Running head: Structural relationships of literacy skills

The unique contribution of handwriting accuracy to literacy skills in Japanese adolescents

Sadao Otsuka*, Toshiya Murai

Department of Psychiatry, Graduate School of Medicine, Kyoto University,
54 Shogoin-kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan

Corresponding author
Sadao Otsuka, PhD
Department of Psychiatry, Graduate School of Medicine, Kyoto University,
54 Shogoin-kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan
Email: o2ka.psycho@gmail.com

This version of the article has been accepted for publication after peer review, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/s11145-023-10433-3. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms

Abstract

There is widespread concern about declining literacy skills in recent young Japanese. The present study investigated how higher-level reading and writing proficiencies are underpinned by basic literacy skills in Japanese adolescents. From a large database of the most popular literacy exams in Japan, we retrospectively analyzed word- and text-level data for middle and high school students who had taken the exams during the same period in the 2019 academic year using structural equation modeling. We extracted main data for 161 students as well as six independent datasets for validation. Our results validated the three-dimensional view of word-level literacy (reading accuracy, writing accuracy, and semantic comprehension) and demonstrated that writing and semantic skills underpinned text writing and reading, respectively. The semantic comprehension of words affected text writing indirectly via text reading; however, it could not replace the direct effect of word writing accuracy. These findings, which were robustly replicated with multiple independent datasets, provided new evidence of dimension-specific relationships between word- and text-level literacy skills and confirmed the unique contribution of word handwriting acquisition to text literacy proficiency. The replacement of handwriting by digital writing (e.g., typing) is a global trend. However, the dual-pathway model of literacy development identified in this study suggests there are advantages in sustaining early literacy education by handwriting for the growth of higher-level language skills in future generations.

Keywords: written language, kanji ability, multidimensionality, PISA, digitalization, replicability

Reading and writing a variety of texts are key skills for participating in a knowledge-based or information society. These important skills have been internationally examined in adolescents aged 15 years using the Programme for International Student Assessment (PISA), which is conducted by the Organization for Economic Cooperation and Development (OECD). The PISA has been implemented every 3 years from 2000. Japan has continuously participated in this assessment since its inception, and consistently shown high scores among the participating countries (Ikeda & Schwabe, 2019). However, despite maintaining its top performance in Mathematics and Science, Japan showed a significant decline in Reading, dropping from 8th among countries participating in PISA 2015 to 15th in 2018 (National Institute for Educational Policy Research, 2019a). The National Institute for Educational Policy Research (2019b) suggested that the decrease in Reading score may be attributable to a combination of factors, including unfamiliarity with reading lengthy texts on a computer screen or digital texts in various forms (e.g., websites, web posts, or emails). However, the mass media widely referred the noticeable decline in text reading skills of Japanese adolescents as the "PISA shock (Chiwaki et al., 2019)." Furthermore, Japanese students scored lower on free-description type questions in the PISA 2018 Reading scale than the OECD average (National Institute for Educational Policy Research, 2019b), which raised concerns about declines in both writing and reading skills.

In addition to the PISA, which assessed text-level literacy, an investigation using a large database of the most popular word-level literacy exam in Japan demonstrated that more basic (i.e., word-level) literacy skills in Japanese young adults had deteriorated in recent decades (Otsuka & Murai, 2020). This finding together with the recent trend in the PISA scores suggested there may be an intrinsic relationship between the declines in word- and text-level literacy in young Japanese. If there is a close relationship between these factors, a further question is "how" they are interrelated.

The above-mentioned study demonstrated a specific decline in word-level literacy, as among the three dimensions of word-level literacy (i.e., reading accuracy, writing accuracy, and semantic comprehension), only writing deteriorated in the 10-year period from 2006 to 2016 (Otsuka & Murai, 2020). The authors speculated that this dimension-specific decline in word literacy may be attributable to the reduction in the frequency of handwriting (Agency for Cultural Affairs, 2013), resulting from the rapid spread of digital writing devices such as computers and smartphones (Ministry of International Affairs and Communications, 2017). A survey by the Ministry of Internal Affairs and Communications (2020) reported that 76.7% of adolescents aged 13–19 years and 35.0% of children aged 6–12 years used smartphones. These environmental changes were presumed to have affected the handwriting habits of Japanese adolescents, which may in turn have reduced the accuracy of word writing in this population. Moreover, reduced word-level writing skills would negatively impact text-level literacy skills, which could explain the decreased Reading score in the PISA 2018. Given the progression of digitalization, it is important to examine the relationships between each dimension of word-level literacy, including handwriting accuracy, and text-level reading/writing to explore policy directions for literacy education.

To date, the relationships between word- and text-level literacy skills, especially the dimension-specificity of those skills, are not well understood. Cross-sectional studies involving English-speaking children reported that word reading ability affected sentence/passage reading comprehension at ages 7 and 9 years (Kim, 2017, 2020) and was correlated with text writing skills at age 5 years (Bourke et al., 2014). In addition, word writing among Chinese children at ages 7 and 8 years was found to longitudinally predict text writing skills at ages 8 and 10 years, respectively (Yeung et al., 2013). These pioneering studies only addressed word-level reading or writing literacy skills, and thus, could not reveal the dimension-specificity of these relationships with text-level literacy. In contrast, some previous studies investigated children's inter-relationships among four literacy skills, i.e., word- and text-level reading and writing skills (Abbott et al., 2010; Ahmed et al., 2014; Kim & Graham, 2022). Among them, a 5-year longitudinal study involving English-speaking children showed the dimension-specific relationships between word- and text-level literacy, as follows. Word reading at ages 9–12 predicts text reading in the following years, but not text writing except at age 7. Word spelling at ages 6–12 predicts text writing in the following years, but not text reading, also except at age 7 (Abbott et al., 2010). In addition, a study of Japanese university students (Otsuka & Murai, 2021) examined the relationships between word-level reading/writing as well as semantic comprehension and text writing based on the three-dimensional view of word-level literacy (Otsuka & Murai, 2020). The authors reported that only the accuracy of word handwriting affected text writing via knowledge acquisition. These dimension-specific relationships between word and text writing skills suggested that handwriting acquisition, which sustained the accumulation and transmission of human wisdom before the advent of digital writing devices, may make an irreplaceable contribution to the development of higher-order language skills. Handwriting itself appears to be less necessary in our digital society. However, to discuss the appropriate use of these devices for education with a view to the healthy language and cognitive development of future generations, it is crucial to determine whether word handwriting skills have a unique effect on text writing as well as reading proficiency, the decline of which is of public concern.

The aim of this cross-sectional study was to clarify the structural relationships among multidimensional literacy skills at both the word level (i.e., word reading accuracy, semantic comprehension, and writing accuracy) and the text level (i.e., text reading and writing) in Japanese adolescents. This retrospective study used data from a large database of the most popular literacy exams in Japan. Against the backdrop of progressive replacement of handwriting by digital writing and resulting concerns about declining literacy skills in adolescents, we tested the hypothesis that accurate word handwriting made an irreplaceable contribution to text literacy proficiency using structural equation modeling (SEM). First, we validated the three-dimensional view of word-level literacy, which was derived from dual-route models of reading/writing

(Coltheart et al., 2001; Iwata, 1984; Rapcsak et al., 2007; Sakurai, 2004; Sakurai et al., 2008). Then, we examined the goodness of fit of two models (Model 2 and 3) in comparison with a model in which word comprehension and writing skills affected both text reading and writing (Model 1). In Model 2, word writing affected text writing but not reading, and semantic comprehension of words did not directly affect text writing but indirectly affected it via text reading. In Model 3, text writing and reading were not affected by word writing but were affected by comprehension skills (see Fig. 1). First, regarding the relationships between reading and writing skills within the same levels of literacy, a unidirectional model from the former to the latter, rather than a bidirectional model, was previously proposed (Ahmed et al., 2014). Second, regarding the relationships across the word- and text-levels of literacy, the dimension-specific relationships between word reading and text reading – as well as word writing (or spelling) and text writing – were demonstrated by the above-mentioned study (Abbott et al., 2010). We constructed our models using those findings as a theoretical foundation, and then, modified them by incorporating the three-dimensional view of word-level literacy proposed in the study of Japanese literacy skills (Otsuka & Murai, 2020).

Methods

Nature of the data

We used data for two exams from a large database of the most popular literacy exams in Japan: the Japan Kanji Aptitude Test (*Nihon Kanji Noryoku Kentei*; Kanken) and the Japan Proficiency Test in Reading and Writing (*Nihon Bunsho Dokkairyoku Sakusei Noryoku Kentei*; Bunshoken). Both exams are administered by the Japan Kanji Aptitude Testing Foundation and taken voluntarily or semi-voluntarily by a large number of Japanese (e.g., 1,831,851 people took the Kanken in the 2019 academic year). The Kanken, which is a word-level literacy exam, started in 1975 and has 12 levels of difficulty from the easiest (Level 10) to the most difficult (Level 1, including Pre-2 and Pre-1). The Bunshoken, which is a text-level literacy exam, was launched in 2013 and has four levels of difficulty, from Level 4 to Level 2 (including Pre-2).

