Philology Matters / ISSN: 1994 – 4233
Uzbekistan State World Languages University
2025 Vol. 52 No. 1
METHODOLOGY
DOI: 10.36078/987655228

Received: January 17, 2025
Accepted: March 20, 2025
Available online: March 25, 2025

Kuniyuki Noto

Doctor of Philosophy in Language and Culture (PhD)
Uzbekistan State World Languages University
Tashkent, Uzbekistan
E-mail: notokuniyuki@gmail.com
ORCID iD: 0009-0002-5609-8441

CAN INDIVIDUAL DIFFERENCES IN WORKING MEMORY CAPACITY PREDICT LANGUAGE APTITUDE: MEASUREMENT USING THE READING SPAN TEST

ABSTRACT

Language teachers, through their experience, may have perceived the presence or absence of what is often referred to as 'Language Aptitude' in their students. This paper reviews previous research on the relationship between Working Memory (WM) and learning outcomes, exploring the potential applications of these findings in educational settings. The WM model has been expected to contribute not only to language acquisition but also to the understanding of cognitive functions, ranging from dementia to developmental disorders, and its potential applications in clinical settings have been widely recognized. However, to date, these insights have not led to revolutionary applications in educational settings.

In this experiment, the Reading Span Test (RST) was used to measure WM in Chinese learners of Japanese. If WM is shown to be a reliable predictor of learning outcomes, it could become a valuable tool, akin to placement tests. The Japanese version of the RST was translated into Chinese, and test materials were adapted accordingly, highlighting several challenges that warrant further investigation. Notably, the relationship between individual differences in memory capacity and language aptitude remains unclear without concurrently addressing what constitutes a valid assessment of language proficiency.

The results suggest that learners with higher WM capacity exhibit remarkably stable

Kuniyuki Noto

Til va madaniyat bo'yicha falsafa doktori (PhD)
O'zbekiston davlat jahon tillari universiteti
Toshkent, O'zbekiston

ISHCHI XOTIRA SIG'IMIDAGI INDIVIDUAL FARQLAR TIL QOBILIYATINI BASHORAT QILA OLADIMI: "O'QISH ORALIG'I" TESTI YORDAMIDA O'LCHASH

ANNOTATSIYA

Ushbu maqolada til o'rgatuvchi pedagoglar o'z tajribalari asosida talabalarda til o'zlashtirish qobiliyatining (language aptitude) mavjudligi yoki yo'qligini sezishlari mumkinligi bilan bog'liq holatlar muhokama qilinadi. Shu munosabat bilan ishchi xotira (IX) va o'rganish natijalari o'rtasidagi bog'liqlik o'rganilgan avvalgi tadqiqotlar tahlil qilinadi hamda ularning ta'limiy amaliyotlarda qo'llanish ehtimollari ko'rib chiqiladi. IX modeli nafaqat til o'zlashtirish, balki demensiya va rivojlanishdagi buzilishlar kabi kognitiv funksiyalarni tushunishda ham muhim omil sifatida e'tirof etilgan. Shunga qaramay, bu tushunchalar hanuz ta'limda sezilarli o'zgarishlarga olib kelmagan.

Tadqiqot doirasida yapon tilini o'rganayotgan xitoylik talabalar ishtirokida IXni baholash uchun "o'qish oralig'i" testi (Reading Span Test) qo'llanildi. Agar IX o'rganish natijalarining ishonchli bashoratchisi sifatida isbotlansa, bu metod darajaviy testlarga o'xshash muhim diagnostik vosita bo'lishi mumkin. Tadqiqot davomida test materiali xitoy tiliga moslashtirildi, ammo bu jarayonda yuzaga chiqqan bir qator muammolar yanada chuqurroq tahlilni talab qiladi. Ayniqsa, IX sig'imidagi individual farqlar va til o'rganish qobiliyati o'rtasidagi bog'liqlikni til kompetensiyasini ishonchli baholamasdan turib yetarlicha asoslangan, deb bo'lmaydi.

Natijalar shuni ko'rsatdiki, yuqori IX sig'imiga ega bo'lgan talabalar test sinovlarida barqaror natijalar ko'rsatgan, past sig'imga ega

performance, contrasting with those with lower WM capacity, who tend to experience significant score drops when encountering unfamiliar test formats. Beyond the superficial scores of paper tests, learners with higher WM capacity may be steadily acquiring Japanese language proficiency.

Accurately and efficiently measuring WM could significantly advance research into its role in second language acquisition. Therefore, further research is needed to refine the methods for measuring WM in learners of Japanese. This study represents an initial step toward achieving that goal.

Key words: working memory, language aptitude, reading span test, second language learning, memory capacity, learning outcomes, individual differences, Chinese version of RST, performance, placement test.

bo‘lganlar esa noodatiy test shakllarida sezilarli ball yo‘qotgan. Bunday kuzatishlar, hatto yuzaki test natijalaridan tashqari, yuqori IXga ega talabalarning til o‘zlashtirish jarayonida barqarorlikni saqlab qolayotganini anglatadi.

