Learning and Individual Differences 54 (2017) 30–40

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

The influence of memory updating and number sense on junior high school math attainment

Miao-Hsuan Yen [a,1], Cheng-Ching Han [b,1], Pei-Chi Yu [c], Tsung-Han Yang [b], Daniele Didino [d,e], Brian Butterworth [b,f,g,2], Nai-Shing Yen [b,g,*,2]

[a] Graduate Institute of Science Education, National Taiwan Normal University, Taipei, Taiwan
[b] Department of Psychology, National Chengchi University, Taipei, Taiwan
[c] Tainan Municipal Yong-Ren Senior High School, Tainan, Taiwan
[d] Department of Economics, Tomsk Polytechnic University, Tomsk, Russia
[e] Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany
[f] Institute of Cognitive Neuroscience, University College London, London, UK
[g] Research Center for Mind, Brain, and Learning, National Chengchi University, Taipei, Taiwan

A R T I C L E I N F O

Article history:
Received 5 February 2016
Received in revised form 5 January 2017
Accepted 8 January 2017
Available online xxxx

Keywords:
Number sense
Working memory
Mathematics achievement
Approximate numerosity system
Memory updating task

A B S T R A C T

The present study investigated the development and influence of working memory abilities (WM) and number sense (NS) on mathematics achievement in junior high school students (grades 7–9, $N = 267$). Math achievement was measured by three sectional examinations in a semester, NS was indicated by an approximate numerosity system task, and WM was assessed by a battery of four tasks. Developmental trends in both WM and NS task scores were observed. Memory updating (MU) in the WM tasks was found to be dominant in predicting math achievement in correlation and regression analyses. A similar pattern was observed for separate analyses across grade levels, except that in grade 7 a significant unique contribution of NS to math was observed after taking MU into account. The findings suggest that WM ability (especially that used in MU task) had greater influence on math achievement than NS.

1. Introduction

Basic mathematical ability is of critical importance for high school students. It is fundamental for learning science in school and poor mathematical ability is a serious handicap in the workplace and in daily life (Bynner & Parsons, 1997; Parsons & Bynner, 2005). A UK study found that poor levels of mathematical ability are a major cost to society (Gross, Hudson, & Price, 2009), and improvements in national levels of mathematics ability promote economic growth (OECD, 2010). The junior high school students (grades 7–9) tested in the present study were required to be competent in a wide range of mathematical topics, such as linear and quadratic equations, formulas for polynomials and square roots, similar geometrical figures, properties of triangles, and probability. Mathematics attainment has been shown to be related to a wide range of cognitive and perceptual abilities including number sense, visuo-spatial ability, and the domain-general ability to maintain and manipulate information, usually labeled working memory (WM) and executive functions (e.g., Chen & Li, 2014; De Smedt, Noël, Gilmore, & Ansari, 2013; Fazio, Bailey, Thompson, & Siegler, 2014; Raghubar, Barnes, & Hecht, 2010; for a review). Nevertheless, there is some evidence of domain-specificity in WM (Butterworth, Cipolotti, & Warrington, 1996; Iuculano, Moro, & Butterworth, 2011), and it is maintaining and manipulating numerical information that is linked to mathematical attainment, at least for 8- and 9-year-old children (Iuculano et al., 2011). However, research about the contribution of number sense (NS) and WM to math performance mainly focuses on preschool and elementary students (e.g., De Smedt et al., 2013; Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013; Halberda & Feigenson, 2008; Iuculano et al., 2011); there is little evidence about which of these factors is of particular importance to attainment in junior high school. The aim of the present study was to investigate the role of NS and aspects of WM on individual differences in math achievement in junior high school students.

1.1. Number sense and mathematics

Number sense (NS) is the ability to represent and manipulate numerical quantities (Dehaene, 2001). It is held to reflect the operation

* Correspondence to: N.S. Yen, Department of Psychology, National Chengchi University, No. 64, Sec 2, Zhih-Nan Rd., Taipei 11605, Taiwan.
E-mail address: nsy@nccu.edu.tw (N.-S. Yen).
[1] These authors contributed equally to this work.
[2] These authors are corresponding authors.

http://dx.doi.org/10.1016/j.lindif.2017.01.012
1041-6080/

Fig. 1. Experimental procedure for the number sense (NS) task (an example based on the paradigm used by Halberda et al. (2008)).

of the Approximate Number System (ANS, sometimes called the Analogue Magnitude System), which makes approximate estimates of the numerosity of a visual display and maps them onto compressed overlapping analog representations. The operation of NS is present in human infants and in many other species (Nieder & Dehaene, 2009). In the literature, NS is typically assessed by the ability to select the larger of two arrays of dots, often with the spreading areas of the dots controlled. One standard procedure requires selecting the more numerous of two intermixed arrays of blue and yellow dots (see Fig. 1). Typically, individual differences are assessed psychometrically in terms of the proportional differences in quantities between two sets of dots an individual can reliably discriminate, as in Halberda, Mazzocco, and Feigenson (2008). Halberda and Feigenson (2008) documented an increasing trend in NS acuity from age 3–6 years to adulthood. Similarly, from a large online assessment from age 11–85 years, Halberda, Ly, Wilmer, Naiman, and Germine (2012) observed developmental improvement peaking around 30 years old.

Halberda et al. (2008) were the first to discover that an individual's psychometric function, called numerical acuity (here expressed as an "internal Weber fraction" to capture the idea that the internal neural representations of numerosity are log compressed), correlates with arithmetical ability. Sixty-four children, aged 14 years and without learning disabilities, participated in their study. Sets of math achievement tests were collected annually from their kindergarten years up to grade 6 (ages 5–11). Positive correlations were observed between numerical acuity assessed at 14 and mathematics test scores in each of the early years, suggesting that NS was highly associated with math achievement. In addition to the retrospective correlation found by Halberda et al. (2008), Libertus, Feigenson, and Halberda (2011) also documented a positive correlation between numerical acuity and math ability in 200 children aged 3–5 years. Similarly, Mazzocco, Feigenson, and Halberda (2011) found that numerical acuity measured in children aged 3 or 4 years can predict math performance two years later (see also Gilmore, McCarthy, & Spelke, 2010).

These studies suggest a stable relationship between NS and school mathematics performance from children aged 3–11 years[3]. However, whether NS is still relevant in the later stages in school has remained unclear. The relationship between NS and mathematics performance was not consistently observed in the literature. For instance, Libertus, Odic, and Halberda (2012) observed a positive correlation between NS acuity and test scores in the Quantitative section of the Scholastic Aptitude Test (SAT). In the large online assessment, Halberda et al. (2012) also found that NS precision correlated with self-reported mathematics performance in school and SAT-Quantitative scores. However, Wei, Yuan, Chen, and Zhou (2012) showed that while spatial abilities correlated with advanced math concepts of undergraduate students, basic numerical processing did not. Several other studies failed to find the significant correlations between approximation skills and math achievement from childhood to adulthood (e.g., De Smedt et al., 2013; Inglis, Attridge, Batchelor, & Gilmore, 2011; Iuculano, Tang, Hall, & Butterworth, 2008).

The inconsistencies in the contribution of NS to math achievement may result from different measures of NS used in particular studies. Inglis and Gilmore (2014) compared four major measures of NS acuity (Weber fraction, accuracy, and numerical ratio effect – consisting of accuracy and RT), and found that accuracy measure was more reliable than Weber fraction, which was better than the two measures for numerical ratio effect. However, in the review of De Smedt et al. (2013), no patterns associated with differences in measures emerged from studies with positive and negative results. In addition, inconsistent findings may stem from the fact that NS was assessed by various types of approximation tasks in different studies. For instance, Xenidou-Dervou, De Smedt, van der Schoot, and van Lieshout (2013) found that the contribution of non-symbolic approximation to math achievement in preschoolers was mediated by symbolic approximation. Mundy and Gilmore (2009) found that the mapping ability between symbolic and non-symbolic numerical representations predicted math attainment in children aged 6–8 years, above and beyond symbolic and non-symbolic tasks alone.

