<table>
<thead>
<tr>
<th>Title</th>
<th>Developmental changes in working memory, updating, and math achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Kerry Lee and Rebecca Bull</td>
</tr>
<tr>
<td>Source</td>
<td><em>Journal of Educational Psychology, 108</em>(6), 869-882</td>
</tr>
<tr>
<td>Published by</td>
<td>American Psychological Association</td>
</tr>
</tbody>
</table>

Copyright © 2016 American Psychological Association

This is an original manuscript of an article published by American Psychological Association in *Journal of Educational Psychology*. Changes introduced as a result of publishing processes such as peer review, copy-editing and formatting may not be reflected in this document.

**Notice**: This article may not exactly replicate the final version published in the APA journal. It is not the copy of record. For a definitive version of this work, please refer to the published source.

Citation: Lee, K., & Bull, R. (2016). Developmental changes in working memory, updating, and math achievement. *Journal of Educational Psychology, 108*(6), 869-882. [http://dx.doi.org/10.1037/edu0000090](http://dx.doi.org/10.1037/edu0000090)
Developmental Changes in Working Memory, Updating, and Math Achievement

Kerry Lee and Rebecca Bull
National Institute of Education, Nanyang Technological University

Author Note

This study was supported by grants from the Office of Educational Research and the Centre for Research in Pedagogy and Practice. Views expressed in this article do not necessarily reflect those of the National Institute of Education. We thank the children who participated in the study and school administrators who provided access and assistance. Special thanks to Ringo Ho who provided advice on growth modelling. We also thank Su Yin Ang, Fannie Khng, Jeremy Ng Lan Kong, and all the research assistants who helped with data collection and project management.

Correspondence concerning this article should be addressed to Kerry Lee, National Institute of Education, 1 Nanyang Walk, Singapore 637616. Email: Kerry.Lee@nie.edu.sg. Telephone +65 6219 6251, Fax: +65 6896 9845.

This is an uncorrected preprint. The authoritative version can be found at http://dx.doi.org/10.1037/edu0000090
Abstract

Children with higher working memory or updating (WMU) capacity perform better in math. What is less clear is whether and how this relation varies with grade. Children ($N = 673$, kindergarten to Grade 9) participated in a four-year cross-sequential study. Data from three WMU (Listening Recall, Mr. X, and an updating task) and a standardised math task (Numerical Operations) showed strong cross-sectional correlations at each of the ten grades, but particularly at Grades 1 and 2. Cross-lagged autoregressive analysis showed invariance in the predictive relations between WMU and subsequent math performance, but the importance of domain-specific knowledge increased with grade. Latent growth modelling showed that higher WMU capacity at kindergarten predicted higher math growth rates, averaged across all grades, but WMU growth rate was invariant across grades. SES, but not gender, explained variance in WMU at kindergarten. Implications for WM training are discussed.

Keywords: executive functioning, academic performance, working memory, updating, math
Developmental Changes in Working Memory, Updating, and Math Achievement

There is now a wealth of information showing that working memory (WM) or updating consistently explain substantial variance in the math performance of both typically achieving children and children with math difficulties (for a review, see Raghubar, Barnes, & Hecht, 2010). In this study, we focused on two issues that require clarification. First, it is unclear whether the magnitude of relations between WM, updating, and math achievement vary with age. With recent efforts to develop WM training that has the aim of improving academic performance, longitudinal information on relations between WM and math performance can potentially help identify when intervention should be administered and what kind of math difficulties are more likely to benefit from WM and updating training. A second issue is the importance of WM and updating relative to prior math achievement. Prior achievement is one of the best predictor of later achievement (Jordan, Kaplan, Ramineni, & Locuniak, 2009) and is particularly important for math where the learning of skills progresses in a hierarchical manner (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004). As highlighted by Fuchs et al. (2006), most studies examining cognitive determinants do not consider how lower order math skills determine subsequent skills, nor do they consider how the inclusion of lower order skills in an explanatory model affect the role of domain-general correlates in higher order math skills. Examining both domain-specific and domain-general skills longitudinally may provide information on the kind of amelioration measures that is appropriate for specific developmental periods.

Working Memory and Updating

WM is involved in short-term memory storage, reasoning, problem solving, and other higher cognitive tasks that require simultaneous representation and manipulation of information. Many studies have used Baddeley’s model (Baddeley, 2000; Baddeley & Hitch, 1974) as the basis for conceptualizing and measuring WM capacity. The latest version of the
model consists of four components: central executive, phonological loop, visual spatial sketchpad, and an episodic buffer. Both the phonological loop and the visual spatial sketchpad are short-term storage systems, responsible for storing and rehearsing auditory and visual spatial information respectively. The episodic buffer facilitates exchange of information between the central executive and long-term memory. The central executive was argued to be responsible for the simultaneous processing and storage of information.

More recent works have conceptualised the ability to process and store information as just one aspect of executive functioning (EF). Miyake et al. (2000) discussed EF in terms of inhibition (overriding of prepotent or dominant responses), shifting (switching flexibly between tasks or mental sets), and updating (monitoring, and the addition or deletion of contents from WM). On the measurement level, WM and updating are closely associated (Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009; St Clair-Thompson & Gathercole, 2006; Wilhelm, Hildebrandt, & Oberauer, 2013), but the two constructs are not conceptually identical (processing and recall versus the selective replacement of information) and may contain different sub-components (Ecker, Lewandowsky, Oberauer, & Chee, 2010). Amongst the three commonly studied EF (i.e., inhibition, shifting, and updating), updating is most consistently related to math achievement, with the relations between other EFs and math often found to be non-significant when WM or updating is entered into explanatory models (Bull & Lee, 2014). For this reason, we focus here on studies that examined directly relations between WM, updating, and math achievement.

**Working Memory, Updating, and Math Performance**

Previous findings can be characterised by one of several patterns of relations between WM, updating, and math performance. A number of cross-sectional studies, especially those conducted with younger children, suggest that WM explains math performance more consistently than does related domain-specific variables. Studies of children in kindergarten
and first grade typically focus on foundation skills or abilities, such as the identification of numerical patterns (Lee et al., 2012), magnitude judgments and number writing (Simmons, Willis, & Adams, 2012), number-line estimation, and the representation and mental combination of precise small quantity representations (Fuchs, Geary, Compton, Fuchs, Hamlett, & Bryant, 2010). Lee et al. (2012) found that whilst performance on numerical patterns was significantly related to arithmetical proficiency, this was rendered non-significant when WM and updating were included in the model. They suggested that relations between different aspects of math may be an artefact of common WM and updating demands. Toll, Van der Ven, Kroesbergen, and Van Luit (2011) found that whilst WM, updating, and preparatory math abilities provided equally good prediction of low math achievement measured 1.5 years later, WM and updating predicted persistence of math difficulties, over and above the predictive value of preparatory math abilities.

WM and updating are needed when children learn new math skills and are likely involved in the process of understanding, interpreting, and integrating new information with existing math knowledge. WM and updating are also needed when children are faced with math problems and have to generate solutions by considering the problems in light of their existing math knowledge. According to these views, abilities to learn or demonstrate new math skills should be tightly intertwined with existing math knowledge. This should be of particular importance in areas of math that have a cumulative or hierarchical nature (e.g., multiplication building on earlier competencies in addition). In other areas of math, there may be less dependency; either because of their foundational nature (e.g., magnitude comparison versus knowledge of number words) or because of the specialist knowledge required for solution (e.g., probability). By definition, for areas where there is relative independence, prior performance will have a weaker influence on subsequent performance. Even amongst areas of math that are logically interconnected, learners may not perceive or understand their
interconnections and treat them as discrete pieces of information. If this is the case, WM and updating, being domain-general, will serve as a more consistent explanatory variable for math performance.

Others suggest that math performance is dependent largely on building up domain specific skills and knowledge. Although a certain level of WM is needed to support the performance of math tasks, performance is largely predicted by domain-specific variables. Focusing on arithmetic fluency, Fuchs, Geary, Compton, Fuchs, Hamlett, and Bryant (2010), found that only domain-specific skills significantly predicted simple arithmetic computation skills in Grade 1; domain-general abilities did not mediate this relation (WM did however explain variance in a word problem version of the computation task). Fuchs et al. argued that arithmetic fluency grows out of the development of meaningful and interconnected knowledge about numbers that may have their foundation in early ability to subitize and estimate magnitude. Zheng, Swanson, and Marcoulides (2011) examined the contribution of WM to 7-10 years’ ability in arithmetic computation and arithmetic word problems. WM was a significant predictor, but its direct relation to word problem solving accuracy was fully mediated by reading and computational skills (see also Fuchs et al., 2006, who found that phonological processing mediated the relationship of WM to arithmetic fact fluency and word problem solving). Träff (2013) focused on 10 - 13 year olds and examined the role of WM, numeracy, and magnitude comparison skills on arithmetic fact retrieval, computation, and word problem solving. Numeracy and magnitude comparison skills, but not WM, uniquely predicted performance in arithmetic fact retrieval; verbal fluency and intelligence, but not WM, predicted arithmetic computation and word problem solving.

