately, than either the Australian or Japanese students. Both the unit-decade compatibility and part-whole congruency effect comparisons converged in suggesting that mental abacus experts were less reliant on decomposed mental representations than the abacus-naïve adults. However, there was no evidence to suggest that the mental abacus experts relied more on holistic magnitude representations.

Mental calculation was also required in the mental addition task of Experiment 4. This task presented pairs of reference and target equations, each containing two-digit operands, and participants were asked to judge whether each pair yielded the same or different totals. Group differences in decomposed mental representations were tested using the *part-whole congruency effect* based on response consistency between the entire equations and their unit digit subsets. For example, the pair  $67+31_{85}+42$  is part-whole congruent (hereafter the *Different Unrelated* condition), because the response ‘different’ is supported by both the equations’ totals (viz. 98 and 127), and the sum of their unit digits (viz. 8 and 7). Conversely, the pair  $67+31_{45}+23$  is part-whole incongruent (hereafter the *Same Unit Total* condition), because the two equations yield different totals (viz. 98 and 68), but their unit digit partial sums are the same (viz. 8). Group differences in holistic mental representations were tested using the *overall distance effect* by manipulating the difference in total between the reference and target equations: in the *Close* condition, totals only differed by 1 (e.g.,  $67+31_{85}+12$ ), substantially less than the average difference of 26.35 between totals in the Different Unrelated condition (e.g.,  $67+31_{85}+42$ ). If the Abacus Experts rely more on holistic mental representations, they should produce a smaller difference in performance between the incongruent and congruent stimuli, but a larger discrepancy between the Close and Different Unrelated conditions, than either the Australian or Japanese students.

Additional conditions were constructed to explore whether the mental abacus experts and non-experts use different calculation strategies to solve mental arithmetic problems. Group differences in calculation strategies were assessed using the *algorithms effect* based on the same total conditions in which the interim sums derived for the reference equation were the operands shown in the target equation. In the *Decomposition* condition, target equation operands were the reference equation interim sums derived using a decomposition algorithm (e.g.,  $67+31_{90}+08$ ), while the *Sequential* condition used the interim sums from a sequential algorithm (e.g.,  $67+31_{97}+01$ ), and the *Same Unrelated* condition used numbers consistent with neither of these algorithms (e.g.,  $67+31_{56}+42$ ). Faster and more accurate judgements for the Decomposition condition were expected for the Australian and Japanese students, but the Sequential condition for the Abacus Experts, because the target equation operands were the interim sums recently computed during mental addition of the reference based on the algorithm usually taught to each group by their arithmetic teachers (e.g., Stigler, 1984).

A final subset of conditions investigated whether different processing strategies were used by the mental abacus experts and non-experts based on perceptual or semantic overlap. Group differences in the *perceptual overlap effect* were tested by varying the number and position of shared digits between the reference and target equations. Conditions in which the target equation operands were *Identical* to its reference (e.g.,  $67+31_{67}+31$ ), *Legally Transposed* within place-value (e.g.,  $67+31_{61}+37$ ), or *Illegally Transposed* between place-values (e.g.,  $67+31_{63}+71$ ) were compared with the Same Unrelated condition (e.g.,  $67+31_{56}+42$ ) and the Different Unrelated condition (e.g.,  $67+31_{85}+42$ ) in which the target shared few if any digits with its reference. These conditions were used to separately evaluate the perceptual overlap effect for the Same and Different Total problems, while the *overall distance effect* was used to estimate the influence of semantic overlap on performance. If the processing strategy used by the Abacus Experts is less sensitive to perceptual overlap and more sensitive to semantic overlap, they should produce a smaller difference in performance between the Identical, Legally Transposed, and Same Unrelated conditions for the Same Total problems, as well as a smaller difference between the Illegally Transposed and Different Unrelated conditions for the Different Total problems, than either the Australian or Japanese students. Conversely, a larger difference in performance was expected for the Abacus Experts between the Close and Different Unrelated conditions.

## Method

### Participants

Altogether, 255 Australian students, 79 Abacus Experts, and 99 Japanese students from Experiment 1 participated in the mental addition task.



## Design and Stimuli

The reference equations for the mental addition task comprised 120 pairs of two-digit operands. Half of these addition problems required a carry from the unit position (hereafter *Unit Carry* problems); the rest did not involve any carry operations (hereafter *No Carry* problems). All reference equations produced a two-digit sum, which excluded decade numbers (e.g., 40) or repeated digits (e.g., 77) in either the operands or solution. Each digit across both operands was unique. Neither the operands' sum,  $M = 86$ ,  $SD = 11.29$ ;  $M = 83.1$ ,  $SD = 10.13$ , nor average problem size,  $M = 43$ ,  $SD = 5.64$ ;  $M = 41.55$ ,  $SD = 5.06$ , differed significantly between the No Carry and Unit Carry problems,  $t(118) = 1.48$ ,  $p = .14$ ;  $t(118) = 1.48$ ,  $p = .14$ . Nine target equations were constructed for each reference equation, one for each condition. As illustrated by the examples in Table 2, five were *Same Total* problems, because each reference and target equation pair yielded the same total for the Identical, Legally Transposed, Decomposition, Sequential, and Same Unrelated conditions. The rest were *Different Total* problems for the Illegally Transposed, Close, Same Unit Total, and Different Unrelated conditions.<sup>8</sup> Comparisons between these Same and Different Total problem conditions yielded the following effects used to infer which mental representations and strategies affect task performance.

**Table 2**

*Examples of the Reference and Target Equations in the Mental Addition Task*

Target	Reference	
	No Carry (67 + 31)	Unit Carry (69 + 23)
<b>Same Total</b>		
Identical	67 + 31	69 + 23
Legally Transposed	61 + 37	63 + 29
Decomposition	90 + 08	80 + 12
Sequential	97 + 01	89 + 03
Same Unrelated	56 + 42	58 + 34
<b>Different Total</b>		
Illegally Transposed	63 + 71	62 + 93
Close	85 + 12	78 + 13
Same Unit Total	45 + 23	48 + 24
Different Unrelated	85 + 42	85 + 74

Within the Different Total problems, two conditions assessed the part-whole congruency effect. Target equations from the *Different Unrelated* condition were defined as part-whole congruent, because the unit digits also yield a different total from the reference equations' unit digits (e.g.,  $67 + 31 \neq 85 + 42$ , and  $7 + 1 \neq 5 + 2$ ). Conversely, target equations from the *Same Unit Total* condition were defined as part-whole incongruent, because these Different Total problems had unit digits that yield the same total as the reference equations' unit digits (e.g.,  $67 + 31 \neq 45 + 23$ , but  $7 + 1 = 5 + 3$ ).

Within the Same Total items, two conditions assessed the perceptual overlap effect: in the *Identical* condition, all digits from the reference equation were repeated in the same position, while in the *Legally Transposed* condition, both digits occupying the unit positions from the reference equation were interchanged. Another two sets of Same Total items assessed the algorithms effect. In the *Decomposition* condition, target equation operands were interim sums created by separately adding pairs of decade and unit digits together from the reference equation. In the *Sequential* condition, target equation operands were intermediary steps in the sequential algorithm described earlier: the first

8) Additional filler Different Total problems were created for two reasons. First, to equate the number of Same and Different Total conditions, and second to control for the appearance of decade (e.g., 70) and single-digit numbers (e.g., 01) in the Same Total problems. These filler problems were created using the same algorithm as the Sequential condition, but the decade digit from one of the operands in the reference equation was incorrectly added to the unit position of the other operand to yield different totals (e.g.,  $67+31 \neq 70+01$ ).

operand was constructed by incrementing the augend by the addend's decade value, while the other operand was the residual unit digit of the addend. Finally, target equation operands in the *Same Unrelated* condition served as a baseline for the Same Total problems to assess the perceptual overlap and algorithms effects: they comprised numbers that excluded interim sums derived using either a decomposition or sequential algorithm and contained fewer than two of the four digits printed in each reference equation.

For the Different Total problems, target equation operands in the *Illegally Transposed* condition were created by swapping the augend's unit digit with the addend's decade digit from the reference equation. In the *Close* condition, totals differed from the reference by  $\pm 1$ . In the *Same Unit Total* condition, sums were identical with the reference in the unit position but differed by at least  $\pm 1$  in the decade position. Finally, target equation operands in the *Different Unrelated* condition served as a baseline for the Different Total problems to assess the perceptual overlap, overall distance, and part-whole congruency effects: they had different digits and yielded sums which differed from the reference equation by more than  $\pm 1$ .

Within each condition, overall distance was defined as the difference between totals, while problem size was calculated as the sum over the reference equation's operands (cf. Ashcraft & Battaglia, 1978; LeFevre, Sadesky, & Bisanz, 1996). Appendix B summarises these attributes for each of the Same and Different Total conditions. Based on all 120 reference equations, target equations from the Identical and Legally Transposed conditions had significantly more digits in common with the reference than either the Decomposition, Sequential or Same Unrelated conditions,  $|t(595)| > 13.17$ ,  $p < .001$ . From the pool of Different Total problems, a subset of 80 target equations (20 per condition) were selected to ensure that overall distance and problem size was matched between the Illegally Transposed, Same Unit Total, and Different Unrelated conditions. These target equations from the Different Unrelated condition did not significantly differ in problem size to the Illegally Transposed, Close or Same Unit Total conditions,  $|t(76)| < 1.54$ ,  $p > .13$ , but they had a significantly larger overall distance between sums than in the Close condition,  $t(76) = 6.31$ ,  $p < .001$ , and they contained fewer digits in common with the reference than in the Illegally Transposed condition. As required for the perceptual overlap and part-whole congruency effects, overall distance was not significantly different between the Different Unrelated and either the Illegally Transposed,  $t(76) = 0.05$ ,  $p = .96$ , or Same Unit Total conditions,  $t(76) = 0.21$ ,  $p = .83$ .