In junior high school, students start learning a wide range of complicated mathematical concepts and dealing with mathematical problems involving problem solving beyond simple calculations (Best, Miller, & Naglieri, 2011). Although a correlation between NS precision for junior high school students and their self-reported math performance in school was observed in the online study of Halberda et al. (2012), the relationship between NS and math performance has mainly been investigated with either younger children or adults as reviewed above (see

[3] It should be noted that in the study of Halberda et al. (2008), NS acuity was assessed at age 14 and correlated with math achievement assessed annually from ages 5–11.

also De Smedt et al., 2013 for a review). Our research was motivated by the fact that there were only a few studies investigating the developmental trend in NS and its role in math performance in adolescence. In the present study, we aimed to bridge the missing gap by examining the role of NS on individual differences in junior high school students' math achievement.

1.2. Working memory and mathematics

The math curriculum in junior high school goes beyond simple numerical tasks, and involves complex multi-step operations on numbers, symbols, and spatial configurations. Thus, the three components of working memory (WM) proposed by Baddeley and Hitch (1974) may be involved. Specifically, phonological loop and visuo-spatial sketchpad are responsible for maintaining verbal and visuo-spatial information, respectively, for further processing. The central executive (CE) component was proposed to allocate cognitive resource and coordinate these two slave systems.

Several studies have found that the CE is a key component related to math attainment (Bull & Scerif, 2001; Cragg & Gilmore, 2014; Iuculano et al., 2011; Van der Ven, Kroesbergen, Boom, & Leseman, 2012). In particular, sub-functions of the CE all play a role in math performance: *inhibiting* irrelevant information or suppressing inappropriate strategies (e.g., Bull & Scerif, 2001; Gilmore et al., 2013; St Clair-Thompson & Gathercole, 2006), *shifting* between mental sets, operations, and strategies (e.g., Bull & Scerif, 2001; Yeniad, Malda, Mesman, van IJzendoorn, & Pieper, 2013), and *updating* information (e.g., Bull & Scerif, 2001; Van der Ven et al., 2012) in WM. Among the three CE sub-functions, updating consistently plays an important role in math performance (e.g., Kolkman, Hoijtink, Kroesbergen, & Leseman, 2013; Lechuga, Pelegrina, Pelaez, Martin-Puga, & Justicia, 2016; Van der Ven et al., 2012). In a meta-analysis comparing tasks probing the two slave systems and the three sub-functions of CE, Friso-van den Bos et al. (2013) found that the correlation with math performance was the highest for updating task with verbal material, followed by updating with visuo-spatial material, visuo-spatial sketchpad, phonological loop, inhibition, and shifting tasks. However, variations existed for different types of mathematics measures. The national curriculum tests, composite measures or teacher ratings had higher correlations with WM components than other mathematics measures such as counting and basic understanding of numerical concepts, simple arithmetic, advanced arithmetic, and word problems. These results are consistent with claims in Best et al. (2011). They proposed that general mathematical problem solving involves strategy formulation and self-monitoring, and depends more on CE functions than calculation.

Differential developmental trends have been observed for different WM components. Gathercole, Pickering, Ambridge, and Wearing (2004) demonstrated that performance of WM tasks improved from years 4 to 15. However, developmental trends in tasks with increasing demands on WM reached asymptote at later ages (e.g., the ability of maintenance and manipulation of multiple spatial units improves until 13–15 years old; self-organization strategy advances until 16–17 years old; Luciana, Conklin, Hooper, & Yarger, 2005). Linares, Bajo, and Pelegrina (2016) further showed age-related difference in sub-functions of CE in 4 age groups (i.e., 8-, 11-, 14-, and 21-year-olds). As WM ability progresses in information processing, the contribution of different WM components to math achievement varies across grade levels for different math problems as well. For example, Holmes and Adams (2006) found that CE had consistent and large contribution on curriculum-based math assessment for both grades 3 and 5 (abbreviated as G3 and G5). Phonological loop did not predict math achievement above and beyond CE and visuo-spatial sketchpad, except for the significant correlation with mental arithmetic for G5, presumably because G5 used subvocal rehearsal for retention of intermediate results. On the other hand, younger children (G3) relied on visuo-spatial sketchpad when solving mathematics problems while older children resorted to it in difficult problems. Similarly, De Smedt et al. (2009) also found that CE predicted math achievement for both G1 and G2, while phonological loop and visuo-spatial sketchpad were better predictors for G2 and G1, respectively. Furthermore, in the meta-analysis of Friso-van den Bos et al. (2013), there was a negative correlation between age of the sample (4–12 years) and effect size of visuo-spatial sketchpad on math performance, while the correlation was positive between age and effect size of updating with visual material. It is possible that the updating ability is more important in solving more complicated math problems. In a recent study (Han, Yang, Lin, & Yen, 2016), memory updating (MU) ability for numerical and spatial operations was found to highly correlate with the accuracy of multi-digit mental multiplication in young adults, especially with the complex problem like a two-digit number multiplied by a two-digit number (e.g., 35×67) rather than an easier problem like a two-digit number multiplied by a one-digit number (e.g., 35×4). Further evidence also showed that, for 14-year-old children, achievements in mathematics and science were strongly correlated with complex WM test scores, such as measures on the CE capacities (backward digit recall and listening recall) and the phonological loop (word list recall and word list matching) (Gathercole, Pickering, Knight & Stegmann, 2004). All of these results support the role of WM ability in complex math performance.

Although the development in WM itself has been investigated from as young as 4 years to early 20s (e.g., Gathercole, Pickering, Ambridge et al., 2004; Linares et al., 2016; Luciana et al., 2005; Swanson, 1996), age-related differences in the contribution of WM to math achievement in most studies were examined before high school education stages (see Raghubar et al., 2010, for a review). In the present study, we further examined the role of different aspects of WM in math achievement of junior high school students aged 13–15 years. Many studies have shown that complex span tasks have higher correlations with measures of higher-order cognition than simple span tasks (e.g., Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Conway & Engle, 1996; Engle, Tuholski, Laughlin, & Conway, 1999; Unsworth & Engle, 2005, 2007; Unsworth, Redick, Heitz, Broadway, & Engle, 2009). Correlation between performances of complex span and standardized mathematics test has also been reported in the literature (e.g., Bayliss, Jarrold, Gunn, & Baddeley, 2003). In addition, as aforementioned, updating was associated with math attainment (e.g., Lechuga et al., 2016; Van der Ven et al., 2012). As these two types of WM tasks demand the CE component of WM, they were found to correlate with each other (Schmiedek, Lövdén, & Lindenberger, 2014; Wilhelm, Hildebrandt, & Oberauer, 2013). Nevertheless, differences exist between these two tasks. In a complex span task, participants have to maintain a set of items (e.g., a series of letters) interleaved with performing another task (e.g., arithmetic judgment in the operation span task, Turner & Engle, 1989). Thus, participants have to switch between memorizing the items and processing an irrelevant task. Conversely, in an updating task, participants keep updating several items (i.e., substituting outdated items with the encoding of new items). In some versions of updating tasks, transformation (e.g., arithmetic operation or mental rotation) is required (Ecker, Lewandowsky, Oberauer, & Chee, 2010). Contrary to the complex span task, there is no unrelated secondary task to be processed. As both tasks are associated with math performance, in the present study, a WM battery (Lewandowsky, Oberauer, Yang, & Ecker, 2010) including both tasks was adopted to investigate the relationship between WM and math performance of junior high school students.

1.3. Number sense versus working memory

Although both NS and WM abilities have been shown to correlate with mathematics achievement, it has been argued that NS was not implicated in number reasoning and arithmetic (Butterworth, 2010). Thus, it remains unclear whether NS is a good predictor for the skills needed for reasoning and manipulation of numbers or spatial configuration.

Xenidou-Dervou, van Lieshout, and van der Schoot (2014) found that performing the NS task demanded WM. In their study, children aged 5–6 years performed the NS task with a secondary WM task (separately for phonological loop, visuo-spatial sketchpad, and CE). The results showed that the CE task had the largest interference, indicating that WM, especially CE function, was involved in NS. In addition, Hassinger-Das, Jordan, Glutting, Irwin, and Dyson (2014) found that CE was a partial mediator between NS assessed during kindergarten and three measures (i.e., applied problems, calculation, and NS) assessed at grade 1. However, with confirmatory factor analyses and structural equation modeling, Xenidou-Dervou et al. (2013) found that the approximation task performance of kindergarteners contributed to math performance (counting and addition) above and beyond WM, and vice versa. In addition, symbolic approximation mediated the effect of non-symbolic approximation and WM on math attainment. Thus, although both NS and WM contribute to math performance (also see Hornung, Schiltz, Brunner, & Martin, 2014), the relative contribution of WM and NS to math performance remains controversial. In addition, there is no evidence for the relationship among WM, NS, and math attainment at high school level.

