Reading in a Foreign Language October 2021, Volume 33, No. 2
ISSN 1539-0578 pp. 141–167

Visuospatial Working Memory and the Hybrid Reading Fluency Measure

Bartolo Bazan
Ryukoku University Heian Junior & Senior High School
Japan

Abstract

In this study, I sought to investigate whether visuospatial working memory (VWM) explains individual differences in reading fluency growth. One hundred and thirty Japanese junior and senior high school students were administered the Mr. Peanut and reversed Corsi tasks, respectively as measures of VWM, followed by three waves of reading fluency assessment. To assess reading fluency, a novel reading fluency measure was constructed by combining the participants' comprehension scores with their reading speeds into a single measure using many-facet Rasch measurement. Latent growth curve analyses revealed that VWM significantly predicted variation in the individual growth trajectories at the initial status but did not significantly predict rate of change. Results of Rasch analyses indicated that the novel reading comprehension-speed measure assessed a unidimensional construct, thus suggesting that the construction of this measure is both a practical and valid approach to estimating reading fluency as well as reading passage difficulty.

Keywords*:* L2 reading fluency, latent growth modeling, many-facet Rasch measurement, the Rasch model, visuospatial working memory

Reading, whether in a first (L1) or second language (L2) involves a myriad of cognitive processes that start with the visual recognition of word forms including letter shapes, chunks of letters, and overall word shape (Grabe, 2009). Eye-movement research has shown that in order to be able to visually recognize words and thereby be able to read, readers perform two types of eye movements: fixations and saccades (Dronjic & Bitan, 2016). Fixations refer to the moments when the eyes come to fixate or rest (although the eyes are not capable of remaining fixed) on a part of the text in order to extract visual information about the printed words. In contrast, saccades are the jumps between fixations that allow readers to assemble the portions of text seen during the fixations into the field of vision. As the visual information of words is extracted, readers activate phonological, syntactic, and semantic information, and access their mental lexicon to retrieve meaning (Grabe, 2009). Simultaneously, readers engage in higher-level processes of text comprehension, such as storing the information from a preceding portion of a given text and using that information to process a subsequent piece of text (Daneman & Carpenter, 1980). Many of these cognitive processes, which underlie reading, are mediated by working memory (WM). In fact, WM capacity has been shown to be a predictor of variance in reading ability. That is, individual differences in WM capacity are predictive of higher- or lower-reading skills. However, as the emphasis of the research up to date has been placed on the phonological and executive components of WM (Pham & Hasson, 2014), little has been investigated about the role, if any, that visuospatial WM

(VWM) plays in reading skills. In particular, it is unknown whether individual differences in VWM can be predictive of different levels of reading fluency growth. Therefore, the primary purpose of this study was to investigate whether individual differences in VWM are associated with L2 reading fluency growth. To explore this issue, two pencil-and-paper versions of the Mr. Peanut (Case, 1985; DeAvila, 1974) and the Corsi block span (Corsi, 1972) tasks were utilized to measure the participants' VWM capacity. The performance on these tasks involves fixations and saccades, respectively. In total, three waves of reading assessments were chronologically administered to investigate whether the instruction that the participants received was value-added with regard to reading ability.

The second motivation for this study was to construct a measure of reading fluency/proficiency that assessed both comprehension and reading speed as a single measure. Despite the fact that reading fluency is defined as the ability to read with velocity and comprehension (Grabe, 2010), research to date has measured speed and comprehension as two separate constructs (e.g., Chang, 2012; Chang & Millet, 2013; Grabe, 2009; Klauda & Guthrie, 2008; Khun & Stahl, 2003; NRP, 2000; Shimono, 2018; Therrien et al., 2006; Vadasy & Sanders, 2008). This study presents a hybrid measure that was developed by anchoring the participants' comprehension scores with their reading speeds using the many-facet Rasch measurement software (MFRM) (Linacre, 2017). The logic behind the measure is that speed and comprehension are two facets of reading that are, to some extent, inter-dependable. In order to comprehend, readers need to hold preceding text information in WM and assemble it with subsequent text information (Daneman & Carpenter, 1980), so a slow reading pace might impact WM negatively. This is because the preceding text information might have to be retained in WM for longer intervals, which could potentially lead to this information being forgotten by the time it is to be merged with the newly-read information, thereby affecting comprehension.

Literature Review

Baddeley's (2000) WM Model

WM refers to the mental workspace responsible for retrieving a limited amount of information from long-term memory and briefly holding and manipulating that information (Baddeley et al., 2002). Although several WM models have been postulated in the cognitive psychology literature (Miyake & Shah, 1999), this study is framed around Baddeley's (2000) multicomponent WM model. This model consists of four subsystems. The principal subsystem is the central executive or executive WM (EWM), as it fulfills an important number of cognitive functions such as controlling attention, regulating the flow of information within working memory, and integrating information. The central executive is accompanied by two storage-only subsidiary subsystems. The first is known as the phonological loop or phonological WM (PWM), which specializes in temporarily storing verbal information. The second is described as the visuospatial sketchpad, which is in charge of briefly holding visual and spatial information. The visuospatial sketchpad is further fractioned into the visual cache (i.e., visual memory) and the inner scribe (i.e., spatial memory). The visual cache is responsible for retaining visual patterns such as the appearance of an object, its shape, and color, whereas the inner scribe is in charge of retaining sequences of movements from one viewpoint to another (Baddeley & Logie, 1999). Both the phonological loop and the visuospatial sketchpad, from here on referred to as VWM, are linked to long-term memory through the episodic buffer, which serves as the interface between WM and long-term memory.

Of particular importance for the purpose of this study is VWM. This WM subsystem has arguably been under-researched and can potentially play a role in L2 comprehension, particularly in reading. This is because, in addition to its main storage function, it has been claimed that VWM serves the function of transforming verbal input into imagery (Gathercole & Baddeley, 1993), which is arguably the final product of all the processes involved in decoding language (Cornoldi & Vecchi, 2003).

Rationale for VWM as a predictor of reading

In the neuroimaging literature, many fMRI (functional magnetic resonance imaging) studies (Bolger et al., 2005; Hamasaki et al., 1995; Sakurai et al., 1994; Tan et al., 2005; Zhu et al., 2014) have compared the brain areas underlying reading in alphabetical and logographic scripts, including Chinese and Japanese kana (e.g., syllabographic) and kanji (e.g., logographic). These studies have revealed that the left hemisphere, which has been associated with phonological processing (Bolger et al., 2005; Tan et al., 2005; Zhu et al., 2014), is uniquely activated in readers of alphabetic writing systems. Conversely, activation is greater in the right hemisphere, which has been associated with VWM, among Chinese and Japanese readers. These findings have been interpreted to suggest that the grapheme-to-phoneme conversion characteristic of alphabetic writing systems is supported by the neural circuits in the left hemisphere whereas the right hemisphere is responsible for visuospatial processing (Baddeley, 2010; De Renzi & Nichelli, 1975; Smith & Jonides, 1997), which is required to process the shape of the kanji characters and the spatial arrangement of the strokes within them (Nelson et al., 2009). Reading studies in Chinese-English bilinguals have shown that the reading network of L2 readers adapts to reading in a new writing system with proficiency and exposure (Nelson et al., 2009; Yokohama et al., 2013). For example, proficient Chinese-English bilinguals develop the ability to use the same reading mechanisms that English readers use, phonological processing, to read in their alphabetic L2. It is thus likely that intermediate-level learners transfer the mechanisms that they predominantly rely on when reading in their first language to reading in their second language as some evidence suggests (Yokohama et al., 2013). In other words, it is possible that intermediate-level Japanese learners of English rely on VWM to a larger extent than they do on phonological processing when reading in English. Hence, in this study, it is hypothesized that VWM should predict reading skills in the current sample.

VWM measures

There are two major VWM assessment paradigms (Wen, 2016) in cognitive psychology. One, the simple span task paradigm, is focused on the measurement of the storage-only capacity of VWM through tasks involving sequences of items presented in different spatial locations, which need to be recalled by participants. These tasks include the Mr. Peanut task (Case, 1985), the Corsi block span task (Corsi, 1972), and the dot memory task (Ichikawa, 1983).

The other, the complex span task paradigm, consists of tasks that involve visuospatial storage supplemented by concurrent processing demands such as the letter rotation and the dot matrix tasks (Miyake et al., 2001). Current research has shown, however, that both simple and complex VWM tasks measure the compound VWM + EWM rather than VWM alone

(Miyake et al., 2001). Therefore, the simple VWM tasks employed in this study, the Mr. Peanut and reversed Corsi block span tasks, measure VWM and EWM as one block.

WM and L2 reading

As the bulk of existing research has focused on the association between individual differences in PWM and EWM and L2 reading (e.g., Harrington & Sawyer, 1992; Geva & Ryan, 1993; Leeser, 2007; Pae & Sevcik, 2011), little empirical work has been conducted on whether differences in VWM account for variation in L2 reading. One of the few studies on how individual differences in EWM and VWM relate to learning to read in an L2, particularly Chinese characters, was that of Kim et al. (2015). Seventy students enrolled in Chinese classes in an American university participated in the investigation. First, the participants' EWM and VWM capacity was assessed through an L1 version of the reading span task and the letter rotation task, respectively. Next, a test was administered where participants were shown 60 words (extracted from the Chinese course textbooks), requiring them to write the corresponding pronunciation of each word using pinyin as well as their English meanings.