We analyzed seven independent datasets for students aged 12–18 who had taken Level 4 to Level 2 of both the Kanken and Bunshoken at their schools during the same period in the 2019 academic year, i.e., October to November in 2019 or January to February in 2020. Multiple dates are offered for Kanken or Bunshoken examinations for a particular certificate (e.g., Kanken Level 3) during the same season for the convenience of schools or examinees themselves. Different versions of the examinations are used for different test dates to maintain the fairness of the examinations. Consequently, the resulting datasets consist of nine versions for the Kanken, for each of four levels, and six versions of the Bunshoken, for each of four levels. From these entire datasets, we first extracted the data where both the Kanken and Bunshoken results were available for the same person. Then, we classified them into datasets depending on the test versions of the Kanken or Bunshoken. Of the resulting datasets (i.e., particular combinations of a Kanken version and a Bunshoken version), we employed the dataset with the largest sample size for the main analysis: that is, with 161 students (mean ± standard deviation [SD] age: 16.55 ± 0.78 years) who had taken particular versions of Level 3 (middle school graduation to high school level) of the Kanken and the Bunshoken.

In addition, to examine the replicability of the SEM results derived from the main analysis, we selected multiple validation datasets from the combinations of the Kanken and Bunshoken versions, with exclusion criteria that included: sample size smaller than 60; the level of both exams differed from that of the main data, i.e., not Level 3; and, mean age of the sample was less than 15 years, considering comparability with the main data. Consequently, five datasets remained. However, as only one of the five contained Level 3 data for both the Kanken and Bunshoken, an additional Level 3-Level 3 dataset (i.e., Validation data 2) with the next largest sample size was included as a validation dataset. Thus, a total of six independent datasets were used for validation. Validation data 1 and 2 included students who had taken Level 3 of both exams, as with the main data (validation data 1: 74 students, mean ± SD age 16.49 ± 0.79 years; validation data 2: 56 students, mean ± SD age: 16.11 ± 0.79 years). Validation data 3 and 4 included students who had taken Level 3 of the Kanken and Level 4 (middle school level) of the Bunshoken (validation data 3: 137 students, mean ± SD age 15.52 ± 1.23 years; validation data 4: 82 students, mean ± SD age 16.06 ± 0.29 years). Finally, validation data 5 and 6 included students who had taken Level Pre-2 (high school level) of the former and Level 3 of the latter (validation data 5: 115 students, mean ± SD age 16.39 ± 0.72 years; validation data 6: 94 students, mean ± SD age 16.70 ± 1.10 years). The samples included students from 5–11 schools. The demographic characteristics and the scores for the Kanken and Bunshoken for each dataset are presented in Table 1. This study employed datasets with characteristics that were consistent with the main data as validation data. Thus, we considered the population of all the datasets as the same group, i.e., Japanese adolescents, and analyzed multiple datasets separately, instead of simultaneously as multiple groups.

The methodological validity of using this data in our study was supported by several characteristics of the Kanken and Bunshoken, as follows. (1) The Kanken comprises 10 subtests and had a three-dimensional structure (word-level reading, writing, and semantic comprehension) in the previous factor analyses of the large data for all ages (Level 2 data for 33,659 people aged 9–106 years in 2006 and 16,971 people aged 9–91 years in 2016, and multiple validation datasets for Levels 2, Pre-2, 3, and 4; Otsuka & Murai, 2020). Furthermore, a previous study demonstrated that all three dimensions were closely related to higher-level language abilities (i.e., acquired knowledge measured using the Vocabulary, Arithmetic, and Information subtests of the WAIS-III; all $r \geq 0.71$).[12] The Bunshoken comprises several subtests that measure text-level reading and writing. (2) Both exams were implemented throughout Japan, and the sample for the main data included students from 10 schools, with one or more schools from six of the seven regions in Japan (i.e., Hokkaido, Tohoku, Kanto, Chubu, Kinki, Chugoku-Shikoku, and Kyushu) thereby reducing region-specific effects. (3) Both exams were multisite, had multiple levels of difficulty, and were conducted around the same period using alternative versions of exam papers, which enabled us to examine the replicability of the analysis results.

All procedures in this study were approved by the Ethics Committee of the Psychological Science Unit at ***** University (protocol number: 29-P-25; study title: Examination of the components of Japanese kanji ability and its relationship to text literacy). This study was conducted in accordance with the Code of Ethics and Conduct of the Japanese Psychological Association. The data used in this study were de-identified before being provided by the Japan Kanji Aptitude Testing Foundation.

Measures

Word-level literacy skills

The three-dimensional structure of the Kanken (word-level reading accuracy, semantic comprehension, and writing accuracy) was established previously (Otsuka & Murai, 2020). The time limit for Level 3 and Pre-2 of the exam was 1 hour and the criterion for certification was a score of 70% or higher (maximum score of 200). Pass rates were 45.0%–45.7% for Level 3 and 29.0%–36.8% for Level Pre-2 in the 2019 academic year.

1. Reading accuracy

   The reading accuracy dimension in Levels 3 and Pre-2 was assessed using the Reading subtest. This subtest requires examinees to write the correct pronunciation of a marked kanji word (i.e., convert it to kana) that appeared in 30 sentences, with consideration of the context. A kanji word can be written alternatively in kana letters, which have highly regular letter–sound correspondences. Therefore, the kanji-to-kana conversion is usually used in literacy education in Japan. Each correct item was awarded 1 point, with a maximum score of 30 for both Level 3 and Pre-2 (0–30; Otsuka & Murai, 2020, 2021).

2. Semantic comprehension

The semantic comprehension dimension in Level 3 was assessed using four subtests. In the Homophones subtest, examinees were required to differentiate between three homophones of kanji characters that were written as the same marked kana letters, each appearing in a sentence, by choosing the correct characters from five options. Five sets of three homophones (items) were prepared and each correct item was awarded 2 points, giving a maximum score of 30 (0–30). The Compounds Completion subtest required examinees to complete five sets of three two-character kanji compounds by choosing the correct character that preceded or followed the three characters in each set (from 10 options). Each correct item was awarded a score of 2, giving a maximum score of 10 (0–10). In the Compounds Structure subtest, examinees were required to sort 10 two-character kanji compounds into five categories based on their structure. The categories included cases where the two characters had similar meanings, opposite meanings, the latter was modified by the former, the latter was an object/complement of the former, and the meaning of the latter was denied by the former. Each correct item was awarded 2 points, giving a maximum score of 20 (0–20). The Radicals subtest required examinees to extract a radical from 10 kanji characters by choosing the correct response from four prepared options. Radicals are the visual components of kanji characters, most of which represent the semantic category. For example, the left part of the kanji 海 (*umi* or *kai*, sea) is regarded as the radical 氵 (*sanzui*), meaning “water” or “fluid.” General dictionaries of Japanese kanji use 214 radicals to classify kanji characters and assign one radical to each kanji. Each correct item was awarded 1 point, giving a maximum score of 10 (0–10).

In Level Pre-2, this dimension was assessed using the three subtests described above (i.e., Compounds Completion, Compounds Structure, and Radicals), along with a Compounds Meaning subtest. This subtest required examinees to choose the option that represented the meaning of five sentences from 10 four-character kanji compounds in the

Compounds Completion subtest. Each correct item was awarded 2 points, giving a maximum score of 10 (0–10). The Compounds Completion subtest for Level Pre-2 was a little more difficult than Level 3. The Level Pre-2 subtest required examinees to complete four-character kanji compounds by choosing one that preceded or followed two-character kanji compounds from kana words and converting the kana words to two kanji characters. There were 10 items and 10 kana word options, as for Level 3. Each correct item was awarded 2 points, giving a maximum score of 20 (0–20; Otsuka & Murai, 2020, 2021).

3. Writing accuracy

The writing accuracy dimension was assessed in Level 3 using five subtests. The Antonyms/Synonyms subtest required examinees to complete five antonyms and five synonyms of two-character kanji compounds by choosing one that preceded the latter or followed the former character from 10 kana letter options and rewriting kana into kanji. Each correct item was awarded 2 points, giving a maximum score of 20 (0–20). In the Kana Suffixes subtest, examinees were required to rewrite marked kana letters in five sentences into correct kanji characters accompanied by a kana suffix. Each correct item was awarded 2 points, giving a maximum score of 10 (0–10). The Four-character Idioms subtest required examinees to complete 10 four-character idioms by rewriting marked kana letters that preceded or followed a two-character kanji compound in 10 sentences into two kanji characters. Each correct item was awarded 2 points, giving a maximum score of 20 (0–20). The Error Correction subtest required examinees to identify homophonic errors in kanji characters in five sentences and rewrite the correct characters. Each correct item was awarded 2 points, giving a maximum score of 10 (0–10). The Writing subtest required examinees to rewrite marked kana letters in 20 sentences into the correct kanji word. Each correct item was awarded 2 points, giving a maximum score of 40 (0–40).