Consistent with this view are findings that WM and updating are not uniformly correlated with math performance, even when prior math performance is not included in the explanatory model. Simmons et al. (2012), for example, found that WM was correlated with
single digit addition at Grade 1, but not with single digit multiplication accuracy in Grade 3. Similarly, Meyer, Salimpoor, Wu, Geary, and Menon (2010) found that WM did not correlate with performance on the Numerical Operations task (Wechsler Individual Achievement Test, WIAT), but did predict performance on WIAT Math Reasoning in 2nd, but not 3rd grade. They suggested that WM may be important in initial learning (e.g., in guiding the use of counting strategies), but its importance diminishes during later stages of learning.

In contrast to findings on arithmetic problems, a number of studies that focused on algebraic problems suggest a more prominent role for WM and updating. Lee, Ng, Ng, and Lim (2004) found WM directly explained variance in algebraic problem solving accuracy, even when literacy and performance IQ were included as mediators. Focussing on children aged 10 years and older, Lee, Ng, and Ng (2009) replicated their earlier finding that WM and updating explained significant variance in algebraic problem solving, but also showed that performance is reliant on ability to decode and assign math relations to quantitative relations specified in the word problems. Of interest was that there were also strong relations between WM, updating and these more foundational math abilities. Khng and Lee (2009) found a more nuanced relation between WM and algebra in 14 year olds. In an exploratory path analysis, they found WM explained variance in a conceptual measure of algebra. For a computational measure of algebra, WM explained performance only indirectly, via a measure of intelligence. The mixed findings from these studies point to a third and perhaps more probable pattern, which suggests that the relation of WM to arithmetic may diminish with age, but that the relation to more complex skills may depend on the type of math task under consideration and the kind of domain-specific or domain-general factors that are included as explanatory variables.

To summarise, there are differing models for the relation between WM, updating, and early math achievement. Some suggest that earlier math skills are more dependent on basic
number abilities (Fuchs, Geary, Compton, Fuchs, Hamlett, Seethaler, et al., 2010). Others have argued that even simple tasks like magnitude comparison and deciding whether a verbally presented word refers to a number, demand both simultaneous and sequential processes of perceiving, coding, interpreting, and comparing information in different modalities (Kolkman, Hoijtink, Kroesbergen, & Leseman, 2013; Noël, 2009). Even a simple arithmetic task will likely rely on counting and other effortful strategies that require children to keep track, update, and manipulate numbers at hand (Monette, Bigras, & Guay, 2011). Few studies have focused on more complex math. Those that have focused on older children solving relatively simple arithmetic problems suggest that their growing knowledge of math and problem solving strategies decrease their reliance on effortful, controlled processes. Performance becomes increasingly dependent on how well they have acquired prior knowledge. However, the acquisition of more complex math skills, acquired via schooling, is still reliant on general cognitive abilities.

**Longitudinal Studies.** The studies reviewed above are cross-sectional in nature. Although they provide some insight on age related variation in the relation between WM, updating and domain specific knowledge, they provide little information on the extent to which performance is predicted by prior math performance versus WM or updating capacities. A number of recent studies have begun to address this issue using either repeated measures or longitudinal designs.

Geary and his colleagues have recently published a number of studies based on data from children tracked from kindergarten to 5th grade. Their findings suggest that the role of WM varies both with children’s experience and the type of math problems that they were asked to solve. Focusing on the use of addition strategies, they concluded that individual differences in WM are particularly important for understanding variation in the rate of learning to use strategies. Once these strategies have been learned, WM largely affects rate of
executing the strategies (Geary, Hoard, & Nugent, 2012). In contrast, their findings on children’s performance on achievement tests show that easier items on the WIAT Numerical Operations test did not require extensive WM engagement; WM became important with successive grades and more difficult test items (Geary, 2011). Geary also found that math cognition measured at Grade 1 predicted math achievement and growth beyond the contribution of domain-general abilities.

In other studies, WM played a more muted role. Focusing on children in Grades 2 and 3, Cowan and Powell (2014) found WM failed to explain variances in either calculation fluency or written arithmetic, but did explain variance in arithmetic word problems. For all three arithmetic skills, numeracy and magnitude knowledge regarding multi-digit numbers accounted for the most variance. Results from De Smedt et al. (2009) revealed that whilst WM was uniquely predictive of math achievement at both Grades 1 and 2, it was no longer a significant predictor of achievement at Grade 2 once Grade 1 math achievement was included as a predictor. Similar patterns of partial or complete mediation were found in studies that focused on more complex math skills. Lee, Ng, Bull, Pe, and Ho (2011) tested the extent to which 10 years olds’ performance on algebraic word problems was related to WM and updating, ability to solve math patterns, and their computational proficiencies. Children were tested first at Grade 4, then one year later. Relations between WM, updating, and algebra were mediated and rendered non-significant by patterns and computational proficiencies.

Fuchs et al. (2012), also focusing on pre-algebra, found WM measured at Grade 2 was not a significant direct predictor of performance, but had a small indirect effect on calculation skill at Grade 3. Seethaler, Fuchs, Star, and Bryant (2011) found Grade 3 computational fluency accounted for the most variance in both whole and rational number skills at Grade 5. Of the WM measures, only backward Digit Recall, not Listening Span, uniquely accounted for a small but significant amount of variance in both whole and rational number skills.
Differences in the predictive power of the various WM tasks suggest that it is ability to manipulate number rather than WM per se that explained variance in number skills.

In contrast to findings suggesting that WM is not as important as domain-specific knowledge at Grades 1 and 2, Van der Ven, Kroesbergen, Boom, and Leseman (2012) found that updating, but not other EFs, explained math achievement at each of four time-points in Grades 1 and 2. Latent growth models showed significant individual differences in the rate at which children improve in math, with minimal variation in growth rate for updating. They also found a significant correlation between attainments in math and updating, and significant correlation in their rates of growth, but attainment was not related to rate of growth in either cases. Van der Ven et al. argued that updating plays an important role in the math learning process, and that they influence each other’s development. Monette et al. (2011) focused on slightly younger children and found that WM at kindergarten predicted performance on math reasoning at Grade 1, even when pre-numeracy abilities were included as a covariate. They argued that the lack of automaticity in simple arithmetic at this age may make performance on math reasoning questions particularly demanding on WM.

**The Present Study**

Previous findings point to three ways in which patterns of relations between WM, updating, and math performance can be characterised. Studies conducted with younger children, in particular, suggest that WM predicts math performance more consistently than does prior math performance. Others suggest that the explanatory power of WM and updating is largely mediated by prior math achievement. A third and more probable model is that the relative importance of WM, updating, and prior math performance vary both with age and the domain of math under consideration.

In the typical course of development, competency in a particular domain of mathematics is likely to improve with experience, with subsequent competency dependent on
earlier performance. In addition to changes within each domain, the topics that are covered in math vary across grades. Some topics are likely to depend more heavily on prior understanding than others. It is difficult to discern grade-related changes from the existing literature partially because of differences in the math tasks used across studies and across grades. In this study, we used the WIAT Numerical Operations task to provide a measure of math achievement that can be put on a common scale, with raw scores that can be compared across all the grades included in this study. The Numerical Operations task includes questions from a wide range of math domains. The breadth of coverage is achieved by sacrificing depth; only a small number of questions are used for each math domain. For this reason, the task does not allow us to examine, with confidence, specific interdependencies between different math domains. However, because children from all grades progress as far in the test as their abilities allow, it allows for individual differences in the questions that are attempted at each grade. This is an important advantage because it provides a more sensitive measure of the range of each child’s knowledge. What we lose in ability to address questions regarding interdependencies between math domains, we gain in sensitivity to address our main question regarding the extent to which later math achievement is dependent on earlier competencies.

We examined the relations between WM, updating, and math performance from three different perspectives. First, we used a cross-sectional approach and examined whether the concurrent relation between WM, updating, and math performance remained stable from Kindergarten to Grade 9. In these initial analyses, prior math performance is not included in the analyses. Second, we examined the extent to which earlier capacities in WM, updating, and math contributed to later development in the two areas. These analyses allowed us to examine specifically the relative predictive power of WM and updating, versus prior math performance. In relation to this, our final research question specifically addresses the extent to which performances in kindergarten affected rates of growth. Few studies have examined
this issue, but findings from Van der Ven et al. (2012) suggest that individual differences in updating growth rates can be expected to be small.

We used a cross-sequential design that covered a span of ten grades: kindergarten to Grade 9. WM and updating capacities were measured using a multi-task latent approach that captured their shared variance. Although there is some data to suggest that visual versus verbal WM tasks may predict math performance differently at different grades (e.g., Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2014), given the conceptual assumption that the processing component of executive WM is amodal (Baddeley, 2000), we thought it important to have a reliable measure of this underlying construct. For this reason, we selected WM and updating tasks that used both visual and verbal stimuli as indicators to minimise the impact of stimulus modality on our latent measure.