As an intervention, the participants received pronunciation instructions on how to read 18 infrequent Chinese characters with each character presented in three conditions (distinctive, enhanced, and normal). The participants saw the items for a few seconds on a computer screen and heard their respective pronunciation over a set of headphones, with this procedure repeated three times. Finally, the participants' pronunciation was assessed through a test in which they were shown the same items as in the intervention and asked to read them aloud. The results of a mixed logit analysis showed that differences in WM capacity reflected differences in reading (estimate = 0.03, $p < .01$), depending on the condition, with EWM related to learning how to read normal characters (estimate = 0.04, $p < .01$) and VWM contributing to learning how to read enhanced characters (estimate = 0.07, $p < .01$).

As in Kim et al. (2015), Tsai (2014) looked at the relationship between EWM and VWM and reading in another logographic language, Japanese. The researcher assessed the EWM and VWM capacity of 30 native-English speaking Hawaiian university students, who were enrolled in a Japanese course, with the same instruments as those employed by Kim et al. (2015). These included a reading span task and a letter rotation task, in addition to a dual 3-back task that was used as an additional measure of VWM. In this latter task, the participants saw four blank squares on the computer screen in which unverbalizable shapes appeared one by one. They were required to remember the locations of the shapes and click a button when they saw the same shape in the same square as they had viewed three screens before. Further, the participants took a Japanese cloze test and self-rated their perceived level of reading proficiency in Japanese. Correlational analyses revealed a weak relationship between EWM ($r = .33$, $p = .07$) and the results of the cloze test, while showing no relationship between VWM and the cloze test (letter rotation, $r = -.01$, $p = .42$ and dual-back ($r = .05$, $p = .81$). The author argued that the cloze did not tap into VWM as expected and that it seemed to test grammar and lexical knowledge instead of reading.

The only study involving VWM and reading in an orthographic language identified during the literature search was conducted in L1 research. In this study involving native-English speaking elementary-school children, Pham and Hasson (2014) set to investigate which component of WM, whether EWM or VWM, had the most predictive power on reading fluency and comprehension. One hundred and fifty-seven children in grades 4 and 5 from elementary schools in the US were recruited and administered an IQ test, an assessment of behavioral inattention, a battery of reading measures, and a battery of WM tests. Specifically, the EWM tests consisted of a forward digit span task, a backward digit, and a word span task. In contrast, VWM was assessed through the reversed Corsi block span task in addition to a symbolic WM task, in which participants were orally given a series of number-letter combinations and were required to point to the locations of those on a card. Reading fluency and reading comprehension were assessed through the reading fluency subtest of the Gray Oral Reading Test, Fourth Edition (GORT-4; Wiederholt & Bryant, 2001). The test required the participants to read aloud a number of short passages that increased in difficulty and answer five multiple-choice questions after each passage. The participants' reading speed and reading accuracy were combined into a single reading fluency score, whereas the comprehension scores were calculated based on the number of correct answers to the multiple-choice questions. To explore the contribution of the WM subcomponents of interest (EWM and VWM) on reading fluency and comprehension, hierarchical regression analyses, controlling for IQ, inattention, and phonological memory, were conducted on the data. With reading fluency as the dependent variable, the model explained 43.00% of the variance, with EWM providing 5.00% of the explanatory power and VWM failing to contribute to the model. However, when the dependent variable was reading comprehension, the model accounted for 50.00% of the variance with EWM contributing 5.00% (F = 15.30, p = .04) to the variance and VWM explaining an additional 4.00% (F = 13.49, p = .04). The researchers interpreted these results to suggest that VWM plays an important role in reading comprehension, but it does not explain reading fluency.

As shown above, little work has been undertaken on the association between individual differences in VWM and reading, and no study has looked into the interplay between VWM and reading growth in particular. Therefore, this investigation contributes to the existing literature by addressing this gap.

Research questions

This study was guided by the following research questions:

1. To what degree will the participants' reading skills (compound speed rate and comprehension) grow over the study period?
2. To what degree is VWM predictive of individual differences in reading fluency performance at the outset of the study?
3. To what degree is VWM predictive of individual differences in reading fluency growth?
4. Is the new comprehension-speed reading measure valid?

Research Question 1 addresses whether the instruction the participants receive at the institution is value-added with regard to reading fluency. That is, whether the instruction is effective in bringing about reading growth. Research Question 2 examines whether individual differences at the level of VWM predict individual differences in reading fluency performance. Research Question 3 aims to investigate whether VWM is a source of individual differences in reading fluency growth. Research Question 4 investigates whether the new hybrid measure is a valid measure of reading fluency.

Methodology

Participants

This investigation took place at a private junior and senior high school located in western Japan. At the institution, students were streamed into high- and low-level classes based on academic performance. The participants were 97 third-grade junior high school students aged between 14 and 15 years old in three intact classes (one high-level and two low-level) and 114 first-year senior high school students aged between 15 and 16 years old in three intact classes (two high-level and one low-level). This made a total sample of 211 participants (N = 211). However, as the study was conducted at the time of the Japanese national entrance examination, a large number of students missed at least one of the measures due to absenteeism to either prepare for or sit these exams. This left a final sample of 130 participants who completed all of the measures (the Mr. Peanut, the reversed Corsi task, and the three reading waves) and 162 with completed reading waves but missing at least one VWM task. English proficiency ranged from Eiken (a standardized test of English proficiency for Japanese learners of English) level five to Eiken level two, which according to the website www.eiken.or.jp, approximately equate to levels A1 and B1 of the Common European Framework of Reference (CEFR), respectively. All participants' first language was Japanese.

The approach to reading instruction at the institution was intensive (Nation & Macalister, 2021). In a typical lesson, students were pre-taught new vocabulary, instructed to read the target text silently, and then asked to answer written comprehension questions in both Japanese and English. Subsequently, they were given a written translation of the texts and were explained the meaning and the grammatical constructions. The reading passages came from the course textbook and were generally short dialogues that ranged between 80 and 120 words in length that were flooded with the target grammar point of the lesson.

Materials and procedure

This study took place in five sessions spread over a three-and-a-half month period. Sessions 1 and 2 comprised the measurement of the participants' VWM capacity, and Sessions 3, 4, and 5 involved the participants' assessment of their reading ability.

Individual differences measures

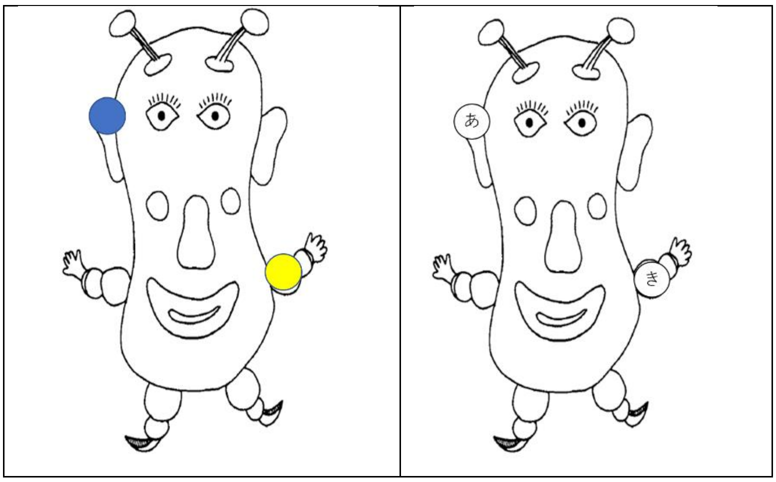
The Mr. Peanut task (Case, 1985; Deavila, 1974). An adaptation of the Mr. Peanut task was employed to tap into participants' VWM capacity. In Case's (1985) version of the task, test-takers were shown a figure of Mr. Peanut with a number of colored dots on different parts of his body. The number of dots varied from two to five, which was mirrored by the exposure time, with one second given per dot. For example, a figure of Mr. Peanut with three dots, illustrated in Figure 1, was shown for three seconds. After the exposure time, the figure was removed and participants were required to place the same color chips on the body parts that were covered with a dot in the stimulus image. Thus, the participants had to remember both the color and location of the dots. The task was administered individually, which reduced its practicality. For this reason, a paper-and-pencil group-version was designed and piloted with a group of 36 second-grade high school students. In this adapted version, the participants were shown the figure of Mr. Peanut with the colored dots on different parts of its body on a large screen. After the presentation, the figure was removed, and the participants were required to use their memory to recall the locations of the colored dots by writing the first hiragana (Japanese syllabary) character of the name of the color over the corresponding location on the test handout. For example, if there was a blue dot on the right ear and a yellow dot on the left elbow, the participants wrote あ (*a*, *ao*, blue) over the ear and き (*ki*, *kiiroi*, yellow) over the elbow (see Figure 1). Based on Miyake et al. (2001)'s work, the figure was presented for 750 ms to keep the participants from employing idiosyncratic strategies or encoding the locations verbally (i.e., body parts). In addition, an effort was made to avoid systematic dot patterns such as a dot on each of the two antennae or the two eyes. There were two initial practice sets of two dots, which were included in the scores, after which difficulty was gradually increased from three to six dots with two sets for every level of complexity (i.e., two sets of two dots, two sets of three, and so forth).