In Level Pre-2, this dimension was assessed using the four subtests described above (i.e., Antonyms/Synonyms, Kana Suffixes, Error Correction, Writing), along with a Homophones subtest. This subtest differed from the multiple-choice task of the same name in Level 3 and required examinees to write accurately. In this subtest, examinees differentially wrote two homophones of kanji words that were written as marked kana letters in five pairs of sentences. Each correct item was awarded 2 points, giving a maximum score of 20 (0–20). In Level Pre-2, the Writing subtest had 25 items (0–50), and the Antonyms/Synonyms subtest was a little more difficult compared with Level 3. This subtest required examinees to choose an antonym or synonym for a two-character kanji compound from kana words and write it correctly in kanji. There were 10 items and 10 kana word options, as in Level 3. Each correct item was awarded 2 points, giving a maximum score of 20 (0–20; Otsuka & Murai, 2020, 2021).

Text-level literacy skills

The Bunshoken was designed to assess text-level Japanese literacy skills, including text reading and writing. The time limit for the Level 3 and 4 exams was 1 hour and the criterion for certification was a score of 70% or higher (maximum score of 200). The overall pass rates for the exams taken by this sample were 65.2%–77.8% for Level 3 and 79.1%–81.1% for Level 4.

1. Text reading

The text reading dimension was assessed in Level 3 using the following two subtests. We used the sum of the scores on these subtests as the score for this dimension (0–60). The Chart Reading subtest required examinees to read a chart and the accompanying text, and then answer three questions by choosing the correct response from three to four options of a passage or sentence. Two were fill-in-the-blank questions, which required examinees to comprehend the chart and text and answer the questions with consideration of the context. The remaining question was a task whereby the examinee had to correctly answer the point or rationale of a sentence or paragraph regarding chart interpretation. Each correct item was awarded 10 points, giving a maximum score of 30. In the Text Reading subtest, examinees were required to read several paragraphs of an essay, and then respond to three questions by choosing the correct response from three to four options of a passage or sentence. For example, the questions asked about the relationship among several paragraphs, the point of the whole text or paragraph, and the meaning of a passage in the context. Each correct item was awarded 10 points, giving a maximum score of 30. In Level 4, the Chart Reading subtest had two questions and the Text Reading subtest had four questions, although the maximum score for this dimension was 60, consistent with Level 3, and the question forms were generally similar.

2. Text writing

The text writing dimension was assessed in Level 3 using two subtests. We used the sum of the scores for these subtests as the score for this dimension (0–110). The Letter Writing subtest required examinees to read a letter and then correct three errors (e.g., typos, honorific errors, or rude expressions), and rewrite the marked sentence into multiple shorter sentences that were more easily understood. Each item for which an error was accurately corrected was awarded 5 points, giving a maximum score of 15. The rewritten sentences were scored on a 25-point scale based on coherent content, consistent meaning, appropriateness of expression for a letter, and absence of typos or grammatical errors. The Opinion Writing subtest required examinees to take one of two positions on a topic, write about the facts (i.e., their experiences or knowledge of the topic) in the first paragraph, express their opinion in the second paragraph, and discuss the reasons for their opinion in the third paragraph, using 500 characters or less. The written text was scored on a 70-point scale based on (1) a three-paragraph structure with facts and reasons tied to opinions, (2) concreteness of facts in the first paragraph, (3) clarity of opinions in the second paragraph, (4) acceptability of reasons in the third paragraph, and (5) absence of errors (e.g., typos, misuse of words, and grammatical errors). In Level 4, the Opinion Writing subtest had a maximum score of 60; therefore, the sum of the maximum scores of both subtests was 100, and both subtests were easier than those in Level 3, although question forms were generally similar.

Statistical analyses

Data were analyzed in four steps. All statistical analyses were conducted using Mplus version 8.7 (Muthén & Muthén, 2017). The statistical tests were two-tailed, unless otherwise noted, and α was set at .05.

Step 1: For the measurement model, the goodness of fit for the three-dimensional model of word-level Japanese literacy, previously established (Otsuka & Murai, 2020), was assessed by confirmatory factor analysis (CFA) with maximum likelihood estimation with robust standard errors (MLR), using the analysis data. In addition to the traditional χ^2 statistic, the indices used to test the model fit were: root mean square error of approximation (RMSEA) with a 90% confidence interval (CI), comparative fit index (CFI), Tucker-Lewis index (TLI), standardized root mean square residual (SRMR), and Akaike's information criterion (AIC). RMSEA values <0.05 indicate a good fit and values <0.08 are considered acceptable. *P*-values were also calculated for the test of the close-fit hypothesis that RMSEA ≤ 0.05. This one-sided null hypothesis (i.e., *p*-close $\geq .05$) was considered to indicate a good fit (Browne & Cudeck, 1993). CFI and TLI values >0.95 and SRMR values <0.08 suggest a good fit, and lower AIC values indicate relatively better fit (Brown, 2015; Hu & Bentler, 1999; Kline, 2016). Furthermore, we examined the validity of the model by the size of the standardized factor loadings as well as testing based on the Wald statistic (Wald test) and modification indices (Lagrange multiplier test). These two univariate statistics follow the χ^2 distribution with one degree of freedom (*df*); that is, Wald test values ≤ 3.84 ($p \geq .05$) suggest an improvement in model fit when the freely estimated parameter is constrained to zero (i.e., when the path was trimmed from the model), whereas Lagrange multiplier test values >3.84 ($p < .05$) suggest an improvement in model fit when the constrained-to-zero parameter was freely estimated (i.e., when the path was added to the model; Kline, 2016). To calculate the former, the variance of each factor was fixed at 1 and the factor loading for the Reading subtest, which was the only subtest loaded by reading accuracy, was also fixed at 1; therefore, these two statistics could not be computed for this subtest. The internal consistency was assessed with the coefficient omega for the composite reliability (McDonald, 1999) of the subtests loaded by each factor after the analysis. For text-level literacy, we used the sum of the subtest scores because model identification in CFA required at least three indicators and there were fewer than three subtests measuring each skill (Brown, 2015).

Step 2: Regarding the structural model, the goodness of fit for the three models (see Fig. 1) was assessed by SEM with MLR using the analysis data. These models mainly differed in assumptions about the effects of word writing on text literacy; that is, word writing affected both text reading and writing (Model 1), affected only text writing (Model 2), or had no effect on text reading/writing (Model 3). In addition to examining the model fit indices as described above, we directly compared the fit of nested models (i.e., Model 1 vs. 2 or 3) using the Satorra-Bentler scaled chi-square difference tests (Satorra & Bentler, 2010; Muthén & Muthén, 2011). When the scaled chi-square difference was not statistically significant, the nested models (i.e., more restrictive models with more *df* than the comparison model; Models 2 and 3 in this case) were retained as constraining the parameters of those models did not worsen the model fit. However, when the difference was statistically significant, the models were not retained as constraining the parameters worsened the model fit. Furthermore, we examined the validity of the models with a good fit using the size of the standardized path coefficients, percentage of explained variance of dependent variables, and modification indices.

Step 3: As both Models 2 and 3 were retained after Step 2, we tested whether the good fit of these two models could be replicated by SEM with MLR using the validation data 2 and 3, where both literacy exams were Level 3 as in the analysis data (see Table 1). In addition to confirming the model fit indices, Satorra-Bentler scaled chi-square difference tests were used to investigate the replicability of the results in that constraining the parameters did not worsen the fit in Models 2 and 3. Next, we examined the standardized path coefficients, explained variance of dependent variables, and modification indices.

Step 4: To examine the replicability of Model 2, which was retained after the analyses in Step 3, SEM with MLR was performed using the validation data 3–6, where the word- or text-level literacy exams were Level 3, as in the analysis data (see Table 1). After confirming the model fit indices, Satorra-Bentler scaled chi-square difference tests were implemented.

Results

Measurement model of word-level literacy

We conducted CFA with MLR to examine whether the three-dimensional model of word-level Japanese literacy, previously validated (Otsuka & Murai, 2020), could be replicated using the analysis data in this study. All model fit indices demonstrated a good fit: the chi-square statistic ($\chi^2(33) = 41.93$, $p = 0.137$, scaling correction factor = 1.03) and RMSEA estimates (RMSEA [90%CI] = 0.04 [0.00, 0.08], p-close = 0.631) were not significant; the CFI (0.99) and TLI (0.99) values were >0.95, and the SRMR (0.03) was <0.08. The estimates of the standardized factor loadings and results of the Wald and Lagrange multiplier tests are shown in Table 2. All Wald statistics were significantly high and the modification indices were not significant, which suggested no improvement in the model fit by free estimation of the parameters previously constrained to zero or vice versa. In addition, the composite reliability coefficients for semantic comprehension (0.81) and writing accuracy (0.90) were adequate in this model.

