Title:

Verbal and visual, STM predicts performance in a multiplication-production task: Evidence from a Malaysian sample

Authors: **Soh** Mei Ling[1], **Garcia Orza** Javier[2], **Mennie** Neil Russel[1], **Estudillo** Alejandro J.[3]

Corresponding author: Soh Mei Ling

Corresponding email: hpxms1@nottingham.edu.my

[1]School of Psychology, University of Nottingham – Malaysia

[2]Department of Basic Psychology, University of Malaga, Spain

[3]Department of Psychology, Bournemouth University, UK

Abstract

Arithmetic requires the use of multiple cognitive processes, such as short-term memory (STM). However, findings on the association between STM and simple multiplication solving are mixed, potentially due to large interindividual differences in multiplication proficiency within and between samples. The present study aims to explore further the relationship between visual and verbal STM and simple multiplication solving with a large Malaysian sample (N = 230). Adults (age = 17-42) completed an online production-based multiplication-solving task, short-term memory measures (verbal and visuospatial STM tasks), and a demographic survey. A mixed model analysis found that verbal STM and visual STM predict multiplication performance, with lower-span participants having longer reaction times during multiplication solving. Interestingly, we also observed the relationship between verbal STM and multiplication was moderated by interference, the impact of verbal STM was stronger in high interference problems, while the visual STM-multiplication relation was moderated by problem size, high visual span participants took more advantage of their visual STM when presented with large size problems. Thus, our findings show that both verbal and visual STM in interaction with problem properties predicts simple multiplication solving in adults. Hypotheses on the concrete mechanisms involved in these relationships are discussed.

Keywords: short-term memory, multiplication fluency, verbal and visual

Public significance statement: Individual differences in solving single-digit multiplications are substantial and remain unexplained. The involvement of verbal and visual short-term memory in solving single-digit multiplications is explored here. We found that a complex relationship exists between visual and verbal short-term memory and the characteristics of the problem in terms of size and interference.

Author note. The data and codes for analysis are available at the Github.

Acquiring basic multiplication facts (e.g., 5 × 7) is a fundamental milestone in children's mathematical learning. Difficulties in reaching fluency with multiplications have been the focus of numerous studies in children. Interestingly, even among literate adults without documented learning difficulties, there are substantial differences in multiplication fluency (Hecht, 1999); however, the specific factors for these variations have received limited attention in research. The present study focuses on the origin of these individual differences in adults.

The individual-differences literature suggests that performance in solving single-digit multiplications can be attributed to individuals' and problems' properties. Among individual properties, the focus has been on general cognitive abilities such as processing speed, phonological awareness, attention, executive functions, and memory (Agostini et al., 2022). In the present study, we focused on verbal and visual short-term memory (STM) – the modal systems responsible for storing phonological and visual information and are components of the working memory system (e.g., Baddeley, 2001) –which has been previously implicated in multiplication solving (van der Ven et al., 2013). As for problem properties, we focused on two key variables consistently shown to affect multiplication-solving: problem size and interference. While the former shows that small problems (e.g., 2 × 3) are solved more efficiently than large problems (e.g., 9 × 8) (see Zbrodoff & Logan, 2005), interference reflects the similarity between the problem and its solution, and other problems and their solutions (De Visscher et al., 2015). In this sense, problems with low interference (e.g., 2 × 6) are more readily stored and retrieved from long-term memory than high interference problems (e.g., 8 × 3) (De Visscher et al., 2015; De Visscher & Noël, 2014).

STM and single-digit multiplications

Studies involving children with mathematical learning disabilities (MD) suggest that short-term memory (STM) plays a role in mathematical competence. For instance, Andersson (2010) observed a positive correlation between performance in the Digit Span Task and accuracy and speed in arithmetic facts retrieval among children. Furthermore, the same study revealed that participants with MD performed more poorly than control participants in the visual matrix span task. More recently, Szucs et al. (2013) reported that children with mathematical learning disabilities performed worse than controls in the Dot Matrix task, a measure of visuo-spatial STM.

The previously reviewed studies investigated the connection between STM and arithmetic skills globally (e.g., Coolen & Castronovo, 2023; Soltanlou et al., 2015), but fewer studies have explored the specific link between STM and single-digit multiplication fluency, particularly in the adult population. For instance, Lee and Kang (2002) conducted a small study with Korean students and observed that reaction times for solving simple multiplications increased only when simultaneous phonological (but not visual) suppression was applied. However, it is important to note that Lee and Kang's study was very limited in sample size (n = 10), and multiplication fluency was assessed by having participants type the solutions on a keyboard without considering typing speed as a potential confounding factor. Interestingly, a different pattern emerged among non-Asian communities in European samples; de Rammelaere et al. (2001) and Seitz and Schumann-Hengsteler (2000) failed to find the effects of articulatory suppression. These results open the question of the relevance of culture and educational systems in shaping the multiplication process. Findings in Korean suggest the prominent role of verbal cues in multiplication fact retrieval among this group.

To make things more complex, Cavdaroglu and Knops (2016) in German students found that both phonological and visuospatial working memory load suppressions influenced multiplication performance, even after controlling for task difficulties. However, it is worth considering that the use of a multiplication verification task with two options in this study, compared to the use of a verification task with just one in the studies by de Rammelaere et al. (2001) and Seitz and Schumann-Hengsteler (2000), may favour approximate calculation procedures that rely more on visuospatial (Andersson, 2010) and verbal strategies.

In conclusion, studies involving MD suggest a potential role for STM in mathematical competence. However, as these studies measure mathematical competence by combining performance in the four basic operations, they only provide indirect evidence for the role of STM in multiplication. Additionally, when exploring this issue with adults without MD, researchers have used various cultural groups and tasks to assess STM links to multiplications of different sizes, making direct comparisons challenging. Therefore, further empirical investigation is necessary to determine whether individual differences in STM are associated with simple multiplication performance.

The role of problem size and interference in multiplication solving

In single-digit multiplication solving, problem properties – problem size and interference – are key performance factors. The problem size effect is marked by increased errors and response times as the numerical size of the operands grows (Ashcraft & Christy, 1995). While its exact origin is debated, it seems to be influenced by the frequency of occurrence and the problem-solving procedures employed (Zbrodoff & Logan, 2005). Smaller problems, typically those with solutions under 25, are often retrieved automatically. In contrast, larger problems, with solutions exceeding 25, are less common and tend to be addressed using procedural methods (Ashcraft & Christy, 1995; Verguts & Fias, 2005). Problem size can be linked to educational practices. For example, while Chinese participants exhibited a stronger problem-size effect in brain regions associated with phonological processing, American participants showed a significant problem-size effect in areas linked to calculation procedures (LeFevre & Liu, 1997; Prado et al., 2013)

Concerning problem interference, research has consistently shown that errors and response times in single-digit multiplication increase as the similarity between the problem and its solution and other problems rises (De Visscher et al., 2015; De Visscher & Noël, 2014; Verguts & Fias, 2005). Problems with low interference appear more readily stored and retrieved from long-term memory than their high-interference counterparts.

Within the previously reviewed literature exploring the relationship between multiplication and short-term memory (STM), it is noteworthy that relatively few studies have considered the specific properties of the multiplication problems used to gauge participants' variability in multiplication fluency. This oversight could potentially introduce variability and limit the power of the studies.

Current study

All in all, the role of STM in simple multiplication solving is unclear. Thus, the present study focuses on the nuanced differentiation between verbal and visual STM and aims to scrutinize their distinct impacts on individual differences in multiplication fluency across the broader adult population. Specifically, we seek to ascertain whether verbal and visual STM performance predicts performance in solving single-digit multiplication. Furthermore, our interest extends to understanding how these STM capabilities interact with multiplication problem properties, such as problem size and interference. We employed a digit span task to evaluate verbal STM and a visual pattern task to gauge visual STM. Participants were tasked with completing a production-based multiplication task to measure multiplication fluency. Unlike verification tasks, this approach provides a more naturalistic assessment of fluency in multiplication solving. Moreover, by presenting multiplication problems with different properties, including varying levels of interference and operand sizes, our task offers a comprehensive assessment, providing a complete picture of not only the effect of STM and multiplication solving but also the potential interactive effects of problem size and problem interference.

Based on past findings, we can formulate two hypotheses. First, if we postulate that automatic retrieval operates independently of STM stores (e.g., de Rammelaere et al., 2001; Simmons et al., 2012; Soltanlou et al., 2015), then we anticipate no significant relationship between STM and multiplication-solving for problems that are retrieved automatically (i.e., small-size problems). On the other hand, if we posit that even automatic retrieval demands verbal STM, then we expect an STM-multiplication relationship in both small-size and large-size problems. Furthermore, given that larger and/or high interference problems might involve the use of strategies: the connection between visual and verbal-STM and multiplication solving appears only for those problems.