## Procedure

Each trial of the mental addition task consisted of three successive displays. The first display, containing two sets of hash marks (#####) situated approximately 8 cm from the top and bottom of a 15" LCD screen, appeared for 300 ms. The second display replaced the upper series of hash marks with the reference equation for 1 second. Finally, the reference equation disappeared, and the target equation replaced the lower series of hash marks for 3 seconds or until response to the third display. Single-digit operands always included an extra 0 in the decade position (e.g., 01), and the larger operand within each equation always appeared on the left. As in the number bisection task, participants saw an equal number of equation pairs within each condition, 22 practice trials preceded the critical items, and the target equations were presented using a Latin Square design, such that each participant saw all 120 reference equations exactly once, but across participants all reference equations were paired with each of their different target equations. Participants were instructed to press one response key with their right hand if the upper and lower equation pair produced the same total, and another key with their left hand if they produced different totals. These instructions were illustrated with one example from the Same Unrelated condition, and another from the Illegally Transposed condition, to encourage decisions based on the reference and target equations' totals rather than digit overlap alone. Participants were not informed about any other relation between the reference and target equations.

## Results

Eighteen Australian students who responded to fewer than 50% of the Same or Different Total problems, two Australian students with an excessive number of anticipatory responses below 100 ms (>10%), and one Australian student who classified all problems as having different totals were excluded from the analyses. Data for the remaining 234 Australian

students, 79 Abacus Experts and 99 Japanese students were cleaned to remove anticipatory RT below 100 ms and responses exceeding 3,000 ms (3.31%, 2.87% and 2.23% of each group's total, respectively).<sup>9</sup>

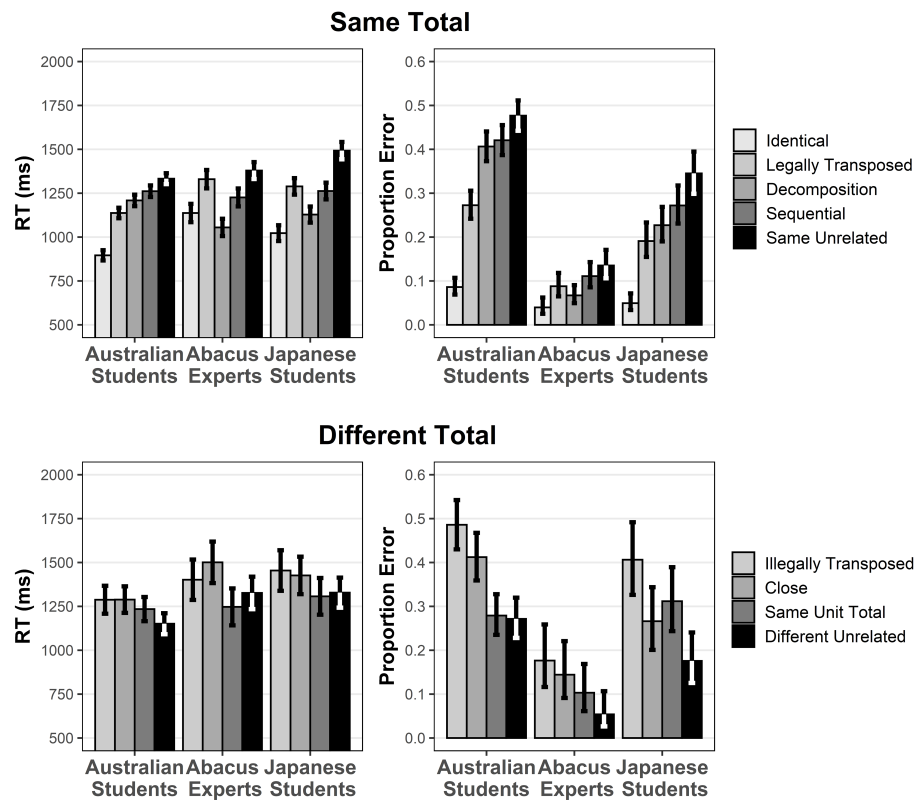
Figure 7 separately presents the mean RTs and error proportions for the Same and Different Total problems according to group and condition. Averaged over group, the GLMMs revealed that performance was affected by part-whole congruency, perceptual overlap, algorithms, and overall distance. For the part-whole congruency effect, significantly more errors were observed in the Same Unit Total than Different Unrelated condition,  $t = -0.16$ ,  $p = .87$ , RT estimate = -6ms;  $z = 2.76$ ,  $p = .01$ , log-odds estimate = 0.50. For the perceptual overlap effect, significantly faster RT and fewer errors were observed in the Identical,  $t = -27.81$ ,  $p < .001$ , RT estimate = -382 ms;  $z = -16.12$ ,  $p < .001$ , log-odds estimate = -1.97, and Legally Transposed,  $t = -11.18$ ,  $p < .001$ , RT estimate = -150 ms;  $z = -9.46$ ,  $p < .001$ , log-odds estimate = -0.72, than Same Unrelated condition, but significantly longer RT and more errors were observed in the Illegally Transposed than Different Unrelated condition,  $t = 2.98$ ,  $p = .01$ , RT estimate = 113 ms;  $z = 6.41$ ,  $p < .001$ , log-odds estimate = 1.14. Direct comparison between the Identical and Legally Transposed equations revealed significantly faster RT and fewer errors in the former than latter condition,  $t = 17.38$ ,  $p < .001$ , RT estimate = 224 ms;  $z = 10.45$ ,  $p < .001$ , log-odds estimate = 1.13. For the algorithms effect, significantly faster RT and fewer errors were observed in the Decomposition,  $t = -19.80$ ,  $p < .001$ , RT estimate = -271 ms;  $z = -8.14$ ,  $p < .001$ , log-odds estimate = -0.55, and Sequential,  $t = -10.99$ ,  $p < .001$ , RT estimate = -152 ms;  $z = -4.26$ ,  $p < .001$ , log-odds estimate = -0.27, than Same Unrelated condition. Direction comparison revealed significantly faster RT and fewer errors in the Decomposition than Sequential condition,  $t = 12.58$ ,  $p < .001$ , RT estimate = 109 ms;  $z = 4.21$ ,  $p < .001$ , log-odds estimate = 0.29. For the overall distance effect, significantly longer RT and more errors were observed in the Close than Different Unrelated condition,  $t = 3.68$ ,  $p < .001$ , RT estimate = 137 ms;  $z = 4.18$ ,  $p < .001$ , log-odds estimate = 0.75, and for problems with a smaller difference in sum between the reference and target equation calculated within conditions,  $t = -2.43$ ,  $p = .01$ , RT estimate = -43 ms;  $z = -2.35$ ,  $p = .02$ , log-odds estimate = -0.17. The problem size effect was not observed, because the reference equations with larger sums did not yield significantly longer RT or more errors for either the Same,  $t = -0.46$ ,  $p = .65$ , RT estimate = -2 ms;  $z = 1.11$ ,  $p = .27$ , log-odds estimate = 0.03, or Different Total problems,  $t = 1.76$ ,  $p = .08$ , RT estimate = 27 ms;  $z = 1.48$ ,  $p = .14$ , log-odds estimate = 0.09.

Averaged over condition, the GLMMs revealed that for both problem types, the Australian students were significantly faster but less accurate than either the Japanese students (Same Total:  $t = 3.59$ ,  $p < .001$ , RT estimate = 72 ms;  $z = -5.93$ ,  $p < .001$ , log-odds estimate = -0.62; Different Total:  $t = 3.75$ ,  $p < .001$ , RT estimate = 138 ms;  $z = -2.80$ ,  $p = .01$ , log-odds estimate = -0.35) or Abacus Experts (Same Total:  $t = -2.72$ ,  $p = .01$ , RT estimate = -58 ms;  $z = 13.13$ ,  $p < .001$ , log-odds estimate = 1.59; Different Total:  $t = -3.28$ ,  $p < .001$ , RT estimate = -130 ms;  $z = 8.63$ ,  $p < .001$ , log-odds estimate = 1.50). On average, the Japanese students were significantly less accurate than the Abacus Experts (Same Total:  $z = 6.97$ ,  $p < .001$ , log-odds estimate = 0.97; Different Total:  $z = 6.04$ ,  $p < .001$ , log-odds estimate = 1.15), but they did not differ significantly in RT (Same Total:  $t = 0.55$ ,  $p = .59$ , RT estimate = 14 ms; Different Total:  $t = 0.21$ ,  $p = .84$ , RT estimate = 10 ms).

9) Another 72 Australian students, 2 Abacus Experts and 13 Japanese students produced mean accuracy below 50% for either the Same or Different Total problems. Excluding those participants reduced the gap in RT and accuracy between the Australian students and Abacus Experts or Japanese students, but did not change any of the effects assessing differences in mental representations or strategy between the groups.