**Method**

**Participants & Design**

We recruited children \((N = 673)\) from four cohorts: Kindergarten II \((M_{age} = 5.72, SD = 0.34)\), Primary 2 \((M_{age} = 7.85, SD = 0.32)\), Primary 4 \((M_{age} = 10.05, SD = 0.30)\), and Primary 6 \((M_{age} = 12.32, SD = 0.29)\). With the exception of the two youngest (i.e., 6 and 7 year olds) and the two oldest age groups (i.e., 14 and 15 year olds), this cross-sequential design provided us with data from two cohorts for each of 10 grades.

Children in the four commencing cohorts were recruited from five kindergartens and five public schools in western Singapore. Most of these schools served families with low to middle SES (measured by monthly household income on a 6 point scale from below $1000 to above $5000 in $1000 gradation, \(M = 3.7, SD = 1.7\)). In Singapore, children enter primary school, or Grade 1, the year they turn 7 years of age. Secondary school starts at Grade 7. Because of the longitudinal nature of the study, at the end of the study the children were spread across 81 primary and secondary schools. 365 children participated in all waves of
data collection and provided full data. Three children contributed only some auxiliary data, but none of the main variables used in this analysis. Of the remaining 670 children, 38 participated in the first year only. Overall, there are 98 different missing patterns: 43 children, for example, were missing for the second year of data collection, but contributed data to some or all subsequent waves.

All children participated with parental consent. Similar to the national distribution, 64% of the children were ethnic Chinese, 9% Indian, and 24% Malay; the remainder was from other ethnicities. The data reported in this study are from a larger study that examined the development of executive functioning and algebraic skills. Details on developmental changes in the factorial structure of executive functioning measures, including the WM and updating measures reported here can be found in Lee, Bull, and Ho (2013).

**Materials and Procedure**

Children were tested annually for four years, once per year. At each time-point, children were administered a large battery of EF (WM and updating, inhibition, and shifting, see Lee, Bull, & Ho, 2013), performance intelligence (PIQ), reading and math tasks. These were divided into five sets, with each set taking 45 to 60 minutes to administer. The tasks were administered in a fixed order, but the order in which the sets were administered was varied. Typically, we administered one set per day, but depending on schools’ schedule and availability, test administration spanned from 2 days to several weeks. All tasks were administered individually to the younger children. For older children, the computerised tasks were administered in small groups.

Children from the four commencing cohorts were tested at different times of the year. To ensure that differences in math performances were not affected by differences in curriculum coverage, children from different cohorts, but who are at the same grade, were tested at the same time of the academic year. The current analysis focussed on the WM,
updating, and the math tasks. The PIQ, inhibition and shifting measures were included as control variables.

**WM and Updating.** The Listening Recall (Alloway, 2007), Mr. X (Alloway, 2007), and Pictorial Updating (Lee et al., 2013) tasks were used to measure children’s WM and updating capacity. In Listening Recall, the children were first administered a series of simple statements (e.g., fish live in the sky), which they were asked to verify as “true” or “false”. The children were also asked to remember the last word in each statement. At the end of each trial (which contained one to six statements), there was a recall test. The dependent measure was the number of words recalled correctly in the order in which the statements were presented. Because the output of this task does not provide item-by-item data, internal reliability could not be computed; Alloway (2007) reported test-retest reliability of .81.

In the Mr. X task, two human figures were presented side-by-side. Each figure held a ball in one of six cardinal positions. One figure was identified as the target. In each trial, the children were asked to do two things: first, to decide whether the two figures were holding the ball on the same side; second, to remember where the target figure held the ball. Similar to Listening Recall, the number of figures presented in each trial varied from one to seven pairs. The dependent measure was the total number of positions recalled correctly in the order in which they were presented (test-retest reliability = .77, Alloway, 2007).

In the Pictorial Updating task, we presented black-and-white sketches of animals to children on a computer screen, one at a time. Children were asked to remember the last few animals that were presented. The number of animals to be remembered (2 – 4) was stated at the beginning of each trial, but the number of animals that would be presented in each trial was not. The number presented in each trial varied quasi-randomly from three in the easiest block, to 11 in the most difficult. Each block contained two practice sets and 12 experimental
trials. The dependent measure was the number of animals recalled correctly (Cronbach’s alpha ranges from .76 - .83, Lee et al., 2013).

**Math.** The Numerical Operations task from the WIAT (2nd ed., Wechsler, 2001) was used to measure math proficiency. Questions for children in kindergarten and Grade 1 measured basic counting knowledge, number discrimination, and transcribing between spoken and written numerals. Questions intended for older children covered arithmetic skills involving the four operators, with or without renaming or regrouping. More advanced questions covered the different number systems, exponents, algebra, and geometry. The dependent measure was the number of questions answered correctly. Children started at different points of the test depending on their grade; in generating a total score, full credit was given for earlier questions that were not administered. Published split-half reliability for the Numerical Operations tasks is .89 (Harcourt Assessment, 2005).

**Control Measures.** Although not a main focus of the study, we examined whether the relations between WMU and math were modified by the inclusion of other domain-general abilities, as measured by three inhibitory and three shifting measures, and a measure of PIQ. The Block Design task was used because it has the highest correlation with the PIQ subscale in the WISC (Sattler, 2001). We administered the task using published instructions (Wechsler, 1991); children were asked to copy abstract shapes using coloured blocks.

Because the inhibitory and shifting tasks are described in detail elsewhere (Lee et al., 2013), we describe them only briefly here. The six inhibitory and shifting measures were generated from four computerised tasks. In the Flanker task (modified from Fan, McCandliss, Sommer, Raz, & Posner, 2002), children were asked to identify whether a fish in the centre of the screen faced left or right. Congruent and incongruent trials were generated by having flanking fish that faced either the same or the opposite direction. The task contained blocks of congruent trials, blocks of incongruent trials, and mixed blocks containing both congruent
and incongruent trials. The Simon task (modified from Davidson, Amso, Anderson, & Diamond, 2006) used a similar procedure with congruent, incongruent, and mixed blocks. Children were asked to guide a butterfly or a frog to their respective homes on either side of the screen. In the congruent trials, the stimuli appeared on the same side as their home. They appeared on the opposite side in the incongruent trials. We also used a Picture-Symbol (based on the Number-letter task, Miyake et al., 2000) and a Mickey task to assess shifting and inhibition respectively. In Picture-Symbol, children were asked to judge whether a bigram contained a depiction of an animal or a number. Instructions were alternated to provide alternate trials that involved a switch in the rule by which the bigrams were to be assessed versus trials that did not involve rule switch. In the Mickey task, a picture of Mickey Mouse was presented on either the right or the left side of the screen. Each trial was preceded by squares that were presented on both sides of the screen, or squares that appeared on either the same or the opposite side as Mickey. Children were asked to indicate the side on which Mickey appeared. In all tasks, reaction time were used in latent structure models in which mean reaction time in the incongruent or switch conditions were regressed on to the congruent or non-switch conditions, respectively. The incongruent and switch data were then used as indicators for the inhibitory and switch latent measures, respectively (Lee et al., 2013).

Analyses

Data were first screened for missing values, outliers, and normality of distribution. We conducted three sets of analyses. To ascertain the relations of WM, updating, and math at each of the ten time points, we conducted a series of cross-sectional analyses. Performances on executive functioning measures, including updating, are typically affected by multiple processes. To obtain separate estimates of true scores and measurement error, we used a latent approach to generate a latent factor (WMU) from the Pictorial Updating, Mr. X, and
Listening Span tasks. Data from these tasks were reported in Lee et al. (2013) to load on the same factor and to exhibit longitudinal factorial invariance. In the second set of analysis, we fitted a cross-lagged autoregressive model to the data. This examined the extent to which earlier capacities contributed to later development. WMU scores were used to predict subsequent performance in both WMU and math; similar predictive relations were specified for the math scores. The final set of analyses examined how growth can be characterised and examined whether performance in kindergarten affected subsequent rates of growth. Latent growth curves were fitted to both WMU and the math data. We examined whether the WMU parameters were predictive of those from the math measures.

For the cross-sectional analysis, because the kindergarten children were recruited from five different centres, we took account of potential cluster effects by regressing the various manifest variables onto dummy variables that corresponded to the kindergartens. With transition to higher grades and the availability of same-grade data from other cohorts, children in other grades were drawn from a larger number of schools. For these grades, we used the cluster sampling algorithm in Mplus to correct for potential underestimations of standard errors. Inclusion of cluster information in the cross-sectional analysis affected the overall pattern of findings only for the kindergarten children. Nonetheless, we took a more cautious approach and corrected for cluster effects for variables that exhibited a large design effect (> 2.4, see Table 1). In both the cross-lagged autoregressive and latent growth models we regressed the affected variables to dummy variables that corresponded to schools from which children were drawn.