Method

Participants

A total of 241 participants were recruited through various social media channels and student association newsletters in Malaysia. Eleven participants were excluded from the analysis due to incomplete performance or low accuracy in the typing speed task (accuracy < 50%). The final sample comprised 230 Malaysians (77.92% females) aged 18 to 42 years (M = 22.78, SD = 3.47). Participants reported no history of neurological or developmental disorders, demonstrated proficiency in English, and had normal-to-corrected vision. The study was conducted in accordance with the ethical guidelines of the [masked_university]

Materials and Procedures

The four experimental tasks were programmed using PsychoPy (Peirce et al., 2019; v.2020.2.10) and hosted online on Pavlovia (URL: pavlovia.org). The stimuli in our tasks were presented using a height-based formatting approach for consistency and adaptability on different devices with varying screen resolutions. Experiments ran in full-screen mode. Participants were asked to complete a Typing Speed Assessment, a Single-Digit Multiplication Production Task, a Digit Span Task evaluating verbal short-term memory, and a Visual Pattern Task assessing visual short-term memory. The initial Typing Speed Assessment was consistently presented first, but the order of the subsequent tasks was counterbalanced across participants to mitigate potential order effects. Before the experimental trials, participants completed three practice trials for each task, with feedback provided to facilitate task familiarization. Those stimuli used during practice trials were not presented again during the actual trials. The study concluded with participants filling out a demographic information questionnaire on the Qualtrics platform.

Typing Speed Assessment. Participants were presented with 28 Arabic numbers (see Appendix A for details); the numbers served as answers to multiplication problems ranging from 2 × 3 to 8 × 9. The stimuli were presented randomly, and the presentation sequence involved a 500ms blank screen followed immediately by the Arabic number stimuli. No time limit was imposed, but participants were instructed to type the digits accurately and rapidly. The participants pressed the 'Enter' key to proceed to the next trial. Reaction times for all trials (i.e., time from presentation of stimuli till the 'Enter' key was pressed) were recorded, and a median typing speed score for correct trials was computed for each participant.

Multiplication Production Task. Sixty single-digit multiplication problems, ranging from 2 × 3 to 8 × 9, were carefully pseudo-randomized into two blocks, each containing 30 problems, to ensure that participants did not encounter three consecutive problems from the same multiplication table during the task. Each multiplication problem appeared twice, once with the smaller digit first and once with the larger digit first, except for tie problems presented once. These problems were: 28 small-sized problems (magnitude < 26), 28 large-sized problems (magnitude > 25), and 4 tie problems (see Appendix B for the details on the stimuli). Participants were encouraged to respond as quickly and accurately and no time limit was imposed. The correct response and the reaction time for each trial (from the presentation of the problem till the pressing of the 'Enter' key) were registered and reflected participants' performance in single-digit multiplication.

Verbal STM (Digit Span Task). We employed two sets of digit stimuli: one derived from the Wechsler Adult Intelligence Scale (WAIS) and the other generated by an online algorithm (see Appendix C for details)[1]. A female voice generated by the Balabolka app (Kutasov & Morozov, 2021; v2.15.0.767) using Microsoft Speech API (SAPI) pronounced the digit lists. Participants were presented with digit sequences and had to repeat them in the same order using the keyboard immediately. The length of digit sequences increased from three to nine items. Participants had two chances at each span length, and if they answered correctly at least one out of two trials, they

[1] A significant positive correlation was observed between the two sets of digit span stimuli (r = 0.620, $p < 0.01$)

proceeded to the next length. The digit span score was computed as the average of the two sets of digit spans, representing the participant's verbal short-term memory capacity.

Visual Pattern Task. A computerized version of the Visual Patterns Test (Della Sala et al., 1999) was used. Participants were presented with a matrix pattern containing 2.5 × 2.5 cm black and white squares which was displayed at the center of the screen for a duration of 2000 milliseconds, followed by a 1000-millisecond blank screen interval. Subsequently, an empty matrix consisting solely of white squares was presented to participants (refer to Appendix D for details). Participants were asked to recall and replicate the original pattern by clicking on the empty matrix. The test included 14 levels of increasing difficulty based on the number of black squares within the matrix, ranging from 2 to 15 black squares per span length. At each span length, participants completed three trials. If they achieved success in at least two out of the three trials, they advanced to the next level. If not, the task terminated. Visual memory span was computed as the average number of black squares correctly recalled by participants in the last three trials.

Results

Data Analysis Plan

Our study utilized mixed-effects models with reaction time (RT) and accuracy as dependent variables, focusing on the relationship between multiplication performance and participant factors (verbal and visual STM scores) and problem properties. We standardized predictors and employed a robust model-fitting approach. Statistical analyses were conducted in R (version 4.3.1) using the lme4 package. Post-hoc analyses were performed to examine significant interactions. For comprehensive data and analysis details, please refer to our supplementary document or GitHub (removed_link).

Results

Descriptive Statistics

Descriptive statistics and group differences across behavioral tasks are presented (see Table 1,2).

Relationships between the tasks

Full correlation analyses on the relationships between the cognitive tasks are available (see **Error! Reference source not found.**). We also ran partial correlations to assess the relationships between verbal and visual STM and multiplication performance while controlling for the influence of typing speed. Verbal (r = -0.234, p <.001) and visual STM (r = -0.168, p = 0.011) still significantly correlated to the multiplication performance (RT) even after controlling for the influence of typing speed.

Predictors of RT in Multiplication Task

To identify the predictors of reaction time (RT) in the multiplication production task, we executed stepwise regression analyses and used ANOVA to assess the relative goodness-of-fit between two linear mixed models. We also used likelihood-based criteria like the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) our model fitness (see Table 4). Our analysis began by considering problem features, specifically problem size, and interference score, owing to their theoretical relevance in numerical cognition. Subsequently, we introduced verbal short-term memory (STM) and visual STM into the models. We ran ANOVA analysis to compare the fitness of each model with previous models. For more detailed information, refer to Tables 1 and 2 of the Supplementary Document.

Including two-levels interactions

We aimed to investigate the potential moderating effects of problem features, specifically problem size and interference score, on the relationships between STM components and multiplication time. We systematically examined all possible two-way interactions among predictors (Models 5-10, Table 1 of Supplementary Document).

For a comprehensive understanding, we also extended our analytical work by including all two-level interactions into a unified model (Model 10) and compared its goodness-of-fit with that of Model 9. Importantly, Model 10, including all two-level interactions, performed significantly worse than Model 9 in model fit ($x^2(1) = 0.523$, $p = 0.77$). A full description of Model 9 is presented in Table 1. Simple slope analyses to clarify the interactions found is presented in Figures 1 and 2.

Discussion

The present study aimed to investigate the role of short-term memory (verbal and visual) in a multiplication production task while considering multiplication problem attributes such as problem size and problem interference. Despite the considerable research on arithmetic and working memory, surprisingly, there is a notable gap in adults' interindividual differences in single-digit multiplication solving while accounting for problem characteristics and using production tasks with non-European population.

With a large sample of participants and the use of hierarchical linear mixed-effects models, the results were clear (see Table 1). Confirming prior research (see respectively, Zbrodoff & Logan, 2005 for review, and De Visscher & Nöel, 2014), we observed positive linear effects for problem size and interference. Furthermore, we observed a significant negative association between verbal STM and multiplication performance (absence of significance in the case of visual STM). Interestingly, we identified a negative interaction between multiplication interference scores and verbal STM capacity: our analyses showed that the detrimental effect of problem interference was stronger for participants with low verbal span. Similarly, presenting larger problems to participants with higher visual span had less impact than presenting them to participants with low span. In sum, interference moderated the verbal-multiplication relation, but problem size moderated the visual-multiplication relation, demonstrating that cognitive resources like verbal and visual STM are more influential in certain problem-solving contexts.

Whereas it is clear that both STM components are related with solving single-digit multiplications, the precise mechanisms underlying these relationships remain elusive. It could be that verbal and visual STM play a role in the automatic retrieval of multiplications (Lee & Kang, 2002; Cadvaroglu & Knops, 2016). As suggested by the triple-code model (Dehaene, 1992), multiplication facts are represented in phonological format, and retrieval demands individuals to engage in the internal rehearsal of number words. Additionally, as interference is related to the phonological similarity (De Visscher & Noël, 2014), it seems that when presented with high interference problems, superior verbal STM capabilities would allow for a more efficient retrieval process. In the case of the visual STM, it can be that, as suggested by Campbell's network interference model (Campbell, 1994), the magnitude system – representation of numerical magnitudes on a spatially oriented number line – helps in activating the correct problem solution. When faced with large problems, individuals with less precise phonological representations may need to rely more on his/her visual representation to retrieve the correct solution from long-term memory. Under such circumstances, individuals with better visual STM may activate the magnitude representations better than those with poor magnitude representation.