1.4. Purposes of the present study

In summary, the inconsistencies for the contribution of WM and NS to mathematic achievement in previous studies (Hassinger-Das et al., 2014; Xenidou-Dervou et al., 2013) may result from the specific tasks used in different research with different ages. In our study, a WM test battery with complex span, updating and spatial memory tasks was used to investigate the relative contribution of WM and NS in junior high school students who are obliged to study reasonably complex mathematics. Our second purpose was to document the developmental trend in WM, NS, and their relative contribution to math performance, which would show us the general pattern among WM, NS, and math of junior high school students and whether there is any specific pattern at a particular grade level. As subject content in the mathematics class becomes more demanding, the reliance of mathematics performance on basic NS might decrease, while the role of other abilities such as visuo-spatial ability, reasoning, and management of strategy use might become more relevant. Thus, a substantial contribution from WM compared to NS might be expected for mathematics attainment of junior high school students.

2. Method

2.1. Participants

Two hundred and sixty-seven students (grades 7–9, abbreviated as G7–G9) in a junior high school in southern Taiwan participated in this experiment. This study was conducted with the school's approval. The homeroom teacher of each class was also informed about the purpose and procedure of this study. In addition, the ethical guidance for educational research established by the Human Research Protection Program of National Science Council (NSC-HRPP) in Taiwan was followed. Originally, four classes were randomly selected from each grade for participation (total, 353 students). Students were excluded from analysis if they did not complete the whole experiment, did not have a complete record of three sectional achievement tests in school (e.g., if they transferred from another school during the semester), or had disabilities or a special demographic background (such as Taiwan aboriginal people and children of denizens, who may have low learning performance due to cultural or family issues). The final sample consisted of 131 females and 136 males: 86 students in grade 7 (female: 46; male: 40), 79 students in grade 8 (female: 38; male: 41), and 102 students in grade 9 (female: 47; male: 55). The range of ages in the whole sample was 12.3–15.3. The mean ages and standard deviations from grades 7–9 were 12.9 (0.54), 13.8 (0.30), and 14.7 (0.30), respectively.

2.2. Design

The aims of the present study were to investigate the developmental trend in WM and NS, as well as their relationship with mathematics achievement of junior high school students. First, we were interested how children in grades 7–9 differed in WM and NS task performance and mathematics achievement. Second, correlational and hierarchical regression analyses were conducted to explore the relationships among these factors.

2.3. Materials

WM and NS tasks will be described in detail in the following sub-sections, and achievement tests will also be presented.

2.3.1. Working memory tasks (WM)

The WM test battery developed by Lewandowsky et al. (2010) was adopted in the present study. This battery was designed to include tasks with both verbal (including numerical) and spatial material. In addition, updating, maintenance while processing unrelated stimuli, and relational integration were assessed with the memory updating (MU), operation span (OS) and sentence span (SS), and spatial short-term memory (SSTM) tasks, respectively. In their study, the battery was validated in three experiments conducted in two languages (English and Chinese; in the present study, we employed the Chinese version), involving >350 participants. The tasks were found to load on a single latent variable, which was found in a subsequent experiment to correlate highly with performance on Raven's matrices test of fluid intelligence (De Lemos & Raven, 1989). The battery consists of four tasks: an MU task, an OS task, an SS task, and an SSTM task; see Fig. 2.

First, in each trial of the MU task, there were three, four or five rectangular frames shown on the screen simultaneously (Fig. 2-A), and a single digit was presented for 1 s in each of the frames in sequence (e.g., "7" or "5"). Participants had to remember the digit for each particular frame. Next was the updating period. An arithmetic operation (such as "-4" or "$+3$") appeared in a particular frame for 1.3 s, and participants had to calculate the result in that frame (e.g., $7-4=3$). The number of updating operations varied from two to six in each trial, and this updating process continued until a question mark "?" appeared in each frame. This was the recall period, in which participants were required to enter the memorized updated results into each frame. Therefore, in this task, participants should remember the original digit and do one-digit addition or subtraction for each frame, then keep updating until reporting the final answer at the end of the trial. There was no time constraint on the response, but participants could not change the answer once it was entered. Before testing, participants completed two practice trials to familiarize themselves with the task procedure, and there were 15 experimental trials in total. Task accuracy was the dependent measure collected.

Second, in each trial of the OS task, a fixation cross was shown at the center of the screen for 1.5 s, then an arithmetic equation (e.g., $2+8=5$) was presented on the screen for 3 s for participants to judge whether it was correct or not. Thereafter, a letter (e.g., H) appeared on the screen for 1 s for participants to memorize. The sequence (arithmetic equation and letter) was repeated between four and eight times in each trial. At the end of each trial, participants had to recall all the letters in the presented order when a question mark "?" appeared on the screen. Therefore, what participants had to do was memorize each letter correctly in the presented sequence (see Fig. 2-B). There was no time constraint on the response, but participants could not change the answer once it was entered. After entering all letters, the next trial proceeded. Before testing, participants completed three practice trials to familiarize themselves with the experimental procedure, and the formal task comprised 15 experimental trials. Since the equation judgment was a distractor, task accuracy was collected from the correctness of the memorized letter sequence.

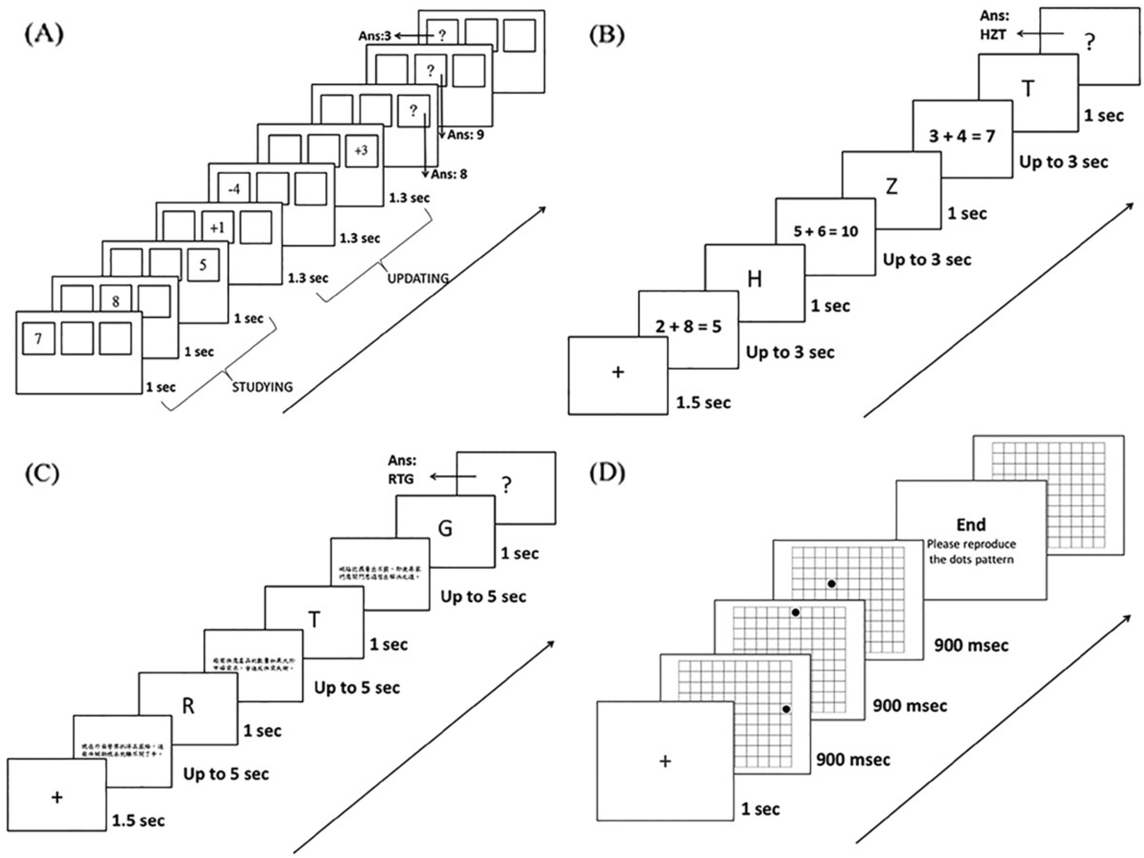

Fig. 2. Experimental procedures for 4 tasks in working memory (WM) test battery: (A) memory updating task (MU), (B) operation span task (OS), (C) sentence span task (SS), and (D) spatial short-term memory task (SSTM).