Shunday ekan, ikkinchi til sifatida yapon tilini o‘zlashtirayotgan talabalar orasida IXni aniq va samarali o‘lchash metodikasini ishlab chiqish, mazkur sohadagi tadqiqotlarni yangi bosqichga olib chiqishi mumkin. Mazkur tadqiqot ushbu yo‘nalishda amalga oshirilgan dastlabki ilmiy qadam sifatida baholanadi. Binobarin, zamonaviy til o‘rganish metodologiyasida individual kognitiv omillarni hisobga olish zarurati ortib bormoqda, bu esa IXni til o‘zlashtirishdagi muhim diagnostik vosita sifatida o‘rganishning dolzarbligini ko‘rsatadi.

Kalit so‘zlar: ishchi xotira, til o‘zlashtirish qobiliyati, “o‘qish oralig‘i” testi, ikkinchi tilni o‘rganish, xotira sig‘imi, o‘rganish natijalari, individual farqlar, O‘OTning xitoycha varianti, natijadorlik, darajaviy test.

INTRODUCTION

Over the past few decades, research on Working Memory (WM) has advanced significantly [Baddeley & Hitch, 1974; Just & Carpenter, 1992; Osaka, 2002; Unsworth & Engle, 2007; Mukoyama, 2009; Futakuchi, 2015]. However, the practical applications of WM that were initially anticipated when it began to gain attention do not seem to have unfolded as expected. Among these, the relationship with language learning was included in early studies, leading to the expectation of a significant impact on educational settings [Baddeley et al., 1998]. Yet, it appears that the research is still in the stage of refining its precision.

The application of WM has shifted, responding to societal needs, from the field of language to broader cognitive issues. In fact, it has been applied in medical settings, particularly in the study of dementia. In advanced countries, where aging societies have become a key issue, there seems to be a higher interest in dementia compared to second language acquisition. However, while the ultimate goal of preventing the onset of dementia was considered, just as with language learning, WM is still primarily being used as one tool in dementia testing. It can be argued that the expectations for WM research to open new paths for second language acquisition were somewhat premature, resulting in an outcome similar to that of the dementia focus. Furthermore, WM research has expanded into exploring its relationship with developmental disabilities, aligning with societal trends [Gathercole & Baddeley, 1990]. Both dementia and developmental disabilities are related to high-level cognitive functions, and it is clear that WM plays a significant role in these areas [Papagno & Vallar, 1995; Baddeley, 2003].

However, despite the vast number of studies, the application of WM remains

hindered by the lack of a unified model that concretely explains the mechanisms of how WM capacity imposes constraints on our activities [e.g., Brown & Hulme, 1995; Friedman & Miyake, 2004]. This difficulty is evident in the fact that, despite numerous studies, there is no consensus on how WM influences our cognitive activities, leading to a cycle of new experiments being conducted without clear interpretations [Takahashi et al., 2006]. A prime example of this is the abundance of WM-related books in bookstores under the guise of "Brain Training".

This paper aims to avoid falling into such pitfalls by first conducting the Reading Span Test (RST), which is used to measure WM capacity, and examining its relationship with language learning outcomes. Although the claim that WM capacity is related to language learning has been consistently made, there are also experimental results that fail to show a clear relationship. The interpretation of experimental results suggests that the effects may vary depending on the stage of learning or the type of skill involved. What the RST scores represent and how they relate to language learning seems to be a complicated maze of hypotheses that have yet to be fully validated. Therefore, the RST will be discussed in detail in the following sections.

Working memory and learning outcomes

Recently, research has shifted away from the dead-end discussions surrounding "memory" to studies based on the model of "attention" [Minamoto et al., 2010; Unsworth et al., 2014]. In N.Osaka, it is argued that "working memory capacity is an indicator of an individual's attentional control abilities, which can be measured by the working memory span test" [Osaka, 2013]. Furthermore, N.Osaka presents the hypothesis that the correlation between WM capacity and reading comprehension does not reflect specific abilities but rather reflects a general attentional control function associated with the frontal lobe. To put it simply, while WM capacity can predict reading comprehension scores, this is not because WM directly supports reading ability, but because WM's role in controlling thinking ultimately impacts reading performance. In M.Yuzawa, the research findings suggest that the relationship between WM capacity and reading comprehension may be partially mediated by off-task thinking [Yuzawa & Yuzawa, 2014]. Interestingly, when the frequency of off-task thinking is considered, the correlation between WM performance and reading comprehension becomes significantly weaker. In other words, it was suggested that WM capacity affects the occurrence of off-task thinking, and this, in turn, can lead to poorer reading performance. Whether in class or during reading, everyone occasionally finds themselves thinking about irrelevant things. Naturally, when this occurs frequently, the content of the lesson is not absorbed, and comprehension tends to lag. In the process of text comprehension, if off-task thinking occurs frequently – if one is not good at controlling their thoughts to stay focused on the task at hand – then reading performance will likely decline. This indicates that the ability to sustain attention on what is necessary, which can simply be called concentration, is related to WM capacity.