An alternative perspective, which remains consistent with the existing framework, suggests that it's essential to reconsider the role of verbal and visual Short-Term Memory (STM) when automatic retrieval is not a feasible option. This typically occurs in situations where high interference or complex problem-solving tasks are involved, and individuals may need to resort to alternative strategies, such as calculation or counting. The application of these strategies often requires the temporary retention of specific pieces of information, such as partial results, in either verbal or visual STM.

For instance, when using calculation procedures during multiplication solving, strategies like retrieving the previous numerical value and performing addition may be employed. As proposed by the triple-code model (Dehaene, 1992), this process necessitates a mental journey along the number line, which in turn involves creating a temporary visual representation that gets stored in our visual STM.

An unexplored possibility we need to consider is that the observed relationship may not be directly tied to the retrieval process itself but rather to the learning process. When individuals possess a higher span during the learning phase, they tend to develop robust mental representations that can be more easily retrieved even at a later stage (Suárez-Pellicioni et al., 2019). Since span is a relatively stable individual characteristic, it tends to correlate with multiplication retrieval abilities in adults.

It's important to note that our data, which reveals an association between verbal STM and multiplication, doesn't definitively distinguish between the various hypotheses. Regardless of the specific mechanisms at play, these hypotheses both imply a connection between multiplication learning and visual and verbal STM during different phases, including the encoding and retrieval stages. STM plays a facilitating role in temporarily storing, manipulating, and rehearsing these facts during the learning phase. Once these facts become strongly associated with long-term memory, the role of STM in the actual retrieval process may diminish, creating the impression that STM is not crucial for the retrieval phase. While admittedly speculative, it's evident that the interplay between these hypotheses underscores the complexity of the cognitive processes involved in solving multiplication problems.

Several limitations in this study warrant discussion. To enhance the interpretation of our results, it would have been valuable to gather information about the strategies employed by participants in each problem through self-reports. Additionally, it's essential to bear in mind that our correlational methodology does not permit us to pinpoint causal mechanisms. Utilizing experimental manipulation techniques, such as phonological suppression paradigms, could help us identify the role of STM in multiplication problem-solving. However, it's crucial to exercise caution in these paradigms and control for problem properties, as they can introduce different cognitive demands.

Conclusion

This study has illuminated the complex interplay between STM components and problem features in the context of multiplication performance. While it's essential for future studies to validate our findings and explore the alternative hypotheses we have put forth, the results emphasize the domain-specific nature of cognitive processes at play in arithmetic.

References

Agostini, F., Zoccolotti, P., & Casagrande, M. (2022). Domain-general cognitive skills in children with mathematical difficulties and dyscalculia: A systematic review of the literature. *Brain Sciences*, *12*(2), 239. https://doi.org/10.3390/brainsci12020239

Andersson, U. (2010). Skill development in different components of arithmetic and basic cognitive functions: Findings from a 3-year longitudinal study of children with different types of learning difficulties. *Journal of Educational Psychology*, *102*(1), 115–134. https://doi.org/10.1037/a0016838

Ashcraft, M. H., & Christy, K. S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in grades 1-6. *Journal for Research in Mathematics Education*, *26*(5), 396–421. https://doi.org/10.2307/749430

Baddeley, A. D. (2001). Is working memory still working? *The American Psychologist*, *56*(11), 851–864. https://doi.org/10.1037/0003-066x.56.11.851

Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition*, *53*(1), 1–44. https://doi.org/10.1016/0010-0277(94)90075-2

Cavdaroglu, S., & Knops, A. (2016). Mental subtraction and multiplication recruit both phonological and visuospatial resources: Evidence from a symmetric dual-task design. *Psychological Research*, *80*(4), 608–624. https://doi.org/10.1007/s00426-015-0667-8

Coolen, I. E. J. I., & Castronovo, J. (2023). How memory counts in mathematical development. *Journal of Cognition*, *6*(1), 1. https://doi.org/10.5334/joc.248

de Rammelaere, S., Stuyven, E., & Vandierendonck, A. (2001). Verifying simple arithmetic sums and products: Are the phonological loop and the central executive involved? *Memory & Cognition*, *29*(2), 267–273. https://doi.org/10.3758/BF03194920

De Visscher, A., Berens, S. C., Keidel, J. L., Noël, M.-P., & Bird, C. M. (2015). The interference effect in arithmetic fact solving: An fMRI study. *NeuroImage*, *116*, 92–101. https://doi.org/10.1016/j.neuroimage.2015.04.063

De Visscher, A., & Noël, M.-P. (2014). The detrimental effect of interference in multiplication facts storing: Typical development and individual differences. *Journal of Experimental Psychology: General*, *143*(6), 2380–2400. https://doi.org/10.1037/xge0000029

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1), 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo–spatial memory. *Neuropsychologia*, *37*(10), 1189–1199. https://doi.org/10.1016/S0028-3932(98)00159-6

Hecht, S. A. (1999). Individual solution processes while solving addition and multiplication math facts in adults. *Memory & Cognition*, *27*(6), 1097–1107. https://doi.org/10.3758/BF03201239

Lee, K.-M., & Kang, S.-Y. (2002). Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition*, *83*(3), B63–B68. https://doi.org/10.1016/S0010-0277(02)00010-0

LeFevre, J.-A., & Liu, J. (1997). The role of experience in numerical skill: Multiplication performance in adults from Canada and China. *Mathematical Cognition*. https://doi.org/10.1080/135467997387470

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Prado, J., Lu, J., Liu, L., Dong, Q., Zhou, X., & Booth, J. (2013). The neural bases of the multiplication problem-size effect across countries. *Frontiers in Human Neuroscience*, *7*. https://www.frontiersin.org/articles/10.3389/fnhum.2013.00189

Seitz, K., & Schumann-Hengsteler, R. (2000). Mental multiplication and working memory. *European Journal of Cognitive Psychology*, *12*(4), 552–570. https://doi.org/10.1080/095414400750050231

Simmons, F. R., Willis, C., & Adams, A.-M. (2012). Different components of working memory have different relationships with different mathematical skills. *Journal of Experimental Child Psychology*, *111*(2), 139–155. https://doi.org/10.1016/j.jecp.2011.08.011

Soltanlou, M., Pixner, S., & Nuerk, H.-C. (2015). Contribution of working memory in multiplication fact network in children may shift from verbal to visuo-spatial: A longitudinal investigation. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/article/10.3389/fpsyg.2015.01062

Suárez-Pellicioni, M., Fuchs, L., & Booth, J. R. (2019). Temporo-frontal activation during phonological processing predicts gains in arithmetic facts in young children. *Developmental Cognitive Neuroscience*, *40*, 100735. https://doi.org/10.1016/j.dcn.2019.100735

Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex*, *49*(10), 2674–2688. https://doi.org/10.1016/j.cortex.2013.06.007

van der Ven, S. H. G., van der Maas, H. L. J., Straatemeier, M., & Jansen, B. R. J. (2013). Visuospatial working memory and mathematical ability at different ages throughout primary school. *Learning and Individual Differences*, *27*, 182–192. https://doi.org/10.1016/j.lindif.2013.09.003

Verguts, T., & Fias, W. (2005). Interacting neighbors: A connectionist model of retrieval in single-digit multiplication. *Memory & Cognition*, *33*(1), 1–16. https://doi.org/10.3758/BF03195293

Zbrodoff, N. J., & Logan, G. D. (2005). What everyone finds: The problem-size effect. In *Handbook of mathematical cognition* (pp. 331–345). Psychology Press. https://doi.org/10.4324/9780203998045

Data Analysis Plan

Our analysis focused on mixed-effects models with reaction time (RT) and accuracy as dependent variables. For RT analysis, we included only correct trials. Within the multiplication production task, we first excluded trials with RTs below 250 milliseconds. Secondly, we also applied a log10 transformation to address the positively skewed distribution of reaction time data.

In the final dataset, all 230 participants provided complete data for all four variables: typing speed, verbal STM, visual STM, and multiplication. We treated the verbal and visual short-term memory (STM) scores as ordinal variables, computed the median typing speed for each participant, and included a total of 12,874 trials for the multiplication task in the final dataset. All performance metrics for cognitive tasks are reported in **Table *1***.