Third, in the SS task, similar in structure to the OS task, in each trial a fixation cross was shown at the center of screen for 1.5 s, and then a Chinese sentence was presented on the screen for 5 s for participants to judge whether it was correct or not. Thereafter, a letter (e.g., R) appeared on the screen for 1 s for participants to memorize. The sequence (sentence and letter) repeated between three and seven times in each trial. At the end of each trial, participants had to recall all the letters in the presented order when a question mark "?" appeared on the screen. Therefore, what participants had to do was memorize each letter correctly in the presented sequence as they did in the OS task (see Fig. 2-C). There was no time constraint on the response, but participants could not change the answer once it was entered. After entering all letters, the next trial proceeded. Before testing, participants completed three practice trials to familiarize themselves with the experimental procedure, and the formal task comprised 15 experimental trials. Again, task accuracy was collected from the correctness of the memorized letter sequence.

Fourth, in each trial of the SSTM task, participants saw a black grid of 10×10 cells presented on a white screen. At the beginning, a cross was shown at the center of screen for 1 s, and then a black dot was shown in one of the cells for 900 ms. Between two and six dots were shown in different cells, successively (Fig. 2-D). Participants had to memorize the relative positions of the dots appearing on the grid in each trial. After the dots had been displayed, "End – Please reproduce the dots pattern" was shown on the screen. Participants had to reproduce the pattern of dots in an empty 10×10 grid on the screen by clicking the dot positions. If needed, they could delete any dot by clicking on it again. Each answer would be regarded as correct if the relative spatial relationships between the dots were reproduced correctly. There was no time constraint on the response. After recalling the relative position of dots in that trial, participants clicked the "next" button on the screen to proceed to the next trial. There were two practice trials for familiarizing the method before testing, and the formal task comprised 30 experimental trials. Task accuracy was the dependent measure collected.

Scoring of the four WM tasks was conducted with the analysis package provided by the WM test battery (Lewandowsky et al., 2010). In MU and the two span tasks, proportional correctness in each trial was calculated (e.g., the score 5/6 represented that there were five correct answers with six items to be memorized in a trial). Next, scores in all trials were averaged. For each participant the maximum score was 1 and the minimum was 0. The scoring in SSTM task was based on similarity. Two points were awarded for each dot if participants clicked on the exact position, while one point was assigned if the reported position was within one grid from the correct position. The final score for each participant was the sum of points received, and divided by the maximal number of points (if all dots were reported correctly). The maximum score for each participant was 1 and the minimum was still 0. The reliability measures (Cronbach's α) for MU, OS, SS, and SSTM tasks were 0.90, 0.89, 0.90, and 0.93, respectively.

2.3.2. Number sense (NS; the approximate numerosity system task)

The task was based on the one used by Halberda et al. (2008). In this task, blue and yellow dots of various sizes were presented simultaneously on a gray screen (Fig. 1). The number of blue dots was more than that of yellow dots in half of the trials. The ratios of quantity for the two dot groups were 2 (2:1), 1.33 (4:3), 1.2 (6:5), and 1.14 (8:7), respectively. In half of the trials, the total pixels occupied by blue and yellow dots were the same (area-controlled trials). In the other half, the average sizes of blue and yellow dots were the same (dot-size-controlled trials.)

Each trial began with a fixation cross for 1–1.5 s, and then the dots appeared on the screen for 200 ms. After a 500 ms interval, a black question mark "?" appeared on the center of screen, reminding participants to press the button N or M on the keyboard to indicate which group of dots had a higher quantity. In the whole task, the correspondence between answer key (i.e., N or M) and quantity (i.e., more blue or more yellow) was counter-balanced. There was no time constraint on the response, and after entering, the next trial proceeded.

There was a practice block before the task, with 12 trials for participants to familiarize themselves with the task procedure (three trials for each of the four ratios). After practicing, participants performed the experimental task, which consisted of 320 trials (120 trials respectively for the ratios 1.14 and 1.2, and 40 trials, respectively, for the ratios 1.33 and 2); all were presented randomly in 10 blocks (32 trials in each block). Participants received accuracy feedback in the practice block, but no feedback appeared in the experimental task. For data analysis, accuracy rates for items with four different ratios were calculated for each participant. The reliability measures (Cronbach's α) of ratios 2, 1.33, 1.2, and 1.14 were 0.82, 0.56, 0.61, and 0.50, respectively.

2.3.3. Mathematics achievement tests (MATH)

The average scores of three sectional mathematical examinations during one semester was the index for mathematics achievement. The topics covered in each grade were as follows. In grade 7, the topics included arithmetic operations of integers and fractions, exponentials, scientific notations, common factors and common multiples, and linear equations with a single unknown. In grade 8, students learned multiplication formulas, formulas for polynomials and square roots, Pythagoras Theorem in geometry, factorization, and quadratic equations with a single unknown. In grade 9, there were similar geometrical figures, properties of circles, properties of triangles, and probability. In each examination, test items generally included 10–14 multiple choice questions, 7–15 fill-in-the-blank questions, and 2–6 calculational questions or word problems. The duration of each examination was 50 min. Students at each grade level took the examination at the same time in their own classrooms. During the exam, teachers monitored the students to ensure that there was no cheating among students. Questions about test items would be clarified without further conceptual explanation. After the exam, scoring was conducted by the instructors. There were standard scoring rubrics for each exam; 3–4 points were rewarded for each multiple choice, fill-in-the blank question, and sub-item in calculational questions or word problems.

The score of each sectional examination in each grade was first normalized. Next, three normalized exam scores were averaged for each student. The mean of the score was set at 0.5 and the standard deviation was set at 0.15. The range of scores was between 0 and 1.

2.4. Procedure

Both the WM test battery and NS task were administered in a computer lab. Each class of students participated in the five tasks as a group. Each task was completed in about 10 min; the total amount of time was approximately 50 min. Half of the students completed the NS task before the WM tasks, and vice versa. The WM tasks were performed following the fixed order MU, OS, SS, and SSTM. The WM and NS tasks were administrated from November 2011 to January 2012; while the three sectional examinations were held during the semester from September 2011 to January 2012. Two consecutive sectional examinations were separated by 6–7 weeks.

3. Results

There were two sets of analyses. First we investigated the developmental trend in WM tasks, NS tasks, and math achievement tests across grades 7–9 in three separate analyses of variance. Second, the correlations among WM tasks, NS tasks, and math achievement tests were analyzed. Then, the relative contribution of background factor (i.e., grade) and observational factors (performance on WM and NS tasks) to math achievement scores was examined by hierarchical regression analyses. Separate correlation and regression analyses were also conducted to reveal differential patterns according to grade levels.

3.1. Analyses of variance

The means and standard deviations in WM tasks, NS tasks, and three sectional examination scores as a function of grade are presented in Table 1.

3.1.1. Working memory tasks

Separate one-way ANOVA was conducted to examine grade effect on each WM task. There were significant main effects of grade on all WM tasks (Fs > 7.053, all $ps < 0.01$). Post hoc analyses revealed that in MU task, G7 and G8 had significantly lower scores than G9 (both $ps < 0.05$), while there was no difference in scores between G7 and G8. In OS and SS tasks, G8 and G9 had significantly higher scores than G7 (all $ps < 0.01$), but there was no difference in scores between G8 and G9. In the SSTM task, the performance increased as a function of grades, and all pairwise comparisons were significant (all $ps < 0.01$). The improvement on performance with grade supports the validity of these tasks as the measures of cognitive abilities. Therefore, as age increases, the cognitive abilities of junior high school students mature.