Structural model for word- and text-level literacy

After determining the measurement model, we conducted SEM with MLR using the analysis data to examine the structural relationships among word- and text-level literacy skills (correlation matrix was provided in Supplement Table 1). As illustrated in Fig. 1, we evaluated

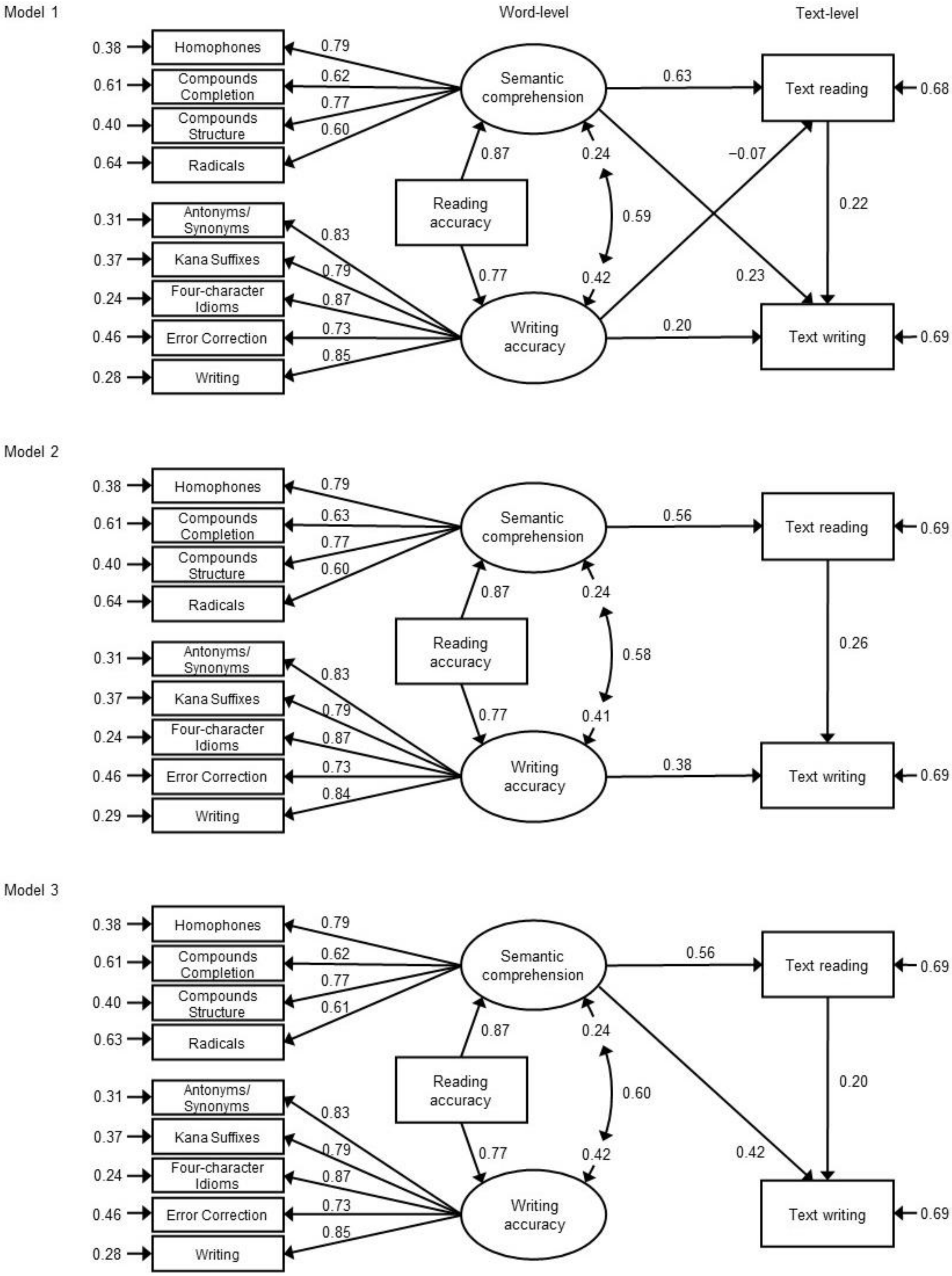

Fig. 1. Illustration of the structural equation modeling results with the analysis data. Numbers on

single-headed arrows indicate standardized path coefficients. Numbers at the bottom of single-headed arrows represent residuals. Numbers on a double-headed arrows indicate correlations.

the goodness of fit of the three models that differed in the effects of word writing on text literacy. The model fit indices are shown in Table 3, and the estimates of standardized path coefficients and the variances explained by the three models are shown in Table 4.

The model fit indices suggested all three models had a good fit with the analysis data (all p-values in $\chi^2 \geq .187$; all RMSEA ≤ 0.03; all CFI ≥ 0.99; all TLI ≥ 0.99; all SRMR ≤ 0.03). Furthermore, the Satorra-Bentler scaled chi-square difference tests demonstrated that when compared with Model 1 (comparison model), constraints of the parameters in Model 2 (i.e., trimming the paths from word writing accuracy to text reading and from word semantic comprehension to text writing: TRd(2) = 1.83, $p = .401$) and Model 3 (i.e., trimming the paths from word writing accuracy to text reading and writing: TRd(2) = 1.85, $p = .397$) did not worsen the fit and both nested models should be retained. The AIC values for the two nested models (Model 2: AIC = 10016.91; Model 3: AIC = 10016.84) were equivalent and comparatively lower than that for Model 1 (AIC = 10019.24).

In addition, all modification indices concerning the structural models were not significant in the three models using the analysis data (all $\chi^2(1) \leq 1.56$, all $p \geq .212$), which suggested no improvement in model fit by free estimation of the parameters previously constrained to zero (i.e., the addition of paths). The estimates of path coefficients showed that while the direct effects from word-level semantic ($p = .220$) and writing skills ($p = .228$) to text-level writing were not significant in Model 1, all estimated direct effects were significant in Models 2 and 3 (all $p \leq .017$). The variances of both text reading (all $R^2 \geq 0.31$) and writing (all $R^2 \geq 0.31$) explained by the three models were equivalent and sufficiently large. These results suggested Models 2 and 3 had an equally good fit for the analysis data.

Replicability of the model fit

To examine the replicability of the good fit of Models 2 and 3 to the analysis data, we used SEM with MLR with validation data 2 and 3 (see Table 1). The model fit indices are shown in Table 3, the modification indices are shown in Table 5, and the estimates of the standardized path coefficients and variances explained by the three models are shown in Table 6. Although the good fit of all three models was not replicated using validation data 1 (all *p*-values in χ^2 tests ≤.009, all RMSEA ≥0.09, all CFI ≤0.92, all TLI ≤0.89, all SRMR ≥0.08), almost all model fit indices suggested a good fit for each of the models with validation data 2 (all *p*-values in χ^2 tests ≥.063, all RMSEA ≤0.08, all CFI ≥0.96, all TLI ≥0.95, all SRMR ≤0.07). In the analyses with validation data 1 and 2, the scaled chi-square difference tests replicated the result that the goodness of fit was not worse in Model 2 (validation data 1: TRd(2) = 4.50, $p = .105$; validation data 2: TRd(2) = 0.48, $p = .786$), but was significantly worse in Model 3 (validation data 1: TRd(2) = 7.33, $p = .026$; validation data 2: TRd(2) = 8.44, $p = .015$) compared with Model 1, suggesting only Model 2 should be retained. Consistent with those results, the lower AIC values for validation data 1 and 2 in Model 2 (validation data 1: AIC = 4567.47; validation data 2: AIC = 3557.10) suggested a relatively better fit compared with Model 1 (validation data 1: AIC = 4568.82; validation data 2: AIC = 3560.55); values for Model 3 were equivalent to or higher than those for Model 1 (validation data 1: AIC = 4568.73; validation data 2: AIC = 3565.97).

In addition, all modification indices in the structural models were not significant in Model 2 or Model 1 with validation data 1 and 2 (all $\chi^2(1) \leq 2.82$, all $p \geq .093$), indicating no improvement in model fit by free estimation of the parameters previously constrained to zero. However, with these two datasets in Model 3, the modification indices suggested that the addition of the paths between word- and text-level writing skills improved the fit of the model (all $\chi^2(1) \geq 3.93$, all $p \leq .047$). Regarding the path coefficients, the significant direct effect from word semantic comprehension to text reading was replicated in Models 2 ($p = .026$) and 3 ($p = .028$) with the two sets of validation data, and that from word- to text-level writing skills was replicated in Model 2 with validation data 2 ($p < .001$) but not with validation data 1 ($p = .185$).