We conducted statistical analyses using the R programming language (version 4.3.1) and the `lme4` package (version 1.1-33). In our study, we employed mixed effects modelling to identify whether there is a relationship between one outcome, multiplication performance, and participant's factors (verbal and visual STM scores with Typing speed as a covariate) and problem properties (size and interference). We standardized all the continuous predictors and employed a binary encoding approach for the categorical predictor, where we used -1 to represent small-sized and +1 to represent large-sized categories. The model was fit using maximum likelihood estimation, which yielded a robust and comprehensive representation of the data. Mixed-effects models are well-suited for analyzing hierarchical data structures, accommodating both fixed and random effects (Bates et al., 2015). These models also account for correlations within repeated measurements, which is crucial for our study (Pinheiro & Bates, 2000). Additionally, mixed-effects models enable us to comprehensively examine main and interaction effects (Barr et al., 2013)

Our model parameters were estimated using maximum likelihood estimation with the bobyga algorithm wrapped by the optimx package (version 2020-4.2; Nash & Varadhan, 2011) as the optimizer. The random structure adheres to the "keep it maximal" rule, fitting the most complex model consistent with the experimental design while removing only necessary terms to allow a non-singular fit (Barr et al., 2013). In pursuing the best model we employed a theoretically based strategy together with a step-wise procedure, so we included first the covariate and the properties of the problem, and then, to test our hypothesis, the STM factors and then the first-order interactions between STM factors and problems factors. More complex interactions were not tested as their interpretation remains too complex and overfitting is also an issue. P-values were obtained through chi-squared likelihood ratio tests (lmerTest package, version 3.1-3, Kuznetsova et al., 2017) for assessing the improvement in model fit by the inclusion of the fixed effect of interest, compared to a significance level of 0.05. Significance testing for fixed effects was conducted through likelihood ratio and Wald tests. Model fit was rigorously evaluated using various diagnostic tools, including visual residual analysis and examination of the variance inflation factor (VIF) (VIF < 5), ensuring robust model selections.

For significant interactions, we conducted post-hoc analyses using the emmeans package (Lenth et al., 2023, version 1.8.7), which provides marginal means and pairwise comparisons. These analyses allowed us to explore how the significant interactions manifested in terms of mean RTs across different levels of the interacting predictors.

The complete raw data and analysis, including data wrangling and mixed modeling analysis, can be found at GitHub (https://github.com/sohmeiling/stm_multiplication)**.**

Predictors of RT in Multiplication Task

To identify the predictors of reaction time (RT) in the multiplication production task, we executed stepwise regression analyses and used ANOVA to assess the relative goodness-of-fit between two linear mixed models. We also used likelihood-based criteria like the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) our model fitness. Lower AIC or BIC values indicate a better fit to the data. Our analysis began by considering problem features, specifically problem size, and interference score, owing to their theoretical relevance in the realm of numerical cognition. Subsequently, we introduced verbal short-term memory (STM) and visual STM into the models (see supplementary document).

Model 0 introduced typing speed as a control variable. **Model 1** included the problem size as predictor into the model. **Model 2** extended the analysis by adding interference score as an additional predictor. Subsequently, **Model 3** introduced verbal STM as another predictor, further enhancing model fit. Finally, we included visual STM in **Model 4**. We ran ANOVA analysis to compare the fitness of each model with previous models. The model fitness comparisons revealed notable distinctions among the various models. Initially, Model 1 displayed a substantially superior fit compared to Model 0, as evidenced by a statistically significant chi-squared statistic, $x^2(1) = 54.711$, $p < 0.001$. Building upon this improvement, Model 2 outperformed Model 1, exhibiting a significantly better fit, with a chi-squared statistic of $x^2(1) = 22.358$, $p < 0.001$. This trend continued with the introduction of Model 3, which introduced additional enhancements in model fit in contrast to Model 2, as reflected by a chi-squared statistic of $x^2(1) = 7.3435$, $p <.001$. Notably, however, the transition from Model 3 to Model 4 did not yield a significant improvement in model fit, as indicated by a chi-squared statistic of $x^2(1) = 2.2865$, $p = 0.131$.

Notably, we considered interactions between verbal STM and problem size, visual STM and problem size, verbal STM and interference score, as well as visual STM and interference score as two-level predictors. Among these models (Models 5-8), we observed statistically significant interactions only in the cases of problem size and visual interactions (Model 6) and interference and verbal interactions (Model 7). To gauge the improvements achieved by Models 6 and 7 over the baseline model (Model 4), we conducted model comparisons. The likelihood ratio tests (LRT) revealed that both Model 6 ($x^2(1) = 6.852$, $p = 0.009$) and Model 7 ($x^2(1) = 7.795$, $p = 0.005$) demonstrated a significantly better fit than Model 4. These results underscore the significance of these interactions in enhancing the overall model performance.

We integrated the two significant interactions into a unified and comprehensive model, denoted as Model 9. We then conducted a comparative assessment of its goodness-of-fit in relation to Models 6 and 7. Strikingly, Model 9 exhibited a notably superior fit. Specifically, the likelihood ratio tests (LRT) indicated that Model 9 outperformed Model 6, with a significant difference in model fit ($x^2(1) = 7.765$, $p = 0.005$). Moreover, Model 9 also demonstrated a statistically significant improvement in model fit when compared to Model 7 ($x^2(1) = 6.821$, $p = 0.009$).

In Model 3, the analysis of fixed effects revealed several significant predictors. Typing speed was a significant predictor ($\beta = 1.050$, SE = 0.005, $t(230.47) = 10.426$, $p < 0.001$), with faster typing speed associated with faster reaction time in multiplication task. Moreover, problem size had a significant positive effect on the reaction time in the multiplication task ($\beta = 1.131$, SE = 0.017, $t(58.66) = 7.267$, $p < 0.001$), demonstrating that a larger problem size incurred a longer reaction time. Similarly, the interference score was a significant predictor ($\beta = 1.046$, SE = 0.009, $t(67.01) = 5.166$, $p < 0.001$), suggesting that as interference of the problem increases, the RT also increases. Conversely, verbal STM span exhibited a significant negative relationship with the RT ($\beta = -0.987$, SE = 0.005, $t(229.34) = -2.767$, $p = 0.006$), suggesting that as verbal span increased, the RT decreased. Detailed analyses are in the Supplementary Document.

The analysis of fixed effects for Model 4 with the inclusion of visual span as a predictor reveals several significant factors. The intercept, typing speed, problem size, and interference remain strong positive predictors of the RT in the multiplication task. However, including visual span decreased the coefficient of the verbal span from -0.987 ($p = 0.006$ in Model 3) to -0.988 ($p = 0.016$ in Model 4). Notably, visual span was not significant ($\beta = -0.992$, $SE = 0.005$, $t\,(229.95) = -1.527$, $p = 0.128$).

We aimed to investigate the potential moderating effects of problem features, specifically problem size and interference score, on the relationships between Short-Term Memory (STM) components and multiplication time. To explore these interactions, we systematically examined all possible two-way interactions among predictors, as summarized in Table 5 of the Supplementary Document. Notably, we considered interactions between verbal STM and problem size, visual STM and problem size, verbal STM and interference score, as well as visual STM and interference score as two-level predictors. Among these models (Models 5-8), we observed statistically significant interactions only in the cases of problem size and visual interactions (Model 6) and interference and verbal interactions (Model 7). To gauge the improvements achieved by Models 6 and 7 over the baseline model (Model 4), we conducted model comparisons. The likelihood ratio tests (LRT) revealed that both Model 6 ($x^2(1) = 6.852$, $p = 0.009$) and Model 7 ($x^2(1) = 7.795$, $p = 0.005$) demonstrated a significantly better fit than Model 4. These results underscore the significance of these interactions in enhancing the overall model performance.

We integrated the two significant interactions into a unified and comprehensive model, denoted as Model 9. We then conducted a comparative assessment of its goodness-of-fit in relation to Models 6 and 7. Strikingly, Model 9 exhibited a notably superior fit. Specifically, the likelihood ratio tests (LRT) indicated that Model 9 outperformed Model 6, with a significant difference in model fit ($x^2(1) = 7.765$, $p = 0.005$). Moreover, Model 9 also demonstrated a statistically significant improvement in model fit when compared to Model 7 ($x^2(1) = 6.821$, $p = 0.009$).