On the other hand, there is Cognitive Load Theory, which has developed relatively independently from the theoretical studies of WM mentioned earlier. This

model posits three types of cognitive load: extraneous load (the load generated by instructional design and the structure of learning materials), intrinsic load (the load arising from the internal processing required by the task), and germane load (the resources needed for task performance). According to this theory, if the total load exceeds working memory capacity, learning efficiency will decline. For example, consider a situation in Japanese language education where we are deciding whether to present learning materials via video or static images. Given the advancements in the convenience of viewing devices, one might think that it would be better to use videos, assuming they provide a richer experience. However, in terms of information volume, videos contain far more content than static images, and for learners with smaller WM capacity, this may actually be detrimental. The external load imposed by the video could exceed their WM capacity, leading to a negative impact on learning. Similarly, even with static images, the impact could be different depending on whether the image contains detailed depictions of backgrounds and people, or if it is simply a stylized, simplified image. Materials that are intended to help learners may, depending on individual differences in WM capacity, have a negative impact.

Learners with larger WM capacities can focus on the necessary parts of the presented information, even when there is a lot of content, so they are less likely to be affected. This leads to differences in learning outcomes between learners. In response, suppose we mark specific areas of the image to guide where the learner should focus. This would reduce the extraneous load for learners with smaller WM capacities, allowing them to process the information within their capacity range, which would enhance their learning (this is referred to as the "attention isolation effect"). However, this approach would now lower the performance of learners with larger WM capacities. For them, the unnecessary markings in the image would increase redundancy and have a negative effect (this is called the "redundancy effect").

What this shows is that the appropriate method for reducing cognitive load varies depending on individual differences in WM capacity. In T.Tanaka, it was revealed that participants with smaller WM capacities tend to use active annotation more frequently when reading texts, which is related to WM's role in attentional control [Tanaka, 2017]. Experienced teachers, intuitively observing learners' responses, likely adjust materials accordingly to improve their lessons. On the other hand, less experienced teachers may feel that their efforts to use materials do not produce the desired results. This could be an example of how theory has not yet fully caught up with practical classroom experience. M.Yuzawa also stated that "the impact of off-task thinking in educational settings is thought to be significant, but practical research on this has only just begun. It is expected to develop further, grounded in theoretical foundations" [Yuzawa & Yuzawa, 2014]. Thus, understanding individual differences in WM can contribute not only to class grouping but also to guiding teaching methods [e.g., Tsuchida & Murohashi, 2017].

The relationship between the Reading Span Test and language proficiency

The Reading Span Test is the most well-known test for measuring working memory capacity [Conway et al., 2005; Tanaka et al., 2014]. It involves a dual-task

where participants read sentences aloud while simultaneously remembering target words in the sentences. This is considered a reasonable method for measuring WM capacity, as it balances both information processing and retention, key functions of WM. According to N.Osaka, one of the strengths of the RST is that it can predict intelligence based on test performance [Osaka, 2013]. Specifically, it has been shown that high scorers on the RST tend to perform well on tasks involving reading comprehension, a cognitive function related to language, as evidenced by the research of M.Daneman and P.A. Carpenter [Daneman & Carpenter, 1980]. Using various types of reading strategies with high frequency may allocate working memory capacity to those strategies, potentially hindering text comprehension [Futakuchi, 2020]. Additionally, individuals with high scores in RST are reported to exhibit better performance in processing relative clauses in a second language [Kashiwagi & Nakayama, 2012].

Building on these findings, there have been numerous studies in Japan that examine the relationship between WM and English language learning. For example, K.Ariji suggests that WM involves both syntactic processing and vocabulary processing, and posits that separate WM systems are required for the first and second languages [Ariji, 2016]. Although it is still unclear exactly what the RST measures, he concludes that the RST score reflects WM capacity specific to vocabulary meaning processing in the second language, and this capacity is a key determinant of proficiency in reading comprehension in English. Aru treats language-related WM as having separate capacities for vocabulary meaning processing and syntactic processing.

On the other hand, K.Sugiura and H.Nakanishi, who also focused on Japanese learners of English, believe that WM has a stronger influence on English comprehension when language processing has not yet been automated [Sugiura & Nakanishi, 2011]. They investigate the correlation between RST scores and vocabulary processing ability or vocabulary knowledge, approaching the issue from a vocabulary perspective. They suggest that teachers should adjust their approach based on the learners' vocabulary knowledge. For learners with a large vocabulary, it might be effective to leverage their existing knowledge, while for learners with a smaller vocabulary, it may be more effective to focus on expanding their vocabulary.

There are also studies that propose dividing WM into long-term and short-term types, similar to how memory is divided into long-term and short-term categories [Morishima, 2013]. Long-term WM plays an important role in the reading comprehension process, but it has been suggested that Japanese learners of English do not fully utilize their long-term WM. In K.Sakakibara, no correlation was found between reading comprehension and RST (WM capacity) among Japanese English learners [Sakakibara, 2013], while M.Satori found that WM is related to listening comprehension ability [Satori, 2012].