**Results**

To attenuate the effects of outliers while preserving their status as extreme values, we replaced accuracy scores that were > 3 SD from the sample mean with values computed at 3 SD from the mean. Screening for multivariate outliers resulted in the deletion of 17 cases
from the total sample of 670. The final sample included in this analysis was as follows: Kindergarten II (N = 181, 94 boys), Primary 2 (N = 167, 85 boys), Primary 4 (N = 158, 74 boys), and Primary 6 (N = 147, 65 boys). In total, 14.02% of data points were missing because participants dropped out of the study, failed to attend the whole set of assessment during a particular wave, or failed to participate in a particular wave of data collection. To avoid bias (Graham, 2009) and reduction in power that result from listwise deletion of data, analyses were conducted using all available data using full information maximum likelihood, as implemented in Mplus 7.11 (Muthén & Muthén 2011). Means, variance, and bivariate correlations can be found in Table 1.

Concurrent Relations

To examine whether relations between WMU and math performance varied from Kindergarten to Grade 9, we conducted a series of correlational analyses. For grades that recruited from two cohorts, data were collapsed across cohorts. WMU was correlated with math performance at each grade. Correlations ratios were transformed into Fisher’s z to examine differences in relations across grades. These relations were found to be the weakest at Kindergarten (r = .327), peaked at Grade 1 and 2 (r = .658 and .627 respectively), became significantly or marginally lower for much of Grades 3 to 7 (.47 < r < .551), before increasing again in Grades 8 and 9 (r = .588 and .547 respectively). Table 2 provides fit indices for each age group, factor loadings for WMU, and the correlation coefficients.

Regarding the control measures, Lee et al. (2013) reported that the six inhibitory/switch measures conformed to a unitary factor from kindergarten to Grade 8 and were distinguishable as separate inhibition and shifting factors only at Grade 9. Here, we used the same data and incorporated the corresponding executive functioning factor into our
previous model such that it was allowed to covary with both WMU and math performance. With the three constructs considered simultaneously, the findings showed that inhibition/shifting and math performance did not correlate at any grade, but correlations between WMU and math remained statistically significant \((r = .256 \text{ for K2, } r > .409 \text{ for Grades 1 – 9})\).

PIQ was modelled as a single indicator latent construct with residuals estimated from published reliability coefficients (Wechsler, 1991) and sample variance, and were covaried with WMU and math performance. PIQ correlated strongly with WMU \((.520 < r < .701)\). Its correlation with math performance was relatively weak at K2 \((r = .211)\), remained weak to moderate across Grades 1 to 5 \((.280 < r < .396)\), but became strong for the upper grades \((.459 < r < .561)\). Relations between WMU and math performance remained little changed from when PIQ was not included \((.211 < r < .652)\). It is notable that, from Grades 1 to 5, though WMU was more strongly correlated with math performance than PIQ was with math, both WMU and PIQ were correlated strongly and similarly with math for the remaining grades.

**Predictive Relations**

To examine whether prior performances or capacities predicted future performances, we fitted a cross-lagged autoregressive model to the data. WMU from each cohort and each wave of data were regressed to the year before. A parallel autoregressive chain was formed using Numerical Operations. Predictive relations between math and WMU were modelled using cross-lagged regressive parameters between the two constructs.

One advantage of our cross-sequential design is that it allows us to examine growth across ten age groups even though children in each cohort were tracked for only four years. To model the complete longitudinal sequence, we constrained autoregressive and cross-lagged coefficients from overlapping grades, originating from different cohorts, to be the same. Intercepts and factor loadings for each WMU manifest measure were also held constant.
across all grades and cohorts (Model 1). In effect, these equality constraints specified that, across cohorts, children of the same grade were expected to exhibit the same pattern of findings.

Fit indices for this model were relatively poor, $\chi^2(596) = 1032.399$, $p < .001$, RMSEA = .067, CFI = .877, SRMR = .126. Much of the misfit originated from the Pictorial Updating task (Kindergarten, Wave 1). The modification indices showed that the Kindergarteners’ range of performance commenced at a lower point than that estimated by the model. Given the age of the children, it seems reasonable that some of them had difficulties with the task. The modification indices also showed that equality constraints placed on 2 of the 48 intercepts (the Pictorial Updating task at Grade 2 and Mr X task at Grade 9) resulted in some misfit. Additional cluster control was also recommended for two data points. These changes were incorporated into a revised model (Model 2). Notably, none of the modification indices suggested that misfit resulted from cohort differences. In fact, a comparison of models in which same grade, across cohort parameters were estimated freely, versus constrained to equality, revealed little differences, $\Delta \chi^2(12) = 13.981$, $p = .302$, $\Delta$CFI = .001.

The revised model exhibited a better fit to the data, $\chi^2(592) = 832.041$, $p < .001$, RMSEA = .05, CFI = .932, SRMR = .107. The only other modification of interest was an auto-regression relation for the Numerical Operations measures between the second and fourth time-point of the Grade 6 cohort. This relation suggests that material learned at Grade 7 have both a proximal effect on performance one year later and a distal effect two years later. Incorporation of this parameter resulted in a final model with a reasonable fit to the data, $\chi^2(591) = 809.329$, $p < .001$, RMSEA = .048, CFI = .939, SRMR = .103.

The revised model showed that in kindergarten and the first two years of formal schooling, performance in math is less dependent on previous achievement in math than it is on WMU. Despite moderate to strong bivariate correlations between the math scores from the
first three time-points, these relations were not significant when the cross-lagged influences of WMU were included in the model. Similar to the cross-sectional findings, cross-lagged relations between WMU and math were the strongest at K2 and Grade 1, but were more moderate for much of the primary school years from Grade 2 - 5. These relations were no longer significant from Grade 7 onwards. The autoregressive relations between WMU at each time-point were very strong. As noted earlier, autoregressive relations between the math measures were not significant between K2 and Grade 2, but were strong in other grades. Notably, math performance did not predict WMU capacity at any point.

To test explicitly the grade-related differences in relations between math and WMU, we compared the previous model against another in which the same relations were held constant across all ten grades (Model 3). This model exhibited only a very small deterioration in fit, $\Delta \chi^2(10) = 10.698$, $p = .220$, $\Delta CFI = .001$, and being more parsimonious, is deemed a more appropriate model for the data. This model (see Figure 1) differs from the prior model in two important ways. First, in specifying the same predictive relations between WMU and math across grades, these relations are now significant even for the two most senior grades. The findings also differ by showing that though the predictive relations between each year’s math performance is lower between Grade 1 and kindergarten ($B = .240$) than other grades ($.425 < B > .719$), they are significant even amongst the youngest grades. To test the constancy of these grade-related differences, we held them constant in a final model (Model 4). This model exhibited a large deterioration in fit, $\Delta \chi^2(9) = 100.968$, $p < .001$, $\Delta CFI = .026$, showing that relations between math performance do vary with grades.

Comparing the relative contributions of WMU and math to subsequent performance in math in Model 3, the pattern was mixed amongst the younger grades. Contributions of WMU and math were similar with the exception of Grades 1 and 4. For those grades,
performances were better predicted by WMU ($\beta = .379$ and 484, respectively) than by math ($\beta = .208$ and .107, respectively). From Grade 5 onwards, math ($0.370 < \beta > 0.705$) had greater influence than did WMU ($0.157 < \beta > 0.329$).

Because the cross-sectional analyses showed that PIQ predicted math performance as well as did WMU for the upper grades, we examined whether the predictive power of WMU is shared with PIQ. We conducted additional analyses in which PIQ was included as an explanatory model. To Model 2, we added a parallel autoregressive chain for PIQ. Predictive relations between PIQ, WMU, and math were modelled using cross-lagged regressive parameters between the three autoregressive chains (Model 5). The model provided a reasonable fit to the data, $\chi^2(907) = 1210.22, p < .001$, RMSEA = .045, CFI = .94, SRMR = .11. The findings showed that math performance was predicted consistently by WMU, but not PIQ.

**Characterizing Growth**

A question we have not yet addressed is whether there are individual differences in children’s rates of growth and whether early attainments influenced their rates of growth. To examine these issues latent growth curve models for WMU and math were first estimated. Intercept and growth parameters from the math model were then regressed onto those from WMU (see Figure 2).

For the math data, we fitted an autoregressive latent trajectory model (Bollen & Curran, 2004), in which the very first time-point (kindergarten) was rendered an extraneous variable and a predictor for a growth curve that described development from Grades 1 to 9. Rates of growth were estimated freely with the constraint that they are equal across grades from different cohorts. As shown in Figure 2 (upper panel), the primary years were characterised by rapid periods of growth punctuated by slower growth between Grades 2 and
3 and between Grades 4 and 5. There was a marked slowing during the final three years, perhaps related to the switch to the secondary school curriculum. There were significant individual differences in performance at Grade 1, $\sigma^2 = .04, p < .001$. Rates of growth also exhibited significant individual differences, $\sigma^2 = .03, p < .001$.