3.1.2. Number sense task

There were four types of items that differed in the ratio between blue and yellow dots in the NS task, and the item difficulty increased as the ratio decreased. Thus, item difficulty was included in the analysis as a within subject factor, and grade was a between subject factor, in a two-way ANOVA. Significant main effects of grade ($F = 11.855$, $p < 0.001$) and difficulty ($F = 819.911$, $p < 0.001$) were observed. G9 had significantly higher scores than G7 and G8 (both $ps < 0.001$) while the latter groups did not differ from each other ($p > 0.999$). Performance decreased as the item difficulty increased, with significant pairwise comparisons (all $ps < 0.001$). There was a significant interaction effect between grade and difficulty ($F = 3.347$, $p < 0.01$). Follow-up interaction contrasts revealed that G7 and G9 differed in the comparisons between ratios 2 vs. 1.2, ratios 2 vs. 1.14, ratios 1.33 vs. 1.2, and ratios 1.33 vs. 1.14; while G8 and G9 differed in comparisons between ratios 2 vs. 1.2 and ratios 2 vs. 1.14 (all $ps < 0.05$). For the difficult items ($R = 1.14$ and 1.2), the improvement from G7 to G9 was smaller than that for the easier items ($R = 1.33$ and $R = 2$). The difference between the two easy and the two difficult items was significantly different between G7 and G9. For G8 and G9, the differential pattern was significant only when compared to the easiest item ($R = 2$). The finding

Table 1
The means and standard deviations (in parentheses) on accuracy of 4 working memory (WM) task scores, NS task scores with 4 different ratios ($R = 2$, 1.33, 1.2, 1.14), and 3 sectional math examination scores as a function of grade.

	G7	G8	G9	All
Working memory tasks				
MU	0.53 (0.19)	0.56 (0.2)	0.63 (0.19)	0.58 (0.20)
OS	0.53 (0.20)	0.62 (0.17)	0.66 (0.17)	0.61 (0.19)
SS	0.54 (0.23)	0.67 (0.20)	0.66 (0.20)	0.63 (0.22)
SSTM	0.78 (0.08)	0.82 (0.08)	0.85 (0.06)	0.82 (0.08)
Number sense task				
$R = 2$	0.83 (0.12)	0.83 (0.15)	0.90 (0.09)	0.86 (0.12)
$R = 1.33$	0.69 (0.10)	0.70 (0.12)	0.75 (0.09)	0.72 (0.11)
$R = 1.2$	0.63 (0.07)	0.62 (0.07)	0.66 (0.07)	0.64 (0.07)
$R = 1.14$	0.60 (0.05)	0.59 (0.07)	0.62 (0.06)	0.61 (0.06)
Overall	0.65 (0.06)	0.64 (0.07)	0.69 (0.06)	0.66 (0.06)
Sectional examinations of math				
Test 1	0.50 (0.15)	0.56 (0.14)	0.54 (0.15)	0.53 (0.15)
Test 2	0.52 (0.16)	0.57 (0.17)	0.54 (0.16)	0.54 (0.16)
Test 3	0.52 (0.17)	0.57 (0.16)	0.54 (0.16)	0.54 (0.16)
Mean	0.51 (0.15)	0.57 (0.15)	0.54 (0.15)	0.54 (0.15)

Note: G7, G8, G9: grades 7, 8, 9 respectively.

that differential patterns between G8 and G9 were fewer than those between G7 and G8 signifies developmental improvement. To sum up, the main finding was that performance in NS task increased with grade levels as WM abilities did. Nevertheless, this decreased when the item difficulty increased. These findings were similar to the findings from Halberda and Feigenson (2008).

3.1.3. Mathematics achievement tests

The examination scores were all normalized, because the content in each sectional examination and between grades was different. To ensure that there were no difference among grade and sections, a 3×3 (grade × section) ANOVA was conducted. In the omnibus analysis, there was no significant main effect of grade ($F = 2.607$, $p = 0.076$). Although the main effect of section was significant ($F = 3.038$, $p = 0.049$), the follow-up pairwise comparisons were not significant (all $ps > 0.07$). The interaction between grade and section was not significant ($F = 1.050$, $p > 0.38$).

3.2. Correlational and regression analyses

3.2.1. Correlational analysis

Table 2 presents correlations among math score, WM and NS tasks. The correlation matrix indicates that all variables were significantly correlated with each other ($rs > 0.3$, all $ps < 0.001$). Overall, MU had the highest correlation coefficient with math score and had higher correlations with other tasks than the remaining pair-wise correlations except that between the span tasks (OS and SS). Since MU may be involved in the correlations of other WM or NS tasks with math score, partial correlations among variables without MU score were calculated. After the MU score was partialled out, the correlation between math achievement score and other WM tasks and NS were no longer significant (OS: $r = 0.115$; SS: $r = 0.030$; SSTM: $r = 0.002$; NS: $r = 0.068$). When OS, with the second largest correlation with math achievement, was partialled out, all of the correlations except SS were still significant (MU: $r = 0.481$, $p < 0.001$; SS: $r = 0.054$, $p > 0.37$; SSTM: $r = 0.122$, $p < 0.05$; NS: $r = 0.178$, $p < 0.01$).

When correlational analyses were conducted for each grade level separately, similar patterns of results were obtained. As shown in Appendix 1, all WM and NS tasks significantly correlated with each other and with math scores except for SS and NS at grade 9 ($r = 0.163$, $p > 0.10$). Analogous to the analysis with the whole data, when MU was partialled out, the correlations with math score was not significant ($|r|s < 0.19$, $ps > 0.09$) except for a significant correlation between NS and math score at grade 7 ($r = 0.226$, $p < 0.05$). When OS was partialled out, the correlation between MU and math score was significant for all grade levels ($rs > 0.45$, $ps < 0.001$); additionally, a partial correlation between NS and math score was also significant for G7 ($r = 0.304$, $p < 0.01$). This pattern of results suggests that MU played a dominant role in math achievement scores, similar to the findings from Han et al. (2016). Although both MU and OS were related to math operations, letter-triad was updated in the OS task while arithmetic results were updated in the MU task. Thus, MU made a unique contribution to math achievement even after OS was partialled out. The significant partial correlation between NS and math score at G7 signifies the role of NS in math achievement above and beyond MU and OS. Similar results were obtained from the following hierarchical regression analysis.

Table 2
Correlations among math achievement, 4 working memory (WM) tasks (MU, OS, SS, SSTM), and NS task.

	Math achievement	MU	OS	SS	SSTM	NS
Math achievement	1	0.631***	0.477***	0.363***	0.348***	0.349***
MU		1	0.648***	0.545***	0.549***	0.479***
OS			1	0.689***	0.541***	0.435***
SS				1	0.527***	0.308***
SSTM					1	0.477***
NS						1

*** $p < 0.001$.

3.2.2. Regression analysis

Two sets of model comparisons through hierarchical regression analyses were conducted (Tables 3 and 4). The dependent variable was the mean score of the three mathematics achievement tests. Before the analysis, collinearity among variables was checked. The variance inflation factors (VIFs) were smaller than 5, indicating that there was no collinearity among the variables.

In the first analysis, according to the decreasing order of correlation coefficients between math score, MU, OS, SS, NS and SSTM scores were entered one-by-one after the background factor (i.e., grade) was entered. The results of R^2 change are shown in Table 3. Overall, the six variables jointly accounted for 41.7% of total variance. Specifically, MU and OS accounted for 39.9% ($p < 0.001$) and 1.1% ($p < 0.05$) of variance, respectively. Other variables did not result in significant changes in R^2.

Our second analysis was to examine other possible variables that may contribute to the total variance, but their effects might be overridden by the two dominant variables (i.e., MU and OS) in the first analysis. Thus, we reversed the order of entering variables in the second analysis. Background factor was still entered into the first model, followed by the five observational factors with a reversed order compared to the first analysis (i.e., SSTM, NS, SS, OS, and MU). All of the observational factors contributed to the total variance by a significant amount when OS and MU were entered in the last steps (Table 4) while grade merely accounted for a negligible amount of variance (0.5%, $p > 0.2$). The results of the two analyses suggest that MU could account for the largest amount of the total variance. In addition, other observational variables also accounted for some of the total variance. However, if MU was entered earlier, then the contribution of other variables reduced substantially. In other words, the influence of other variables was smaller compared with that of MU.