However, the significant direct effect from word semantic comprehension to text writing in Model 3 was not replicated with either of the validation datasets. Although the variance of text reading by Model 1 with validation data 1 was small, that of text reading/writing explained by Model 2 (text reading: $R^2 \geq 0.07$; text writing: $R^2 \geq 0.25$), which were estimated using the validation data, were equivalent to that explained by Model 1 (text reading: $R^2 \geq 0.10$; text writing: $R^2 \geq 0.26$). However, the variance in text writing explained by Model 3 ($R^2 = 0.20$) was relatively smaller than that explained by Model 1 ($R^2 = 0.34$), which was estimated using validation data 2, whereas the other variances explained by Models 1 and 3 (text reading: $R^2 \geq 0.07$; text writing: $R^2 = 0.25$) with the two datasets were almost equivalent. The SEM using validation data 1 and 2 generally replicated the good fit for Model 2 shown with the analysis data, but not for Model 3.

Finally, to further confirm the replicability of the good fit of the models 2, we conducted SEM with MLR using the validation data 3–6 and compared the fit with Model 1 (see Table 1). The model fit indices are shown in Table 7. Almost all model fit indices suggested Model 2 had a good fit with validation data 3–5 (all p-values in χ^2 tests ≥.025, all RMSEA ≤0.06, all CFI ≥0.97, all TLI ≥0.96, all SRMR ≤0.05). In addition, the scaled chi-square difference tests with those datasets replicated the retained good fit of Model 2 in comparison with Model 1 (all TRd(2) ≤ 3.84, all $p \geq .147$). However, only the analysis with validation data 6 suggested that the fit of Model 2 was worse (TRd(2) = 22.19, $p < .001$), although the RMSEA (0.07, p-close = .153) and SRMR (0.07) showed an acceptable fit. Consistent with those results, the AIC values for Model 2 (AIC = 5723.23) with validation data 6 suggested a worse fit compared with Model 1 (AIC = 5714.59), but those for validation data 3–5 (data 3: AIC = 8438.19; data 4: AIC = 5014.30; data 5: AIC = 7058.13) indicated a better or equivalent fit compared with the comparison model (data 3: AIC = 8441.21; data 4: AIC = 5014.10; data 5: AIC = 7060.70). Overall, the good fit for Model 2 shown using the analysis data was repeatedly replicated by the independent datasets for validation.

Discussion

There is widespread concern about decline of reading and writing skills in the contemporary population in Japan. This cross-sectional study used word- and text-level data for the most popular Japanese literacy exams to test whether word handwriting skills made an irreplaceable contribution to text literacy proficiency in Japanese adolescents. Our SEM results validated the three-dimensionality of word-level Japanese literacy skills and demonstrated that word handwriting skills affected text writing. Word comprehension skills did not replace this direct effect but had an indirect effect via text reading (see Model 2, Fig. 1). These robust findings, which were repeatedly replicated with multiple independent datasets for validation, supported an inter-relationship between the reported declines in word handwriting and text literacy skills in recent young Japanese. Thus, our findings imply that the declines may be partly explained by the less frequent use of handwriting, which has resulted from the replacement of handwriting with digital writing. To our knowledge, these findings represent the first reported evidence of dimension-specific relationships between multiple dimensions of word- and text-level literacy skills, with both direct and indirect effects.

Three-dimensionality of word-level Japanese literacy

The good fit of the three-dimensional model of Japanese kanji abilities, which was previously reported based on CFA using Kanken data for all ages (Otsuka & Murai, 2020), was replicated in this study focused on adolescents. This result suggested the robustness of this model, and supported the methodological validity of using the Kanken as a measure of word-level Japanese literacy. This is because Japanese kanji abilities, which comprise reading/writing and semantic dimensions, are considered to rely on phonological and orthographic lexicons and the lexico-semantic system in the lexical route for processing written language at the word level, as proposed by the dual-route hypotheses (Coltheart et al., 2001; Iwata, 1984; Rapcsak et al., 2007; Sakurai, 2004; Sakurai et al., 2008). On the other hand, reading/writing at the sub-word level by the non-lexical route simply depends on grapheme-to-phoneme or phoneme-to-grapheme conversion based on letter-sound correspondence.

The three dimensions of word-level Japanese literacy skills reflect phonological, orthographic, and semantic processing, which are required to master Japanese kanji characters that have multiple pronunciations, visual complexity, and different meanings. The multidimensional nature of word literacy, based on dimension-specific and common cognitive underpinnings (Otsuka & Murai, 2021), implies educational or therapeutic strategies may be necessary depending on the dimensions and causes of difficulties in literacy acquisition.

Effect of word handwriting on text literacy

Although the SEM with the analysis data supported both Models 2 and 3, the analyses with the independent datasets for validation only replicated the good fit of Model 2 (see Fig. 1). Our results suggested that word-level comprehension and writing skills acquired during school years underpinned the development of text-level reading and writing proficiency, respectively and dimension-specifically, and the former underlies the latter. This was consistent with accumulated knowledge (Abbott et al., 2010; Ahmed et al., 2014; Bourke et al., 2014; Kim, 2017, 2020; Kim & Graham, 2022; Otsuka & Murai, 2021; Yeung et al., 2013). Furthermore, the significantly large value of the modification index in Model 3 shown with validation data 2 clearly indicated that the contribution of word handwriting skills to text writing proficiency cannot be replaced by the semantic comprehension of words. However, in line with a previous report (Abbott et al., 2010), the direct path from word writing to text reading, which was included in Model 1, was not supported. Further investigation is required to determine the causes of the decreased Reading score in the PISA 2018 (National Institute for Educational Policy Research, 2019a, 2019b) and the replicability of this decline, which is a matter of public concern.

Our final model (i.e., Model 2) showed that individual differences in text writing proficiency were not only explained by text reading skills, which are based on the accumulation of semantic knowledge of words, but also by the acquisition of accurate word handwriting. In the handwriting of texts, the simple ability to write the correct form of Japanese kanji words with visual complexity or English words with irregular spelling is undoubtedly important. However, this finding may imply more than that. Supplementary SEM using the analysis data showed that the effect from word to text writing was significant even when using the scores of either the Letter Writing ($\beta = 0.46$, $p < .001$) or Opinion Writing subtests ($\beta = 0.27$, $p = .001$) as the outcome measure in Model 2. The Letter Writing subtest did not require examinees to write kanji words that were not in the text provided for this task: therefore, the score would not simply reflect whether the words could be correctly handwritten. In addition, a previous investigation demonstrated that word writing skills measured using the Kanken affected text writing proficiency, as measured from essays typed by university students, via knowledge acquisition (Otsuka & Murai, 2021). These findings suggested that the effect of word handwriting acquisition was not confined to the process of handwriting *per se* but could be generalized to higher-level writing skills. The orthographic ability to reproduce the accurate words in a semantic context could be a fundamental component of the writing skills necessary to produce appropriate and creative text. The significant correlation between the residuals of word comprehension and writing skills appeared to support this notion. The dual-pathway model from word reading to text writing skills, as identified in this study (see Model 2, Fig. 1), may also be applicable to people using other orthographies. This model of literacy development was based on the three-dimensional view of word-level literacy (Otsuka & Murai, 2020) and the finding of dimension-specificity of the relationships between word- and text-level literacy skills (Abbott et al., 2010). The former was derived from the dual-route hypotheses of reading/writing in Japanese (Iwata, 1984; Sakurai, 2004; Sakurai et al., 2008) and alphabetic orthographies (Coltheart et al., 2001; Rapcsak et al., 2007). The latter was provided by the study of English-speaking children (Abbott et al., 2010). The present study replicated this finding for non-English speaking adolescents, and adds new evidence of indirect effects: i.e., word reading and comprehension skills do not directly affect text writing but affect it indirectly via text reading. These indirect effects and the irreplaceable direct effect of word writing on text writing proficiency in non-Japanese people should be examined in further studies.

The replacement of handwriting by digital writing (e.g., typing) is a global trend, and early signs of this can be seen in basic literacy education. For example, some schools in Scandinavia, where digitalization is well underway, have begun to teach digital writing using computers or tablets ahead of handwriting (Gamlem et al., 2020). In Japan, although basic literacy education is currently provided using a paper and pencil format, the Ministry of Education has launched the GIGA school program to ensure “one computer per student,” which was accelerated by school closures because of the COVID-19 pandemic (Ministry of Education, Culture, Sports, Science, and Technology, 2020). Considering the possible impact on basic handwriting acquisition and the development of higher-order literacy skills, the controversial issue of whether these technologies should be applied in basic literacy education should be carefully discussed (Mangen & Balsvik, 2016; Wollscheid et al., 2016). On one hand, the ease of typing or other supportive functions of digital tools may benefit literacy learning, particularly for children with undeveloped motor skills (Genlott & Grönlund, 2013) or reading/writing difficulties (Morphy & Graham, 2012). A recent report from Norway describes first grade children who received parallel literacy instruction using both paper-and-pencil and digital devices from the time they entered school (Spilling et al., 2022). That study showed that the quality of narratives written by the children three months after entering school did not differ depending on the modality used. On the other hand, the coupling of motor action and perception in the process of handwriting may facilitate literacy acquisition, based on evidence from experimental (Longcamp et al., 2008),[32] neuroimaging (Longcamp et al., 2008), and intervention studies (Kiefer et al., 2015). The dimension-specific association of basic handwriting acquisition with written language proficiency shown in this study represents evidence supporting the latter view. This suggests there may be advantages of sustaining early literacy education by handwriting for the development of higher-level language skills.