In the pursuit of a more comprehensive understanding, we extended our investigation by including all two-level interactions into a unified model (Model 10) and compared its goodness-of-fit with that of Model 9. Importantly, the Model 10 with the inclusion of all two-level interactions performed significantly worse than Model 9 in terms of model fit ($x^2(1) = 0.523$, $p = 0.77$).

In Model 9, the interaction between interference and verbal STM demonstrates a significant negative effect ($\beta = -0.992$, $SE = 0.003$, $t(228.51) = -2.811$, $p = 0.005$). Additionally, the interaction between problem size and visual STM reveals a significant negative relationship ($\beta = -0.989$, $SE = 0.004$, $t(222.51) = -2.632$, $p = 0.009$). Simple slope analyses revealed significant trends for verbal and visual span, indicating a greater slope increase for high verbal and visual STM at high interference and large problem respectively (see Figures 1 and 2 in the main document).

Table 1.

Hierarchical Linear Models for Reaction time (RT) in Multiplication Task

Variable	β (std)	β (exp)	SE	df	t	p	sig
Model 0: log10_rt ~ typing_RT + (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.313	1.368	0.016	82.59	19.84	<.001	***
Typing RT	0.051	1.053	0.005	230.50	10.86	<.001	***
Model 1a: log10_rt ~ typing_RT + problem_size + (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.232	1.261	0.013	76.06	17.32	<.001	***
Typing RT	0.051	1.053	0.005	230.70	10.86	<.001	***
Problem size (large)	0.167	1.182	0.018	60.41	9.37	<.001	***
Model 2a: log10_rt ~ typing_RT + problem_size + interference_score + (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.259	1.296	0.013	100.60	20.27	<.001	***
Typing RT	0.051	1.053	0.005	230.80	10.86	<.001	***
Problem size (large)	0.123	1.131	0.017	58.65	7.27	<.001	***
Interference score	0.045	1.046	0.009	67.01	5.17	<.001	***
Model 3a: log10_rt ~ typing_RT + problem_size + interference_score + verbal_span + (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.259	1.295	0.013	98.33	20.39	<.001	***
Typing RT	0.049	1.050	0.005	230.47	10.43	<.001	***
Problem size (large)	0.123	1.131	0.017	58.66	7.27	<.001	***
Interference score	0.045	1.046	0.009	67.01	5.17	<.001	***
Verbal STM span	-0.013	-0.987	0.005	229.34	-2.77	0.006	**
Model 4a: log10_rt ~ typing_RT + problem_size + interference_score + verbal_span + visual_span + (problem_size + interference_score | participant) + (1 | problem)							

(Intercept)	0.259	1.295	0.013	98.01	20.41	<.001	***
Typing RT	0.047	1.048	0.005	230.76	9.71	<.001	***
Problem size (large)	0.123	1.131	0.017	58.66	7.27	<.001	***
Interference score	0.045	1.046	0.009	67.02	5.17	<.001	***
Verbal STM span	-0.012	-0.988	0.005	229.63	-2.43	0.016	*
Visual STM span	-0.008	-0.992	0.005	229.95	-1.53	0.128	
Model 5a: log10_rt ~ typing_RT + interference_score + visual_span + problem_size * verbal span (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.259	1.295	0.013	97.930	20.41	<.001	***
Typing RT	0.047	1.048	0.005	230.721	9.70	<.001	***
Problem size (large)	0.123	1.131	0.017	58.635	7.27	<.001	***
Interference score	0.045	1.046	0.009	67.022	5.16	<.001	***
Verbal STM span	-0.012	-0.988	0.005	230.285	-2.56	0.011	*
Visual STM span	-0.008	-0.992	0.005	229.918	-1.53	0.128	
Problem size * Verbal STM span	-0.005	-0.995	0.004	221.568	-1.11	0.270	
Model 6a: log10_rt ~ typing_RT + interference_score + verbal_span + problem_size * visual span (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.259	1.295	0.013	97.911	20.41	<.001	***
Typing RT	0.047	1.048	0.005	230.715	9.70	<.001	***
Problem size (large)	0.123	1.131	0.017	58.470	7.27	<.001	***
Interference score	0.045	1.046	0.009	67.028	5.16	<.001	***
Verbal STM span	-0.012	-0.988	0.005	229.592	-2.42	0.016	*
Visual STM span	-0.009	-0.991	0.005	231.093	-1.87	0.063	
Problem size * Visual STM span	-0.011	-0.989	0.004	222.426	-2.65	0.009	**
Model 7a: log10_rt ~ typing_RT + problem_size + visual_span + interference_score * verbal span (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.259	1.295	0.013	97.181	20.46	<.001	***
Typing RT	0.047	1.048	0.005	230.728	9.70	<.001	***
Problem size (large)	0.123	1.131	0.017	58.670	7.27	<.001	***

Interference score	0.045	1.046	0.009	66.517	5.17	<.001	***
Verbal STM span	-0.025	-0.975	0.007	234.794	-3.71	<.001	***
Visual STM span	-0.008	-0.992	0.005	229.928	-1.53	0.127	
Interference score * Verbal STM span	-0.009	-0.992	0.003	228.453	-2.82	0.005	**
Model 8a: log10_rt ~ typing_RT + problem_size + verbal_span + interference_score * visual span (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.259	1.295	0.013	97.775	20.42	<.001	***
Typing RT	0.047	1.048	0.005	230.714	9.71	<.001	***
Problem size (large)	0.123	1.131	0.017	58.666	7.27	<.001	***
Interference score	0.045	1.046	0.009	66.922	5.17	<.001	***
Verbal STM span	-0.012	-0.988	0.005	229.590	-2.43	0.016	*
Visual STM span	-0.014	-0.986	0.007	238.680	-2.01	0.046	*
Interference score * Visual STM span	-0.004	-0.996	0.003	227.735	-1.31	0.191	
Model 9a: log10_rt ~ typing_RT + problem_size + verbal_span + problem_size * Visual span + interference_score * verbal span (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.259	1.295	0.013	97.123	20.46	<.001	***
Typing RT	0.047	1.048	0.005	230.725	9.70	<.001	***
Problem size (large)	0.123	1.131	0.017	58.474	7.27	<.001	***
Interference score	0.045	1.046	0.009	66.521	5.17	<.001	***
Verbal STM span	-0.025	-0.975	0.007	234.857	-3.70	<.001	***
Visual STM span	-0.009	-0.991	0.005	231.109	-1.87	0.063	
Problem size * Visual STM span	-0.011	-0.989	0.004	222.506	-2.63	0.009	**
Interference score * Verbal STM span	-0.008	-0.992	0.003	228.509	-2.81	0.005	**
Model 10a: log10_rt ~ typing_RT + problem_size * verbal_span + problem_size * Visual span + interference_score * verbal span + interference_score * visual span + (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.259	1.295	0.013	97.090	20.46	<.001	***
Typing RT	0.047	1.048	0.005	230.724	9.70	<.001	***
Problem size (large)	0.123	1.131	0.017	58.469	7.27	<.001	***
Interference score	0.045	1.046	0.009	66.496	5.17	<.001	***
Verbal STM span	-0.025	-0.976	0.007	227.920	-3.56	<.001	***

Visual STM span	-0.012	-0.988	0.007	234.840	-1.73	0.085	
Problem size * Verbal STM span	-0.002	-0.998	0.004	221.742	-0.42	0.673	
Problem size * Visual STM span	-0.010	-0.990	0.004	222.345	-2.43	0.016	*
Interference score * Verbal STM span	-0.008	-0.992	0.003	228.766	-2.57	0.011	*
Interference score * Visual STM span	-0.002	-0.998	0.003	228.193	-0.59	0.556	

Note.To address the positively skewed nature of reaction time data, we applied a logarithmic transformation. Consequently, the β (std) coefficients presented in the table pertain to the log-10 of reaction times. For a more intuitive interpretation, please refer to the exp(β) values, which represent reaction times in seconds. This transformation facilitates easier comprehension of the effect sizes of predictors on reaction times.

Standardized coefficients (β (std)) and exponentiated coefficients (β (exp)) are presented. SE = Standard Error, df = Degrees of Freedom, t = t-statistic, p = p-value, sig = Significance level (* $p < .05$, ** $p < .01$, *** $p < .001$).

Table 2.