When the unique contribution of the WM and NS tasks to math score was analyzed for each grade level separately (shown in Appendices 2 and 3), in the first analysis, MU accounted for a 41.5, 34.9, and 48.1% of variance respectively for grades 7, 8, 9 (all $ps < 0.001$). With the first entering order, NS accounted for a significant amount of variance (2.9%, $p < 0.05$) after MU was entered in the model for G7. Other variables did not significantly contribute to math score after MU. This pattern of results was similar to that in the correlational analysis. With the reversed order, all variables except SS at G7 (0.9%, $p > 0.31$) contributed to math score. In summary, separate analyses at each grade level provided convergent evidence that MU played a dominant role in math achievement. The finding was not only shown in junior high school students in the present study, but in university students in the study of Han et al. (2016). Nevertheless, it should be noted that NS also contributed to the math achievement at G7.

4. Discussion

The purpose of the present study was to document the developmental trend in WM abilities and NS in junior high school students, and to investigate their influence on math attainment.

First, there were developmental trends in WM abilities and NS. In WM tasks, G9 had better performance in the MU task than G7 and G8. G9 and G8 had better performance than G7 in both OS and SS span tasks. Similarly, in the SSTM task, there was an improvement in which all pairwise comparisons were significant. This is consistent with findings in the literature, with the general trend of cognitive improvement at least until adolescence (e.g., Gathercole, Pickering, Ambridge et al., 2004; Linares et al., 2016; Luciana et al., 2005; Swanson, 1996). Furthermore, there was also a developmental trend in the NS task. Overall, G9 had better performance than G7 and G8. The NS task showed the

Table 3
Model comparisons through hierarchical regression analysis with the entering order of grade, MU, OS, SS, NS, and SSTM. The unstandardized (b) and standardized (β) coefficients were obtained from the full model with all variables[a].

Variable	R	R^2	ΔR^2	F change	p value (F change)	b	β	p value (coefficient)
Grade	0.068	0.005	0.005	1.246	0.265	−0.018	−0.098	0.060
MU	0.635	0.403	0.399	176.290	0.000	0.425	0.549	0.000
OS	0.644	0.414	0.011	5.067	0.025	0.122	0.152	0.044
SS	0.644	0.415	0.001	0.303	0.582	−0.022	−0.031	0.646
NS	0.646	0.417	0.002	0.959	0.328	0.131	0.056	0.327
SSTM	0.646	0.417	0.000	0.020	0.886	−0.017	−0.009	0.886

[a] Unstandardized (b) and standardized (β) coefficients from the final model with all variables are also reported in Tables 3 and 4. It should be noted that model comparisons with hierarchical regression analysis examined whether a particular variable uniquely contributed to math score above and beyond other variables in the previous model. Conversely, regression coefficients for each variable in the final model revealed the relationship between that variable and math score. These two measures may not coincide with each other. According to our research question, ΔR^2 was the main focus for discussion.

pattern that scores were lower for more difficult items (i.e., lower ratio between blue and yellow dots), and this interacted with grade such that the similarity of pattern in improvement across item difficulty between G8 and G9 was higher than that between G7 and G9. The improvement of NS accuracy from G7 to G9 was consistent with the results from the large online assessment conducted by Halberda et al. (2012). Most studies in the literature addressed the developmental trend in NS in the early stage of life; however, the present study tracked and observed the trend in young adolescents.

Second, working memory ability, especially that probed by the MU task, played a dominant role in math achievement. In the correlational analysis, MU had the highest correlation coefficient with math achievement scores, and it also had high correlation coefficients with NS and other WM task scores. Although all WM and NS task scores significantly correlated with math achievement scores, the partial correlation was negligible after MU had been partialled out. The correlation between MU and math achievement remained when OS that had second largest correlation with math scores was partialled out. This pattern of results suggests that MU made a higher contribution to math achievement. The finding was consistent with the subsequent regression analyses. With the model comparisons through hierarchical regression analysis, when MU was entered first, other task scores made a negligible contribution to math achievement (changes in R^2 were <11%) except that the unique contribution of OS was significant. However, when the entering order of variables was reversed, all task scores made a significant unique contribution to math achievement scores. In both regression analyses, MU still had the largest unique contribution to math scores (39.9 and 14.8%). Analogous to the correlational analyses, MU made a dominant contribution to math score, which masked the contribution from other task scores. The influences of other task scores were negligible when that of MU was partialled out, and their influences were observable only when MU was entered into regression model in the last step. The pattern of the correlational and regression analyses almost remained across the three grade levels. However, at grade 7, NS significantly correlated with math scores after MU and OS were partialled out. In the regression analysis, NS also had significant contribution to math performance after MU had been included in the regression model.

In the present study, MU made more contribution to math achievement than other WM tasks and NS task. In the literature, it has been documented that both WM (especially CE) and NS made a unique contribution to math above and beyond the other variable (e.g., Xenidou-Dervou et al., 2013). In addition, both mediation directions (from WM to NS and the reverse) were observed (Hassinger-Das et al., 2014; Xenidou-Dervou et al., 2013). The inconsistent findings indicated the complicated relationship among WM, NS, and math achievement. The tasks, measures, and mathematic topics are influential factors. Both MU and OS used in the present study involved operating on numbers (see the next paragraph for further discussion), while, for example, those used in the study of Xenidou-Dervou et al. (2013) involved recalling words backward and to remember the locations of sets of boxes that contained a unique shape, and thus had no numerical content. Moreover, the math attainment tests used in the present study covered a variety of topics in the first semester of grades 7–9, while those used in the study of Xenidou-Dervou et al. were counting and exact symbolic addition for preschoolers. Thus, the relative contribution of working memory to mathematics performance may change with cognitive development and curriculum. Studies about preschoolers and elementary students demonstrated the involvement of CE; however, the contribution of phonological loop and visuo-spatial sketchpad varies across grade levels (De Smedt et al., 2009; Holmes & Adams, 2006). With progression into higher grade levels, advanced mental operations are necessary. Updating, a CE sub-function, involves assessing information that may not be within attentional focus, transformation and substitution (Ecker et al., 2010). These processes might be involved in solving advanced math questions in the high school curriculum. All other WM components, such as maintaining verbal and visual information, as well as NS, still engage in math operations but are relatively less essential than MU. It is also possible that the tasks we adopted were not comparable in terms of complexity (e.g., SSTM might involve less mental operation than other tasks.) Nevertheless, it is intriguing that NS plays an important role (although moderate) in math achievement at grade 7 above and beyond MU. It is possible that seventh graders still rely on innate ability during their first semester with formal operations. Further investigation can be devoted to the mechanism of transition

Table 4
Model comparisons through hierarchical regression analysis with the entering order of grade, SSTM, NS, SS, OS, and MU. The unstandardized (b) and standardized (β) coefficients were obtained from the full model with all variables.

Variable	R	R^2	ΔR^2	F change	p value (F change)	b	β	p value (coefficient)
Grade	0.068	0.005	0.005	1.246	0.265	−0.018	−0.098	0.060
SSTM	0.355	0.126	0.121	36.566	0.000	−0.017	−0.009	0.886
NS	0.413	0.171	0.045	14.338	0.000	0.131	0.056	0.327
SS	0.459	0.210	0.039	13.084	0.000	−0.022	−0.031	0.646
OS	0.519	0.269	0.059	20.986	0.000	0.122	0.152	0.044
MU	0.646	0.417	0.148	66.124	0.000	0.425	0.549	0.000

from basic math operations in elementary school to formal operations in high school.