Limitations

First, the data used in our retrospective analyses were not gathered for research purposes, and the sample was not randomly extracted from the general population. Differences in motivation for taking the exam might have affected scores. Second, this cross-sectional study cannot rigorously examine the longitudinal causal relationship between basic literacy acquisition during elementary school years and higher literacy proficiency in high school students. Third, the good fit of Model 2 was not exactly replicated by analyses using validation data 1 and 6. Regarding these datasets, the mean total scores for the Kanken and Bunshoken were relatively high and the variability was small. This may be why the correlations between their total scores (validation data 1: $r = 0.35$; validation data 6: $r = 0.30$) were relatively weak compared with those in the analysis data ($r = 0.57$) or validation data 2 ($r = 0.56$), and the explained variances of text reading were small. When examining differences in the relationships among literacy skills by literacy achievement or age group, a multi-group analysis should be performed using independent datasets with comparable variability in scores and distinct differences in either area (e.g., level of the Kanken and the Bunshoken, or mean age). Finally, the models examined in this study did not incorporate writing skills with digital devices. In a further study, it will be necessary expected to clarify the relationships among word- and text-level literacy skills, including both handwriting and digital writing at the word-and text-levels. To determine the impact of learning to write with paper-and-pencil or with digital devices on literacy proficiency, intervention studies with rigorous methodologies are needed.

Conclusion

This study clarified the structural relationships among multidimensional word- and text-level literacy skills, and provided new evidence of the unique contribution of handwriting accuracy to literacy proficiency in Japanese adolescents. Our results revealed that word-level comprehension skills could not replace the direct effect of word-level handwriting on text-level writing skills. This implies the replacement of handwriting by digital writing in our society may partly explain the reported decline in literacy skills in the contemporary population, which is a matter of public concern in Japan. Our findings warrant further research on the effect of handwriting practices on higher-level language and cognitive skills in children and adults acquiring Japanese or other orthographies, with or without linguistic difficulties.

Acknowledgments: We thank Audrey Holmes, MA, from Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

Author contributions: Both authors contributed to the study concept and design. S.O. managed the data, performed statistical analysis and interpretation, created the tables and figure, and drafted the manuscript. T.M. was responsible for the project and provided critical revisions. Both authors approved the final version of the manuscript for submission.

Data availability: The data analyzed in this study are available from the corresponding author upon reasonable request.

Statements and Declarations

Competing interests: This work was supported by the Japan Kanji Aptitude Testing Foundation (*Nihon Kanji Noryoku Kentei Kyokai*). T.M. received research funding from this foundation, and S.O. was employed by Kyoto University with this funding.

References

Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology*, 102, 281–298. https://doi.org/10.1037/a0019318

Agency for Cultural Affairs. (2013). *Public opinion poll on Japanese language in 2012 (Heisei-24-nendo-kokugo-ni-kansuru-yoron-chosa; in Japanese)*. Retrieved from http://www.bunka.go.jp/tokei_hakusho_shuppan/tokeichosa/kokugo_yoronchosa/pdf/h24_chosa_kekka.pdf

Ahmed, Y., Wagner, R. K., & Lopez, D. (2014). Developmental relations between reading and writing at the word, sentence, and text levels: a latent change score analysis. *Journal of Educational Psychology*, 106, 419–434. https://doi.org/10.1037/a0035692

Bourke, L., Davies, S. J., Sumner, E., & Green, C. (2014). Individual differences in the development of early writing skills: testing the unique contribution of visuospatial working memory. *Reading and Writing*, 27, 315–335. https://doi.org/10.1007/s11145-013-9446-3

Brown, T. A. (2015). *Confirmatory factor analysis for applied research (second edition)*. New York, NY: Guilford.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Chiwaki, K., Narita, Y., Mito, K., & Izawa, T. (2019). International student assessment: reading comprehension sharply declines, “PISA shock” again (Kokusai-gakuryoku-tesuto-dokkairyoku-kyuraku-“pisa-shokku”-futatabi; in Japanese). *The Mainichi Newspapers*, morning newspaper, 2019-12-4. The Mainichi Newspapers Digital, https://mainichi.jp/articles/20191204/ddm/003/100/107000

Coltheart, M., Rastle, K., Perry, C., Langton, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204–256. https://doi.org/10.1037//0033-295X.108.1.204

Gamlem, S. M., Rogne, W. M., Rønneberg, V., & Uppstad, P. H. (2020). Study protocol: DigiHand – the emergence of handwriting skills in digital classrooms. *Nordic Journal of Literacy Research*, 6, 25–41. http://dx.doi.org/10.23865/njlr.v6.2115

Genlott, A. A., & Grönlund, Å. (2013). Improving literacy skills through learning reading by writing: the iWTR method presented and tested. *Computer & Education*, 67, 98–104. https://doi.org/10.1016/j.compedu.2013.03.007

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55. https://doi.org/10.1080/10705519909540118

Ikeda, M., & Schwabe, M. (2019). *Japan – Country note – PISA 2018 results*. Retrieved from https://www.oecd.org/pisa/publications/PISA2018_CN_JPN.pdf

Iwata, M. (1984). Kanji versus Kana: neuropsychological correlates of the Japanese writing system. *Trends in Neurosciences*, 7, 290–293. https://doi.org/10.1016/S0166-2236(84)80198-8

Kiefer, M., Schuler, S., Mayer, C., Trumpp, N. M., Hille, K., & Sachse, S. (2015). Handwriting or typewriting? The influence of pen- or keyboard-based writing training on reading and writing performance in preschool children. *Advances in Cognitive Psychology*, 11, 136–146. https://doi.org/10.5709/acp-0178-7

Kim, Y.-S. G. (2017). Why the simple view of reading is not simplistic: unpacking component skills of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading*, 21, 310–333. https://doi.org/10.1080/10888438.2017.1291643

Kim, Y.-S. G. (2020). Hierarchical and dynamic relations of language and cognitive skills to reading comprehension: testing the direct and indirect effects model of reading (DIER). *Journal of Educational Psychology*, 112, 667–684. https://doi.org/10.1037/edu0000407

Kim, Y.-S. G., & Graham, S. (2022). Expanding the direct and indirect effects model of writing

(DIEW): reading-writing relations, and dynamic relations as a function of measurement/dimensions of written composition. *Journal of Educational Psychology*, 114, 215–238. https://doi.org/10.1037/edu0000564

Kline, R. B. (2016). *Principles and practice of structural equation modeling (fourth edition)*. New York, NY: Guilford.

Longcamp, M., Zerbato-Poudou, M.-T., & Veley, J.-L. (2005). The influence of writing practice on letter recognition in preschool children: a comparison between handwriting and typing. *Acta Psychologica*, 119, 67–79. https://doi.org/10.1016/j.actpsy.2004.10.019

Longcamp, M., Boucard, C., Gilhodes, J.-C., Anton, J.-L., Roth, M, Nazarian, B. *et al*. (2008). Learning through hand- or typewriting influences visual recognition of new graphic shapes: behavioral and functional imaging evidence. *Journal of Cognitive Neuroscience*, 20, 802–815. https://doi.org/10.1162/jocn.2008.20504

McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mangen, A, & Balsvik, L. (2016). Pen or keyboard in beginning writing instruction? Some perspectives from embodied cognition. *Trends in Neuroscience and Education*, 5, 99–106. https://doi.org/10.1016/j.tine.2016.06.003

Ministry of Internal Affairs and Communications. (2017). *Communications usage trend survey in 2016* (*Heisei-26-nen-tsushin-riyo-dokotyosa-hokokusyo-setai-hen*; in Japanese). Retrieved from http://www.soumu.go.jp/johotsusintokei/statistics/pdf/HR201600_001.pdf

Ministry of Internal Affairs and Communications. (2020). *The results of communications usage trend survey in 2019* (*Reiwa-gan-nen-tsushin-riyo-dokotyosa-no-kekka*; in Japanese). Retrieved from https://www.soumu.go.jp/johotsusintokei/statistics/data/210618_1.pdf

Ministry of Education, Culture, Sports, Science, and Technology. (2020). *Education in Japan beyond the crisis of COVID-19: leave no one behind*. Retrieved from https://www.mext.go.jp/en/content/20200904_mxt_kouhou01-000008961_1.pdf