Because the WMU data were multivariate in nature, growth was modelled using a curve-of-factors approach (T. E. Duncan, Duncan, & Strycker, 2013) in which a growth curve was modelled on WMU latent factors taken at each of the four time-points. Similar to the model for math scores, rates of growth were estimated freely with equality constraints placed across same grade cohorts. The model showed significant growth in WMU capacity across age groups, $M = .812, p < .001$, and significant differences in kindergarten children’s WMU capacity, $\sigma^2 = .258, p < .001$. However, rates of growth did not differ across individuals. As illustrated in Figure 2 (lower panel), the slope coefficients suggest steady growth with signs of slowing during the last two grades.

The overall model showed a moderate fit to the data, $\chi^2(619) = 871.958, p < .001$, RMSEA = .050, CFI = .929, SRMR = .135. The WMU intercept predicted the math intercept ($\beta = .369, p < .001$). Because the math growth function was estimated with performance at Grade 1 as the intercept, this replicates the finding from the predictive analyses showing that WMU capacity at kindergarten predicted math performance at Grade 1. Average rate of growth in math was also predicted by the children’s WMU capacity at kindergarten ($\beta = .312, p < .001$). Kindergarteners with higher WMU capacity improved in math faster than did those with lower WMU capacity. In contrast, children with poorer math achievement at Grade 1 exhibited more improvement in math than did those with better achievement ($r = -.858, p < .001$).

To examine possible causes of variation, we added gender and SES as covariates for the intercept and growth terms of math and WMU. The findings showed that higher SES was
related with higher WMU at kindergarten ($\beta = .233$). SES did not explain variance in the math intercept, but explained differences in math growth for the Grade 4 cohorts ($\beta = .281$).

Boys in the youngest cohort exhibited a marginally lower rate of growth in math than did the girls ($\beta = -.200$, $p = .053$), but no gender related differences were found in the other cohorts.

**Discussion**

The current findings add to the literature by examining patterns of growth in WM, updating, and math over an extensive period from early childhood to the mid-adolescence years. The concurrent, cross-sectional, findings are generally in support of the hypothesis that their relations vary with grade. However, the variation is more notable across kindergarten and Grade 2. Although some of the higher grades exhibited lower correlations relative to the peak at Grade 1, correlations were generally strong and there was relatively little variation across Grades 3 to 9. Similar to findings from a recent review of the literature on relations between updating, inhibition, shifting, and math performance (Bull & Lee, 2014), we found no significant correlation between inhibition, shifting, and math performance, nor did they affect the correlation between WM, updating, and math. Although our measure of performance intelligence was correlated with math performance, its inclusion in the cross-sectional model did not reduce the strength of correlation between WM, updating, and math. These findings demonstrate that the role of WM and updating is not an artefact of either performance intelligence or other aspects of executive functioning.

Similar to findings from Geary (2011), the concurrent findings suggest that the relation between WM, updating, and math is weaker in kindergarten than in other grades. The majority of our kindergarten children were able to answer the questions on basic number recognition, writing, counting, and sequencing, with just under half also advancing to the simplest single digit arithmetic questions. Given these observations, it is likely that these basic numeracy skills are well practiced amongst our kindergarten children and imposed less
demand on WM and updating resources than they would have on children who are less skilled. However, as discussed later, when we examined relations from a longitudinal perspective, WM and updating capacity at kindergarten is important in another way; specifically, it predicts averaged math growth rate from Grades 1 to 9.

The heavy dependence between WM, updating, and math at Grades 1 and 2 may be due to the immaturity of number bonds and arithmetic strategies. An inspection of children’s performance in the Numerical Operation task showed that only half of the children in Grade 1 were successful on questions involving single digit addition and subtraction. Around one third of the children were successful on the multi-digit addition and subtraction questions. Thus, although many children were doing well, a large number were still having difficulties even with simple arithmetic questions. Indeed, experimental data suggest that, even for adults, multi-digit addition is demanding on WMU resources (Fürst & Hitch, 2000).

**Predictive Relations.** It is important to note that the concurrent findings did not take account of the contributions of prior math achievement. In the cross-lagged autoregressive models, which tested the contributions of both prior math achievement and WM capacity, the latter was found to predict math performance at all grades. Of note is that the magnitude of these predictive relations remains the same across all grades. On the face of it, the concurrent and predictive findings seem contradictory. Analytic techniques aside, these differences are perhaps reflective of differences in resources that are needed for executing computation or problem solving, versus resources needed to support the learning of new concepts and procedures, respectively. Our predictive findings suggest that reliance on WM and updating resources for learning do not vary across grades. Fluctuations observed in the concurrent findings may reflect changes in resources needed for executing computation, with reductions resulting from automaticity that comes with additional practice. This is perhaps the reason for the slight reduction in cross-sectional correlation at Grade 6. In the local context, a sizeable
proportion of time in that grade is devoted to consolidation in preparation for a high stake
examination held at the end of the grade. Furthermore, the finding that in the higher grades (6
– 9), correlations between our performance intelligence measure and math became as strong
as those between WM and math suggest that there is an increasing reliance on reasoning or
visualisation skills in solving problems in those grades.

The predictive analyses also show that there is significant variation in the extent to
which math performance is reliant on prior math performance. Specifically, math
performance at Grade 1 is noticeably less reliant on prior performance at kindergarten than
are relations found in most other grades. A comparison of the standardised regression
coefficients also suggests that doing well in math at Grade 1 depends somewhat more on WM
and updating capacity (β = .379) than on children’s math performance the previous year (β
= .208). This finding seems inconsistent with the extant findings on the importance of early
numeracy skills (De Smedt et al., 2009; Geary, 2011). However, bearing in mind that our
kindergarten children were performing at high levels on the numeracy questions, it may be
the case that though basic numeracy is important (Fuchs, Geary, Compton, Fuchs, Hamlett, &
Bryant, 2010), it contributes less to performance in arithmetic computation once competency
is past a certain threshold (Meyer et al., 2010).

Disparity between the predictive effects of WM, updating, and prior math
performance is most notable for Grade 4, in which performance is predicted only by WM and
updating. One explanation for this finding is that it is in this grade that most children would,
for the first time, encounter decimal and fractions in the Numerical Operations task.

Children’s initial difficulties with fractions are well documented (e.g., Mack, 1995).
Although we need to be cautious in interpreting our findings because the questions attempted
by most of the Grade 4 children involve not only fractions, the present finding suggests that
earlier competencies in whole numbers provide little predictive power for how well children
will perform when they first encounter fractions (DeWolf & Vosniadou, 2015; Hartnett & Gelman, 1998; c.f., Siegler, Thompson, & Schneider, 2011).

Unlike the earlier grades, the predictive power of WM and updating, relative to prior math performance, was noticeably weaker for the more senior grades (5 - 9). As children progress through the Numerical Operations task, some of the items become more complex (simple algebra in Grade 7, exponentials in Grade 8). In the same sense as complexity was defined in the intelligence literature (e.g., Stankov, 2000), more complex tasks require a broader range of math knowledge for solution. With items that are often more complex (and more difficult), one would expect them to place greater, not less demands on WM and updating resources. In contrast, our finding suggests that the increase in complexity places more demands on domain specific knowledge, but no additional demands on WM or updating resources. One possible explanation for these findings is that although the more complex items require knowledge of decimals, exponentials, and algebra, they are no more demanding in terms of the number of steps or the kind of arithmetic computation needed for solution. Because of the reliance on domain specific knowledge, only children who have the requisite knowledge will likely be successful on the more difficult items.

**Characterising Growth.** A surprising finding was that though significant individual differences in WM and updating capacity were found at kindergarten, the rate at which capacity increased, when averaged across the ten grades, did not exhibit individual differences. We know of only two other studies that examined this issue. Hackman et al. (2014) found significant variation in rate of WM growth in participants aged 10 to 18 years, and Van der Ven et al. (2012) found significant individual differences in rates of WM growth in children aged 7-8 years. However, an inspection of the model used in Van der Ven et al. (2012) suggests that a non-significant term that specifies no individual differences would have yielded a more parsimonious model. Given the differences in age range and material, it
is difficult to pinpoint the variables that resulted in the differences in findings. It should be noted that in the present study, variance in rate of growth was estimated over the entire age range. This does not preclude the possibility of localised changes within a narrower range.