A partial credit scoring system, in which test takers received one point per correctly recalled color and location, was used to score the test. For example, in a set of four items, if test-takers wrote the name of two colors over the correct body parts, they were given two points for the set. If test-takers marked the location of the dot without writing the name of the color, they were given no credit. Logit Rasch measures were computed for each participant on the Winsteps software (Linacre, 2018).

Part of the rationale for using the Mr. Peanut task is that performance on this task resembles eye fixations in reading as it involves keeping the eyes aligned with the figure image to extract the locations of the static-colored dots. Thus, participants who score highly on this test are not only hypothesized to have a higher visual memory span, but are also more likely to be able to see a greater number of words per eye fixation when reading. This is because the visual cache (Baddeley & Logie, 1999), which is the component of VWM that this task is intended to measure, could be one factor that contributes to variance in the number of words per fixation. Although determining a precise range is beyond the scope of this study, as research has shown that 0.88 is the average word span for L2 English readers (Oller & Tullius, 1973), participants with lower visual memory are theorized to see less than this average, while those with higher visual memory might perform above this average. In any case, high performers are likely to have word-recognition spans below 1.60, which is the average span for college native English speakers (Gray, 1969).

Figure 1

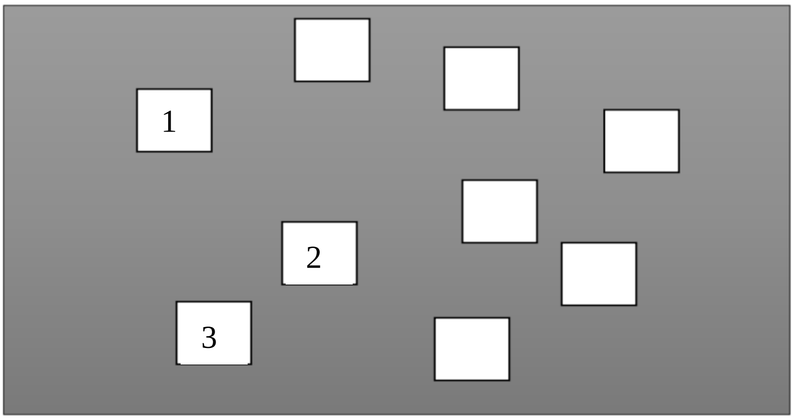
Mr. Peanut Sample Item With the Stimuli on the Left and the Recall Figure on the Right



The reversed Corsi block span task (Corsi, 1972). To obtain more reliable estimates of VWM, the participants' VWM capacities were assessed by means of a second VWM task, namely the reverse Corsi block span task. This task is intended to measure the inner scribe or the spatial component of VWM (Baddeley & Logie, 1999). In the original Corsi task, test takers saw an arrangement of blocks on a board. Subsequently, the researcher tapped a number of blocks sequentially and the test-takers were asked to remember the sequence and repeat it by tapping the blocks in the same order. The sequences of taps increased in length from two to seven and the task was administered individually. For the purposes of practicality, a paper-and-pencil version that could be administered collectively was developed. This adaptation involved test takers looking at a screen with nine white squares on a grey background from which several squares illuminated in a sequence. Each square in the sequence was lit up (i.e., turned yellow) for a short period of time (approximately 200 ms) to discourage the participants from engaging in the use of idiosyncratic strategies. Additionally, an attempt was made to control the configurations of dots to avoid systematic patterns. The participants had to number, from memory, the squares in reverse order on their test handouts. The participants were presented with increasingly longer sequences of flashing squares from two to eight with two sets of every length (i.e., two sets of two flashing squares, two sets of three, etc.) including two practice sets of three squares at the beginning of the test. Figure 2 illustrated a completed sequence of three squares in which the block on the left corner lit up, followed by the block to its upper right, and subsequently, the block in the top left corner.

Figure 2

Completed Set of Three Items in the Reversed Corsi Block Span Task



The task was scored using a partial credit system adopted from Bazan (2020), in which one point was awarded to each item recalled in the correct order until memory failure (i.e., until the order of the string of items was broken). The items recalled correctly after memory failure were considered residual and were awarded a score of zero points. For example, if in a set of five items, test takers correctly recall items 1, 2, and 3, then fail to recall item 4, but successfully recall item 5, they will receive a score of "3." The participants' raw scores were again transformed into Rasch logit measures using Winsteps (Linacre, 2018).

The rationale for using the reversed Corsi task rather than the forward Corsi task was that the former version yielded a wider spread of scores than the latter version when they were trialed with two groups of second-grade high school students.

As described, this task entails a dynamic component as participants are required to visually follow a sequence of movements, which is analogous to performing saccades. Therefore, it is likely that participants with high scores on this task can be considered as more capable of connecting fixations, which is hypothesized to facilitate reading.

English reading measures

Reading fluency/proficiency was gauged using six timed 400-word excerpts of easy-starts graded readers from the Penguin collection accompanied by a set of five multiple-choice questions. The vocabulary and text profiles of the passages are shown in Table 1.

Table 1

Vocabulary and Text Profiles of the Stories

		Vocabulary Profile			Text Profile		
Story	Title	WC	Coverage	TTR	AWPS	FRES	FKGL
A	*The Cup in the Forest*	400	95.00%	0.33	5.50	98.10%	0.90
B	*Tinkers Island*	400	98.00%	0.47	7.10	93.10%	2.00
C	*The Pearl Girl*	401	95.00%	0.39	6.30	100.00%	0.70
D	*Who wants to be a star?*	402	98.00%	0.39	7.00	92.30%	2.10
E	*The Long Road*	401	95.00%	0.42	8.10	77.40%	4.40
F	*Hannah and the Hurricane*	403	95.00%	0.35	8.20	94.90%	2.00

Note. Vocabulary profile based on BNC-COCA 1-25K; WC = word count; TTR = type-token ratio; AWPS = average word number per sentence; FRES = Flesch reading ease; FKGL = Flesch-Kincaid grade level.

The reading assessment was conducted over three waves of data collection separated by a month (three months in total) with each involving the administration of two consecutive reading tests. The data were collected in parallel with both the junior high school participants and the high school participants. In Wave 1, Session 4, Class 1 took Tests A and B back-to-back; Class 2 took Tests B and C; and Class 3 completed Tests C and D. One month later, Class 1 was administered Tests D and E; whereas Class 2 was administered Tests A and E; and Class 3 was administered Tests B and E. Finally, after another month, Class 1 was given Tests C and D; Class 2 was given Tests B and E; and Class 3 was given Tests A and F. This design is visually displayed in Table 2.

Table 2

Visual Representation of the Design

High school	Junior high school	Wave 1 (Session 4)	Wave 2 (Session 5)	Wave 3 (Session 6)
Class 1	Class 1	Test A	Test D	Test C
		Test B	Test E	Test F
Class 2	Class 2	Test B	Test A	Test D
		Test C	Test E	Test F
Class 3	Class 3	Test C	Test B	Test A
		Test D	Test E	Test F

The rationale underlying this design was to control for the difficulty of the multiple-choice items, and by extension, text difficulty, by conducting common-item linking (Bond & Fox, 2015) through the MFRM software (Linacre, 2017). For example, in Session 4, participants in Classes 1 and 2 encountered common items (in Test B) and the same was true for participants in Classes 2 and 3, who encountered the items of Test C (see Table 2). As each participant took all three tests, this procedure allowed for test scores to be put on a common scale. This was carried out by anchoring the item difficulty of the subsequent administrations of the tests (Waves 2 and 3) on the item difficulty of the Wave 1 tests. Anchoring items not only allows for the establishment of more valid comparisons across the classes, but it also effectively tracks growth over time (Bond & Fox, 2015).

In each reading ability test session, the participants were given a test handout including instructions in Japanese as well as the corresponding two texts together with their accompanying sets of five multiple-choice questions (see Appendices A and B for the two tests). A timer was projected on a large screen and participants were asked to record their reading time for each text before moving to the multiple-choice questions. The questions were printed on the back of the texts and the participants were instructed not to look at them until they had read the text once. The participants were not allowed to return to the text. Thus, they had to respond from memory. A maximum of 20 minutes was given to read the two texts and answer the multiple-choice questions. The participants were asked to read at a speed at which they could maintain comprehension.