Figure 7

Mean RT and Error for Each Group and Condition in the Mental Addition Task Separately Calculated for the Same (Top Row) and Different (Bottom Row) Total Problems

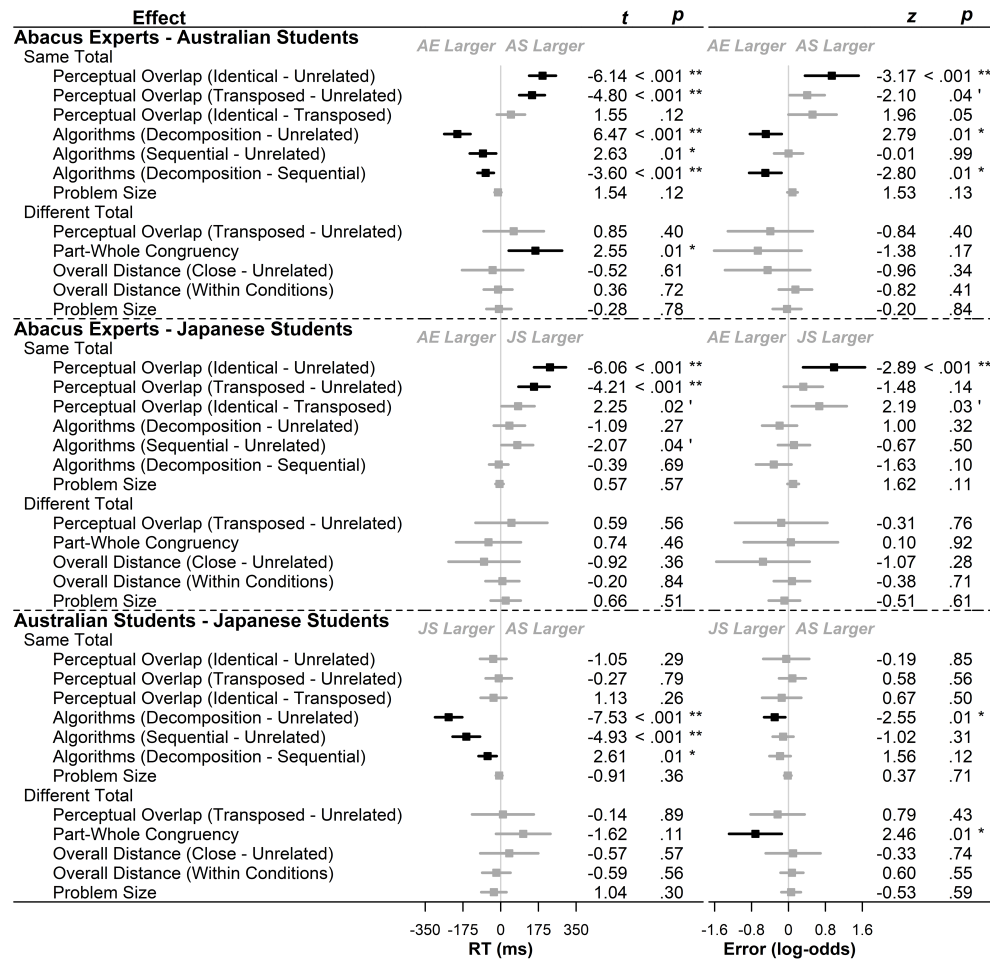


*Note.* The means for each condition were averaged over carry status. Error bars represent 95% confidence intervals estimated from GLMMs based purely on fixed effect uncertainty.

The descriptive estimates and test statistics from the group comparisons for each effect are presented in Figure 8. The perceptual overlap effects based on comparisons between the Identical and Legally Transposed conditions with the Same Unrelated baseline were significantly weaker for the Abacus Experts than either the Australian or Japanese students. The algorithms effect based on comparisons between the Decomposition and Sequential conditions with the Same Unrelated baseline were significantly weaker for the Australian students than either the Abacus Experts or Japanese students. The faster RT and fewer errors in the Decomposition than Sequential condition was significantly more pronounced for both the Abacus Experts and Japanese students than the Australian students. The part-whole congruency effect on accuracy was significantly more pronounced for the Japanese than Australian students, and it also differed significantly between the Australian students and Abacus Experts, because the Australian students produced slower RT in the Same Unit Total than Different Unrelated condition, but the opposite was observed among the Abacus Experts. Neither the overall distance nor problem size effects differed significantly according to group.

Figure 8

The GLMM Group Comparisons for Each Effect in the Mental Addition Task



Note. Error bars represent 95% confidence intervals for each comparison. AS = Australian students; AE = Abacus Experts; JS = Japanese students. \* $p < .05$ . \*\* $p < .017$ . \*\*\* $p < .001$ .

## Discussion

In contrast to the number bisection task, the Abacus Experts' advantage in overall performance was confined to accuracy. Responses to both the Same and Different Total problems were significantly faster, but less accurate, among the Australian students than either the Japanese students or Abacus Experts. The Abacus Experts produced significantly fewer errors than the Japanese students, but there was no RT difference between those two groups. This discrepancy between the RT and accuracy data patterns suggests that a speed-accuracy trade-off affected the abacus-naïve participants. Given the relatively short time available to compute the reference equation's sum, the Australian and Japanese students apparently relied on the number of shared digits between the reference and target equations, as indexed by their significantly larger perceptual overlap effects. Conversely, the smaller perceptual overlap effects for the Abacus Experts indicates that they were relatively immune to this variable, suggesting that they computed the reference equation's sum more quickly.

The results also suggested that all groups preferred using a decomposition than sequential algorithm. However, the better performance in the Decomposition than Sequential or Same Unrelated conditions was significantly more marked for the Abacus Experts and Japanese students than Australian students. Such findings might reflect greater emphasis

on decomposition algorithms in Japanese classrooms, but they deviate from the prediction that Abacus Experts would prefer using a sequential algorithm.

Finally, the part-whole congruency effect results indicated that the Australian students were less reliant on decomposed mental representations than the Japanese students. There was also a significant discrepancy between the Abacus Experts and Australian students in the part-whole congruency effect due to opposite RT differences between the Same Unit Total and Different Unrelated condition. We also observed a significant overall distance effect which did not differ significantly between the groups, suggesting greater reliance on holistic than decomposed magnitude representations than in the preceding experiments. However, Nuerk, Kaufmann, Zoppoth, and Willmes (2004) suggested an alternative interpretation for this observation. Totals from the reference and target equations were identical in the decade and unit positions for the Close (e.g.,  $67+31\_85+12 = 98\_97$ ) and Same Unit Total (e.g.,  $67+31\_45+23 = 98\_68$ ) conditions, respectively. Thus, rather than reflecting holistic representations, worse performance in the Close condition might arise from decomposed representations in which participants focus entirely, or predominantly, on the decade digits' totals (cf. the rationale provided for the Same Unit Total condition). The combination of a large overall distance effect but a negligible part-whole congruency effect shown by the Australian students may therefore reflect a tendency to respond 'same' when the decade digits totals matched and 'different' when they mismatched – a strategy that would only produce an average error rate of 15%. Such a strategy would also align with the salience of decades in both English and Japanese following translation of the reference equation into verbal working memory for later comparison with the target equation (Macizo & Herrera, 2010; Miura, 1987; Nuerk, Weger, & Willmes, 2005).

## General Discussion

The present research compared a large sample of mental abacus experts with two groups of abacus-naïve adults on performance in three domain-general measures of working memory (memory updating, operation span, and spatial memory) and three domain-specific numerical tasks (magnitude judgement, number bisection, and mental addition). The results indicated that the Abacus Experts only performed better when the task required mental arithmetic – for example, calculation of midpoints or totals for subsequent report (memory updating), or comparison with each trial's medial number (number bisection) or target equation total (mental addition). The Abacus Experts did not outperform the Japanese students in the operation span or magnitude judgement tasks. Although those tasks also required decisions about numerical stimuli, directly retrieved solutions or quantities from long-term memory can be used for single-digit equation verification in the operation span task (e.g.,  $4 + 2 = 7$ , true or false?), or identifying the larger number in the magnitude judgement task (e.g., 4 or 2, which is bigger?). Overall, this pattern suggests that mental abacus expertise is associated with a *domain-specific* advantage in mental calculation, rather than a general processing advantage for numerical stimuli, at least for the range of numbers tested in this research (either single-digit in the working memory tasks, or double-digits in the numerical tasks).

Apart from the overall group comparisons summarised above, stimuli for the working memory and numerical tasks were selected to investigate whether mental abacus expertise or socio-cultural factors are associated with augmented central executive and visuo-spatial working memory, increased reliance on holistic than decomposed mental representations of two-digit numbers, preference for a sequential than decomposition algorithm, and greater sensitivity to semantic than perceptual overlap between numbers. The group differences relevant to each hypothesis are discussed in the following sections, followed by a consideration of their broader implications for numerical cognition and the study of expert performance.