It should be noted that although the updating task adopted in the present study involves numerical operation, *updating* per se rather than number processing might be more critical for math achievement. In the study of Han et al. (2016), the relationship between various versions of updating tasks and multi-digit mental multiplication by undergraduates was investigated. Modified from the MU task used in the present study (nominated as MUcalc in their study), the MUSpace task only required participants to update the final positions of three, four, or five dots after several 90° clockwise and counterclockwise transformations. Thus, the information about assessing, transformation, and substitution were applied to non-numerical material. In their third task (MUNumber), participants had to remember the larger number in a pair of digits presented in each frame. Only substitution was required, because each pair presented in the same frame was independent from the preceding and following pairs. Similarly, the larger animal in the pair presented as written words was updated in the fourth task (MUWord). Among these tasks, MUclac and MUSpace significantly correlated with the performance in difficult items (four-digit number multiplied by one-digit number) rather than that in easy items (two-digit number multiplied by one-digit number). Although digit was updated in the MUNumber task, it did not correlate with math performance. Thus, it was updating per se (especially transformation) rather than numbers that predicted the math performance. This may explain the much smaller contribution of OS than that of MU to math achievement in this study. The arithmetic task involved in the OS task was the distractor rather than the target to be memorized. It is also possible that when students solve mathematics problems, they do not switch away to process an irrelevant task as they do in the complex span task (e.g., the OS task), whereas updating temporary results are involved in problem solving.

Although NS may be fundamental to concepts of number (Feigenson, Dehaene, & Spelke, 2004), much more is involved in succeeding in learning mathematics in grades 7–9. In particular, mathematics involves complex tasks that relate different elements together in order to understand the concepts and to reach solutions. Therefore, this may load more heavily on working memory (especially the updating ability) than simple tasks that require little more than understanding numbers and their relationships in the four arithmetical operations (see Best et al., 2011). This suggests that students identified as having low working memory capacity could be helped by breaking mathematical operations into manageable pieces and using scaffolding to train procedures that reduce working memory load. The finding that working memory ability had greater influence on math achievement than NS is important for math learning in school. It is not easy to find an effective way to train the NS ability, but there are many ways we can help students to improve their working memory, which may enhance their math performance (e.g., Holmes, Gathercole, & Dunning, 2009).

The present study documented the developmental trend in WM and NS, as well as their contribution to math achievement of junior high school students. Although age-related improvement in WM and NS through adolescence has been investigated in some studies, their relative contribution to math performance was missing in the literature. The findings from the present study let us understand the relative role of NS and aspects of WM on individual differences in math achievement of junior high school students. The influence of NS on math performance was smaller than WM. Moreover, among various WM tasks, MU was found to play a dominant role in advanced math operations. However, in the present study, math achievement was assessed by the sectional examination tests during the semester in high school; despite its ecological validity, it is not easy to scrutinize a specific predictor for each topic covered in one sectional exam. Furthermore, many other WM components could be assessed depending on the adopted theoretical framework. To conclude, the present study demonstrated the importance of MU compared to NS in math achievement of junior high school students.

Acknowledgements

This work was partially supported by grants (NSC 99-2511-S-004-001-MY3) from the National Science Council of Taiwan. Daniele Didino was supported by the project "Evaluation and enhancement of social, economic and emotional wellbeing of older adults", under the agreement No. 14.Z50.31.0029, Tomsk Polytechnic University.

Appendix A

Appendix 1
Correlations among math achievement, 4 working memory (WM) tasks (MU, OS, SS, SSTM) and NS task as a function of grade.

	Math achievement	MU	OS	SS	SSTM	NS
Grade 7						
Math achievement	1	0.645***	0.494***	0.367***	0.419***	0.417***
MU		1	0.635***	0.585***	0.541***	0.402***
OS			1	0.774***	0.480***	0.340**
SS				1	0.599***	0.346**
SSTM					1	0.361***
NS						1
Grade 8						
Math achievement	1	0.590***	0.459***	0.396***	0.306**	0.391***
MU		1	0.565***	0.497***	0.551***	0.549***
OS			1	0.581***	0.599***	0.531***
SS				1	0.515***	0.413***
SSTM					1	0.540***
NS						1
Grade 9						
Math achievement	1	0.694***	0.472***	0.293**	0.324***	0.319**
MU		1	0.689***	0.526***	0.468***	0.391***
OS			1	0.615***	0.400***	0.387***
SS				1	0.351***	0.163
SSTM					1	0.418***
NS						1

** $p < 0.01$.
*** $p < 0.001$.

Appendix 2

Model comparisons through hierarchical regression analysis with the entering order of MU, OS, SS, NS, and SSTM as a function of grade. The unstandardized (b) and standardized (β) coefficients were obtained from the full model with all variables.

Variable	R	R^2	ΔR^2	F change	p value (F change)	b	β	p value (coefficient)
Grade 7								
MU	0.645	0.415	0.415	59.691	0.000	0.387	0.494	0.000
OS	0.654	0.427	0.012	1.745	0.190	0.202	0.263	0.059
SS	0.663	0.439	0.012	1.717	0.194	−0.166	−0.256	0.072
NS	0.685	0.469	0.029	4.496	0.037	0.459	0.176	0.055
SSTM	0.690	0.476	0.008	1.168	0.283	0.209	0.116	0.283
Grade 8								
MU	0.590	0.349	0.349	41.216	0.000	0.367	0.486	0.000
OS	0.610	0.372	0.023	2.819	0.097	0.164	0.187	0.160
SS	0.613	0.375	0.003	0.418	0.520	0.075	0.100	0.400
NS	0.613	0.376	0.001	0.104	0.748	0.149	0.073	0.539
SSTM	0.625	0.390	0.014	1.693	0.197	−0.311	−0.164	0.197
Grade 9								
MU	0.694	0.481	0.481	92.818	0.000	0.581	0.712	0.000
OS	0.694	0.481	0.000	0.012	0.911	0.025	0.029	0.795
SS	0.700	0.489	0.008	1.508	0.222	−0.079	−0.106	0.272
NS	0.701	0.491	0.002	0.348	0.557	0.129	0.049	0.568
SSTM	0.701	0.491	0.000	0.002	0.968	−0.008	−0.003	0.968

Appendix 3

Model comparisons through hierarchical regression analysis with the entering order of SSTM, NS, SS, OS, and MU as a function of grade. The unstandardized (b) and standardized (β) coefficients were obtained from the full model with all variables.

Variable	R	R^2	ΔR^2	F change	p value (F change)	b	β	p value (coefficient)
Grade 7								
SSTM	0.419	0.176	0.176	17.932	0.000	0.209	0.116	0.283
NS	0.507	0.257	0.081	9.060	0.003	0.459	0.176	0.055
SS	0.516	0.266	0.009	1.037	0.312	−0.166	−0.256	0.072
OS	0.594	0.353	0.087	10.880	0.001	0.202	0.263	0.059
MU	0.690	0.476	0.123	18.814	0.000	0.387	0.494	0.000
Grade 8								
SSTM	0.306	0.094	0.094	7.968	0.006	−0.311	−0.164	0.197
NS	0.407	0.166	0.072	6.577	0.012	0.149	0.073	0.539
SS	0.468	0.219	0.054	5.142	0.026	0.075	0.100	0.400
OS	0.512	0.262	0.042	4.229	0.043	0.164	0.187	0.160
MU	0.625	0.390	0.129	15.422	0.000	0.367	0.486	0.000
Grade 9								
SSTM	0.324	0.105	0.105	11.756	0.001	−0.008	−0.003	0.968
NS	0.382	0.146	0.041	4.752	0.032	0.129	0.049	0.568
SS	0.426	0.181	0.035	4.218	0.043	−0.079	−0.106	0.272
OS	0.506	0.256	0.075	9.760	0.002	0.025	0.029	0.795
MU	0.701	0.491	0.235	44.314	0.000	0.581	0.712	0.000

References

Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory. Vol. 8*. (pp. 47–90). New York: Academic Press.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*(1), 71–92. http://dx.doi.org/10.1037/0096-3445.132.1.71.

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, *21*(4), 327–336. http://dx.doi.org/10.1016/j.lindif.2011.01.007.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*(3), 273–293. http://dx.doi.org/10.1207/S15326942 D N1903_3.

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences*, *14*(12), 534–541. http://dx.doi.org/10.1016/j.tics.2010.09.007.

Butterworth, B., Cipolotti, L., & Warrington, E. K. (1996). Short-term memory impairment and arithmetical ability. *Quarterly Journal of Experimental Psychology*, *49A*, 251–262. http://dx.doi.org/10.1080/713755603.

Bynner, J., & Parsons, S. (1997). *Does numeracy matter?* London: The Basic Skills Agency.