Hierarchical Linear Models of Reaction Time (including the accuracy as a covariate)

Variable	β (std)	β (exp)	SE	df	t	p	sig
Model 0b: log10_rt ~ typing_RT + correctness + (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.353	1.423	0.017	112.8	20.565	<.001	***
Typing RT	0.052	1.053	0.005	229.5	10.952	<.001	***
Correctness	-0.039	0.962	0.007	12600	-5.767	<.001	***
Model 1b: log10_rt ~ typing_RT + correctness + problem_size + (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.272	1.312	0.015	117.2	18.066	<.001	***
Typing RT	0.052	1.053	0.005	229.6	10.955	<.001	***
Correctness	-0.039	0.962	0.007	12600	-5.768	<.001	***
Problem size (large)	0.168	1.182	0.018	61.1	9.312	<.001	***
Model 2b: log10_rt ~ typing_RT + correctness + problem_size + interference_score + (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.299	1.348	0.015	152.5	20.454	<.001	***
Typing RT	0.052	1.053	0.005	229.7	10.957	<.001	***
Correctness	-0.039	0.962	0.007	12600	-5.755	<.001	***
Problem size (large)	0.125	1.133	0.017	58.41	7.231	<.001	***
Interference score	0.044	1.045	0.009	65.52	4.952	<.001	***
Model 3b: log10_rt ~ typing_RT + correctness + problem_size + interference_score + verbal_span + (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.299	1.348	0.015	150	20.566	<.001	***
Typing RT	0.050	1.051	0.005	229.6	10.499	<.001	***
Correctness	-0.039	0.962	0.007	12600	-5.774	<.001	***
Problem size (large)	0.125	1.133	0.017	58.42	7.231	<.001	***

Interference score	0.044	1.045	0.009	65.54	4.953	<.001	***
Verbal STM span	-0.014	0.986	0.005	229.4	-2.883	0.004	**
Model 4b: log10_rt ~ typing_RT + correctness + problem_size + interference_score + verbal_span + visual_span + (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.299	1.348	0.015	149.5	20.586	<.001	***
Typing RT	0.048	1.049	0.005	229.6	9.743	<.001	***
Correctness	-0.039	0.962	0.007	12600	-5.776	<.001	***
Problem size (large)	0.125	1.133	0.017	58.42	7.232	<.001	***
Interference score	0.044	1.045	0.009	65.54	4.953	<.001	***
Verbal STM span	-0.012	0.988	0.005	229.5	-2.513	0.013	*
Visual STM span	-0.008	0.992	0.005	229.4	-1.691	0.092	
Model 5b: log10_rt ~ typing_RT + correctness + interference_score + visual_span + problem_size * verbal span (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.299	1.348	0.015	149.400	20.59	<.001	***
Typing RT	0.048	1.049	0.005	229.600	9.74	<.001	***
Correctness	-0.039	0.962	0.007	12600.000	-5.77	<.001	***
Problem size (large)	0.125	1.133	0.017	58.380	7.23	<.001	***
Interference score	0.044	1.045	0.009	65.540	4.95	<.001	***
Verbal STM span	-0.013	0.987	0.005	230.200	-2.67	0.008	**
Visual STM span	-0.008	0.992	0.005	229.400	-1.69	0.092	
Problem size * Verbal STM span	-0.005	0.995	0.004	229.000	-1.11	0.269	
Model 6b: log10_rt ~ typing_RT + correctness + interference_score + verbal_span + problem_size * visual span (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.299	1.348	0.015	149.400	20.59	<.001	***
Typing RT	0.048	1.049	0.005	229.600	9.74	<.001	***
Correctness	-0.039	0.962	0.007	12600.000	-5.78	<.001	***
Problem size (large)	0.125	1.133	0.017	58.260	7.24	<.001	***
Interference score	0.044	1.045	0.009	65.550	4.95	<.001	***

Verbal STM span	-0.012	0.988	0.005	229.500	-2.51	0.013	*
Visual STM span	-0.010	0.990	0.005	231.000	-2.07	0.040	*
Problem size * Visual STM span	-0.010	0.991	0.004	229.000	-2.32	0.021	*
Model 7b: log10_rt ~ typing_RT + problem_size + visual_span + interference_score * verbal span (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.299	1.348	0.014	148.700	20.61	<.001	***
Typing RT	0.048	1.049	0.005	229.600	9.74	<.001	***
Correctness	-0.039	0.962	0.007	12600.000	-5.75	<.001	***
Problem size (large)	0.125	1.133	0.017	58.430	7.23	<.001	***
Interference score	0.044	1.045	0.009	65.140	4.96	<.001	***
Verbal STM span	-0.024	0.976	0.007	236.000	-3.56	<.001	***
Visual STM span	-0.008	0.992	0.005	229.400	-1.69	0.092	
Interference score * Verbal STM span	-0.007	0.993	0.003	229.400	-2.53	0.012	*
Model 8b: log10_rt ~ typing_RT + problem_size + verbal_span + interference_score * visual span (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.299	1.348	0.015	149.200	20.60	<.001	***
Typing RT	0.048	1.049	0.005	229.600	9.74	<.001	***
Correctness	-0.039	0.962	0.007	12600.000	-5.77	<.001	***
Problem size (large)	0.125	1.133	0.017	58.430	7.23	<.001	***
Interference score	0.044	1.045	0.009	65.450	4.96	<.001	***
Verbal STM span	-0.012	0.988	0.005	229.500	-2.51	0.013	*
Visual STM span	-0.014	0.987	0.007	240.300	-1.97	0.050	*
Interference score * Visual STM span	-0.003	0.997	0.003	229.3	-1.082	0.280	
Model 9b: log10_rt ~ typing_RT + correctness + problem_size + verbal_span + problem_size * Visual span + interference_score * verbal span (problem_size + interference_score | participant) + (1 | problem)							
(Intercept)	0.299	1.348	0.014	148.600	20.62	<.001	***
Correctness	0.048	1.049	0.005	229.600	9.74	<.001	***
Typing RT	-0.039	0.962	0.007	12600.000	-5.75	<.001	***
Problem size (large)	0.125	1.133	0.017	58.270	7.24	<.001	***

Interference score	0.044	1.045	0.009	65.150	4.96	<.001	***
Verbal STM span	-0.024	0.976	0.007	235.900	-3.59	<.001	***
Visual STM span	-0.010	0.990	0.005	231.000	-2.07	0.039	*
Problem size * Visual STM span	-0.010	0.990	0.004	229.000	-2.36	0.019	*
Interference score * Verbal STM span	-0.008	0.992	0.003	229.5	-2.565	0.011	*

Table 3.

Model fitness for all models with correctness as covariate.

Model	AIC (weights)	AICc (weights)	BIC (weights)	R2 (cond.)	R2 (marg.)	ICC
m_0b	-10245.9 (<.001)	-10245.9 (<.001)	-10163.8 (<.001)	0.501	0.059	0.47
m_1b	-10298.3 (<.001)	-10298.2 (<.001)	-10208.7 (0.002)	0.537	0.195	0.425
m_2b	-10317.0 (<.001)	-10317.0 (<.001)	-10220.0 (0.664)	0.558	0.22	0.433
m_3b	-10323.0 (0.010)	-10323.0 (0.010)	-10218.5 (0.318)	0.555	0.226	0.425
m_4b	-10323.8 (0.016)	-10323.8 (0.016)	-10211.9 (0.011)	0.554	0.227	0.423
m_5b	-10323.0 (0.010)	-10322.9 (0.010)	-10203.6 (<.001)	0.555	0.23	0.422
m_6b	-10327.1 (0.080)	-10327.0 (0.080)	-10207.7 (0.001)	0.557	0.234	0.421
m_7b	-10328.0 (0.129)	-10328.0 (0.130)	-10208.6 (0.002)	0.558	0.24	0.419
m_8b	-10322.9 (0.010)	-10322.9 (0.010)	-10203.5 (<.001)	0.556	0.232	0.421
m_9b	-10331.5 (0.744)	-10331.5 (0.744)	-10204.7 (<.001)	0.561	0.247	0.417

Note. AIC – Akaike Information Criterion, AICc – Corrected Akaike Information Criterion, BIC – Bayesian Information Criterion, R2 (Cond.) – the conditional R-squared, R2 (marg.) – the marginal R-squared, ICC – Intraclass Correlation Coefficient, RMSE – Root Mean Square Error, Sigma (σ) – the standard deviation of the residuals in a statistical model.

Figure 1.