### Augmented Central Executive and Visuo-Spatial Working Memory?

The results from Experiment 1 yielded no evidence to suggest that mental abacus experts have superior central executive and visuo-spatial working memory, because they only performed better than both non-expert groups in the memory updating task. One interpretation for the Abacus Experts' better performance in the memory updating task, but worse performance (than the Japanese students) on the operation span task, is that training cultivates a particular functional aspect of working memory: updating, rather than shifting of attention between processing and

storage (Miyake et al., 2000). Alternatively, better performance in the memory updating task might simply reflect the Abacus Experts' proficiency with mental arithmetic problems. To adjudicate between these possibilities, future research might compare experts and non-experts using tasks without numerical stimuli that specifically target each of the three functional aspects of central executive working memory (updating, shifting and inhibition; Miyake et al., 2000).

Contrary to Lee et al. (2007), and in line with Barner et al. (2016), the Abacus Experts did not show better visuo-spatial working memory. Instead, performance in the spatial memory task was not significantly different between all three groups. However, this result should be interpreted with caution, not only because a ceiling effect was observed, but also because the mental abacus experts were substantially younger than both non-expert groups. If performance in visuo-spatial working memory tasks generally favours older participants (e.g., Pickering, 2001), the expected positive relationship between mental abacus expertise and visuo-spatial working memory in the younger Abacus Experts might be counteracted to yield no group difference in the spatial memory task. Nonetheless, the Australian and Japanese students' performance in the spatial memory task was unsurpassed by the Abacus Experts. This result is unlikely to be due to a lack of statistical power or skill, because the sample of Abacus Experts recruited for this research was larger than most previous studies ( $N = 79$ , with 19% awarded the highest rank of 10<sup>th</sup> dan by the League for Soroban Education in Japan). Consistent with this conclusion, Srinivasan, Wagner, Frank, and Barner (2018) proposed that the mental abacus engages cognitive and perceptual abilities prevalent in novices. Focusing on the physical abacus as a springboard for mental abacus calculations (see also Stigler, 1984), Frank and Barner (2012) hypothesised that the device is optimised for visual processing, because its rectilinear structure with four beads per column permits users to store and manipulate magnitudes using two phylogenetically ancient cognitive mechanisms: the approximate number system for large numbers (e.g., Feigenson, Dehaene, & Spelke, 2004), and the object tracking system for small numbers (e.g., Mou & vanMarle, 2014). Both systems spontaneously operate when abacus-naïve adults estimate quantity over multiple sets in parallel (Halberda, Sires, & Feigenson, 2006). Research with mental abacus experts and non-experts in visual cognition therefore led Frank and Barner (2012) to conclude that the experts devoted "existing visual resources ... to represent large exact numerosities" (p. 136, emphasis mine).

## Less Decomposed and More Holistic Mental Representation of Two-Digit Numbers?

All three numerical tasks were constructed to explore whether mental representations of two-digit numbers differed between the mental abacus experts and non-experts. The results successfully replicated the unit-decade compatibility and overall distance effects for the Australian and Japanese students in the magnitude judgement task of Experiment 2, suggesting that both holistic and decomposed mental representations affected these non-experts' responses. Similar findings were observed in the number bisection task by extending the unit-decade compatibility effect to the problems' outer numbers, and introducing a novel variant – the part-whole congruency effect – based on conditions which manipulated response consistency between the entire stimulus and subset of unit digits. Results from the mental addition task also implied that both representation types are involved for the non-experts. On average, the part-whole congruency effect was significant, indicating that performance was affected by decomposed mental representations. Holistic magnitude representations also seemed to affect responses, because performance was sensitive to the within conditions measure of overall distance between the reference and target equation totals. These findings support Moeller, Huber, Nuerk, and Willmes' (2011) hybrid model for two-digit numbers by providing converging evidence across three different numerical tasks using the unit-decade compatibility and part-whole congruency effects.

The group comparisons revealed that the unit-decade compatibility and part-whole congruency effects were significantly weaker for the Abacus Experts than non-experts in the number bisection task. These findings suggest that mental representations are less decomposed following extensive practice with Hindu-Arabic numbers, as predicted by Tzelgov et al. (2015). Significant group differences in the part-whole congruency effect were also observed in the mental addition task, but the smaller unit-decade compatibility effect shown by the Abacus Experts in the magnitude judgement task did not significantly differ from either of the abacus-naïve groups. This discrepancy suggests that group differences in reliance on decomposed mental representations are more readily observed in arithmetic tasks than simple magnitude judgements.

Contrary to Tzelgov et al.'s prediction, the Abacus Experts showed equivalent or smaller overall distance and problem size effects than the abacus-naïve groups across all tasks, contradicting the expected positive relationship between those variables and the strength of an individual's holistic magnitude representations. This inconsistency might be resolved by examining the origins of the symbolic distance effect for single-digit numbers. According to [Holloway and Ansari \(2009\)](#), the distance effect results from noisy mappings between Hindu-Arabic numerals and the magnitudes they represent. Extensive practice accessing the meaning of a Hindu-Arabic numeral increases the distinctiveness of one magnitude over those in the surrounding neighbourhood. These increasingly precise representations, reflecting stronger connections between symbols and referents, reduce the amount of mental overlap between quantities to yield smaller distance effects for children with higher mathematical achievement test scores ([Holloway & Ansari, 2009](#)), and for literate than illiterate adults who differ in the amount of formal education received ([Zebian & Ansari, 2012](#)). Extending this idea to two-digit numbers, stronger connections between sequences of Hindu-Arabic numerals and their associated holistic magnitude would also produce smaller rather than larger overall distance and problem size effects. Observing this pattern for the Abacus Experts in the present research suggests that their mental representations of two-digit numbers are simultaneously less decomposed and more precise than the abacus-naïve adults.

### Preference for a Sequential Algorithm?

Several conditions from the mental addition task were designed to evaluate whether a sequential algorithm was preferred by the mental abacus experts, while a decomposition algorithm was preferred by the non-experts. Such a difference would be indexed by faster and more accurate performance in the Sequential than Decomposition or Same Unrelated conditions for the Abacus Experts, but superior performance in the Decomposition than Sequential or Same Unrelated conditions for both the Australian and Japanese students. This prediction was not supported, because the better performance in the Sequential than Same Unrelated condition did not significantly differ between the groups, and the superior performance in the Decomposition than Sequential or Same Unrelated condition was significantly more pronounced for the Abacus Experts and Japanese students than Australian students. These results suggest that both Japanese samples preferred using a decomposition algorithm, but they also flexibly used a sequential algorithm like the Australian students because all groups performed better in the Sequential than Same Unrelated condition.

Flexible algorithm use for all groups might have arisen from spontaneous development or formal education about multiple arithmetic procedures. For example, [Heirdsfield and Cooper \(2004\)](#) reported that four Australian Grade 3 students from a sample of six had invented procedures to solve arithmetic problems by themselves, including a sequential (e.g.,  $28 + 35 = 28 + 30$  followed by  $58 + 5$ ) and shortcut (e.g.,  $45 + 19 = 45 + 20 - 1$ ) algorithm not taught in the children's classrooms. Formal education might instead explain the Abacus Experts and Japanese students' preference, because they are taught how to add and subtract using a decomposition algorithm from as early as Grade 2 ([Takahashi et al., 2008](#)). Early and sustained exposure to a decomposition algorithm could have promoted fluency or dominance for this procedure, especially for 'easier' problems which only contain two-digit operands. Another explanation for these unexpected results acknowledges subjectivity in how mental abacus proficiency was defined in this research using ranks awarded by the League for Soroban Education in Japan. Extraordinary mental arithmetic performance can be achieved without mental abacus training ([Pesenti, 2005](#)), and variability in mental abacus proficiency probably exists within each awarded rank. These issues with group definition might therefore have limited the extent of group differences in the calculation strategies they applied. Regarding the Japanese students' flexible algorithm use, [McIntosh, Nohda, Reys, and Reys \(1995\)](#) describe Japanese textbooks which present multiple procedures, such as "subtract then add" (e.g.,  $13 - 9 = (10 - 9) + 3 = 4$ ) and "subtract then subtract" (e.g.,  $13 - 9 = 13 - 3 - 6 = 4$ ). Formal exposure to multiple procedures could therefore have reinforced elements of both decomposition and sequential algorithms for the Japanese students over time ([Sowder, 1990](#)).

### Reduced Sensitivity to Perceptual Overlap?

The perceptual overlap effects based on comparisons between the Identical and Legally Transposed conditions with the Same Unrelated baseline were significantly larger for the Australian and Japanese students than Abacus Experts in the mental addition task of Experiment 4. These results suggest a processing strategy by the Abacus Experts that is less



sensitive to perceptual overlap, while both groups of non-experts increased their reliance on visual characteristics, such as the number of shared digits, to achieve satisfactory performance.

Such reliance on perceptual overlap by the Australian and Japanese students is consistent with Stanovich's (1980) interactive-compensatory model. Originally developed for reading comprehension, the model proposes that individuals simultaneously use different sources of knowledge when processing written information (e.g., shape of letters, word frequency, syntax and context), and that any knowledge source can be used by an individual to compensate for deficits in other areas. For example, "a reader with poor word recognition skills may actually be prone to a greater reliance on contextual factors because these provide additional sources of information" (p. 36). Viewed from this perspective, compensation was used to explain larger regularity effects in speeded pronunciation for slower readers (Brown, Lupker, & Colombo, 1994), stronger frequency effects in lexical decision among those with less print exposure (Chateau & Jared, 2000), and heightened sensitivity to orthographic or phonological neighbourhood size when students with inferior vocabulary perform speeded pronunciation or lexical decisions (Yap, Balota, Sibley, & Ratcliff, 2012).