As shown above, there are various perspectives and conflicting results regarding the relationship between RST-measured WM capacity and second language proficiency. In addition to examining what the RST actually measures, the complexity of defining reading comprehension ability itself also complicates the interpretation of the results. For example, in a study with Malaysian learners of Japanese, no correlation was found

between WM and Japanese reading comprehension [Yoshikawa & Zoraida, 2017]. The issue of what specific skills are needed for reading comprehension still remains unresolved. Although the RST is an effective tool for measuring WM capacity, clarifying the relationship between WM and language proficiency becomes difficult without a clear method for measuring the other aspect of language proficiency.

Nonetheless, among many studies focusing on Japanese learners of English, the research involving foreign learners of Japanese, where WM capacity was measured using the Malay version of the RST, is a valuable contribution. While it should not directly affect WM capacity whether the test is done in the native language or the target language, reading aloud in the target language requires a certain amount of learning time and proficiency. This study focuses on Chinese learners of Japanese and uses Mandarin, the common language among the participants, for the RST.

Based on the various interpretations and research findings reviewed above, the following characteristics are generally noted among high scorers on the RST: A. They are able to hold words in memory for a longer time while reading. B. They can instantly utilize syntactic structures and meaning-related information. C. They are able to devise various strategies to handle the RST tasks. D. They can form a focus and suppress unnecessary information. E. They are skilled in updating their attention and changing their focus as needed. F. They are effective in self-monitoring to check whether they are responding appropriately to the task.

METHODS

Objective

The aim of this study is to investigate whether there is a relationship between Japanese language learning outcomes and individual differences in WM capacity (measured by the Reading Span Test). In this experiment, we aim to verify the validity of the Japanese version of the RST, translated by N.Osaka, as an effective method for measuring WM capacity in Chinese learners of Japanese.

Participants

35 Japanese government scholarship students (18 students in 2017, 9 males and 9 females, and 17 students in 2018, 9 males and 8 females). The participants had completed 10 months of Japanese language education before coming to Japan, where they were pursuing doctoral degrees at Japanese graduate schools. Although the participants came from various regions of China, they were all in their mid-to-late 20s and held master's degrees. The participants were selected from those who had started learning Japanese from a beginner level. The experiment was conducted six months after they began their studies, at a stage when they had reached an intermediate level of Japanese proficiency.

Procedure

The 2012 version of the RST by N.Osaka was translated into Chinese by five students from the Japanese Department of Northeast Normal University, who were either students or graduate students. This was followed by a review and feedback process from other students to refine the translation. The final version of the Chinese

RST used in this experiment is presented in Appendix 1.

On a computer screen, one sentence at a time was visually presented. The participants read the sentence aloud while trying to memorize the target words, which were underlined in red. After each trial, they were asked to recall the target words. After reading a sentence, the experimenter pressed the Enter key to present the next sentence, and the procedure was repeated. A blank page indicated the end of a trial, at which point the participants began recalling the target words. There were five trials for each condition, ranging from two-sentence to five-sentence conditions. The recall was written on a response sheet, not spoken aloud. The participants were told they could use Pinyin if they couldn't recall the characters.

In terms of reading speed, the participants were instructed to read at a natural pace, similar to how they would read a book, avoiding deliberately slowing down to emphasize the target words or rushing through other parts of the sentence. If they made a mistake while reading, they were instructed to continue reading without correcting the error.

When recalling the target words, they were instructed to reproduce them in the order they had memorized, ensuring that the last target word was not recalled first. Once they had completed the recall or could not remember, they moved on to the next trial.

The progress in Japanese language proficiency was measured through regular exams conducted approximately once a month. In the beginner level, exams were administered after every 10 lessons of the main textbook, totaling six exams for 60 lessons. In the intermediate level, two exams were conducted, one for the first half and one for the second half of the course. Thus, a total of eight exams were used for regular assessment. The exam format included questions on grammar, vocabulary, Kanji, listening comprehension, and reading comprehension, presented in both written and multiple-choice formats. There was no speaking test. Since failing the final (eighth) exam would result in the cancellation of the student's study abroad program, learner motivation was high.

Score calculation method

There are several methods for calculating the Reading Span Test scores [Conway et al., 2005]. In this experiment, three different methods were used: (1) span score, (2) total number of correct responses, and (3) the number of correct responses within the successfully completed trials.

Span score

The span score was calculated by the number of sentences correctly recalled in at least 3 out of the 5 trials. If 2 trials were completed correctly, 0.5 points were awarded. For example, if a participant successfully recalled the target words in only 2 trials out of 3-sentence conditions, their score would be 2.5. Since this experiment involved trials up to the 5-sentence condition, the score range was from 0 to 5.

Total number of correct responses

This method scores the total number of target words correctly recalled, regardless

of whether the trial was completed or not. The score range for this method is from 0 to 70.

Number of correct responses in completed trials

For this method, the number of correct target words was counted in the trials that the participant completed successfully. The score range is also from 0 to 70.

The second method, the total number of correct responses, is not directly related to the increasing load on WM as the number of sentences increases. Therefore, the span score (Method 1) is considered to reflect the participant's WM capacity more accurately. However, since the span score has only seven levels (increments of 0.5), it does not capture much variability. In contrast, the third method, which counts the number of correct responses in successful trials, better reflects individual differences in the span score, offering a more detailed measure of performance.