For the purpose of this study, a new hybrid comprehension-speed reading measure was developed by combining the participants' comprehension scores and reading times into a single Rasch performance using the MFRM software. To do this, the participants' scores for each individual multiple-choice question (30 in total), which had been dichotomously scored (0 = incorrect and 1 = correct), as well as the participants' reading times, were first entered into an Excel spreadsheet. As MFRM uses integral digits from zero to nine, the participants' reading times were rounded, and times above nine minutes were given a value of nine. For example, reading times from 00 to 29 seconds were rounded down (e.g., 2 minutes and 25 seconds was entered into the spreadsheet as 2 minutes), while reading times with the seconds from 31 to 59 seconds were rounded up (e.g., 2 minutes and 45 seconds was entered into the spreadsheet as 3 minutes) Next, the spreadsheet was exported into the MFRM software through which an anchored comprehension-speed Rasch measure for each participant at each reading wave was calculated giving equal weight to both the comprehension scores and the reading times. The Rasch anchored person measures were extracted from each wave for further analyses on the SPSS and AMOS statistical software, respectively.

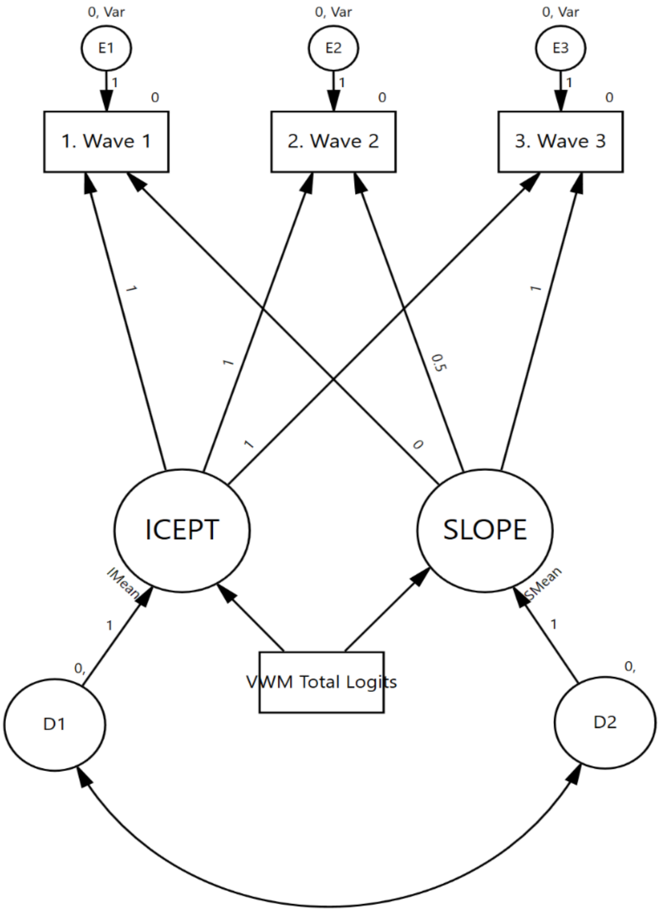
Analyses

Research Question 1 asked whether the study participants experienced growth in reading fluency over the duration of the study. To investigate this question, a repeated-measures ANOVA (RM-ANOVA) was conducted in SPSS (version 24) with time as the independent variable and the comprehension-speed Rasch measures as the dependent variable. The RM-ANOVA was followed by the construction of a basic latent growth curve (LGC) model with no predictor specified. The logic of this analysis was to measure individual changes in reading proficiency over time, because RM-ANOVAs focus on the average group change, and to obtain evidence of variation across the individual change trajectories. This latter point is important in relation to Research Question 2 because it would make little sense to include VWM as a predictor of differences in individual change if there are no differences in change across the individuals in the first place. The model included three waves of repeated measures of reading comprehension-fluency (Wave 1, Wave 2, and Wave 3), the intercept (the reading proficiency level at the start), the slope (the rate of change or growth), and a covariance (correlation) between the intercept and the slope as indicated by a curved double-headed arrow. The small circles with the one-way arrows pointing down towards the waves represent the measurement error and the circles D1 and D2 represent a residual related to the intercept and slope, which become dependent variables with the inclusion of the growth predictor VWM. The regression paths (the single-headed arrows) from the intercept to the three comprehension-speed reading measures were fixed at 1.00 to establish a starting point, whereas the regression paths from the slope were set to 0.00, 0.50, and 1.00, respectively to depict a hypothesized linear incremental growth in reading comprehension-speed over time. The fit of the model to the data was evaluated by examining three of the most widely used fit indices; the $\chi 2$ metric, the comparative fit index (CIF), and the root mean squared error of approximation (RMSEA) (Byrne, 2010). A perfect fit to the data is shown by a $\chi 2$ of 0.00, therefore, the closer to 0.00 the $\chi 2$ is, the better fit the observed model has. The second model fit criterion, the CFI, is considered good if its value is above .90. The third fit criterion, the RMSEA, constitutes a reasonable fit its value is of .08 or below.

Research Questions 2 and 3 concerned the degree to which individual differences in VMW were predictive of differences in reading performance at the outset of the study and differences in reading growth, respectively. These questions were investigated by including VWM, which was calculated as the addition of the participants' logit measures on the Mr. Peanut and the reversed Corsi task, with a regression path going into the intercept and another regression path going into the slope, as a growth predictor in the previous basic latent growth model (see Figure 3). To answer Research Questions 1 through 3, the 130 completed (all five measures) cases were used.

Figure 3

Study Latent Growth Curve Model



Research Question 4, regarding the validity of the new reading fluency/proficiency measure, was answered through an examination of the item and person infit and outfit mean-square (MNSQ) values as well as their respective separation and reliability coefficients, which are provided by the MFRM output. The Rasch fit statistics are quality-control indicators that are useful to evaluate the degree to which the data meet the model's expectations. In addition, they indicate if the facets comprehension items and speed adhere to the measurement of a unidimensional construct (Bond & Fox, 2015). Infit MNSQ is a weighted unstandardized form of fit whereas outfit MNSQ is a non-weighted standardized fit statistic that is sensitive to outliers (Linacre, 2002; Bond & Fox, 2015). As the outfit statistic is affected by outliers (unexpected performance of participants), infit MNSQ tends to be the statistic that guides the assessment of fit (Bond & Fox, 2015). In this study as well, decisions about fit were made based on infit MNSQ, with concerning outfit values (> 2.00) also explored. Based on Linacre's (2002; 2007) guidelines, fitting items and persons were defined as those with fit values of between 0.50 and 1.50, marginal fitting was defined as values between 1.51 and 2.00, and misfit as values greater than 2.00.

The item separation index indicates the number of speed-comprehension difficulty levels into which the items can be divided, whereas the person separation index indicates the number of reading speed-comprehension levels into which participants are separable. These indices are accompanied by reliability coefficients, which reveal the degree to which the replicability of the person and item hierarchy is possible if the three waves of reading tests are administered to a similar sample. In other words, the reliability values reveal the extent to which the separation of items and persons into distinct levels of reading difficulty and reading ability is reproducible in future administrations. The higher the reliability value, the more confidence can be placed in obtaining a similar separation across samples. To address Research Question 4, the data from the 162 participants who completed the three reading waves was employed. All of the above-mentioned indices provide validity evidence for the measure (Bond & Fox, 2015).

Results

Preliminary analyses

Prior to running the RM-ANOVA, a preliminary analysis of the descriptive statistics, which are presented in Table 3, was conducted. The distribution of the comprehension-fluency measures for each wave was normally distributed as revealed by their box plots, whose boxes and whiskers were symmetrical around the median (Larson-Hall, 2016). As seen in Table 2, the visual evidence of normality was corroborated by the skewness and kurtosis values for each wave, all of which were below 1.00 (Porte, 2002). Three statistical outliers were identified in Wave 1 and two in Wave 2, which were removed from further analyses, reducing the sample to 125 participants. The table includes the average reading fluency index (Brown et al., 1944) in raw scores for comparison with the Rasch hybrid reading fluency index.

Table 3

Descriptive Statistics for Each Data Wave (N =130)

	MRM	MRFI	*SD*	Skew	*SES*	Kurt	*KES*
Wave 1	0.04	47.01	0.90	0.28	.21	0.16	.42
Wave 2	0.30	57.36	0.88	-0.09	.21	0.05	.42
Wave 3	0.51	61.38	1.30	0.72	.21	0.01	.42

Note. MRM = Mean Rasch measure; MRFI = Mean Reading Fluency Index; SES = Skewness standard error; KES = kurtosis standard error. All statistics with the exception of the average RFI are based on Rasch logits and correspond to the Rasch hybrid reading fluency measure.