Translating Stanovich's model from reading comprehension to mental arithmetic, deficits in retrieval or algorithm execution might lead some individuals to compensate by increasing sensitivity to stimulus features that will allow them to apply task-specific strategies (e.g., parity or multiplicativity in the number bisection task: Nuerk et al., 2002; magnitude of the proposed solution, "split", when verifying simple equations: Ashcraft & Battaglia, 1978). Such deficits typically manifest as slower RT and poorer accuracy (e.g., LeFevre & Bisanz, 1986; Moeller, Pixner, Zuber, Kaufmann, & Nuerk, 2011). However, variability in strategies chosen across individuals can sometimes disguise or magnify these differences in task performance (e.g., Shrager & Siegler, 1998; Torbeyns & Verschaffel, 2013). For example, Campbell and Xue (2001) found that Canadian adults were more likely to compensate for deficits in retrieval than Chinese adults by using slower error-prone back-up strategies (e.g., counting) when solving single-digit addition problems. Deficits for the Canadian participants were reflected in their slower RT and lower accuracy, while the compensatory strategy used was inferred from the problem size effect, which was larger for the Canadian than Chinese participants. These results comparing 'arithmetically superior' Chinese with 'arithmetically inferior' Canadian participants reinforce Stanovich's (1980) idea that less able individuals compensate by drawing upon other knowledge sources, including the magnitude of operands or alternative algorithms at their disposal. The Abacus Experts' smaller perceptual overlap effects is consistent with Stanovich's framework because it predicts that skilled individuals would be less sensitive to the visual characteristics of the stimulus, such as the number of shared digits, if they can already compute the reference equation's sum before its target equation is presented. The same framework assuming deficits in algorithm execution can therefore explain the Australian and Japanese students' larger perceptual overlap effects and worse performance in the mental addition task of Experiment 4.

## Mental Abacus Expertise or Socio-Cultural Factors?

Three groups were compared in this research to distinguish the effects of mental abacus expertise and socio-cultural factors on performance in the working memory and numerical tasks. The Australian students differed from the Abacus Experts on socio-cultural background and abacus exposure, while the Japanese students only differed from the Abacus Experts on the latter variable. The results indicated fewer and smaller overall differences in performance between the Australian and Japanese students than either group with the Abacus Experts. For example, performance in the numerical tasks only differed between the Australian and Japanese students by an average of 135 ms on RT and 0.36 log-odds on accuracy, compared to 501 ms or 591 ms on RT and 0.86 or 1.06 log-odds on accuracy between either group and the Abacus Experts. This outcome suggests that mental abacus expertise, rather than socio-cultural factors, was responsible for the Abacus Experts' better overall performance. Comparisons between both non-expert groups revealed a small but significant contribution of socio-cultural factors on mental arithmetic performance, because the Japanese students consistently outperformed the Australians in the number bisection, mental addition, memory updating and operation span tasks. They also revealed that mental representations of two-digit numbers, and strategies used in the mental addition task, were affected by socio-cultural factors, because the Australian students produced a larger overall distance effect in the magnitude judgement task, and smaller differences in performance between the Decomposition

or Sequential conditions and Same Unrelated baseline in the mental addition task, than either the Japanese students or Abacus Experts.

Several aspects of mental abacus training might have contributed to the Abacus Experts' better overall performance, less decomposed mental representations, and decreased reliance on perceptual overlap. For example, mental abacus experts often use and memorise sets of complementary numbers (e.g., pairs which sum to 10: 1-9, 2-8, 3-7, etc; Donlan & Wu, 2017), because it reduces the cognitive load associated with carries and borrows (e.g.,  $8 + 3 = 8 + 10 - 7$ ), thus yielding better mental arithmetic performance. Mental abacus experts also tend to gesture while solving arithmetic problems (Hatano et al., 1977), actions which enhance their mental arithmetic performance (Frank & Barner, 2012) by providing "a second code" to augment the representations stored in their visuo-spatial working memory (Brooks, Barner, Frank, & Goldin-Meadow, 2018). Extensive practice solving arithmetic problems could also have contributed to the Abacus Experts' better performance by obviating the need for compensatory strategies like perceptual overlap (Stanovich, 1980) or reliance on decomposed mental representations (Tzelgov et al., 2015), and by automatising processes like algorithm execution or magnitude judgements (e.g., Gebuis, Cohen Kadosh, de Haan, & Henik, 2009; Logan, 1985; Pesenti, 2005).

The Australian students' worse performance in the mental arithmetic tasks, less precise mental representations, and relative insensitivity to decomposition or sequential algorithms might also reflect various socio-cultural factors. First, rote learning techniques are commonly used in Asian countries both at school and at home (e.g., Zhang & Zhou, 2003) to ensure fast and accurate mental arithmetic performance, but they are often discouraged in Western countries such as Australia (e.g., Imbo & LeFevre, 2009). Second, the structure and pedagogy of Japanese and Western maths classrooms differ in emphasis on developing speed and fluency of number fact retrieval, and the value assigned to homework (Chen & Stevenson, 1989; Stigler & Perry, 1988). Third, Asian academic success has sometimes been attributed to a range of motivational factors and attitudes towards maths (e.g., Chen & Stevenson, 1995; Whang & Hancock, 1994), particularly the view from Japanese parents that "everyone, under the right circumstances and with enough hard work, can learn to do math" (Nisbett, 2003, p. 189). Fourth, reliance on electronic calculators is more strongly encouraged in Western than Asian classrooms (Campbell & Xue, 2001; LeFevre & Liu, 1997), which may disguise conceptual and procedural deficits concerning multi-digit algorithms, thus contributing to the poorer arithmetic performance of Western children.

Finally, the Australian students' discrepant results might be associated with English, a language characterised by less transparent number words than Japanese (e.g., Miura, 1987). For example, the Japanese number word "ten-two" shares a transparent relationship with the place-value sequence "12", because the powers of 10 are explicitly represented using a more regular counting system than its English equivalent "twelve" (Miura et al., 1994). English, the less transparent language with irregular number words, might therefore impede place-value understanding among its speakers. Support for this hypothesis can be found in Miura and Okamoto (1989), who reported lower scores in a place-value test for monolingual children who spoke English than Japanese. Because the Japanese children had not yet received lessons on place-value when they were tested, the authors concluded that language rather than education was responsible for their precocious understanding, knowledge which Miura and Okamoto found to be positively correlated with mathematics achievement in both monolingual groups.

## Implications for Numerical Cognition, Cognitive Training, and Expert Performance

The systematic comparisons across three numerical tasks that distinguished this investigation of mental abacus experts offer new insights into the development of numerical cognition across individuals. The results suggested the presence of both holistic and decomposed magnitude representations for two-digit numbers as indexed by the overall distance, problem size, unit-decade compatibility and part-whole congruency effects. However, these representations were less decomposed and more precise for the Abacus Experts than non-experts. Such findings extend Nuerk, Kaufmann, et al.'s (2004) theory regarding the developmental progression of magnitude representations for two-digit numbers: from sequential to parallel processing of decade and unit digits with increasing age. However, beyond parallel processing, another stage was suggested by the present research – one characterised by diminished reliance on decomposition and greater precision for holistic magnitudes with increasing experience with Hindu-Arabic numbers (e.g., Tzelgov et al., 2015). This further stage was attained by the Abacus Experts beyond the parallel processing stage demonstrated by both non-expert groups.

Turning to comparisons between the working memory tasks that yield insight into the effects of training on cognitive abilities, the Abacus Experts' visuo-spatial working memory was not superior, nor were their central executive processes consistently better than the non-experts. These results challenge research (e.g., Dong et al., 2016; Lee et al., 2007) claiming that mental abacus training significantly enhances domain-general abilities. Instead, they are consistent with the literature on cognitive transfer (e.g., Redick et al., 2013) which indicates that it is rare for domain-general abilities to be affected by extensive training on unrelated tasks (see also Barner et al., 2016).

Finally, turning to implications for the study of expert performance, the results supported a combination of flexible and inflexible behaviour. For example, the Abacus Experts flexibly used a decomposition and sequential algorithm in the mental addition task, but their responses were relatively immune to perceptual overlap between equations. Insensitivity to perceptual overlap suggests that the Abacus Experts inflexibly calculated the totals for each pair of reference and target equations. This tension between flexible and inflexible behaviour corresponds to a division between “adaptive” and “routine” experts (e.g., Baroody & Dowker, 2003; Hatano, 1982). Although the latter is often used to describe mental abacus experts (e.g., Kojima, 1954), evidence of flexibility in the present research (see also the ‘super’ chess experts in Bilalić, McLeod, & Gobet, 2008) suggests that this cognitive disadvantage does not apply to highly skilled mental abacus experts.

## Conclusion

The central findings from the present research were that the Abacus Experts only performed better than both non-expert groups when the task involved mental arithmetic, and that their superiority was principally due to mental abacus expertise than socio-cultural factors. The results also showed that mental representations of two-digit numbers are more precise for the Abacus Experts, and that perceptual overlap played a smaller role in the processing strategy used by the experts during mental addition. These revelations suggest that mental abacus training can be used to support arithmetic instruction or increase mental arithmetic proficiency (see also Barner et al., 2016). However, there was no evidence to suggest that training improves domain-general processes like working memory or indiscriminately augments numerical processing efficiency. The efficacy of mental abacus training for socio-cultural contexts differing in pedagogy, calculator use, rote learning practices, or attitudes towards maths, will also need to be addressed in future research.