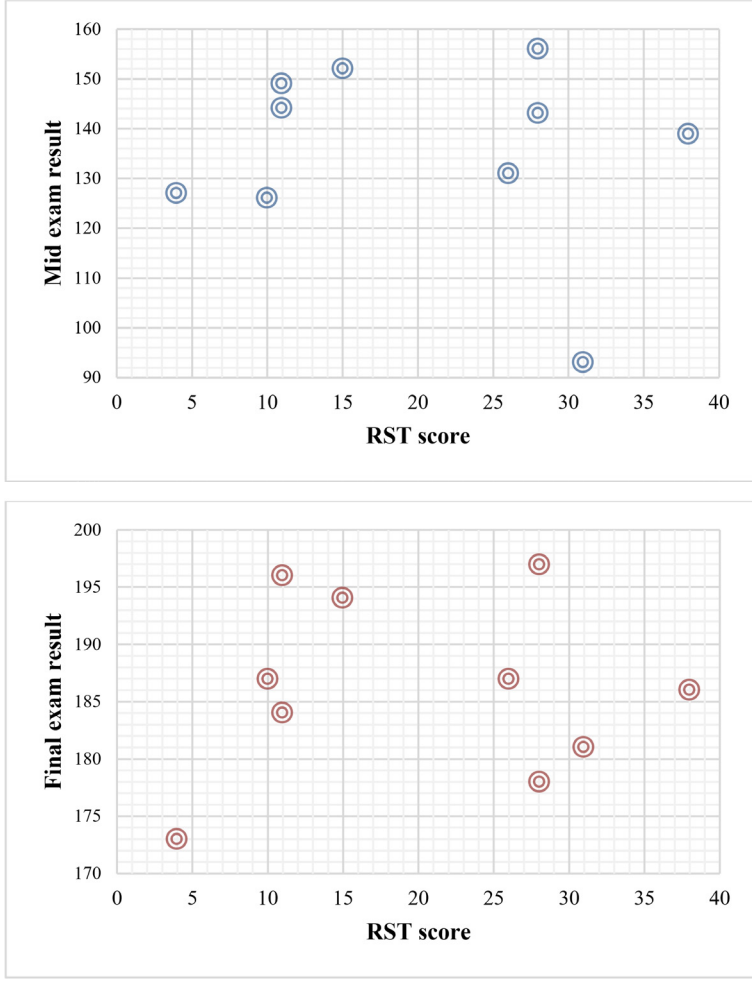
RESULTS AND DISCUSSION

The results for each RST score are shown in Table 1 (See Table 1). No correlation was found between score scores and performance on the Japanese language test, either on the midterm exam (at the end of the elementary level) or on the final exam, Figure 1 (correlation coefficients: midterm: r=−0.14, final: r=0.08) (See Figure 1).

Table 1. *Scoring of RST using three methods*

		Span score	Total number	Total in completed			Span score	Total number	Total in completed
2017	M	4.5	58	36	2018	F	3.5	54	38
2017	F	4.5	54	36	2018	M	3	48	31
2017	F	4.5	50	36	2018	M	3	48	28
2017	M	3.5	53	32	2018	F	3	50	28
2017	F	3	46	23	2018	M	3.5	52	26
2017	F	3	46	23	2018	F	3	49	26
2017	F	3	44	22	2018	F	2.5	39	25
2017	M	3	41	21	2018	M	3	31	20
2017	M	3	40	19	2018	F	3	44	20
2017	F	3	35	19	2018	M	2	40	15
2017	M	3	45	19	2018	M	2.5	41	14
2017	F	3.5	40	18	2018	F	2.5	40	14
2017	M	2.5	31	16	2018	F	2.5	31	14
2017	M	2	41	15	2018	F	2	33	11
2017	F	2.5	30	14	2018	M	2	34	11
2017	M	2	28	8	2018	M	2	33	10
2017	M	2	34	6	2018	M	1.5	26	4
2017	F	2	32	6					

Figure 1. *Correlation between RST scores and Japanese language test scores*



The top graph illustrates the results of the midterm exams, while the bottom graph presents the results of the final exams. In both cases, no significant correlation with RST score was observed.

Issues with the method of measuring Japanese proficiency

The Japanese language tests conducted in this study were primarily achievement tests, which may not have been suitable for highlighting differences in learners' individual proficiency levels. Therefore, these tests were not ideal materials for examining the relationship between WM capacity and Japanese language learning outcomes. Due to procedural constraints, we used the existing regular exams in the curriculum. However, instead of merely assessing whether learners have mastered the content taught in class, it would be more desirable to conduct a separate test that evaluates their current level of Japanese language proficiency.