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–172. http://dx.doi.org/10.1016/j.actpsy.2014.01.016.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–183. http://dx.doi.org/10.1016/S0160-2896(01)00096-4.

Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory*, *4*, 577–590. http://dx.doi.org/10.1080/741940997.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63–68. http://dx.doi.org/10.1016/j.tine.2013.12.001.

Dehaene, S. (2001). Précis of the number sense. *Mind & Language*, *16*(1), 16–36. http://dx.doi.org/10.1111/1468-0017.00154.

De Lemos, M. M., & Raven, J. C. (1989). *Standard progressive matrices.* Australian Council for Educational Research Limited.

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology*, *103*(2), 186–201. http://dx.doi.org/10.1016/j.jecp. 2009. 01.004.

De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior.

Trends in Neuroscience and Education, *2*(2), 48–55. http://dx.doi.org/10.1016/j.tine.2013.06.001.

Ecker, U. K., Lewandowsky, S., Oberauer, K., & Chee, A. E. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 170. http://dx.doi.org/10.1037/a0017891.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331. http://dx.doi.org/10.1037/0096-3445.128.3.309.

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*, 53–72. http://dx.doi.org/10.1016/j.jecp.2014.01.013.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. http://dx.doi.org/10.1016/j.tics.2004.05.002.

Friso-van den Bos, I., van der Ven, S. H., Kroesbergen, E. H., & van Luit, J. E. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, *10*, 29–44. http://dx.doi.org/10.1016/j.edurev. 2013.05.003.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177. http://dx.doi.org/10.1037/0012-1649.40.2.177.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16. http://dx.doi.org/10.1002/acp.934.

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., ... Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE*, *8*(6), e67374. http://dx.doi.org/10.1371/journal.pone.0067374.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, *115*(3), 394–406. http://dx.doi.org/10.1016/j.cognition.2010.02.002.

Gross, J., Hudson, C., & Price, D. (2009). The long term costs of numeracy difficulties. Retrieved from London: www.everychildachancetrust.org

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457–1465. http://dx.doi.org/10.1037/a0012682.

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences*, *109*(28), 11116–11120. http://dx.doi.org/10.1073/pnas.1200196109.

Halberda, J., Mazzocco, M. l. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with math achievement. *Nature*, *455*, 665–668. http://dx.doi.org/10.1038/nature07246.

Han, C. -C., Yang, T. -H., Lin, C. -Y., & Yen, N. -S. (2016). Memory updating and mental arithmetic. *Frontiers in Psychology*, *7*, 1–14. http://dx.doi.org/10.3389/fpsyg.2016.00072.

Hassinger-Das, B., Jordan, N. C., Glutting, J., Irwin, C., & Dyson, N. (2014). Domain-general mediators of the relation between kindergarten number sense and first-grade mathematics achievement. *Journal of Experimental Child Psychology*, *118*, 78–92. http://dx.doi.org/10.1016/j.jecp.2013.09.008.

Holmes, J., & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology*, *26*(3), 339–366. http://dx.doi.org/10.1080/01443410500341056.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*, F9–F15. http://dx.doi.org/10.1111/j.1467-7687.2009.00848.x.

Hornung, C., Schiltz, C., Brunner, M., & Martin, R. (2014). Predicting first-grade mathematics achievement: The contributions of domain-general cognitive abilities, nonverbal number sense, and early number competence. *Frontiers in Psychology*, *5*, 272. http://dx.doi.org/10.3389/fpsyg.2014.00272.

Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review*, *18*(6), 1222–1229. http://dx.doi.org/10.3758/s13423-011-0154-1.

Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, *145*, 147–155. http://dx.doi.org/10.1016/j.actpsy.2013.11.009.

Iuculano, T., Moro, R., & Butterworth, B. (2011). Updating working memory and arithmetical attainment in school. *Learning and Individual Differences*, *21*(6), 655–661. http://dx.doi.org/10.1016/j.lindif.2010.12.002.

Iuculano, T., Tang, J., Hall, C., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science*, *11*(5), 669–680. http://dx.doi.org/10.1111/j.1467-7687.2008.00716.x.

Kolkman, M. E., Hoijtink, H. J., Kroesbergen, E. H., & Leseman, P. P. (2013). The role of executive functions in numerical magnitude skills. *Learning and Individual Differences*, *24*, 145–151. http://dx.doi.org/10.1016/j.lindif.2013.01.004.

Lechuga, M. T., Pelegrina, S., Pelaez, J. L., Martin-Puga, M. E., & Justicia, M. J. (2016). Working memory updating as a predictor of academic attainment. *Educational Psychology*, *36*(4), 675–690. http://dx.doi.org/10.1080/01443410.2014.950193.

Lewandowsky, S., Oberauer, K., Yang, L. -X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods*, *42*(2), 571–585. http://dx.doi.org/10.3758/BRM.42.2.571.

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*(6), 1292–1300. http://dx.doi.org/10.1111/j.1467-7687.2011.01080.x.

Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*, *141*(3), 373–379. http://dx.doi.org/10.1016/j.actpsy.2012.09.009.

Linares, R., Bajo, M. T., & Pelegrina, S. (2016). Age-related differences in working memory updating components. *Journal of Experimental Child Psychology*, *147*, 39–52. http://dx.doi.org/10.1016/j.jecp.2016.02.009.

Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The development of non-verbal working memory and executive control processes in adolescents. *Child Development*, *76*(3), 697–712. http://dx.doi.org/10.1111/j.1467-8624. 2005.00872.x.

Mazzocco, M. l. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PloS One*, *6*(9), e23749. http://dx.doi.org/10.1371/journal.pone.0023749.

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology*, *103*(4), 490–502. http://dx.doi.org/10.1016/j.jecp.2009.02.003.

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*, 185–208. http://dx.doi.org/10.1146/annurev.neuro.051508.135550.

Organization for Economic Co-operation and Development (2010). The high cost of low educational performance. The long-run economic impact of improving educational outcomes. Retrieved from https://www.oecd.org/pisa/44417824.pdf

Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* London: National Research and Development Centre for Adult Literacy and Numeracy, Institute of Education.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*, 110–122. http://dx.doi.org/10.1016/j.lindif.2009.10.005.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2014). A task is a task is a task: Putting complex span, n-back, and other working memory indicators in psychometric context. *Frontiers in Psychology*, *5*, 1475. http://dx.doi.org/10.3389/fpsyg.2014.01475.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, *59*, 745–759. http://dx.doi.org/10.1080/17470210500162854.

Swanson, H. L. (1996). Individual and age-related differences in children's working memory. *Memory & Cognition*, *24*, 70–82. http://dx.doi.org/10.3758/BF03197273.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. http://dx.doi.org/10.1016/0749-596X(89)90040-5.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between operation span and raven. *Intelligence*, *33*, 67–81. http://dx.doi.org/10.1016/j.intell.2004.08.003.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order ability. *Psychological Bulletin*, *133*, 1038–1066. http://dx.doi.org/10.1037/0033-2909.133.6.1038.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635–654. http://dx.doi.org/10.1080/09658210902998047.

Van der Ven, S. H., Kroesbergen, E. H., Boom, J., & Leseman, P. P. (2012). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology*, *82*(1), 100–119. http://dx.doi.org/10.1111/j.2044-8279.2011.02035.x.

Wei, W., Yuan, H., Chen, C., & Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics. *British Journal of Educational Psychology*, *82*(1), 157–181. http://dx.doi.org/10.1111/j.2044-8279.2011.02049.x.

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 433. http://dx.doi.org/10.3389/fpsyg.2013.00433.

Xenidou-Dervou, I., van Lieshout, E. C. D. M., & van der Schoot, M. (2014). Working memory in nonsymbolic approximate arithmetic processing: A dual-task study with preschoolers. *Cognitive Science*, *38*(1), 101–127. http://dx.doi.org/10.1111/cogs.12053.

Xenidou-Dervou, I., De Smedt, B., van der Schoot, M., & van Lieshout, E. C. (2013). Individual differences in kindergarten math achievement: The integrative roles of approximation skills and working memory. *Learning and Individual Differences*, *28*, 119–129. http://dx.doi.org/10.1016/j.lindif.2013.09.012.

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences*, *23*, 1–9. http://dx.doi.org/10.1016/j.lindif.2012.10.004.