**SES and Gender Differences.** Though not a central focus of the study, we capitalised on the availability of SES and gender data and examined whether they explained some of the variance in the data. We found minimal gender differences, but significant SES related differences. In perhaps the first longitudinal prospective study on the relation between SES and WM, Hackman et al. (2014) found their relation to remain unchanged from childhood to adolescence. Similar to Noble, Norman, and Farah (2005), they argued that variation in early childhood family environment is the most likely cause of WM differences. Although we do not have specific data that speak to this issue, our finding is consistent with this interpretation. Regarding math performance, SES predicted higher rates of growth, but only for the Grade 4 cohort. The reason for the lack of a more wide-spread pattern of influence is not clear. One explanation is that we followed this cohort through a major transition: from primary to secondary school. At Grade 6, all local children sit for a high-stake national examination that determines how they are streamed in secondary school. It is plausible that our finding reflects differences in the resources that family can muster in helping children prepare for the examination. However, this explanation is counter-indicated by the lack of a similar finding from the Grade 6 cohort, who also underwent the same transition. Unlike the Grade 6 children from the younger cohort, who exhibited the highest bivariate correlation between SES and math performance ($r = .401$), these children exhibited only a weak correlation ($r = .190$). Furthermore, there is evidence to suggest that children from lower SES families do not perform more poorly in the national examination (Ng, 2011). Given the likelihood that SES effects will differ depending on societal affordances, more local research is needed.
Implications for WM Training and Pedagogical Practice. If difficulties associated with the lack of transfer and maintenance of training effects can be overcome, our findings suggest that the early years (kindergarten to Grade 2) provide a better time for WM intervention. Although the strength of concurrent correlations between WM and math performance remains high through to Grade 9, the early years see a somewhat lower dependence on prior math achievement as compared to the senior grades. This increases the likelihood that improved WM and updating capacity alone will produce better math performance. This recommendation is qualified by two important findings from the growth models. These models show that WM and updating capacity at kindergarten is predictive of averaged rates of growth in math. They also show that the rate at which WM and updating develop is invariant across individuals. One inference that can be drawn from these findings is that if children’s WM and updating capacity can be optimised by the time they are at kindergarten, they will likely improve in math more quickly over the next nine grades. However, this optimism is countered by the finding of invariance. This finding suggests that, in spite of the vicissitudes of experiences and upbringing that would have been encountered by the children, WM and updating capacity is likely resistant to transient changes in the environment. Recent reviews have raised fundamental concerns about whether there is sufficient data to show that WM or updating capacity can be improved (e.g., Shipstead, Redick, & Engle, 2012). Our finding of invariance suggests that training stands a better chance of success if administered at or prior to kindergarten. The possibility of training success notwithstanding, with findings that individual differences in executive functioning are attributable to genetic factors (Friedman et al., 2008) and perhaps resistant to change, more efforts are needed to examine how pedagogical practices can be made to accommodate WM limitations.
From a curricular viewpoint, our findings show that math performance at Grade 7 predicts performance one year later, but it also has a more distal relation with math performance two years later. In some ways, these findings are not surprising because Singapore uses a spiral curriculum in which material taught in earlier years are revisited and built upon in later years. More noteworthy is that this was the only distal relation that was significant. Questions answered by Grade 9 participants dealt largely with fraction division and computation involving symbolic algebra. The distal relation may reflect the importance of attaining a high degree of proficiency in basic computation involving symbolic algebra, when it was first introduced in Grade 7.

Caveats

A strength as well as a weakness of this study is that growth in math performance was indexed by only the WIAT Numerical Operations task. The use of a standardised measure allowed us to put performance on a common metric and to examine growth across the 10 grades: a central aim of this study. Though our findings provide limited insight on the role of WM and updating at the processing level, or identify differences in strategies that children use for solution, children who are faced with an end-of-term achievement test will likely need to make use of a mix of different processes and knowledge. Our findings, based on the Numerical Operations task, provide an appropriate platform to examine the role of WM and in those common, but important school assessments.

In relation to the issue of task selection, some recent studies have suggested that visuospatial versus verbal WM tasks may have different explanatory power for math performance at different grades (e.g., Van de Weijer-Bergsma et al., 2014). We used WM and updating tasks that used both visual and verbal stimuli to maximise the likelihood that commonality in variance between tasks is resulted from relations with the same underlying construct. A reliable examination of the contributions of visuospatial versus verbal WM will
require multiple measures of each construct. This is perhaps an issue that can be addressed in future studies.

Regarding generalizability to other populations, it should be noted that though our sample was recruited from schools serving low to middle SES neighbourhoods, it is within an educational system where math performance is generally high (Mullis, Martin, Foy, & Arora, 2012). As noted earlier, our kindergarten children generally did well and exhibited less variance on the early numeracy items. This may have depressed the predictive power of early math performance on later achievement. Indeed, in a meta-analyses of large scale studies conducted in Canada, the US, and the UK, G. J. Duncan et al. (2007) found early math skills were more predictive than attention skills. However, even in that study, the relative importance of early math versus attentive skills was found to differ across data sets.

Conclusions

Our findings replicate and extend our knowledge regarding relations between WM, updating, and math performances in several ways. First, they show that from a cross-sectional perspective, the correlations between WM, updating, and math performance vary across grades; the strongest relations were observed in Grades 1 and 2. Second, from a longitudinal or predictive perspective, the explanatory power of WM and updating do not vary with grade. However, its importance relative to prior math achievement varies: prior achievement has a more important role in the senior grades. These findings are consistent with the view that relations between WM, updating, and math performance vary with grades and the domain of math under consideration. Of note is that this variation is driven not by changes in the predictive power of WM and updating; these relations are statistically invariant across the ten grades. Instead, variation only arose when the predictive power of WM and updating was considered in contrast to that of prior math achievement. The latter varied considerably across grades. Third, differences between the cross-sectional versus the predictive findings suggest
that there are grade related differences in the role of WM and updating in problem solving, versus in supporting the learning of new concepts and procedures.

An important contribution of this study is the finding that the rates at which WM and updating capacity grow do not exhibit individual differences. SES predicts WM and updating capacity at kindergarten; this, in turn, predicts rates of growth in math from Grade 1 to 9. Considered together, these findings further reinforce the importance of kindergarten and pre-kindergarten experiences in influencing both domain-general and domain-specific abilities.
References


investigation of the relationships between proficiencies in patterns, computation, executive functioning, and algebraic word problems. *Journal of Educational Psychology, 103*(2), 269-281. doi: 10.1037/a0023068


2011 Committee of Supply Debate]. Retrieved from
http://www.moe.gov.sg/media/speeches/2011/03/07/fy-2011-committee-of-supply-de-
1.php.

socioeconomic status in kindergarten children. *Developmental Science, 8*(1), 74-87. doi:
10.1111/j.1467-7687.2005.00394.x

Noël, M.-P. (2009). Counting on working memory when learning to count and to add: a
preschool study. *Developmental Psychology, 45*(6), 1630-1643. doi: 10.1037/a0016224

A review of developmental, individual difference, and cognitive approaches. *Learning
and Individual Differences, 20*(2), 110-122. doi: 10.1016/j.lindif.2009.10.005

J.M. Sattler.

Complex span versus updating tasks of working memory: The gap is not that deep.

Seethaler, P. M., Fuchs, L. S., Star, J. R., & Bryant, J. (2011). The cognitive predictors of
computational skill with whole versus rational numbers: An exploratory study. *Learning


number and fractions development. *Cognitive Psychology, 62*(4), 273-296. doi:
http://dx.doi.org/10.1016/j.cogpsych.2011.03.001

Updating & Math

memory have different relationships with different mathematical skills. *Journal of Experimental Child Psychology, 111*(2), 139-155. doi: 10.1016/j.jecp.2011.08.011


Wilhelm, O., Hildebrandt, A. H., & Oberauer, K. (2013). What is Working Memory

Table 1
Means, variance, and correlations for all working memory, updating, and math measures
### Kindergarten cohort