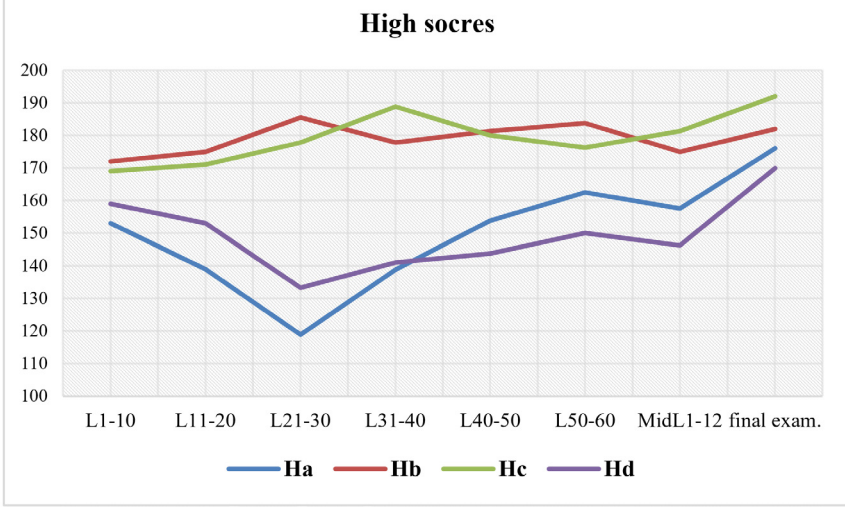
Issues with the translated RST materials

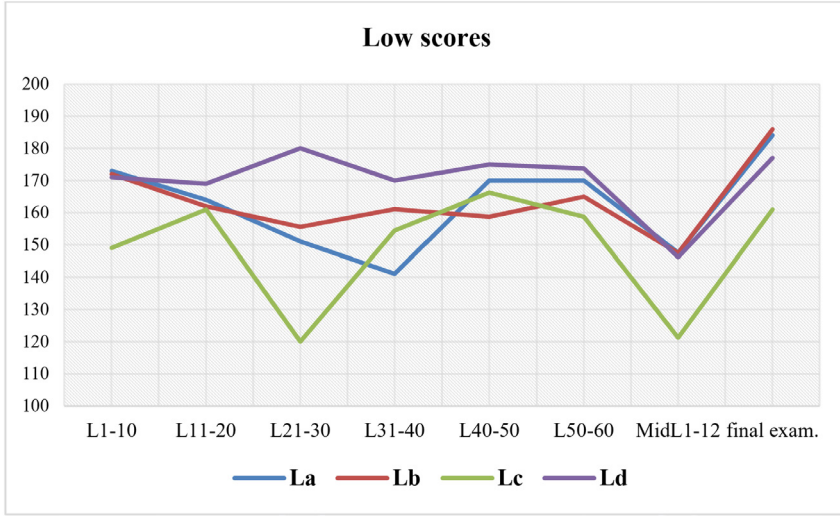
Additionally, a notable issue with the translated version of the Reading Span Test is the significant variation in sentence length. The number of characters, or sentence length, directly affects the cognitive load on memory span. Therefore, we attempted to minimize the differences in character count between sentences. Despite these efforts, the shortest sentence had 13 characters and the longest had 25 characters, resulting in nearly a twofold difference in reading time (as the number of characters corresponds to the number of syllables in Chinese). This disparity is not ideal. The Japanese version of the sentences, sourced from middle school language textbooks, maintains a consistent level of difficulty. For the Chinese version, it might be more appropriate to use sentences that are not merely translations from Japanese but are instead selected from texts that middle school students can comfortably read.

The tendencies observed in the low-scoring group

On the other hand, an interesting trend was observed. The high-scoring RST group showed predictable results in the tests. They either maintained their high performance or showed consistent improvement. In contrast, the low-scoring group exhibited a wide range of scores that were difficult to predict (See Figure 2). This suggests that the high-scoring group was steadily acquiring Japanese, regardless of the exam format, while the low-scoring group's results seemed more dependent on luck, reflecting variations in their test performance. This was particularly evident when the test scores of the low-scoring group dropped significantly when they entered the intermediate-level exam, while the high-scoring group remained unaffected. This drop was likely due to the difference in the format of the intermediate-level exam compared to previous tests. Despite changes in the test content and question format, the high-scoring group was able to achieve high scores, whereas the low-scoring group, who had not fully mastered the language, were unable to score well.

Figure 2. *Progression of Japanese test scores for the high-score group and low-score group in RST*





CONCLUSION

The results of this study revealed no clear relationship between WM capacity and Japanese language learning outcomes. This finding is consistent with previous research on reading comprehension abilities in Malaysian Japanese learners. One possible reason for this may be due to the Chinese-language version of the RST created for this experiment. Although I am also considering creating an RST for Uzbek, a language with syntactic structures similar to Japanese, the current experiment offers important insights that will contribute to the improvement of experimental materials.

Moreover, when examining the learning process, high scorers on the RST tended to show stable performance. In some cases, short-term memorization strategies, such as memorizing vocabulary intensively, may have contributed to improved test scores. Low scorers, on the other hand, were more susceptible to these short-term influences, while high scorers seemed to have consistently internalized the learning content, which allowed them to perform well across different test formats.

In any case, this study has highlighted the numerous challenges that remain in applying the findings of WM research to the field of language learning, from the methods of measuring WM to the assessment of language proficiency. It is still unclear what specific language abilities the RST scores are capturing. Consequently, previous studies have shown diverse results and interpretations. The current stage appears to require a foundational research framework to explore how the theoretical concept of WM can be applied in practical settings.