An interaction plot of predicted values for visual STM

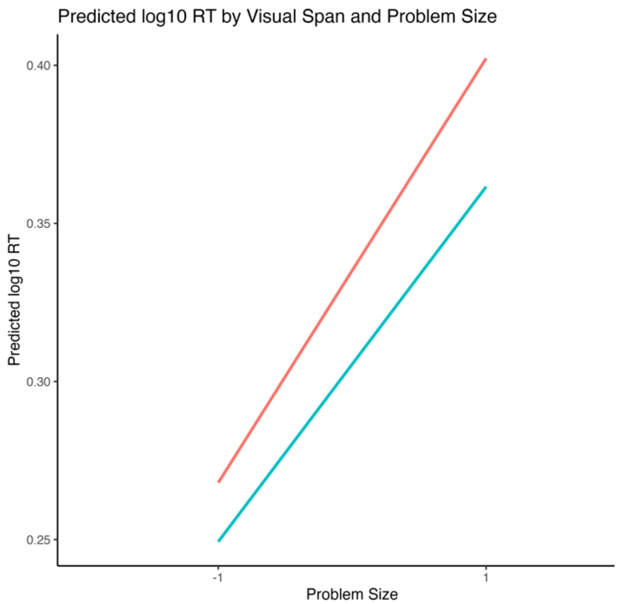
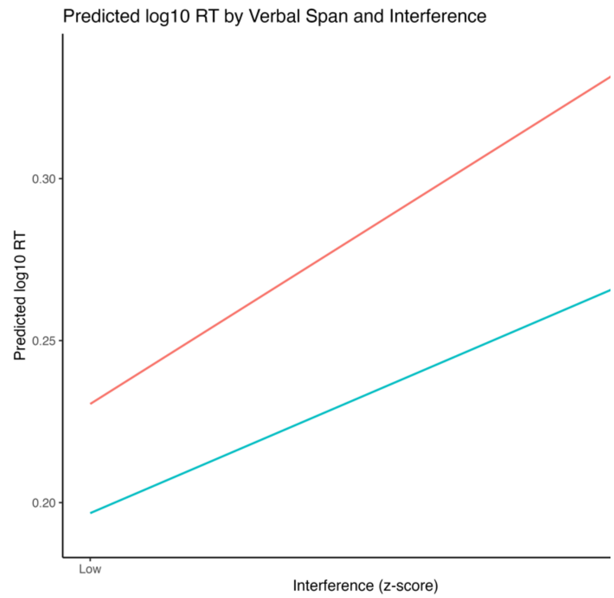


Figure 2.

An interaction plot of predicted values for verbal STM



Marginal Effects (emtrends) Output: The marginal effects of visual STM were assessed with respect to problem size using the mixed-effects model. For participants with small problem size (-1), the estimated trend for visual STM was -0.00933 (SE = 0.00506, 95% CI [-0.0193, 0.000638]). Conversely, for participants with large problem size (1), the estimated trend was -0.02030 (SE = 0.00697, 95% CI [-0.0340, -0.006561]). The trends reflect the change in the outcome variable visual STM as problem size increases.

Contrasts Output: An analysis of the differences in slopes between small and large problem size was conducted. The estimated contrast between the two levels of problem size was 0.011 (SE = 0.00417, 95% CI [0.002265, 0.0197]). The t-statistic (t.ratio) for this contrast was 2.627, with 231 degrees of freedom. The p-value for the test of slope differences was 0.0092, indicating a statistically significant contrast between the two levels of problem size. These results provide evidence of a significant difference in the slopes of visual STM between problems with small and large sizes.

Table 4.

Comparison of Hierarchical Linear Regression results for STM and problem features interactions

	Log-10 Reaction Time						
	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Predictors							
(Intercept)	1.295 ***	1.295 ***	1.295 ***	1.295 ***	1.295 ***	1.295 ***	1.295 ***
Typing RT	1.048 ***	1.048 ***	1.048 ***	1.048 ***	1.048 ***	1.048 ***	1.048 ***
Problem size	1.131 ***	1.131 ***	1.131 ***	1.131 ***	1.131 ***	1.131 ***	1.131 ***
Interference	1.046 ***	1.046 ***	1.046 ***	1.046 ***	1.046 ***	1.046 ***	1.046 ***
Verbal STM	-0.988 (0.016*)	0.988 (0.011*)	-0.988 (0.016*)	-0.975 ***	-0.988 (0.016*)	-0.975 ***	-0.976 ***
Visual STM	-0.992 (0.128)	0.992 (0.128)	-0.991 (0.063)	-0.992 (0.127)	-0.986 (0.046*)	-0.991 (0.063)	-0.988 (**0.085)**
Interactions							
Size × Verbal		0.995 (0.270)					-0.998 (0.673)
Size × Visual			-0.989 (0.009**)			-0.989 (0.009**)	-0.990 (0.016*)
Interference × Verbal				-0.992 (0.005**)		-0.992 (0.005**)	-0.992 (0.011*)
Interference × Visual					-0.996 (0.191)		-0.998 (0.556)
Model Comparison							
AIC	-10588.264	-10578.330	-10583.975	-10584.270	-10578.177	-10579.952	-10557.688
Δ AIC		9.934	4.289	3.994	10.087	8.312	30.576
BIC	-10484.480	-10467.134	-10472.778	-10473.073	-10466.980	-10461.342	-10424.252
Δ BIC		17.346	11.702	11.407	17.500	23.138	60.228
*R*2	0.562	0.562	0.564	0.566	0.563	0.568	0.570
Δ *R*2		0.000	0.002	0.004	0.001	0.006	0.008
χ2		1.181	6.852	7.795	1.668	14.616	15.139
P		0.277	0.009	0.005	0.197	0.001	0.004

Note. Model 5 to Model 10 were compared to the baseline Model 4 to assess improvements in model fit. The likelihood ratio tests (LRT) revealed that Model 9 exhibited a notably superior fit when compared to Model 6, with a significant difference in model fit ($x^2(1) = 7.765$, $p = 0.005$). Furthermore, Model 9 also demonstrated a statistically significant enhancement in model fit relative to Model 7 ($x^2(1) = 6.821$, $p = 0.009$). These results highlight the superiority of Model 9 in explaining the variance in the dependent variable when compared to alternative models. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; R^2 (cond.) = Conditional R-squared; Lower values of AIC, AICc, and BIC indicate better model fit. Higher R^2 values indicate greater explained variance.

Appendix A

Typing stimuli

Problems				Typing stimuli
2 x 3 =	3 x 2 =			6
2 x 4 =	4 x 2 =			8
3 x 3 =				9
2 x 5 =	5 x 2 =			10
2 x 6 =	3 x 4 =	4 x 3 =	6 x 2 =	12
2 x 7 =	7 x 2 =			14
3 x 5 =	5 x 3 =			15
2 x 8 =	4 x 4 =	8 x 2 =		16
2 x 9 =	3 x 6 =	6 x 3 =	9 x 2 =	18
4 x 5 =	5 x 4 =			20
3 x 7 =	7 x 3 =			21
3 x 8 =	4 x 6 =	6 x 4 =	8 x 3 =	24
3 x 9 =	9 x 3 =			27
4 x 7 =	7 x 4 =			28
5 x 6 =	6 x 5 =			30
4 x 8 =	8 x 4 =			32
5 x 7 =	7 x 5 =			35
4 x 9 =	9 x 4 =			36
5 x 8 =	8 x 5 =			40
6 x 7 =	7 x 6 =			42
5 x 9 =	9 x 5 =			45
6 x 8 =	8 x 6 =			48
7 x 7 =				49
6 x 9 =	9 x 6 =			54
7 x 8 =	8 x 7 =			56
7 x 9 =	9 x 7 =			63
8 x 8 =				64
8 x 9 =	9 x 8 =			72

Typing Speed Assessment. Participants were presented with 28 Arabic numbers (see Appendix A); the numbers served as answers to multiplication problems ranging from 2 × 3 to 8 × 9. All digits were displayed in black font (Arial, height unit: 0.08) on a white background in slightly middle of the screen (0, 0.15). The stimuli were presented randomly within a single block. The presentation sequence involved a 500ms blank screen followed immediately by the Arabic number stimuli. No time limit was imposed, but participants were instructed to type the digits accurately and rapidly. The participants pressed the ‘Enter’ key to proceed to the next trial. This task served as a measure of general processing speed and aimed to control for potential confounding effects of individual typing skills on subsequent single-digit multiplication performance. Reaction times for all trials (i.e., time from presentation of stimuli till the ‘Enter’ key was pressed) were recorded, and a median typing speed score for correct trials was computed for each participant.