---

**Funding:** This research is based on the first author's doctoral dissertation, and it was supported by the Australian Government through the Australian Postgraduate Awards, the University of Sydney through the Campbell Perry International Travelling Scholarship, and Rikkyo University through the Short-Term Early Career Researcher Invitation Program to Steson Lo.

---

**Acknowledgments:** We thank the following people for enthusiastically hosting Steson Lo's visits to Japan: Kazuyuki Takayanagi, Chieko Takayanagi, Kazuma Takayanagi, and Akemi Yatani from Soroban USA; Yasushi Hino from Waseda University; and Mariko Nakayama from Rikkyo University. Our gratitude also extends to Masahiro Yoshihara, Mariko Nakayama, Fumie Kato, Hiroko Komatsu, Nerida Jarkey and Chun-fen Shao for translation assistance, Eiko Ohara and Noriko Saito for administrative help, and the keen interest shown by the parents and students from Soroban USA for this project.

---

**Competing Interests:** The authors have declared that no competing interests exist.

---

## References

- Amaiwa, S., & Hatano, G. (1989). Effects of abacus learning on 3rd-graders' performance in paper-and-pencil tests of calculation. *The Japanese Psychological Research*, 31(4), 161–168. <https://doi.org/10.4992/psycholres1954.31.161>
- Andrews, S., & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(1), 152–163. <https://doi.org/10.1037/a0024953>
- Andrews, S., & Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, 68(3), 279–296. <https://doi.org/10.1016/j.jml.2012.12.001>
- 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), 527–538. <https://doi.org/10.1037/0278-7393.4.5.527>

- Ashcraft, M. H., & Stazyk, E. H. (1981). Mental addition: A test of three verification models. *Memory & Cognition*, 9(2), 185–196. <https://doi.org/10.3758/BF03202334>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Baddeley, A. D. (1986). *Working memory*. Oxford, United Kingdom: Oxford University Press.
- Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829–839. <https://doi.org/10.1038/nrn1201>
- Barner, D., Alvarez, G., Sullivan, J., Brooks, N., Srinivasan, M., & Frank, M. C. (2016). Learning mathematics in a visuospatial format: A randomized, controlled trial of mental abacus instruction. *Child Development*, 87(4), 1146–1158. <https://doi.org/10.1111/cdev.12515>
- Baroody, A. J., & Dowker, A. (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise*. Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bellos, A. (2010). *Alex's adventures in Numberland*. London, United Kingdom: Bloomsbury.
- Bilalić, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition*, 108(3), 652–661. <https://doi.org/10.1016/j.cognition.2008.05.005>
- Blöte, A. W., Klein, A. S., & Beishuizen, M. (2000). Mental computation and conceptual understanding. *Learning and Instruction*, 10(3), 221–247. [https://doi.org/10.1016/S0959-4752\(99\)00028-6](https://doi.org/10.1016/S0959-4752(99)00028-6)
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135. <https://doi.org/10.1016/j.tree.2008.10.008>
- Brooks, N. B., Barner, D., Frank, M., & Goldin-Meadow, S. (2018). The role of gesture in supporting mental representations: The case of mental abacus arithmetic. *Cognitive Science*, 42(2), 554–575. <https://doi.org/10.1111/cogs.12527>
- Brown, P., Lupker, S. J., & Colombo, L. (1994). Interacting sources of information in word naming: A study of individual differences. *Journal of Experimental Psychology: Human Perception and Performance*, 20(3), 537–554. <https://doi.org/10.1037/0096-1523.20.3.537>
- Campbell, J. I., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, 130(2), 299–315. <https://doi.org/10.1037/0096-3445.130.2.299>
- Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, 28(1), 143–153. <https://doi.org/10.3758/BF03211582>
- Chen, C., & Stevenson, H. W. (1989). Homework: A cross-cultural examination. *Child Development*, 60(3), 551–561. <https://doi.org/10.2307/1130721>
- Chen, C., & Stevenson, H. W. (1995). Motivation and mathematics achievement: A comparative study of Asian-American, Caucasian-American, and East Asian high school students. *Child Development*, 66(4), 1215–1234. <https://doi.org/10.2307/1131808>
- Cheng, D., Ma, M., Hu, Y., & Zhou, X. (2021). Chinese kindergarteners skilled in mental abacus have advantages in spatial processing and attention. *Cognitive Development*, 58, Article 101046. <https://doi.org/10.1016/j.cogdev.2021.101046>
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44(1-2), 1–42. [https://doi.org/10.1016/0010-0277\(92\)90049-N](https://doi.org/10.1016/0010-0277(92)90049-N)
- Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. New York, NY, USA: Oxford University Press.
- Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 16(3), 626–641. <https://doi.org/10.1037/0096-1523.16.3.626>
- Dong, S., Wang, C., Xie, Y., Hu, Y., Weng, J., & Chen, F. (2016). The impact of abacus training on working memory and underlying neural correlates in young adults. *Neuroscience*, 332, 181–190. <https://doi.org/10.1016/j.neuroscience.2016.06.051>
- Donlan, C., & Wu, C. (2017). Procedural complexity underlies the efficiency advantage in abacus-based arithmetic development. *Cognitive Development*, 43, 14–24. <https://doi.org/10.1016/j.cogdev.2017.02.002>
- Dowker, A. (2005). *Individual differences in arithmetic: Implications for psychology, neuroscience and education*. Hove, United Kingdom: Psychology Press.

- Ecker, U. K., Lewandowsky, S., Oberauer, K., & Chee, A. E. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 170–189. <https://doi.org/10.1037/a0017891>
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. <https://doi.org/10.1016/j.tics.2004.05.002>
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116–124. <https://doi.org/10.3758/BF03195503>
- Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, *141*(1), 134–149. <https://doi.org/10.1037/a0024427>
- Fuson, K. C., & Briars, D. J. (1990). Using a base-ten blocks learning/teaching approach for first-and second-grade place-value and multidigit addition and subtraction. *Journal for Research in Mathematics Education*, *21*(3), 180–206. <https://doi.org/10.2307/749373>
- Ganor-Stern, D., Pinhas, M., & Tzelgov, J. (2009). Comparing two-digit numbers: The importance of being presented together. *Quarterly Journal of Experimental Psychology*, *62*(3), 444–452. <https://doi.org/10.1080/17470210802391631>
- Gebuis, T., Cohen Kadosh, R., de Haan, E., & Henik, A. (2009). Automatic quantity processing in 5-year olds and adults. *Cognitive Processing*, *10*(2), 133–142. <https://doi.org/10.1007/s10339-008-0219-x>
- Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA, USA: Harvard University Press.
- Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, *17*(7), 572–576. <https://doi.org/10.1111/j.1467-9280.2006.01746.x>
- Hanakawa, T., Honda, M., Okada, T., Fukuyama, H., & Shibasaki, H. (2003). Neural correlates underlying mental calculation in abacus experts: A functional magnetic resonance imaging study. *NeuroImage*, *19*(2), 296–307. [https://doi.org/10.1016/S1053-8119\(03\)00050-8](https://doi.org/10.1016/S1053-8119(03)00050-8)
- Harness, A., Jacot, L., Scherf, S., White, A., & Warnick, J. E. (2008). Sex differences in working memory. *Psychological Reports*, *103*(1), 214–218. <https://doi.org/10.2466/pr0.103.1.214-218>
- Hatano, G. (1982). Cognitive consequences of practice in culture specific procedural skills. *The Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, *4*(1), 15–18.
- Hatano, G., Miyake, Y., & Binks, M. G. (1977). Performance of expert abacus operators. *Cognition*, *5*(1), 47–55. [https://doi.org/10.1016/0010-0277\(77\)90016-6](https://doi.org/10.1016/0010-0277(77)90016-6)
- Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. *Cognition*, *15*(1-3), 95–110. [https://doi.org/10.1016/0010-0277\(83\)90035-5](https://doi.org/10.1016/0010-0277(83)90035-5)
- Hatta, T., Hirose, T., Ikeda, K., & Fukuhara, H. (1989). Digit memory of soroban experts: Evidence of utilization of mental imagery. *Applied Cognitive Psychology*, *3*(1), 23–33. <https://doi.org/10.1002/acp.2350030104>
- Hatta, T., & Miyazaki, M. (1989). Visual imagery processing in Japanese abacus experts. *Imagination, Cognition and Personality*, *9*(2), 91–102. <https://doi.org/10.2190/43JU-8CBU-1LTY-RY6W>
- Heirdsfield, A. M., & Cooper, T. J. (2004). Factors affecting the process of proficient mental addition and subtraction: Case studies of flexible and inflexible computers. *The Journal of Mathematical Behavior*, *23*(4), 443–463. <https://doi.org/10.1016/j.jmathb.2004.09.005>
- Hishitani, S. (1990). Imagery experts: How do expert abacus operators process imagery? *Applied Cognitive Psychology*, *4*(1), 33–46. <https://doi.org/10.1002/acp.2350040104>
- Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17–29. <https://doi.org/10.1016/j.jecp.2008.04.001>
- Ifrah, G. (2001). *The universal history of numbers*. New York, NY, USA: Wiley.
- Imbo, I., & LeFevre, J. A. (2009). Cultural differences in complex addition: Efficient Chinese versus adaptive Belgians and Canadians. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1465–1476. <https://doi.org/10.1037/a0017022>
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>
- Jia, X., Zhang, Y., Yao, Y., Chen, F., & Liang, P. (2021). Neural correlates of improved inductive reasoning ability in abacus-trained children: A resting state fMRI study. *PsyCh Journal*, *10*(4), 566–573. <https://doi.org/10.1002/pchj.439>