However, if we can develop more accurate predictors of language proficiency and strengths or weaknesses – beyond relying solely on teachers' experience and intuition – there would be significant value in using such objective indicators to select more suitable teaching methods for each learner. Moving forward, it is hoped that we can better understand each learner's characteristics, identify where they are likely to

struggle, determine the most effective ways of explaining concepts, and pinpoint the different emphases needed between classroom learning and self-study.

To achieve this, it is crucial to continue refining the RST and conducting more precise research on the role WM plays in language acquisition. It is essential to actively integrate findings from the field of psychology into Japanese language education.

Appendix 1. *Chinese version of RST*

Practice 1

因为没赶上电车，母亲便开车送了我。
他虽然很粗鲁，但我认为他本性很好。

Practice 2

两个小孩在碧绿的湖泊旁玩耍。
要关注自己在事物发生变化时的内心波动。

2 sentence condition

Trial 1 它摇晃着，如水银一般闪闪发光向上飘去。
在公园午休的时候，我被大蜜蜂蛰了。

Trial 2 祖母沉默着，眼睛似乎在眺望着屋外。
干冰刚好可以用来给冷冻食品保温。

Trial 3 人类是经过多次的冰期和间冰期进化而来的物种。
这个颜色实际上是从樱花树皮上提取出来的颜色。

Trial 4 表面和侧面看起来像青黑色的金属。
这个和现在世界上发生的事件相同。

Trial 5 棒球最初传入日本是在明治5年之时。
这个技术水平远远超过了非专业的领域。

3 sentence condition

Trial 1 最小的弟弟转动着炯炯有神的眼睛问道。
运用各种各样的办法，努力学习西洋语言。
他为了回报大家的信任，昼夜工作。

Trial 2 农民们期待水稻和小麦可以丰收。
那个男人在会议上通过激烈的演说发出警告。
他得了感冒在出租房里休息，但是听到通知便立刻跳了起来。

Trial 3 那个孩子瞪大眼睛，做出迷惑的表情。
虽然落在地上的雨水会流向大海，但是雪会堆积起来。
即便懂得语法，也不一定精通英语。

Trial 4 父亲在给女儿的信中写道要她好好学习。
在看到科学调查的结果后，他也无可辩驳了。
在严寒中，我回到了阔别20年的故乡。

Trial 5 在用语之中，也有通过汉字翻译成日语的词汇。
妹妹回来那天，我和弟弟把家里菜园里的南瓜全部收割了。
我们平时会遇到各种各样的问题。

4 sentence condition

Trial 1 孩子打开装特产的纸袋子后，一下子惊呆了。
明信片上画了不计其数的圆圈，多到就像快要溢出来。

	教师每个人都选择了可以提出自我独特见解的话题。
	我听到那句话，身体一瞬间颤抖，被一种奇怪的感觉所侵袭。
Trial 2	我看见超市冰箱里整齐摆放着很多大虾。
	老人让我坐在他旁边，告诉我了一些古怪的话。
	他从五年级开始就一直做着天气预报的记录。
	警察在广场中间用尖锐的声音高声叫喊。
Trial 3	少年看见一只向着天空滑行起飞的海鸥。
	长期以来和家人居住的房子已经拆毁了。
	那个学生习惯睡觉前躺在床上看书。
	听者想要知道对方谈话的内容而侧耳倾听。
Trial 4	沿着小路穿过村庄，便来到了可以俯视大海的悬崖边。
	预测日语学习人数超过了一百万人。
	我用橡皮圈把相片捆起来，暂时保存了起来。
	被逼到绝境的人们将一封条约书带给了他。
Trial 5	我从一个人那里得到一个铃铛，把它挂在了椅子下面。
	父亲从东京给孩子们带来了点心。
	意识到的时候小船已经被冲上了海滩。
	去世的父亲是个勤动笔的人，经常寄信给我。

5 sentence condition

Trial 1	转校生在和她目光相对之时，便感到可以和她成为朋友。
	那个人给我看了一件用色彩漂亮的线织成的和服。
	少女在那里看到的是一件难以置信的事情。
	从附近的车站到那个城市的车站，乘坐特快需要花费3小时。
	那个飞行员以前喜欢在晚上仰望星空。
Trial 2	从那天开始，通过凹凸不平的道路便成为了他的乐趣。
	水手将猫带回了位于山丘上的自己的家里。
	孩子们因为明月皎皎而集体来到外面。
	祖父在一个月后便永远地合上了眼睛。
	这时我的脑海里突然浮现出孩童时的画面。
Trial 3	那天久违的从早上到傍晚都一直在下雨。
	吃过午饭后，我在那边悠闲地散步。
	坐在起居间里的父亲，光着脚跑了出来。
	由于雨水连绵不断，池塘的堤坝瞬间崩塌。
	据说现在世界上有2000多种语言。
Trial 4	他没有妻子，和内向的妹妹二个人生活。
	那天早晨一大早，我便站在我家门前。
	母亲第一次看到信封的时候，感到十分震惊。
	那天，包括牧羊人在内谁也没有来山中小屋。
	他缓缓地骑着白色自行车环绕了体育场一圈。
Trial 5	男人提醒孩子今天不要出海比较好。
	想要学习日本的外国人的存在是难能可贵的。
	小学生们精神百倍地度过了暑假的每一天。
	突如其来的通知让两人说不出话来，呆呆地坐着。
	从丈夫坐上轮椅起，已经过了12年。