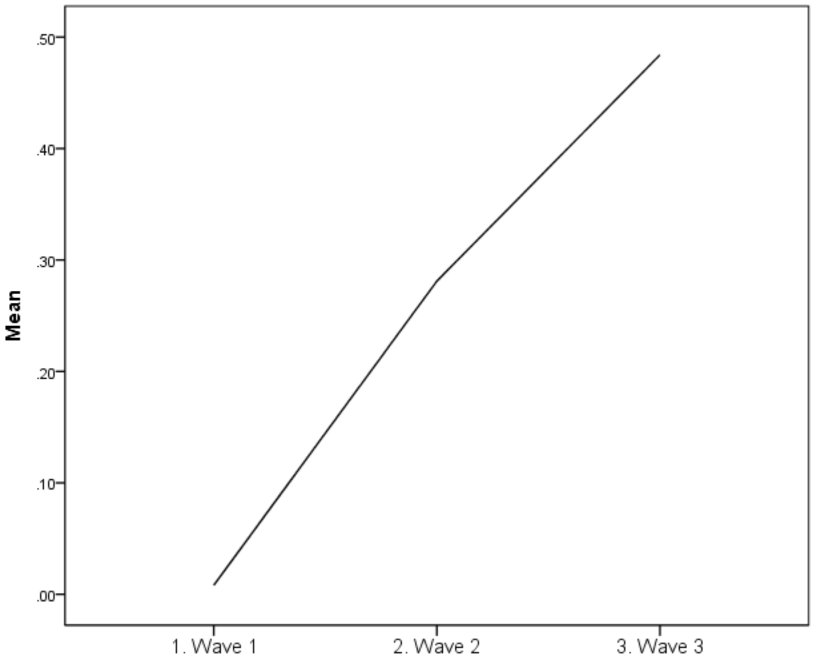
Although the assumption of normality underlying RM-ANOVAs was met, the sphericity assumption was violated, $\chi2(25.6) = 2$, $p < .05$. Because the data did not meet the sphericity assumption, a Greenhouse-Geisser correction was applied in the RM-ANOVA, following Field's (2009) recommendation.

Research Question 1

An RM-ANOVA was conducted with the effect of time as the within-subjects factor and the comprehension-speed Rasch measures as the dependent variable. The results of the ANOVA showed a significant time effect, F (1.73, 223.88) = 11.18, $p < .01$, partial $\eta2 = .10$, indicating that, on average, there was reading proficiency change over time. Figure 4 graphically displays the results of the ANOVA.

Figure 4

Average Change in Reading Proficiency Over Time



Subsequent pairwise comparisons revealed significant mean differences of $p < .05$ between Wave 1 and Wave 2, and Wave 2 and Wave 3 (see Table 4), which indicate a steady linear growth in reading comprehension-speed. Additional evidence of linear change was provided by the results of the polynominal contrasts, which showed a significant linear effect F (1, 125) = 24.34, $p < .05$, partial η2 .16. All in all, these results suggest that, on average, the participants became more proficient readers over time.

Table 4

Pairwise Comparisons Among the Three Reading Waves (N = 125)

(I) growth	(J) growth	*M* Difference	*SE*	*p*	95% *CI* for Difference Lower Bound	95% *CI* for Difference Upper Bound
1	2	-.27*	.06	.00	-.40	-.13
	3	-.47*	.09	.00	-.66	-.29
2	1	.27*	.06	.00	.13	.40
	3	-.20*	.09	.00	-.39	-.02
3	1	.47*	.09	.00	.29	.66
	2	.20*	.09	.00	.02	.39

Note. * = The mean difference is significant at the .05 level. All statistics are based on Rasch logits.

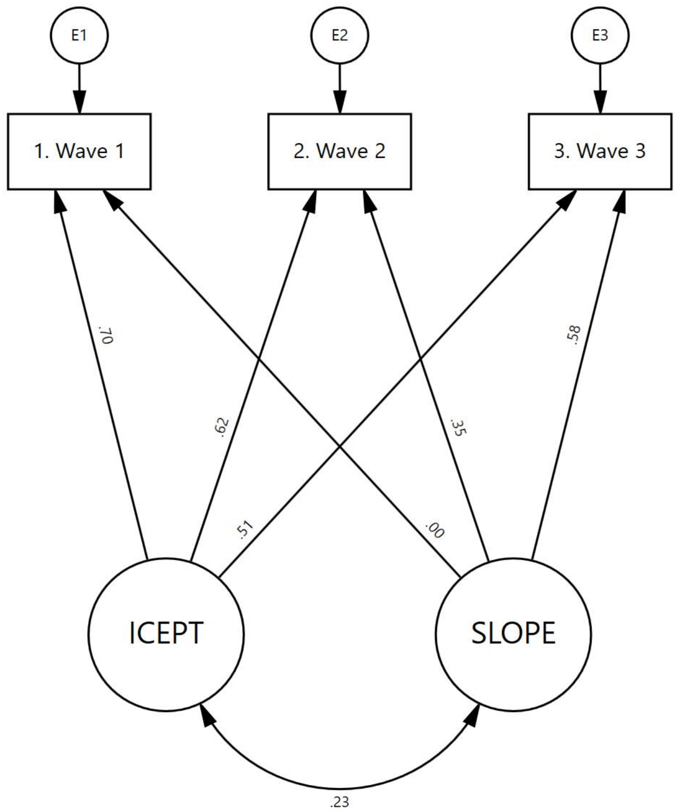
To examine if the average group change was also true for the individual cases, further investigation of the phenomenon was undertaken by constructing a latent growth model. The fit indices for the model were acceptable. Concretely, the model produced a low and random χ2 value (χ2= 22.53, df = 3, $p < .05$), indicating that the model provided a good match to the data. In addition, the CFI value (.82) was close to the criterion of .90, which further suggests an approximate fit to the data. In contrast, the RMSEA (.23), was above the fit criterion of

.08. However, because two of the three fit criteria were met, the model was considered an acceptable fit to the data.

The parts of the model in Figure 5 that reveal individual change with respect to the reading comprehension-speed outcome are the factor loadings attached to the regression paths. As seen, the paths going from the slope to the repeated measures have positive factor loadings which increase over time. Specifically, the path going from the slope into Wave 1 produced a factor loading of .00; the path going into Wave 2 yielded a factor loading of .35 (which suggests growth in reading proficiency relative to Wave 1); and the factor loading to Wave 3 was larger than that of Wave 2 (.58). These results provide evidence of a steady linear growth of reading comprehension-speed, corroborating the results of the RM-ANOVA and suggesting that the instruction at the institution positively impacted the students' reading skills.

Figure 5

Results of the Basic LGC Model for Reading Proficiency



To obtain evidence to support the further investigation of individual variability through the incorporation of the VWM predictor into the model, the variance table of the AMOS output file, displayed in Table 5, was examined. This table shows that the variances related to the intercept and slope are statistically significant at $p < .05$. These results reveal individual comprehension-speed differences in both the initial reading performance in Wave 1 and in its change over time, as the participants moved through the data-collection waves. Consequently, exploring if VWM was the source of such differences was worthwhile.

Table 5

Variances

	Estimate	*SE*	*CR*	*p*
Intercept	0.34	.09	3.81	***
Slope	0.45	.17	2.60	.01
E1	0.36	.04	7.90	***
E2	0.36	.04	7.90	***
E3	0.36	.04	7.90	***

Note. *** = probability < .00. All statistics are based on Rasch logits.

Research Questions 2 and 3

With evidence of heterogeneity across the individual reading growth trajectories, VWM was incorporated into the model as a reading growth predictor. The fit indices resulting from this predictor model were similar to that of the previous basic model (in which the RMSEA criterion was not met and the CFI was marginally below .90) and revealed a fitting model ($\chi2 = 23.91$, $df = 4$, $p < .05$; CFI = .83, RMSEA = .20). As shown in Figure 6, the factor loadings next to the arrows linking VWM to the intercept (.28) and slope (.13) are positive, suggesting an association between VWM and reading comprehension-speed at both the outset and in the rate of change. However, the results of the regression weights, summarized in Table 6, revealed that VWM was a significant predictor of reading comprehension-speed at the initial status (estimate = 0.19, $p < .05$), but not of rate of change as indicated by its nonsignificant estimate (0.10, $p > .05$). It is important to note that although significance was not found, the results related to VWM and the rate of change pointed to the expected direction. These findings can be interpreted as meaning that differences in VWM explain differences in reading comprehension-speed at the initial performance, but they do not significantly explain differences in reading growth over time.

Figure 6

Results of the LGC Model with VWM as a Predictor of Change in Reading Proficiency

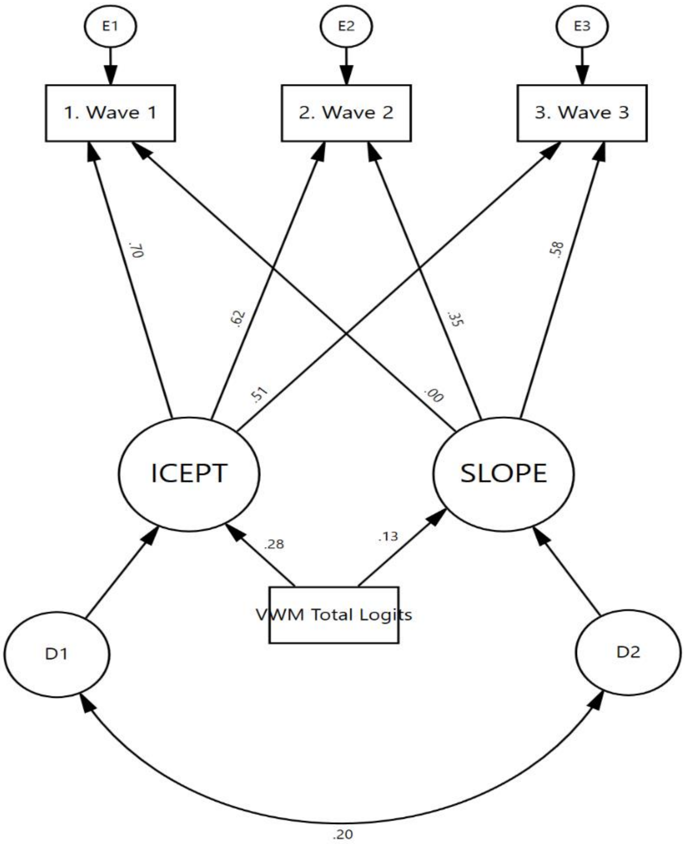


Table 6

Regression Weights

	Estimate	*SE*	*CR*	*p*
ICEPT ← VWMTotalLogits	0.19	.08	2.34	.02
SLOPE ← VWMTotalLogits	0.10	.11	0.89	.37

Note. ICEPT = Intercept. All statistics are based on Rasch logits.