Morphy, P., & Graham, S. (2012). Word processing programs and weaker writers/readers: a meta-analysis of research findings. *Reading and Writing*, 25, 641–678. https://doi.org/10.1007/s11145-010-9292-5

Muthén & Muthén. (2011). *Chi-square difference testing using the Satorra-Bentler scaled chi-square*. Retrieved from https://www.statmodel.com/chidiff.shtml

Muthén, L., & Muthén, B. (2017). *Mplus user's guide, 8th edition*. Los Angeles, CA: Muthén & Muthén. https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

National Institute for Educational Policy Research. (2019a). *Summary of PISA 2018 results* (*OECD-seito-no-gakushu-totatsudo-chosa-PISA-2018-nen-chosa-kokusai-kekka-no-yoyaku*; in Japanese). Retrieved from https://www.nier.go.jp/kokusai/pisa/pdf/2018/03_result.pdf

National Institute for Educational Policy Research. (2019b). *Key Features of OECD Programme for International Student Assessment 2018 (PISA 2018)*. Retrieved from https://www.nier.go.jp/kokusai/pisa/pdf/2018/01_point-eng.pdf

Otsuka, S., & Murai, T. (2020). The multidimensionality of Japanese kanji abilities. *Scientific Reports*, 10, 3039. https://doi.org/10.1038/s41598-020-59852-0

Otsuka, S., & Murai, T. (2021). Cognitive underpinnings of multidimensional Japanese literacy and its impact on higher-level language skills. *Scientific Reports*, 11, 2190. https://doi.org/10.1038/s41598-021-81909-x

Rapcsak, S. Z., Henry, M. L., Teague, S. L., Carnahan, S. D., & Beeson, P. M. (2007). Do dual-route models accurately predict reading and spelling performance in individuals with acquired alexia and agraphia? *Neuropsychologia*, 45, 2519–2524. https://doi.org/10.1016/j.neuropsychologia.2007.03.019

Sakurai, Y. (2004). Varieties of alexia from fusiform, posterior inferior temporal and posterior occipital gyrus lesions. *Behavioural Neurology*, 15, 35–50. https://doi.org/10.1155/2004/305194

Sakurai, Y., Mimura, I., & Mannen, (2008). T. Agraphia for kanji resulting from a left posterior middle temporal gyrus lesion. *Behavioural Neurology*, 19, 93–106. https://doi.org/10.1155/2008/393912

Satorra, A. & Bentler, P.M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, 75, 243–248. https://doi.org/10.1007/s11336-009-9135-y

Spilling, E. F., Rønneberg, V., Rogne, W. M., Roeser, J., & Torrance, M. (2022). Handwriting versus keyboarding: does writing modality affect quality of narratives written by beginning writers?. *Reading and Writing*, 35, 129–153. https://doi.org/10.1007/s11145-021-10169-y

Wollscheid, S., Sjaastad, J., & Tømte, C. (2016). The impact of digital devices vs. Pen(cil) and paper on primary school students' writing skills: a research review. *Computer & Education*, 95, 19–35. https://doi.org/10.1016/j.compedu.2015.12.001

Yeung, P., Ho, C. S., Chan, D. W., Chung, K. K. (2013). Modeling the relationships between cognitive-linguistic skills and writing in Chinese among elementary grades students. *Reading and Writing*, 26, 1195–1221. https://doi.org/10.1007/s11145-012-9411-6

Table 1. Demographic characteristics and scores in the sample of datasets

	Analysis	Validation data					
	Data	1	2	3	4	5	6
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Kanken level	3	3	3	3	3	Pre-2	Pre-2
Bunshoken level	3	3	3	4	4	3	3
Number of schools (n)	10	8	8	10	5	11	9
Sample size (n)	161	74	56	137	82	115	94
Sex (% male)	68.32	48.65	57.10	54.75	34.15	61.74	44.68
Mean age (years)	16.55 (0.78)	16.49 (0.79)	16.11 (0.79)	15.52 (1.23)	16.06 (0.29)	16.39 (0.72)	16.70 (1.10)
Word-level literacy: total score	105.37 (36.02)	123.09 (26.31)	105.77 (37.25)	112.91 (30.51)	104.62 (36.14)	116.54 (27.39)	105.85 (26.83)
Reading accuracy	21.39 (4.97)	23.46 (3.70)	22.48 (4.92)	21.98 (3.94)	20.26 (5.71)	23.63 (4.30)	21.87 (4.01)
Semantic comprehension							
Homophones (multiple-choice)	21.42 (6.07)	24.54 (4.18)	23.39 (5.41)	23.26 (4.92)	22.22 (4.70)		
Compounds Completion	6.31 (2.54)	8.08 (1.66)	6.50 (2.76)	6.29 (2.37)	5.78 (2.93)	7.57 (3.95)	3.70 (3.91)
Compounds Structure	10.58 (5.02)	12.81 (4.05)	12.14 (5.44)	12.04 (4.54)	10.27 (4.38)	12.49 (4.17)	12.21 (3.71)
Radicals	7.09 (2.01)	7.23 (1.37)	7.43 (2.18)	7.64 (1.75)	6.83 (1.30)	6.51 (1.21)	5.79 (1.12)
Compounds Meaning						6.54 (2.63)	5.85 (2.69)
Writing accuracy							
Antonyms/Synonyms	6.87 (5.22)	9.78 (4.61)	5.57 (5.35)	7.58 (4.74)	7.27 (5.64)	9.43 (4.54)	7.85 (4.58)
Kana Suffixes	3.73 (2.71)	3.76 (2.65)	3.79 (2.77)	4.04 (2.66)	3.95 (2.67)	6.05 (2.41)	5.28 (2.56)
Four-character Idioms	7.83 (4.79)	9.22 (3.94)	7.32 (5.44)	8.32 (4.70)	7.46 (5.34)		
Error Correction	3.54 (2.89)	4.22 (3.01)	2.68 (2.81)	3.64 (2.81)	3.85 (2.98)	5.18 (2.85)	4.94 (2.57)
Writing	16.61 (9.08)	20.00 (9.18)	14.46 (10.54)	18.12 (8.38)	16.73 (9.32)	26.21 (9.04)	27.55 (8.61)
Homophones (writing)						12.94 (4.22)	10.81 (4.17)
Text-level literacy: total score	97.01 (41.68)	117.45 (34.40)	101.66 (41.33)	114.99 (29.24)	113.52 (36.94)	125.52 (29.78)	122.03 (28.24)
Text reading	36.89 (14.33)	41.22 (13.65)	45.36 (14.14)	45.47 (11.27)	45.85 (14.98)	45.22 (10.07)	39.47 (13.79)
Text writing	60.11 (33.38)	76.23 (25.62)	56.30 (34.07)	69.51 (23.65)	67.67 (26.37)	80.30 (24.98)	79.83 (20.85)

Note. Kanken = word-level Japanese literacy exam, Bunshoken = text-level Japanese literacy exam.

Table 2. Confirmatory factor analysis for the three-dimensional model of word-level literacy with the analysis data

Factor	Subtest	Loading		Wald statistic		Modification indices χ^2		
		Est.	S.E.	χ^2		Reading	Semantic	Writing
Reading accuracy	Reading	1.00	0.07					
Semantic comprehension	Homophones	0.79	0.04	93.72	***	3.59		2.94
	Compounds Completion	0.63	0.05	73.62	***	0.69		0.17
	Compounds Structure	0.77	0.03	206.50	***	2.49		1.69
	Radicals	0.60	0.07	35.24	***	1.74		1.10
Writing accuracy	Antonyms/Synonyms	0.83	0.03	295.84	***	0.73	0.15	
	Kana Suffixes	0.79	0.04	227.65	***	0.64	1.23	
	Four-character Idioms	0.87	0.02	299.19	***	0.19	0.13	
	Error Correction	0.73	0.04	160.53	***	0.94	1.52	
	Writing	0.85	0.03	233.54	***	0.02	3.63	

* $p < .05$, ** $p < .01$, *** $p < .001$

Note. $n = 161$. Est. = estimates of standardized factor loadings, S.E. = standard error, χ^2 = chi-square statistic.