REFERENCES

1. Ariji, K. (2016). The role of working memory for vocabulary meaning in text comprehension: The divergence point between experts and non-experts in reading comprehension. *Shinshu University Journal of Humanities and Social Sciences, 10*, 92–107.
2. Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews. Neuroscience, 4*(10), 829–839. https://doi.org/10.1038/nrn1201.
3. Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation,* 8, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1.
4. Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158.
5. Brown, G.D.A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed?. *Journal of Memory and Language, 34*(5), 594–621. https://doi.org/10.1006/jmla.1995.1027.
6. Conway, A.R.A., Kane, M.J., Bunting, M.F. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review,* 12, 769–786. https://doi.org/10.3758/BF03196772.
7. Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6.
8. Friedman, N.P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51*(1), 136–158. https://doi.org/10.1016/j.jml.2004.03.008.
9. Futakuchi, W. (2015). Working memory capacity and second language proficiency in second language acquisition: A focus on language aptitude as individual differences. *Studies in Language Sciences: Journal of the Japanese Society for Language Sciences,* 14, 142–162.
10. Futakuchi, W. (2020). Individual differences in working memory capacity and reading strategy in Japanese reading as a second language. *Bulletin of Kaichi International University,* 19, 65–74.
11. Gathercole, S.E., & Baddeley, A.D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection?. *Journal of Memory and Language. 29*(3), 336–360. https://doi.org/10.1016/0749-596X(90)90004-J.
12. Just, M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149. https://doi.org/10.1037/0033-295X.99.1.122.
13. Kashiwagi, A. & Nakayama, M. (2012). Processing relative clauses and working memory in L2 Japanese. *Second Language,* 11, 27–45.
14. Minamoto, T., Osaka, M., & Osaka, N. (2010). Individual differences in working memory capacity and distractor processing: Possible contribution of top–down inhibitory control, *Brain Research,* 1335*,* 63-73. https://doi.org/10.1016/j.brainres.2010.03.088.
15. Morishima, Y. (2013). The role of long-term working memory in second language comprehension. *Educational Studies, 55*, 71–78.
16. Mukoyama, Y. (2009). The impact of learner aptitude on language learning outcomes in second language acquisition: Focusing on linguistic analytical ability, phonological short-term memory, and working memory. *Japanese Linguistics,* 25, 67–90.
17. Osaka, M. (2002). *The brain's notepad: Working memory*. Tokyo: Shinyosha.
18. Osaka, N. (Ed.). (2013). *The brain that controls attention: Information selection and integration from the perspective of neuroattention science* (Social Brain Series 3). Tokyo: Shinyosha.
19. Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology Section A, 48*(1), 98–107. https://doi.org/10.1080/14640749508401378.

20. Sakakibara, K. (2013). The effects of focusing, proficiency, and working memory capacity on text comprehension of Japanese EFL learners: A psycholinguistic study based on eye movement data. *International Cultural Studies,* 26, 135–156.
21. Satori, M. (2012). Working memory as a factor affecting L2 listening comprehension sub-skills. *Kumamoto University Journal of Social and Cultural Studies*, 10, 119–128.
22. Sugiura, K., & Nakanishi, H. (2011). The contribution of lexical processing ability to efficient functioning of working memory in Japanese EFL learners. *JACET Chubu Branch Bulletin*, 125–140.
23. Takahashi, M., Shimizu, H., Saito, S., & Tomoyori, H. (2006). One percent ability and ninety-nine percent perspiration: A study of a Japanese memorist. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1195–1200. https://doi.org/10.1037/0278-7393.32.5.1195.
24. Tanaka, T. (2017). The impact of working memory on active annotation during reading. *Bulletin of the Graduate School of Education, Kyoto University,* 63, 581–591.
25. Tanaka, T., Saito, S., & Kikuchi, S. (2014). The role of sentence information in reading span performance: An examination of the recall reconstruction hypothesis. *Psychologia, 57*(3), 164– 176. https://doi.org/10.2117/psysoc.2014.164.
26. Tsuchida, Y., & Murohashi, H. (2017). The relationship between working memory and learning strategies. *Journal of Child Development and Clinical Research,* 9, 47–55.
27. Unsworth, N., & Engle, R.W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104–132. https://doi.org/10.1037/0033-295X.114.1.104.
28. Unsworth, N., Fukuda, K., Awh, E., & Vogel, E.K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology,* 71, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003.
29. Yoshikawa, T., & Zorayda, M. (2017). The relationship between working memory and Japanese reading comprehension ability in intermediate-level Malaysian learners of Japanese. *Bulletin of the General Education Center, Saga University,* 5, 77–87.
30. Yuzawa, M., & Yuzawa, M. (Eds.). (2014). *Working memory and education*. Kyoto: Kitaohji Shobo.