Research Question 4

Table 7 shows that the items (A1-F5) are within the fit parameters (0.50-1.50), indicating the multiple-choice questions functioned adequately to measure comprehension. Although the outfit column revealed a marginal fit index for item C5 (outfit = 1.61), an inspection of the unexpected responses table in the MFRM output showed that this was caused due to several participants' unexpected responses. In addition, the item's infit MNSQ was of 1.16, which provides evidence that the item is a well-fitting item.

As seen, the reading time measures (RTA-RTF) show good fit to the Rasch model in this comprehension-speed combined context, which suggests that the comprehension and speed-reading items function well together to measure a unidimensional construct that is hypothesized to be reading fluency/proficiency. These findings provide validity evidence for the use of the new hybrid measure.

Table 7

Item Statistics for the Speed-Comprehension Reading Measure

Item	Measure	*SE*	Infit MNSQ	Outfit MNSQ
A1	1.52	0.20	1.18	1.42
A2	-1.08	0.19	1.12	1.20
A3	0.75	0.17	1.05	1.06
A4	-0.03	0.17	1.07	1.09
A5	-0.84	0.18	0.91	0.84
B1	-0.90	0.18	1.12	1.09
B2	-0.42	0.17	1.16	1.16
B3	0.09	0.17	0.93	0.91
B4	-0.12	0.17	0.96	0.96
B5	-0.15	0.17	1.01	0.99
C1	0.99	0.18	1.36	1.48
C2	-1.12	0.19	0.94	0.89
C3	-0.75	0.18	1.09	1.05
C4	-0.89	0.18	1.03	1.17
C5	1.64	0.20	1.16	1.61
D1	-1.50	0.21	1.02	0.99
D2	-1.14	0.19	1.02	1.00
D3	0.00	0.17	0.90	0.90
D4	-1.29	0.20	0.99	0.94
D5	0.41	0.17	1.23	1.28
E1	-1.13	0.19	0.97	0.92
E2	-1.70	0.22	0.96	0.83
E3	-1.29	0.20	0.96	0.90
E4	-0.11	0.17	1.06	1.04
E5	-0.16	0.17	1.06	1.03
F1	-0.64	0.18	0.91	0.86
F2	1.06	0.18	0.93	0.90
F3	-0.41	0.17	0.93	0.90
F4	1.40	0.19	1.07	1.20
F5	0.39	0.17	1.05	1.04
RTA	1.00	0.08	0.90	0.91
RTB	1.54	0.08	1.37	1.39
RTC	0.94	0.08	1.38	1.36
RTD	1.26	0.08	0.70	0.70
RTE	1.89	0.08	1.21	1.25
RTF	0.75	0.08	1.14	1.16
MODEL *RSME* 0.17	TRUE *SD* 1.00	SEPARATION 5.84	RELIABILITY .97	

Note. RT = reading time (RTA = reading time for test A, RTB = reading time for test B, and so forth). All statistics are based on Rasch logits.

Table 7 also reports the item separation and reliability indices. Based on Linacre's (2002, 2007) recommendations, item separation values of 3.00 or above and reliability estimates of at least .80 are desirable. As shown in the table, the comprehension-speed measure yielded a separation of 5.84 and a reliability estimate of .97, which implies that the measure contains six levels of comprehension-speed difficulty and that the probability of finding a similar item separation in similar samples is high. These findings reveal that the measure can define a hierarchy of comprehension-speed items along the hypothesized latent trait, which provides additional evidence of validity to support the use of the combined measure.

With respect to person fit, the majority of the sample performed in accordance with the model expectations (83.00% fitted well and 12.00% marginally fitted) as revealed by their infit indices (see Table 7). Five of these participants had, however, misfitting outfit values (2.24, 2.11, 0.48, 0.44, and 0.42), but they were classified as fitting because person performance is, in general less consistent than that of items and thus, the fit criteria should be more leniently applied to persons (Linacre, 2007). In addition, in this study, the participants' fit indices were based on repeated measures, which means that performing well on two waves but poorly on another would make the person misfit. Only 5.00% of the cases were found to misfit (see Table 8), meeting the cut-off value of 95.00% of fitting or marginally fitting participants (McNamara, 1996). These results indicate that the persons fitted the model well overall.

The Rasch person separation index was estimated at 2.57, indicating that the measure separated participants into three levels of reading fluency (high, average, and low). Furthermore, the reliability index was calculated at .87 (see Table 7), which suggests a relatively high probability of reproducing the separation of participants' reading ability into three different levels if the measure is taken by a sample of similar characteristics.

Table 8

Person Fit Statistics (N = 162)

Number of participants	% of the sample	Infit MNSQ range	*Min*	*Max*
4	2.50%	0.00-0.49	0.39	0.49
135	83.00%	0.50-1.50	0.52	1.50
19	12.00%	1.51-2.00	1.52	1.92
4	2.50%	>2.00	2.16	2.45
MODEL *RSME* 0.30	TRUE *SD* 0.78	SEPARATION 2.57	RELIABILITY	.87

Note. All statistics are based on Rasch logits.

Discussion

The goals of this study were two-fold. The primary goal of this investigation was to examine the degree to which the instruction that the participants received promoted the development of reading fluency and if VWM capacity accounted for the differences among the participants' reading growth trajectories. The results of a RM-ANOVA revealed that, on average, the participants progressed in their reading fluency development in a linear way from Wave 1 through Wave 3. An LGC model corroborated these results by indicating an increase in reading fluency over time at the individual level as well. These findings suggest that the instruction was value-added as it helped the participants improve reading fluency over the duration of the study.

An alternative explanation for the participants' growth in reading fluency, however, is that the reading assessment served as a fluency-development intervention. The assessment pushed the participants to read 800 words (more than they would normally have read in the classroom) three times over the duration of the study and, although the participants were asked to read at a speed at which they could maintain comprehension, they were aware that they were being timed and were required to record their reading times, which might have placed pressure on them to read more quickly.

Nonetheless, VWM capacity was found to be a significant predictor of reading fluency in the initial reading performance, but not a significant predictor of rate of change over time, even though its effects were in the expected direction. These findings suggest that VWM explains individual differences in reading comprehension-speed in one-shot performance, but it does not explain why individuals differ in their reading fluency development. The significant finding for VWM as a predictor of reading fluency contradicts the results obtained by Pham and Hassam (2014) in the L1 context, who found no relation between VWM and reading fluency. These contradicting results suggest that VWM is involved to a lesser extent in L1 reading where people have mastered the characteristics of the language such as the orthographic system and the visual lexical forms. This hypothesis might need to be explored by future research. The results of this study also provide counter evidence for the lack of correlation between VWM and reading found by Tsai (2014) for Japanese as an L2. Tsai observed, however, that the cloze measure used to tap reading skills assessed grammatical knowledge rather than reading. Regardless, these different results provide further motivation for examining the relationship between VWM and other languages.

These VWM findings can also be interpreted as suggesting an increase in skill proceduralization (DeKeyser, 2007). Evidence from cognitive psychology indicates that the proceduralization of a skill, such as reading, is reflected in a decrease in EWM involvement (Poldrack et al., 2005). That is, novice performance on a task requires cognitive effort, but with practice the task becomes automatized and thus can be performed with less effortful cognitive control. This is a plausible hypothesis because VWM measures of the type employed in this study are closely tied to EWM (Miyake et al., 2001). As the initial reading represented a novel task, it placed demands on the participants' executive cognitive resources, hence the resulting significant path from VWM in the LGC model. However, because the assessments in and of themselves functioned as reading practice, they made the participants more skilled readers over time, which resulted in less effortful reading performance in the subsequent waves as shown by the non-significant paths.

Another plausible explanation for the non-significant predictive power of VWM on rate of reading growth might be that the duration of the study was not sufficient to detect a predictor of the rate of change in reading comprehension-speed. Despite this, the paths to the intercept in the LGC model suggest some evidence of VWM influence on rate of change over time. A longer study, perhaps with an intervention, might have produced greater changes in comprehension-fluency, thus increasing the explanatory power of VWM as a growth predictor.