Appendix B

Multiplication stimuli – Block A

Problem	Answer	Presentation Order	Problem size	Interference score
2 x 3 =	6	small first	small-sized	0
4 x 2 =	8	large first	small-sized	1
7 x 3 =	21	large first	small-sized	13
4 x 4 =	16	tie	small-sized	5
3 x 6 =	18	small first	small-sized	8
8 x 6 =	48	large first	large-sized	11
6 x 2 =	12	large first	small-sized	3
9 x 6 =	54	large first	large-sized	13
8 x 3 =	24	large first	small-sized	13
3 x 9 =	27	small first	large-sized	9
5 x 7 =	35	small first	large-sized	7
2 x 8 =	16	small first	small-sized	7
6 x 7 =	42	small first	large-sized	22
8 x 7 =	56	large first	large-sized	9
5 x 6 =	30	small first	large-sized	6
3 x 4 =	12	small first	small-sized	10
5 x 3 =	15	large first	small-sized	2
7 x 9 =	63	small first	large-sized	17
9 x 8 =	72	large first	large-sized	19
2 x 5 =	10	small first	small-sized	0
2 x 7 =	14	small first	small-sized	4
8 x 8 =	64	tie	large-sized	19
5 x 4 =	20	large first	small-sized	8
8 x 5 =	40	large first	large-sized	9
9 x 4 =	36	large first	large-sized	9
5 x 9 =	45	small first	large-sized	6
4 x 6 =	24	small first	small-sized	12
7 x 4 =	28	large first	large-sized	17
9 x 2 =	18	large first	small-sized	7
4 x 8 =	32	small first	large-sized	25

Appendix B

Multiplication stimuli (cont.) – Block B

Problem	Answer	Presentation Order	Problem size	Interference score
2 x 9 =	18	small first	small-sized	7
5 x 2 =	10	large first	small-sized	0
8 x 2 =	16	large first	small-sized	7
8 x 9 =	72	small first	large-sized	19
2 x 6 =	12	small first	small-sized	3
2 x 4 =	8	small first	small-sized	1
6 x 4 =	24	large first	small-sized	12
7 x 6 =	42	large first	large-sized	22
6 x 5 =	30	large first	large-sized	6
6 x 3 =	18	large first	small-sized	8
7 x 7 =	49	tie	large-sized	7
9 x 5 =	45	large first	large-sized	6
4 x 9 =	36	small first	large-sized	9
6 x 9 =	54	small first	large-sized	13
3 x 3 =	9	tie	small-sized	0
4 x 5 =	20	small first	small-sized	8
9 x 3 =	27	large first	large-sized	9
3 x 5 =	15	small first	small-sized	2
8 x 4 =	32	large first	large-sized	25
7 x 5 =	35	large first	large-sized	7
3 x 7 =	21	small first	small-sized	13
7 x 8 =	56	small first	large-sized	9
9 x 7 =	63	large first	large-sized	17
4 x 3 =	12	large first	small-sized	10
4 x 7 =	28	small first	large-sized	17
3 x 2 =	6	large first	small-sized	0
6 x 8 =	48	small first	large-sized	11
5 x 8 =	40	small first	large-sized	9
3 x 8 =	24	small first	small-sized	13
7 x 2 =	14	large first	small-sized	4

Multiplication Production Task. Sixty single-digit multiplication problems, ranging from 2 × 3 to 8 × 9, were carefully pseudo-randomized into two blocks, each containing 30 problems, to ensure that participants did not encounter three consecutive problems from the same multiplication table during the task. Each multiplication problem appeared twice, once with the smaller digit first and once with the larger digit first, except for tie problems presented once. These problems were categorized into three groups: 28 small-sized problems (magnitude < 26), 28 large-sized problems (magnitude > 25), and 4 tie problems (see Appendix B for the stimuli). The product of each problem ranged from 6 to 72. The procedure was the same as that in the Typing Speed Task, but here multiplication problems were presented, and participants were encouraged to respond as quickly and accurately as possible by typing in the correct answers and then pressing the "Enter" key to proceed to the next trial after providing their response. No time limit was imposed. The Multiplication task generated two primary outcome scores: the correct response and the reaction time for each trial (response time for each trial was computed from the presentation of the problem till the pressing of the ‘Enter’ key). These scores collectively reflected participants’ performance in single-digit multiplication.

Appendix C

Stimuli for Digit Span Task

Block A

Span length	**Stimuli**
2	46
2	82
3	157
3	397
4	6352
4	9184
5	59281
5	64738
6	396172
6	594631
7	8241973
7	2814367
8	53827164
8	75241739
9	682594316
9	142963758

Block B

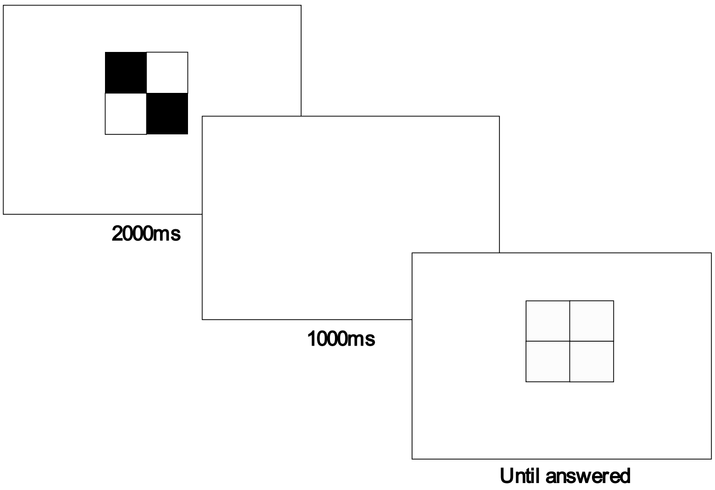
Span length	**Stimuli**
2	97
2	63
3	582
3	694
4	7286
4	6439
5	42731
5	75836
6	392487
6	619473
7	6917428
7	4179386
8	38296174
8	58132647
9	275863194
9	713942568

Verbal STM (Digit Span Task). For assessing verbal short-term memory, we employed two sets of digit stimuli: one derived from the Wechsler Adult Intelligence Scale (WAIS) and the other generated by an online algorithm (Egner et al., 2016) [URL: www.researchgate.net/publication/303985006_Digit-Span_Number_Generator] (see Appendix C). These sets of stimuli adhered to specific rules to ensure consistency. The digit strings consisted of a minimum of three single digits. Sequences of identical digits were permitted only after at least two different digits, ensuring diversity (e.g., 1-3-5-1), and neighbouring digit pairs had to differ by at least two digits (e.g., 5-7-9). The digit strings were presented in a non-sequential order, avoiding simple ascending or descending patterns (e.g., 3-5-7 or 8-6-4), and each adjacent string had different starting and ending digits. To minimize familiarity effects, specific sequences of digit strings (e.g., 5-1-3) were excluded if they appeared as neighbouring number pairs.

A female voice generated by the Balabolka app (Kutasov & Morozov, 2021; v2.15.0.767) using Microsoft Speech API (SAPI) [URL: https://www.cross-plus-a.com/index.htm] pronounced the digit lists. The sound for the task was managed using the PTB sound engine in PsychoPy, with latency mode set to 4 for optimal performance, ensuring minimal auditory latency and precise presentation timings (region of 5ms lag and maybe 1ms precision). Prior to the auditory task, a sequence of pure tones was played to ensure that the sound level was comfortably intelligible. Participants were presented with the digit sequences and were required to repeat them immediately in the same order using the keyboard. The length of the digit sequences progressively increased, starting from three digits and reaching a maximum of nine items. In this task, participants had two opportunities at each span length. If they answered correctly in at least one out of the two attempts, they proceeded to the next span length immediately, otherwise, the task was terminated. A significant positive correlation was observed between the two sets of digit span stimuli (r = 0.620, $p < 0.01$, N = 230). The digit span score was computed as the average of the two sets of digit spans, representing the participant's verbal short-term memory capacity.

Appendix D

Schematic Diagram for Visual Pattern Test



Visual Pattern Task. Visual memory was assessed using the validated Visual Patterns Test (Della Sala et al., 1999). Participants were presented with a matrix pattern containing 2.5 × 2.5 cm black and white squares which was displayed at the center of the screen (coordinates 0,0) for a duration of 2000 milliseconds. This pattern presentation was followed by a 1000-millisecond blank screen interval. Subsequently, an empty matrix consisting solely of white squares was presented to participants (refer to Appendix D for an illustration). Participants were then tasked with recalling and replicating the original pattern by clicking on the empty matrix. The Visual Patterns Test included 14 levels of increasing difficulty based on the number of black squares within the matrix, ranging from 2 to 15 black squares per span length. At each span length, participants completed three trials. If they achieved success in at least two out of the three trials, they advanced to the next level. If not, the task terminated. All visual stimuli were presented on a laptop screen. Visual memory span was computed as the average number of black squares correctly recalled by participants in the last three trials. This score served as an indicator of their visual short-term memory capacity, with higher scores reflecting greater proficiency in this domain.