<table>
<thead>
<tr>
<th></th>
<th>NO1</th>
<th>NO2</th>
<th>NO3</th>
<th>NO4</th>
<th>LR1</th>
<th>MX1</th>
<th>PU1</th>
<th>LR2</th>
<th>MX2</th>
<th>PU2</th>
<th>LR3</th>
<th>MX3</th>
<th>PU3</th>
<th>LR4</th>
<th>MX4</th>
<th>PU4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ms</td>
<td>0.762</td>
<td>1.265</td>
<td>2.095</td>
<td>2.412</td>
<td>1.906</td>
<td>2.532</td>
<td>3.907</td>
<td>3.517</td>
<td>3.669</td>
<td>6.945</td>
<td>4.538</td>
<td>4.711</td>
<td>7.46</td>
<td>5.198</td>
<td>5.464</td>
<td>7.676</td>
</tr>
<tr>
<td>NO1</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO2</td>
<td>0.314</td>
<td>0.055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO3</td>
<td>0.319</td>
<td>0.494</td>
<td>0.091</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO4</td>
<td>0.409</td>
<td>0.45</td>
<td>0.629</td>
<td>0.103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR1</td>
<td>0.141</td>
<td>0.208</td>
<td>0.122</td>
<td>0.257</td>
<td>2.326</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX1</td>
<td>0.172</td>
<td>0.206</td>
<td>0.143</td>
<td>0.183</td>
<td>0.178</td>
<td>2.174</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU1</td>
<td>0.048</td>
<td>0.21</td>
<td>0.079</td>
<td>0.162</td>
<td>0.296</td>
<td>0.213</td>
<td>1.115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR2</td>
<td>0.17</td>
<td>0.251</td>
<td>0.266</td>
<td>0.382</td>
<td>0.201</td>
<td>0.149</td>
<td>3.115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX2</td>
<td>0.132</td>
<td>0.276</td>
<td>0.262</td>
<td>0.209</td>
<td>0.163</td>
<td>0.385</td>
<td>0.218</td>
<td>0.233</td>
<td>2.416</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU2</td>
<td>0.181</td>
<td>0.343</td>
<td>0.311</td>
<td>0.224</td>
<td>0.058</td>
<td>0.285</td>
<td>0.198</td>
<td>0.136</td>
<td>0.19</td>
<td>0.975</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR3</td>
<td>0.224</td>
<td>0.281</td>
<td>0.316</td>
<td>0.328</td>
<td>0.419</td>
<td>0.28</td>
<td>0.262</td>
<td>0.395</td>
<td>0.309</td>
<td>0.234</td>
<td>2.892</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX3</td>
<td>0.122</td>
<td>0.329</td>
<td>0.365</td>
<td>0.285</td>
<td>0.195</td>
<td>0.392</td>
<td>0.25</td>
<td>0.234</td>
<td>0.534</td>
<td>0.127</td>
<td>0.261</td>
<td>3.192</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU3</td>
<td>0.211</td>
<td>0.324</td>
<td>0.399</td>
<td>0.384</td>
<td>0.166</td>
<td>0.099</td>
<td>0.201</td>
<td>0.142</td>
<td>0.217</td>
<td>0.32</td>
<td>0.337</td>
<td>0.289</td>
<td>0.948</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR4</td>
<td>0.285</td>
<td>0.369</td>
<td>0.312</td>
<td>0.304</td>
<td>0.476</td>
<td>0.236</td>
<td>0.231</td>
<td>0.442</td>
<td>0.197</td>
<td>0.275</td>
<td>0.632</td>
<td>0.192</td>
<td>0.333</td>
<td>2.834</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX4</td>
<td>0.137</td>
<td>0.154</td>
<td>0.232</td>
<td>0.146</td>
<td>0.082</td>
<td>0.341</td>
<td>0.172</td>
<td>0.123</td>
<td>0.293</td>
<td>0.101</td>
<td>0.231</td>
<td>0.498</td>
<td>0.234</td>
<td>0.066</td>
<td>4.075</td>
<td></td>
</tr>
<tr>
<td>PU4</td>
<td>0.297</td>
<td>0.336</td>
<td>0.412</td>
<td>0.393</td>
<td>0.221</td>
<td>0.193</td>
<td>0.2</td>
<td>0.226</td>
<td>0.404</td>
<td>0.292</td>
<td>0.437</td>
<td>0.32</td>
<td>0.548</td>
<td>0.389</td>
<td>0.309</td>
<td>1.124</td>
</tr>
</tbody>
</table>

### Grade 2 cohort

<table>
<thead>
<tr>
<th></th>
<th>NO1</th>
<th>NO2</th>
<th>NO3</th>
<th>NO4</th>
<th>LR1</th>
<th>MX1</th>
<th>PU1</th>
<th>LR2</th>
<th>MX2</th>
<th>PU2</th>
<th>LR3</th>
<th>MX3</th>
<th>PU3</th>
<th>LR4</th>
<th>MX4</th>
<th>PU4</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO1</td>
<td>0.074</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO2</td>
<td>0.551</td>
<td>0.098</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO3</td>
<td>0.426</td>
<td>0.402</td>
<td>0.078</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO4</td>
<td>0.452</td>
<td>0.402</td>
<td>0.588</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR1</td>
<td>0.264</td>
<td>0.221</td>
<td>0.309</td>
<td>0.361</td>
<td>3.279</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX1</td>
<td>0.292</td>
<td>0.207</td>
<td>0.273</td>
<td>0.275</td>
<td>0.375</td>
<td>2.552</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU1</td>
<td>0.432</td>
<td>0.383</td>
<td>0.434</td>
<td>0.417</td>
<td>0.364</td>
<td>0.223</td>
<td>0.888</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NOTES:**
- NOS indicate the percentage of students that met or exceeded expectations.
- The **LR** (Learning Readiness) and **MX** (Mathematics) columns represent the percentage of students that scored within the expected range for each subtest.
- The **PU** (Performance) columns indicate the percentage of students who performed at or above the proficient level.

**M:** Mean score for the cohort.
<table>
<thead>
<tr>
<th>LR2</th>
<th>0.245</th>
<th>0.143</th>
<th>0.292</th>
<th>0.141</th>
<th>0.410</th>
<th>0.070</th>
<th>0.364</th>
<th>6.915</th>
</tr>
</thead>
<tbody>
<tr>
<td>MX2</td>
<td>0.191</td>
<td>0.219</td>
<td>0.177</td>
<td>0.342</td>
<td>0.139</td>
<td>0.428</td>
<td>0.263</td>
<td>0.175</td>
</tr>
<tr>
<td>PU2</td>
<td>0.375</td>
<td>0.462</td>
<td>0.385</td>
<td>0.458</td>
<td>0.485</td>
<td>0.328</td>
<td>0.590</td>
<td>0.268</td>
</tr>
<tr>
<td>LR3</td>
<td>0.147</td>
<td>0.177</td>
<td>0.322</td>
<td>0.277</td>
<td>0.536</td>
<td>0.216</td>
<td>0.249</td>
<td>0.316</td>
</tr>
<tr>
<td>MX3</td>
<td>0.311</td>
<td>0.166</td>
<td>0.203</td>
<td>0.261</td>
<td>0.335</td>
<td>0.396</td>
<td>0.164</td>
<td>0.311</td>
</tr>
<tr>
<td>PU3</td>
<td>0.363</td>
<td>0.342</td>
<td>0.377</td>
<td>0.405</td>
<td>0.462</td>
<td>0.156</td>
<td>0.540</td>
<td>0.413</td>
</tr>
<tr>
<td>LR4</td>
<td>0.145</td>
<td>0.177</td>
<td>0.339</td>
<td>0.307</td>
<td>0.502</td>
<td>0.222</td>
<td>0.318</td>
<td>0.314</td>
</tr>
<tr>
<td>MX4</td>
<td>0.318</td>
<td>0.308</td>
<td>0.260</td>
<td>0.272</td>
<td>0.282</td>
<td>0.409</td>
<td>0.182</td>
<td>0.159</td>
</tr>
<tr>
<td>PU4</td>
<td>0.337</td>
<td>0.304</td>
<td>0.297</td>
<td>0.329</td>
<td>0.378</td>
<td>0.152</td>
<td>0.495</td>
<td>0.280</td>
</tr>
</tbody>
</table>

### Grade 4 cohort

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NO1</td>
<td>0.085</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO2</td>
<td>0.658</td>
<td>0.161</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO3</td>
<td>0.660</td>
<td>0.678</td>
<td>0.363</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO4</td>
<td>0.638</td>
<td>0.625</td>
<td>0.805</td>
<td>0.424</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR1</td>
<td>0.271</td>
<td>0.309</td>
<td>0.364</td>
<td>0.329</td>
<td>3.209</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX1</td>
<td>0.252</td>
<td>0.305</td>
<td>0.281</td>
<td>0.263</td>
<td>0.3</td>
<td></td>
<td></td>
<td>4.549</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU1</td>
<td>0.380</td>
<td>0.422</td>
<td>0.463</td>
<td>0.365</td>
<td>0.383</td>
<td>0.354</td>
<td>1.122</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR2</td>
<td>0.352</td>
<td>0.253</td>
<td>0.320</td>
<td>0.198</td>
<td>0.497</td>
<td>0.248</td>
<td>0.382</td>
<td>2.866</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX2</td>
<td>0.321</td>
<td>0.252</td>
<td>0.234</td>
<td>0.178</td>
<td>0.217</td>
<td>0.431</td>
<td>0.189</td>
<td>0.319</td>
<td>5.241</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU2</td>
<td>0.328</td>
<td>0.426</td>
<td>0.468</td>
<td>0.485</td>
<td>0.430</td>
<td>0.365</td>
<td>0.571</td>
<td>0.387</td>
<td>0.242</td>
<td>1.075</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR3</td>
<td>0.356</td>
<td>0.433</td>
<td>0.308</td>
<td>0.344</td>
<td>0.485</td>
<td>0.341</td>
<td>0.413</td>
<td>0.410</td>
<td>0.247</td>
<td>0.418</td>
<td>4.915</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX3</td>
<td>0.147</td>
<td>0.260</td>
<td>0.255</td>
<td>0.281</td>
<td>0.268</td>
<td>0.273</td>
<td>0.235</td>
<td>0.161</td>
<td>0.584</td>
<td>0.329</td>
<td>0.328</td>
<td>7.683</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU3</td>
<td>0.382</td>
<td>0.353</td>
<td>0.4</td>
<td>0.449</td>
<td>0.384</td>
<td>0.397</td>
<td>0.498</td>
<td>0.416</td>
<td>0.450</td>
<td>0.595</td>
<td>0.452</td>
<td>0.409</td>
<td>0.491</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR4</td>
<td>0.323</td>
<td>0.368</td>
<td>0.339</td>
<td>0.352</td>
<td>0.473</td>
<td>0.304</td>
<td>0.408</td>
<td>0.405</td>
<td>0.317</td>
<td>0.477</td>
<td>0.492</td>
<td>0.381</td>
<td>0.513</td>
<td>3.589</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX4</td>
<td>0.316</td>
<td>0.381</td>
<td>0.292</td>
<td>0.310</td>
<td>0.259</td>
<td>0.431</td>
<td>0.293</td>
<td>0.212</td>
<td>0.4</td>
<td>0.265</td>
<td>0.329</td>
<td>0.358</td>
<td>0.373</td>
<td>0.382</td>
<td>7.763</td>
<td></td>
</tr>
<tr>
<td>PU4</td>
<td>0.265</td>
<td>0.319</td>
<td>0.393</td>
<td>0.433</td>
<td>0.425</td>
<td>0.360</td>
<td>0.627</td>
<td>0.387</td>
<td>0.250</td>
<td>0.541</td>
<td>0.439</td>
<td>0.320</td>
<td>0.632</td>
<td>0.530</td>
<td>0.241</td>
<td>0.426</td>
</tr>
</tbody>
</table>