Nonetheless, the positive results in VWM give support to the use of the newly-developed versions of the Mr. Peanut and reversed Corsi tasks in combination to measure VWM capacity. These tasks measure two separate notions of VWM: visual and spatial WM (Baddeley et al., 2002; Cornoldi & Vecchi, 2003), respectively. That is, the Mr. Peanut task assesses the visual component as its focus is on a static object appearance, whereas the reversed Corsi task assesses the spatial component, specifically the dynamic sequences of movements from one location to another. Together, they might provide a more complete assessment of VWM than either task alone, which may explain why this study, unlike Pham and Hasson's (2014) and Tsai's (2014), found a relationship between VWM and reading. Pham and Hasson (2014) used the reversed Corsi task, while Tsai (2014) used the letter rotation and 3-back tasks. These two tasks measure the dynamic and static properties of VWM in isolation, respectively. This lends credence to the possibility that using the two Mr. Peanut and reversed Corsi tasks in combination increased the sensitivity of the measurement to tap into VWM.

The second goal of this study was to address weaknesses of previous reading fluency measures by developing and validating a novel measure, which combined the two facets of reading fluency, comprehension and speed, into a single construct. A hybrid comprehension-speed reading measure was constructed by transforming the participants' reading-speed times onto Rasch measures and combining these with the test comprehension items into an interval scale comprehension-speed using the MFRM software. Rasch analyses provided validity evidence for the measure. Specifically, the analysis of fit revealed that the comprehension and reading speed items functioned well together to form a unidimensional measure with six levels of difficulty. In addition, the person fit measures showed that the participants performed consistently enough to justify the use of the measure in group analyses in a particular context, such as junior and senior high schools. The Rasch person separation and reliability indices indicated that the measure separated participants into three levels of reading fluency (high, average, and low) and that the probability of replicating this separation in future administrations to similar samples was moderately high. All in all, these findings provide evidence that modeling comprehension together with reading time as a single measure is a valid and practical approach to estimating reading passage difficulty and individual fluency performance.

Conclusion

The overarching purposes of this study were to investigate whether VWM was a predictor of individual differences in reading fluency growth as well as to develop and test the validity of a reading fluency measure designed to assess comprehension and reading speed in a single unidimensional measure. The findings indicated that VWM significantly predicted variation in the individual growth trajectories of the participants at the initial status, but did not significantly explain heterogeneity in the rate of change over time, leaving this relationship inconclusive. It is possible that the duration of the study was not sufficient to create greater changes in reading fluency, thus making it difficult to detect a predictor of the rate of change. Although the relationship was not significant, the results are promising as they point to the predicted direction, providing motivation for future research to explore whether VWM is a significant predictor of reading fluency growth. In sum, these findings lead to the conclusion that VWM explains differences in reading fluency in one-shot performance, at least for L2 learners from logographic written systems such as Japanese.

The current study provides validity evidence for the new reading comprehension-speed measure. The two sets of items, comprehension and reading speed, had adequate fit to the Rasch model and the measure was found to reliably separate participants into three levels of reading fluency. Modeling comprehension together with reading speed through MFRM appears, thus, as a valid approach to the measurement of reading fluency. It is hoped that the new approach to test-building presented here can be used by researchers to construct multifaceted measures.

References

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423. https://doi.org/10.1016/s1364-6613(00)01538-2
Baddeley, A. D. (2010*). Working memory, thought, and action*. Oxford University Press.
Baddeley, A. D., Kopelman, M. D., & Wilson, B. A. (2002). *The handbook of memory disorders* (2nd ed.). John Wiley & Sons, Ltd.
Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.005
Bazan, B. (2020). A Rasch-validation study of a novel speaking span task. *Shiken 24*(1), 1–21. Retrieved from http://teval.jalt.org/node/95
Bolger, D.J., Perfetti, C.A., & Schneider, W. (2005). Cross-cultural effect on the brain revisited: Universal Structures plus writing systems variation. *Human Brain Mapping*, *25*(1), 92–104. https://doi.org/10.1002/hbm.20124
Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd. ed.). Erlbaum.
Brown, L., Lauer A. R., & Uhl, E. (1944) A study of the improvement of reading rate and comprehension. *Proceedings of the Iowa Academy of Science*, *51*(1), 367–370. https://scholarworks.uni.edu/pias/vol51/iss1/41
Byrne, B. M. (2010).*Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge.
Case, R. (1985). *Intellectual development: Birth to adulthood.* Academic Press.
Cornoldi, C., & Vecchi, T. (2003). *Visuo-spatial working memory and individual differences*. Psychology Press.
Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. [Unpublished doctoral dissertation, McGill University] eScholarship@McGill Database. https://escholarship.mcgill.ca/downloads/4m90dw30g
Chang, A. C-S. (2012). Improving reading rate activities for EFL students: Timed reading and repeated oral reading. *Reading in a Foreign Language*, *24*, 56–83. Retrieved March 23, 2018, from https://files.eric.ed.gov/fulltext/EJ1015754.pdf
Chang, A. C-S., & Millet, S. (2013). Improving reading rates and comprehension through timed repeated reading. *Reading in a Foreign Language*, *25*, 126–148. Retrieved March 23, 2018, from https://nflrc.hawaii.edu/rfl/item/364
Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6
DeAvila, E. (1974). *Children's transformations of visual information according to nonverbal syntactical rules.* [Unpublished doctoral dissertation]. York University.
DeKeyser, R. M. (2007). Study abroad as foreign language practice. In R. M. Dekeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 208–226). Cambridge University.
De Renzi, E., & Nichelli, P. (1975). Verbal and nonverbal short-term memory impairment following hemispheric damage. *Cortex*, *11*, 341–353. https://doi.org/10.1016/S0010-9452(75)80026-8
Dronjic, V., & Bitan, T. (2016). Reading, brain, and cognition. In B. X. Chen, B. X. Dronjic, & R. Helms-Park, (Eds.), *Reading in a Second Language: Cognitive and Psycholinguistic Issues* (pp. 33–69). Routledge.
Field, A. (2009). *Discovering Statistics Using SPSS (3rd ed.)* Sage.

Gathercole, S. E., & Baddeley, A. D. (1993). *Essays in cognitive psychology. Working memory and language.* Erlbaum.

Geva, E., & Ryan, E. (1993). Linguistic and cognitive correlates of academic skills in first and second languages, *Language Learning*, *43*, 5–42. https://doi.org/10.1111/j.1467-1770.1993.tb00171.x

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice.* Cambridge University Press.

Grabe, W. (2010). Fluency in reading: Thirty-five years later. *Reading in a Foreign Language*, *22*, 71–83. http://www2.hawaii.edu/~readfl/rfl/April2010/articles/grabe.pdf

Gray, W. (1969). *The teaching of reading and writing*. UNESCO.

Hamasaki, T., Yasojima, K., Kakita, K., Masaki, H., Ishino, S., & Murakami, M. (1995). Alexie-agraphie pour l'ecriture kanji apres lesion temporale postero-inferieur gauche. *Rev. Neurol.*, *155*, 16–23.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, *14*, 25–38. https://doi.org/10.1017/S0272263100010457

Ichikawa, S. (1983). Verbal memory span, visual memory span, and their correlations with cognitive tasks. *Japanese Psychological Research, 25,*173–180. https://doi.org/10.4992/psycholres1954.25.173

Kim, S-A., Christianson, K., & Packard, J. (2015). Working memory in L2 character processing: The case of learning to read Chinese. In E. Z. Wen, M. Mota, & A. McNeill (Eds.), *Working Memory in Second Language Acquisition and Processing* (pp. 85-104). Multilingual Matters.

Klauda, S., & Guthrie, J. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology, 100*, 310–321. https://doi.org/10.1037/0022-0663.100.2.310

Kuhn, M., & Stahl, S. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95,* 3–21. https://doi.org/10.1037/0022-0663.95.1.3

Larson-Hall, J. (2016). *A guide to doing statistics in second language research in SPSS and R* (2nd ed.). Routledge.

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning, 57*(2), 229–270. https://doi.org/10.1111/j.1467-9922.2007.00408.x

Linacre, J. M. (2002). What do infit, outfit, mean-square, and standardized mean? *Rasch Measurement Transactions, 16(2),* 878.

Linacre, J. M. (2007). *A user's guide to WINSTEPS: Rasch-model computer program*. MESA.

Linacre, J. M. (2017). Facets computer program for many-facet Rasch measurement (Version 3.80.0). Winsteps.com. htpps://winsteps.com

Linacre (2018). Winsteps computer program (Version 4.0.0). Winsteps.com. htpps://winsteps.com

McNamara, T. (1996). *Measuring second language performance*. Longman.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*(4), 621–640. https://doi.org/10.1037/0096-3445.130.4.621

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.