- Klein, A. S., Beishuizen, M., & Treffers, A. (1998). The empty number line in Dutch second grades: Realistic versus gradual program design. *Journal for Research in Mathematics Education*, 29(4), 443–464. <https://doi.org/10.2307/749861>
- Kliegl, R., Masson, M. E., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition*, 18(5), 655–681. <https://doi.org/10.1080/13506280902986058>
- Kojima, T. (1954). *The Japanese abacus: Its use and theory*. Rutland, VT, USA: Tuttle.
- Lee, Y.-s., Lu, M.-j., & Ko, H.-p. (2007). Effects of skill training on working memory capacity. *Learning and Instruction*, 17(3), 336–344. <https://doi.org/10.1016/j.learninstruc.2007.02.010>
- LeFevre, J. A., & Bisanz, J. (1986). A cognitive analysis of number-series problems: Sources of individual differences in performance. *Memory & Cognition*, 14(4), 287–298. <https://doi.org/10.3758/BF03202506>
- LeFevre, J. A., & Liu, J. (1997). The role of experience in numerical skill: Multiplication performance in adults from Canada and China. *Mathematical Cognition*, 3(1), 31–62. <https://doi.org/10.1080/135467997387470>
- 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), 216–230. <https://doi.org/10.1037/0278-7393.22.1.216>
- Lemaire, P., & Callies, S. (2009). Children’s strategies in complex arithmetic. *Journal of Experimental Child Psychology*, 103(1), 49–65. <https://doi.org/10.1016/j.jecp.2008.09.007>
- Lewandowsky, S., Oberauer, K., Yang, L. X., & Ecker, U. K. (2010). A working memory test battery for MATLAB. *Behavior Research Methods*, 42(2), 571–585. <https://doi.org/10.3758/BRM.42.2.571>
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 6, Article 1171. <https://doi.org/10.3389/fpsyg.2015.01171>
- Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology*, 39(2), 367–386. <https://doi.org/10.1037/h0080066>
- Mabuchi, K. (1993). *Gojūonzu no Hanashi* [in Japanese]. Taishūkan Shoten. [ISBN 4-469-22093-0].
- Macizo, P., & Herrera, A. (2010). Two-digit number comparison: Decade-unit and unit-decade produce the same compatibility effect with number words. *Canadian Journal of Experimental Psychology*, 64(1), 17–24. <https://doi.org/10.1037/a0015803>
- McIntosh, A., Nohda, N., Reys, B. J., & Reys, R. E. (1995). Mental computation performance in Australia, Japan and the United States. *Educational Studies in Mathematics*, 29(3), 237–258. <https://doi.org/10.1007/BF01274093>
- Menninger, K. (2013). *Number words and number symbols: A cultural history of numbers*. (P. Broneer, Trans.). New York, NY, USA: Dover Publications.
- Miura, I. T. (1987). Mathematics achievement as a function of language. *Journal of Educational Psychology*, 79(1), 79–82. <https://doi.org/10.1037/0022-0663.79.1.79>
- Miura, I. T., & Okamoto, Y. (1989). Comparisons of US and Japanese first graders’ cognitive representation of number and understanding of place value. *Journal of Educational Psychology*, 81(1), 109–114. <https://doi.org/10.1037/0022-0663.81.1.109>
- Miura, I. T., Okamoto, Y., Kim, C. C., Chang, C. M., Steere, M., & Fayol, M. (1994). Comparisons of children’s cognitive representation of number: China, France, Japan, Korea, Sweden, and the United States. *International Journal of Behavioral Development*, 17(3), 401–411. <https://doi.org/10.1177/016502549401700301>
- Miura, I. T., Okamoto, Y., Kim, C. C., Steere, M., & Fayol, M. (1993). First graders’ cognitive representation of number and understanding of place value: Cross-national comparisons: France, Japan, Korea, Sweden, and the United States. *Journal of Educational Psychology*, 85(1), 24–30. <https://doi.org/10.1037/0022-0663.85.1.24>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Moeller, K., Huber, S., Nuerk, H. C., & Willmes, K. (2011). Two-digit number processing: Holistic, decomposed or hybrid? A computational modelling approach. *Psychological Research*, 75(4), 290–306. <https://doi.org/10.1007/s00426-010-0307-2>
- Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H. C. (2011). Early place-value understanding as a precursor for later arithmetic performance—A longitudinal study on numerical development. *Research in Developmental Disabilities*, 32(5), 1837–1851. <https://doi.org/10.1016/j.ridd.2011.03.012>
- Mou, Y., & vanMarle, K. (2014). Two core systems of numerical representation in infants. *Developmental Review*, 34(1), 1–25. <https://doi.org/10.1016/j.dr.2013.11.001>

- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519–1520. <https://doi.org/10.1038/2151519a0>
- Nisbett, R. E. (2003). *The geography of thought*. New York, NY, USA: The Free Press.
- Nuerk, H. C., Geppert, B. E., van Herten, M., & Willmes, K. (2002). On the impact of different number representations in the number bisection task. *Cortex*, *38*(5), 691–715. [https://doi.org/10.1016/S0010-9452\(08\)70038-8](https://doi.org/10.1016/S0010-9452(08)70038-8)
- Nuerk, H. C., Kaufmann, L., Zoppoth, S., & Willmes, K. (2004). On the development of the mental number line: More, less, or never holistic with increasing age? *Developmental Psychology*, *40*(6), 1199–1211. <https://doi.org/10.1037/0012-1649.40.6.1199>
- Nuerk, H. C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*, *82*(1), B25–B33. [https://doi.org/10.1016/S0010-0277\(01\)00142-1](https://doi.org/10.1016/S0010-0277(01)00142-1)
- Nuerk, H. C., Weger, U., & Willmes, K. (2004). On the perceptual generality of the unit-decade compatibility effect. *Experimental Psychology*, *51*(1), 72–79. <https://doi.org/10.1027/1618-3169.51.1.72>
- Nuerk, H. C., Weger, U., & Willmes, K. (2005). Language effects in magnitude comparison: Small, but not irrelevant. *Brain and Language*, *92*(3), 262–277. <https://doi.org/10.1016/j.bandl.2004.06.107>
- Nys, J., & Content, A. (2010). Complex mental arithmetic: The contribution of the number sense. *Canadian Journal of Experimental Psychology*, *64*(3), 215–220. <https://doi.org/10.1037/a0020767>
- OECD. (2016). *PISA 2015 Results (Volume I): Excellence and equity in education*. Paris, France: PISA, OECD Publishing. <https://doi.org/10.1787/9789264266490-en>
- OECD. (2019). *PISA 2018 Results (Volume I): What students know and can do*. Paris: PISA, OECD Publishing. <https://doi.org/10.1787/5f07c754-en>
- Pesenti, M. (2005). Calculation abilities in expert calculators. In Jamie I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 413–430). New York, NY, USA: Psychology Press.
- Piaget, J. (1952). *The child's conception of number*. London, United Kingdom: Routledge and Kegan Paul.
- Pickering, S. J. (2001). The development of visuo-spatial working memory. *Memory*, *9*(4–6), 423–432. <https://doi.org/10.1080/09658210143000182>
- Poltrock, S. E., & Schwartz, D. R. (1984). Comparative judgments of multidigit numbers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 32–45. <https://doi.org/10.1037/0278-7393.10.1.32>
- R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria.
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*(2), 359–379. <https://doi.org/10.1037/a0029082>
- Sarnecka, B. W., Goldman, M. C., & Slusser, E. B. (2015). How counting leads to children's first representations of exact, large numbers. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 45–66). Oxford, United Kingdom: Oxford University Press.
- Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*(5), 405–410. <https://doi.org/10.1111/1467-9280.00076>
- Shwalb, D., Sugie, S., & Yang, C. (2004). Motivation for abacus studies and school mathematics. In D. W. Shwalb, J. Nakazawa, & B. J. Shwalb (Eds.), *Applied developmental psychology in Japan* (pp. 109–135). Greenwich, CT, USA: Information Age Publishing.
- Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain and Education*, *3*(3), 143–150. <https://doi.org/10.1111/j.1751-228X.2009.01064.x>
- Smith, S. B. (1983). *The great mental calculators: The psychology, methods, and lives of calculating prodigies; past and present*. New York, NY, USA: Columbia University Press.
- Sowder, J. T. (1990). Mental computation and number sense. *The Arithmetic Teacher*, *37*(7), 18–20. <https://doi.org/10.5951/AT.37.7.0018>
- Srinivasan, M., Wagner, K., Frank, M. C., & Barner, D. (2018). The role of design and training in artifact expertise: The case of the abacus and visual attention. *Cognitive Science*, *42*(S3), 757–782. <https://doi.org/10.1111/cogs.12611>
- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*(1), 32–71. <https://doi.org/10.2307/747348>
- Statistics Bureau of Japan. (2015, May 31). *Statistical handbook of Japan*. Retrieved from <http://www.stat.go.jp/english/data/handbook/index.htm>