### Grade 6 cohort
### Updating & Math

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NO1</td>
<td>0.382</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO2</td>
<td>0.751</td>
<td>0.381</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO3</td>
<td>0.651</td>
<td>0.777</td>
<td>0.389</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO4</td>
<td>0.7</td>
<td>0.831</td>
<td>0.858</td>
<td>0.339</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR1</td>
<td>0.214</td>
<td>0.258</td>
<td>0.183</td>
<td>0.209</td>
<td>4.475</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX1</td>
<td>0.369</td>
<td>0.41</td>
<td>0.289</td>
<td>0.323</td>
<td>0.33</td>
<td>5.631</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU1</td>
<td>0.202</td>
<td>0.316</td>
<td>0.257</td>
<td>0.297</td>
<td>0.226</td>
<td>0.33</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR2</td>
<td>0.297</td>
<td>0.296</td>
<td>0.291</td>
<td>0.33</td>
<td>0.517</td>
<td>0.212</td>
<td>0.207</td>
<td>2.974</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX2</td>
<td>0.289</td>
<td>0.292</td>
<td>0.201</td>
<td>0.223</td>
<td>0.371</td>
<td>0.542</td>
<td>0.231</td>
<td>0.342</td>
<td>5.368</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU2</td>
<td>0.279</td>
<td>0.398</td>
<td>0.315</td>
<td>0.374</td>
<td>0.15</td>
<td>0.326</td>
<td>0.415</td>
<td>0.378</td>
<td>0.274</td>
<td>0.521</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR3</td>
<td>0.209</td>
<td>0.266</td>
<td>0.26</td>
<td>0.26</td>
<td>0.516</td>
<td>0.2</td>
<td>0.173</td>
<td>0.544</td>
<td>0.318</td>
<td>0.253</td>
<td>5.952</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX3</td>
<td>0.296</td>
<td>0.354</td>
<td>0.299</td>
<td>0.307</td>
<td>0.407</td>
<td>0.567</td>
<td>0.174</td>
<td>0.232</td>
<td>0.659</td>
<td>0.172</td>
<td>0.373</td>
<td>7.428</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU3</td>
<td>0.342</td>
<td>0.438</td>
<td>0.464</td>
<td>0.448</td>
<td>0.267</td>
<td>0.261</td>
<td>0.412</td>
<td>0.357</td>
<td>0.26</td>
<td>0.556</td>
<td>0.338</td>
<td>0.267</td>
<td>0.371</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR4</td>
<td>0.38</td>
<td>0.442</td>
<td>0.452</td>
<td>0.428</td>
<td>0.522</td>
<td>0.407</td>
<td>0.347</td>
<td>0.635</td>
<td>0.39</td>
<td>0.411</td>
<td>0.529</td>
<td>0.459</td>
<td>0.458</td>
<td>3.239</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX4</td>
<td>0.307</td>
<td>0.347</td>
<td>0.276</td>
<td>0.217</td>
<td>0.408</td>
<td>0.47</td>
<td>0.269</td>
<td>0.307</td>
<td>0.507</td>
<td>0.248</td>
<td>0.411</td>
<td>0.607</td>
<td>0.222</td>
<td>0.499</td>
<td>7.872</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU4</td>
<td>0.295</td>
<td>0.447</td>
<td>0.492</td>
<td>0.463</td>
<td>0.287</td>
<td>0.349</td>
<td>0.528</td>
<td>0.391</td>
<td>0.283</td>
<td>0.529</td>
<td>0.304</td>
<td>0.255</td>
<td>0.634</td>
<td>0.41</td>
<td>0.239</td>
<td>0.381</td>
<td></td>
</tr>
</tbody>
</table>

*Note. NO = Numerical Operations; LR = Listening Recall; MX = Mr. X; PU = Pictorial Updating. The numeral at the end of each variable name denotes the wave at which data were collected. The diagonals refer to variances. The raw Listening Recall, Mr. X, and Pictorial Updating scores were divided by 2, 2.5, and 10, respectively, to bring their variance onto the same range. Maximum raw scores were 18, 16.8, 21.6, and 54 for the Listening Recall, Mr. X, Pictorial Updating, and math scores respectively.*
Table 2
Concurrent relations between updating and performance on the Numerical Operations task

<table>
<thead>
<tr>
<th>Grades</th>
<th>K2</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. observations</td>
<td>180</td>
<td>159</td>
<td>325</td>
<td>298</td>
<td>303</td>
<td>289</td>
<td>292</td>
<td>248</td>
<td>121</td>
<td>115</td>
</tr>
<tr>
<td>Fit Indices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.059</td>
<td>0.087</td>
<td>0.036</td>
<td>0.081</td>
<td>0.12</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFI</td>
<td>0.979</td>
<td>0.971</td>
<td>0.994</td>
<td>0.975</td>
<td>0.931</td>
<td>0.911</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRMR</td>
<td>0.015</td>
<td>0.024</td>
<td>0.007</td>
<td>0.019</td>
<td>0.005</td>
<td>0.027</td>
<td>0.043</td>
<td>0.054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>3.245</td>
<td>1.728</td>
<td>6.867</td>
<td>0.232</td>
<td>2.767</td>
<td>0.196</td>
<td>5.259</td>
<td>5.472</td>
<td>8.657</td>
<td></td>
</tr>
<tr>
<td>Factor loadings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Recall</td>
<td>0.508</td>
<td>0.375</td>
<td>0.544</td>
<td>0.394</td>
<td>0.582</td>
<td>0.586</td>
<td>0.513</td>
<td>0.682</td>
<td>0.511</td>
<td>0.794</td>
</tr>
<tr>
<td>Mr. X</td>
<td>0.411</td>
<td>0.453</td>
<td>0.497</td>
<td>0.325</td>
<td>0.47</td>
<td>0.404</td>
<td>0.649</td>
<td>0.523</td>
<td>0.482</td>
<td>0.568</td>
</tr>
<tr>
<td>Pictorial Updating</td>
<td>0.424</td>
<td>0.46</td>
<td>0.628</td>
<td>0.846</td>
<td>0.713</td>
<td>0.652</td>
<td>0.609</td>
<td>0.667</td>
<td>0.634</td>
<td>0.528</td>
</tr>
</tbody>
</table>

Correlation coefficients

| NO with WMU | 0.327*** | 0.658*** | 0.627*** | 0.501*** | 0.47*** | 0.551*** | 0.495*** | 0.484*** | 0.588*** | 0.547*** |

Note. NO = Numerical Operations; WMU = working memory and updating latent factor; AIC = Akaike information criterion; RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standardized root mean residual. ** < .01; *** < .001; † significantly different from Grade 1.
Figure 1. Cross-lagged autoregressive model of relations between WMU and performance on the Numerical Operations task for each cohort. Reported values are unstandardized. NO = Numerical Operations; LR = Listening Recall; MX = Mr. X; PU = Pictorial Updating. The letter or numeral at the end of each variable name denotes the grade at which data were collected. On account of the longitudinal nature of the data, residuals from the same measure were allowed to covary across time-points within each cohort. To improve clarity, they are not depicted here. Regression paths to school clusters are not shown.
Figure 2. Estimated slope coefficients for the Numerical Operations task (lower left panel) for WMU (upper left panel), and a schematic for the growth model used to estimate the coefficients (right). Residuals from the same measure were allowed to covary across time-points within each cohort, but are not depicted here. Regression paths to school clusters are not shown.