Nation, I. S. P., & Macalister, J. (2021). *Teaching ESL/EFL reading and writing* (2nd ed.). Routledge.
National Reading Panel (NRP). (2000). *Reports of the subgroups: National Reading Panel*. National Institute of Child Health and Development.
Nelson, J. R., Liu, Y., Fiez, J., & Perfetti, C. A. (2009) Assimilation and accommodation patterns in ventral occipitotemporal cortex in learning a second writing system. *Human Brain Mapping, 30*(3), 810–820. https://doi.org/10.1002/hbm.20551
Oller, J. W., & Tullius, J. R. (1973). Reading skills of non-native speakers of English. *International Review of Applied Linguistics*, *11*, 69–80.
Pae, H. K., & Sevcik, R. A. (2011). The role of verbal working memory in second language reading fluency and comprehension: A comparison of English and Korean. *International Electronic Journal of Elementary Education, 4*(1), 47–65. https://www.iejee.com/index.php/IEJEE/article/view/213
Pham, A. V., & Hasson, R. M. (2014). Verbal and visuospatial working memory as predictors of children's reading ability. *Archives of Clinical Neuropsychology*, *29*(5), 467–477. https://doi.org/10.1093/arclin/acu024
Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y., & Knowlton, B. J. (2005). The neural correlates of motor skill automaticity. *The Journal of Neuroscience, 25*(22), 5356–5364. https://doi.org/10.1523/JNEUROSCI.3880-04.2005
Porte, G. K. (2002). *Appraising research in second language learning: A practical approach to critical analysis of quantitative research*. John Benjamins.
Sakurai, Y., Sakai, K., Sakuta, M., & Iwata, M. (1994): Naming difficulties in alexia with agraphia for kanji after a left posterior inferior temporal lesion. *J Neurol Neurosurg Psychiatry, 57*, 609–613. https://doi.org/10.1136/jnnp.57.5.609
Shimono, T. R. (2018). L2 reading fluency progression using timed reading and repeated oral reading. *Reading in a Foreign Language*, *30(1)*, 152–179. https://nflrc.hawaii.edu/rfl/item/403
Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33*, 5–42. https://doi.org/10.1006/cogp.1997.0658
Tan, L. H., Laird, A., Li, K., & Fox, P.T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Human Brain Mapping*, *25*, 83–91. https://doi.org/10.1002/hbm.20134
Therrien, W., Wickstrom, K., & Jones, K. (2006). Effects of a combined repeated reading and question generation intervention on reading achievement. *Learning Disabilities Research and Practice, 21*, 89–97. https://doi.org/10.1111/j.1540-5826.2006.00209.x
Tsai, A. M. (2014). The role of visuospatial and verbal working memory in L2 Japanese reading proficiency. *Second Language Studies*, *32*(2), 76–113. http://www.hawaii.edu/sls/wp-content/uploads/2014/08/Tsai-Aurora1.pdf
Vadasy, P., & Sanders, E. (2008). Repeated reading intervention: Outcomes and interactions with readers' skills and classroom instruction. *Journal of Educational Psychology, 100*, 272–290. https://doi.org/10.1037/0022-0663.100.2.272
Wen, Z. (2016). *Working memory and second language learning: Towards an integrated approach*. Multilingual Matters.
Wiederholt, J. L., & Bryant, B. R. (2001). *Gray oral reading test* (4th ed.). Pro-Ed.
Yokohama, S., Kim, J., Uchida, S., Miyamoto, T., Yoshimoto, K., & Kawashima, R. (2013). Cross-linguistic influence of first language writing systems on brain responses to second language word reading in late bilinguals. *Brain and Behavior*, *3*(5), 272–290. https://doi.org/10.1002/brb3.153

Zhu, L., Nie, Y., Chang, C., Gao, J.-H., & Niu, Z. (2014). Different patterns and development characteristics of processing written logographic characters and alphabetic words: An ALE meta-analysis. *Human Brain Mapping*, *35*(6), 2607–2618. https://doi.org/10.1002/hbm.22354

Appendices

Appendix A

Read

Kate Grant is fifteen and comes from Canada. She is visiting Holland with her mother and father. On June 8th, the Grants are at a museum in The Hague. "This is very good", Mr. Grant says. He is looking at a picture of an old man.
"Yes," Mrs. Grant says. "What do you think, Kate?"
"It's OK," Kate says. She looks at her watch. It is 4.45. Two minutes later she sees a picture of a young girl with an earring. Kate smiles. She loves earrings. She looks at the girl's face. "Who are you?" she thinks. "What are you looking at?"
"That's a very famous picture," Mrs. Grant says. A sign next to it says, "Girl with a Pearl Earring." "It's beautiful," Kate says.

At 4.58, the Grants are drinking coffee near the museum. "Julia, do you have the guidebook?" Mr. Grant asks. "No, Carl. You have it," Mrs. Grant says.
They look for the guidebook in their bags, then Mr. Grant says. "Oh no, I remember now. It's on a chair near 'Girl with a Pearl Earring'. My passport is in that book!" Kate gets up quickly. "Stay here," she says. "I can find it."

Kate runs to the museum. A guard is closing the door. "Stop!" Kate says. "My father's passport is in the museum."
"I'm sorry, but we close at five o'clock," the guard says. A sign on his shirt says, "Paul Van Dyck."
"Please," Kate says. "It's very important." The guard looks at his watch, then at Kate. "OK," he says and smiles. "Come with me."

Kate goes into the museum with Paul Van Dyck. It is dark and there aren't any people in the rooms.
"Where is the passport?" Paul asks. "Do you know?"
"Yes. It's on a chair near that picture of a girl with an earring," Kate says.
They walk for a minute, then Paul Van Dyck stops. "Sshh!" he says. "I can hear voices"

Kate looks at Paul. He isn't smiling now. Then, she hears voices, too. "Be very, very quiet," Paul says. He takes a phone from his coat pocket, but suddenly there are two men in the room. One is tall and has a big, black bag in his hands. The short man has a gun. They look at Kate and Paul. Kate and Paul look at the men. Time stops.
"Put your phone on the floor," the short man says.

Comprehension Questions

1. 1. What is the story about?
   a) About a girl in a museum.
   b) About a family visiting Holland.
   c) About a museum guard.

2. Which picture does Kate like?
   a) The picture of the old man.
   b) The picture of the girl with the earring.
   c) She likes both pictures.

3. What is in the guidebook?
   a) Kate's passport.
   b) Kate's father's passport.
   c) The family's passports.

4. Where is the guidebook?
   a) It's in the restaurant near the museum.
   b) It's in Kate's father's bag.
   c) It's on a chair near a picture.

5. Who do Kate and Paul see in the museum?
   a) Two men with a phone.
   b) Two thieves.
   c) Two men with bags.

Appendix B

<u>Read</u>

Per and his girlfriend, Nina, live in a big town in Norway. They are on holiday in the country. They are driving on the road near a big forest. "It's beautiful here," Per says. "Let's stop the car and walk in the forest." But the forest is very dark and quiet. "I don't like this place," Nina says. "I'm cold. Let's go back to the car."

Suddenly, Per falls over a tree. "Are you OK?" Nina asks. "Yes," Per answers. Then he sees some old stones. "These are very strange stones," he says. He moves them with his hands. They are very heavy. "There's a hole under the stones," he says. He puts his hand into the hole. "Be careful," Nina says. "Maybe an animal lives in there."

Per takes his hand from the hole. He has an old cup in his hand. "Look at this old cup, Nina!" he says. "It's very interesting. But why is it here? And look! There are some strange words on it." Nina looks at the cup, too. "It's beautiful," she says. She looks again. "And very expensive!" she thinks.

The forest is very dark and cold. "Let's go now," Nina says. "I don' like this forest." They walk back to their car. Suddenly, Per stops. "What is it?" Nina asks. Per looks into the trees. "There's a man there," he says. Nina looks too. "I can't see a man," she says. Per looks again. "You're right," he says. "It's only a tree."

Per and Nina drive to a small town. "It's getting dark," Per says. "Let's stay in this town." They find a small hotel. The people are very friendly. "I like this hotel," Nina says. "Let's stay here." Near the hotel is a museum. "We can take the cup to the museum in the morning," Per says. "We can ask the museum keeper about it."

Later, Per can't sleep. He goes to the window of his room and looks into the hotel garden. A strange man is standing there. He doesn't move. "Who are you?" Per calls. "What do you want?" Then Nina comes into the room. "There's a strange man in the garden," Per says. "What man?" Nina asks. "I can't see a man."

In the morning, Per and Nina take the cup to the museum. The museum keeper is an old man. 'Do you know about this cup?' Per asks.

Comprehension Questions

1. What is the story about?
   a) About a magic forest.
   b) About a boy and a girl.
   c) About some strange stones.

2. Why do Per and Nina walk into the forest?
   a) Because they see a man.
   b) Because the forest is beautiful.
   c) Because their car doesn't have gasoline.

3. Where is the cup?
   a) It is in a hole in the ground.
   b) It is under a tree.
   c) It is on some strange stones.

4. What does the cup have on it?
   a) It has some strange words.
   b) It has some strange stones.
   c) It has some expensive diamonds.

5. Who sees the strange man?
   a) Per and Nina do.
   b) Only Per does.
   c) Only Nina does.

About the Authors

Bartolo Bazan is an English teacher at Ryukoku University Heian Junior & Senior High School. He holds a M.S.Ed. in TESOL and is currently a Ph.D. candidate in the eighth cohort at Temple University (Japan) for Applied Linguistics. His research interests include working memory, skill automaticity, and assessment. E-mail: bazanlinkin2@gmail.com