- Stigler, J. W. (1984). "Mental abacus": The effect of abacus training on Chinese children's mental calculation. *Cognitive Psychology*, 16(2), 145–176. [https://doi.org/10.1016/0010-0285\(84\)90006-9](https://doi.org/10.1016/0010-0285(84)90006-9)
- Stigler, J. W., Chalip, L., & Miller, K. F. (1986). Consequences of skill: The case of abacus training in Taiwan. *American Journal of Education*, 94(4), 447–479. <https://doi.org/10.1086/443862>
- Stigler, J. W., & Perry, M. (1988). Mathematics learning in Japanese, Chinese, and American classrooms. *New Directions for Child Development*, 1988(41), 27–54. <https://doi.org/10.1002/cd.23219884104>
- Takahashi, A., Watanabe, T., & Yoshida, M. (2008). *English translation of the Japanese mathematics curricula in the course of study*. Madison, NJ, USA: Global Education Resources.
- Thomson, S., Wernert, N., O'Grady, E., & Rodrigues, S. (2016). *TIMSS 2015: A first look at Australia's results*. Camberwell, Australia: Australian Council for Educational Research (ACER).
- Torbeyns, J., & Verschaffel, L. (2013). Efficient and flexible strategy use on multi-digit sums: A choice/no-choice study. *Research in Mathematics Education*, 15(2), 129–140. <https://doi.org/10.1080/14794802.2013.797745>
- Tzelgov, J., Ganor-Stern, D., Kallai, A. Y., & Pinhas, M. (2015). Primitives and non-primitives of numerical representations. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 45–66). Oxford, United Kingdom: Oxford University Press.
- Venables, W., & Ripley, B. (2002). *Modern applied statistics using S*. New York, NY, USA: Springer.
- Wang, C. (2020). A review of the effects of abacus training on cognitive functions and neural systems in humans. *Frontiers in Neuroscience*, 14, Article 913. <https://doi.org/10.3389/fnins.2020.00913>
- Wang, C., Geng, F., Yao, Y., Weng, J., Hu, Y., & Chen, F. (2015). Abacus training affects math and task switching abilities and modulates their relationships in Chinese children. *PLoS One*, 10(10), Article e0139930. <https://doi.org/10.1371/journal.pone.0139930>
- Whang, P. A., & Hancock, G. R. (1994). Motivation and mathematics achievement: Comparisons between Asian-American and non-Asian students. *Contemporary Educational Psychology*, 19(3), 302–322. <https://doi.org/10.1006/ceps.1994.1023>
- Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1), 53–79. <https://doi.org/10.1037/a0024177>
- Zebian, S., & Ansari, D. (2012). Differences between literates and illiterates on symbolic but not nonsymbolic numerical magnitude processing. *Psychonomic Bulletin & Review*, 19(1), 93–100. <https://doi.org/10.3758/s13423-011-0175-9>
- Zhang, H., & Zhou, Y. (2003). The teaching of mathematics in Chinese elementary schools. *International Journal of Psychology*, 38(5), 286–298. <https://doi.org/10.1080/00207590344000097>

## Appendices

### Appendix A

All GLMMs statistically controlled for the following trial-level variables on RT and accuracy: trial number, RT of the previous response, and accuracy of the previous response. Whether the previous trial was within- or between-decades was included in the GLMMs for Experiment 2. The position of the largest number, and whether the medial number was above or below the midpoint, were also included as trial-level variables in the GLMMs for Experiment 3. Carry status was also included in the GLMMs for Experiment 4. The effects on RT and accuracy by those trial-level variables were as follows.

The GLMMs for the magnitude judgement task of Experiment 2 revealed significantly slower RT and lower accuracy following an error ( $t = 15.52, p < .001; z = 2.24, p = .03$ ), as the number of trials increased ( $t = 5.28, p < .001; z = 2.95, p = .003$ ), or when the previous trial was within-decades ( $t = -3.22, p = .001; z = -2.80, p = .005$ ). Significantly faster RT and lower accuracy was also observed following shorter previous responses ( $t = 20.14, p < .001; z = -4.01, p < .001$ ).

The GLMMs for the number bisection task of Experiment 3 revealed that for both correctly and incorrectly bisected problems, RT was significantly faster as the number of trials increased ( $t = -3.64, p < .001; t = -1.92, p = .05$ , respectively), or for shorter previous responses ( $t = 7.31, p < .001; t = 6.26, p < .001$ , respectively). Neither predictor significantly influenced accuracy for either problem type ( $|z| < 1.88, p > .06$ ), nor did the spatial position of the larger number on the left or right side of the screen significantly influence RT ( $t = 1.26, p = .21$ ) or accuracy ( $z = 1.07, p = .28$ ) for the incorrectly bisected problems. However, RT was significantly faster when the largest number was on the right for the correctly bisected problems ( $t = 3.18, p = .001; z = -0.27, p = .79$ ). RT for the incorrectly bisected problems was significantly faster ( $t = 2.91, p = .004$ ) and accuracy significantly higher ( $z = 2.27, p = .02$ ) when the medial number was below rather than above the arithmetic midpoint. However, this effect was modulated by spatial position, being



significantly more pronounced in both RT and accuracy when the larger number appeared on the right than left ( $t = -2.41, p = .02; z = -1.99, p = .05$ ).

The GLMMs for the mental addition task of Experiment 4 revealed that for both the Same and Different Total problems, RT was significantly faster for the No Carry than Unit Carry problems ( $t = 11.77, p < .001; t = 2.27, p = .02$ , respectively), or when the previous response was faster ( $t = 24.93, p < .001; t = 9.11, p < .001$ , respectively). Trial number and preceding accuracy also influenced performance for the Same Total problems, with RT significantly decreasing with an increasing number of trials ( $t = -10.02, p < .001$ ), and errors significantly less likely when the previous response was correct ( $z = 4.60, p < .001$ ). Neither carry status ( $z = 0.62, p = .54$ ), trial number ( $z = -0.83, p = .41$ ), speed ( $z = 1.31, p = .19$ ) nor accuracy ( $z = 1.42, p = .16$ ) of participants' previous response significantly correlated with accuracy for the Different Total problems.

## Appendix B: Mean Stimulus Attributes for the Problems Shown in This Research

**Table B1**

*Mean Stimulus Attributes for the Outer Numbers From the Number Bisection Task of Experiment 3*

Attribute	Congruent		Incongruent	
	Compatible	Incompatible	Compatible	Incompatible
Absolute Overall Distance	34.00	34.13	33.88	34.00
Absolute Decade Distance	2.88	3.88	2.88	3.88
Absolute Unit Distance	5.25	4.63	5.13	4.75
Problem Size	58.63	58.06	58.69	58.44

*Note.* Decade distance is necessarily larger for the incompatible than compatible pairs when matched on overall distance because they cross a (another) decade boundary (see Nuerk, Weger, & Willmes, 2004 for a detailed discussion).

**Table B2**

*Example Stimuli in the Number Bisection Task of Experiment 3 According to Part-Whole Congruency, Unit-Decade Compatibility and Bisector Type*

Condition	Correctly Bisectioned	Incorrectly Bisectioned	
		Close	Far
<b>Part-Whole Congruent</b>			
Compatible	45_67_89	45_68_89	45_74_89
Incompatible	47_65_83	47_64_83	47_71_83
<b>Part-Whole Incongruent</b>			
Compatible	42_58_74	42_53_74	42_63_74
Incompatible	45_58_71	45_53_71	45_63_71

**Table B3**

*Mean Values for Each Stimulus Attribute According to Condition for the Same Total Problems of Experiment 4*

Attribute	Identical	Legally			
		Transposed	Decomposition	Sequential	Same Unrelated
Operand Average	42.28	42.28	42.28	42.28	42.28
Reference/Target Operand Average	42.28	42.28	42.28	42.28	42.28
Reference/Target Sum	84.55	84.55	84.55	84.55	84.55
Absolute Difference Between Sums	0	0	0	0	0
Number of Shared Digits	4	4	0.72	2	1.02
Number of Digits in Shared Positions	4	2	0.23	2	0.19

*Note.* The values were averaged over Carry Status.

**Table B4**

*Mean Values for Each Stimulus Attribute According to Condition for the Different Total Problems of Experiment 4*

<b>Attribute</b>	<b>Illegally Transposed</b>	<b>Close</b>	<b>Same Unit Total</b>	<b>Different Unrelated</b>
Operand Average	44.70	42.83	41.05	43.74
Reference Operand Average	39.30	43.00	42.68	42.55
Reference Sum	78.60	86.00	85.35	85.10
Target Operand Average	50.10	42.65	39.43	44.93
Target Sum	100.20	85.30	78.85	89.85
Absolute Difference Between Sums	26.55	1	25.50	26.35
Number of Shared Digits	4	1.05	1.05	0
Number of Digits in Shared Positions	2	0.50	0.45	0

*Note.* The values were averaged over Carry Status.



*Journal of Numerical Cognition* (JNC) is an official journal of the Mathematical Cognition and Learning Society (MCLS).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.