Table 3. Model fit indices obtained from structural equation modeling with maximum likelihood estimation with robust standard errors using analysis data and validation data 1 and 2

Models	χ^2	*df*	*p*	SCF	TRd	RMSEA (90%CI)	*p*	CFI	TLI	SRMR	AIC
Analysis data (*N* = 161)											
Model 1: TR/TW on WA	57.62	49	.187	1.00		0.03 (0.00, 0.06)	.794	0.99	0.99	0.03	10019.24
Model 2: TW on WA	59.49	51	.194	0.99	1.83	0.03 (0.00, 0.06)	.810	0.99	0.99	0.03	10016.91
Model 3: No effect of WA	59.53	51	.193	0.99	1.85	0.03 (0.00, 0.06)	.809	0.99	0.99	0.03	10016.84
Validation data 1 (*N* = 74)											
Model 1: TR/TW on WA	75.20	49	.009	0.88		0.09 (0.04, 0.12)	.078	0.92	0.89	0.08	4568.82
Model 2: TW on WA	79.25	51	.007	0.87	4.50	0.09 (0.05, 0.12)	.065	0.91	0.89	0.08	4567.47
Model 3: No effect of WA	80.90	51	.005	0.87	7.33 *	0.09 (0.05, 0.12)	.051	0.91	0.88	0.08	4568.73
Validation data 2 (*N* = 56)											
Model 1: TR/TW on WA	57.56	49	.188	0.91		0.06 (0.00, 0.11)	.414	0.98	0.97	0.06	3560.55
Model 2: TW on WA	57.58	51	.245	0.92	0.48	0.05 (0.00, 0.10)	.494	0.98	0.98	0.06	3557.10
Model 3: No effect of WA	62.28	51	.063	0.92	8.44 *	0.08 (0.00, 0.12)	.205	0.96	0.95	0.07	3565.97

* $p < .05$, ** $p < .01$, *** $p < .001$

Note. χ^2 = chi-square statistic, *df* = degrees of freedom, SCF = scaling correction factor for maximum likelihood estimation with robust standard errors, TRd = Satorra-Bentler scaled chi-square difference test statistic, RMSEA = root mean square error of approximation, CI = confidence interval, CFI = comparative fit index, TLI = Tucker-Lewis index, SRMR = standardized root mean square residual, AIC = Akaike's information criterion, TR = text-level reading, TW = text-level writing, WA = word-level writing accuracy.

Table 4. Total, direct, and indirect effects and explained variances for the three models estimated using the analysis data

	Model 1: TR/TW on WA			Model 2: TW on WA			Model 3: No effect on WA		
	Est.	S.E.		Est.	S.E.		Est.	S.E.	
Effects from RA to SC	0.87	0.03	***	0.87	0.03	***	0.87	0.03	***
Effects from RA to WA	0.77	0.04	***	0.77	0.04	***	0.77	0.04	***
Effects from RA to TR	0.49	0.06	***	0.49	0.06	***	0.49	0.06	***
Total indirect effects	0.49	0.06	***	0.49	0.06	***	0.49	0.06	***
Via SC	0.55	0.14	***	0.49	0.06	***	0.49	0.06	***
Via WA	−0.05	0.12							
Effects from RA to TW	0.46	0.06	***	0.42	0.05	***	0.47	0.06	***
Total indirect effects	0.46	0.06	***	0.42	0.05	***	0.47	0.06	***
Via SC	0.20	0.16					0.37	0.08	***
Via WA	0.15	0.13		0.29	0.06	***			
Via SC and TR	0.12	0.06	*	0.13	0.04	**	0.10	0.04	*
Via WA and TR	−0.01	0.03							
Effects from SC to TR	0.63	0.15	***	0.56	0.06	***	0.56	0.06	***
Effects from SC to TW	0.36	0.16	*	0.14	0.05	**	0.53	0.06	***
Direct effect	0.23	0.19					0.42	0.08	***
Indirect effect via TR	0.14	0.07	*	0.14	0.05	**	0.11	0.05	*
Effects from WA to TR	−0.07	0.16							
Effects from WA to TW	0.18	0.17		0.38	0.08	***			
Direct effect	0.20	0.17		0.38	0.08	***			
Indirect effect via TR	−0.02	0.04							
Effects from TR to TW	0.22	0.09	*	0.26	0.08	**	0.20	0.09	*
R^2 of SC	0.76	0.06	***	0.76	0.06	***	0.76	0.06	***
R^2 of WA	0.59	0.06	***	0.59	0.06	***	0.59	0.06	***
R^2 of TR	0.32	0.07	***	0.31	0.07	***	0.31	0.07	***
R^2 of TW	0.31	0.06	***	0.31	0.07	***	0.31	0.07	***

* $p < .05$, ** $p < .01$, *** $p < .001$

Note. $n = 161$. Est. = estimate of standardized path coefficient, S.E. = standard error, RA = word-level reading accuracy, SC = word-level semantic comprehension, WA = word-level writing accuracy, TR = text-level reading, TW = text-level writing, R^2 = variance explained by the model.

Table 5. Modification indices from structural equation modeling for the three models with validation data 1 and 2

	Validation data 1			Validation data 2		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
RA → TR (TR on RA)	0.15	0.59	0.54	0.19	0.19	0.26
RA → TW (TW on RA)	0.97	0.50	0.54	0.01	0.21	0.00
SC → TW (TW on SC)		0.01			0.50	
WA → TR (TR on WA)		2.82	2.76		0.07	0.19
WA → TW (TW on WA)			1.41			9.07 **
TR → SC (SC on TR)	0.15	1.04	1.01	0.19	0.26	0.48
TR → WA (WA on TR)	0.15	2.18	2.15	0.19	0.04	0.15
TW → SC (SC on TW)	0.53	0.15	0.00	0.00	0.15	2.39
TW → WA (WA on TW)	0.53	1.75	3.93 *	0.00	0.33	8.24 **
TW → TR (TR on TW)		1.04			0.08	

* $p < .05$, ** $p < .01$, *** $p < .001$

Note. RA = word-level reading accuracy, TR = text-level reading, TW = text-level writing, SC = word-level semantic comprehension, WA = word-level writing accuracy.

Table 6. Estimates of total, direct, and indirect effects from word- to text-level literacy calculated using validation data 1 and 2

	Validation data 1			Validation data 2		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Effects from RA to SC	0.81***	0.84***	0.84***	0.82***	0.82***	0.82***
Effects from RA to WA	0.66***	0.66***	0.66***	0.62***	0.62***	0.62***
Effects from RA to TR	0.24*	0.22*	0.23*	0.41**	0.41**	0.41***
Total indirect effects	0.24*	0.22*	0.23*	0.41**	0.41**	0.41***
Via SC	0.06	0.22*	0.23*	0.44*	0.41**	0.41***
Via WA	0.18			−0.03		
Effects from RA to TW	0.21	0.21*	0.19	0.34**	0.37***	0.34**
Total indirect effects	0.21	0.21*	0.19	0.34**	0.37***	0.34**
Via SC	−0.01		0.09	−0.11		0.26*
Via WA	0.11	0.11		0.34**	0.29**	
Via SC and TR	0.03	0.10	0.10*	0.11	0.09	0.09
Via WA and TR	0.08			−0.01		
Effects from SC to TR	0.07	0.27*	0.27*	0.53**	0.50***	0.50***
Effects from SC to TW	0.02	0.12*	0.23	0.00	0.11	0.41**
Direct effect	−0.01		0.10	0.13		0.31
Indirect effect via TR	0.03	0.12*	0.12*	−0.13	0.11	0.10
Effects from WA to TR	0.27			−0.05		
Effects from WA to TW	0.29	0.16		0.54**	0.47***	
Direct effect	0.17	0.16		0.55**	0.47***	
Indirect effect via TR	0.12			−0.01		
Effects from TR to TW	0.44***	0.45***	0.46***	0.25	0.21	0.21
R^2 of SC	0.66***	0.70***	0.70***	0.67***	0.67***	0.68***
R^2 of WA	0.44***	0.44***	0.44***	0.38**	0.38**	0.38**
R^2 of TR	0.10	0.07	0.07	0.25	0.25	0.25*
R^2 of TW	0.26**	0.25**	0.25**	0.34**	0.33**	0.20*

* $p < .05$, ** $p < .01$, *** $p < .001$

Note. $n = 161$. Effect = estimate of standardized path coefficient, RA = word-level reading accuracy, SC = word-level semantic comprehension, WA = word-level writing accuracy, TR = text-level reading, TW = text-level writing, R^2 = variance explained by the model.

Table 7. Model fit indices obtained from the structural equation modeling with maximum likelihood estimation with robust standard errors using and validation data 3–6

Models	χ^2	*df*	*p*	SCF	TRd	RMSEA (90%CI)	*p*	CFI	TLI	SRMR	AIC
Validation data 3 (*N* = 137)											
Model 1: TR/TW on WA	71.53	49	.020	1.01		0.06 (0.02, 0.09)	.310	0.97	0.95	0.05	8441.21
Model 2: TW on WA	72.68	51	.025	1.01	1.03	0.06 (0.02, 0.08)	.354	0.97	0.96	0.05	8438.19
Validation data 4 (*N* = 82)											
Model 1: TR/TW on WA	62.10	49	.099	1.01		0.06 (0.00, 0.10)	.375	0.98	0.97	0.05	5014.10
Model 2: TW on WA	66.04	51	.077	1.01	3.84	0.06 (0.00, 0.10)	.331	0.97	0.97	0.05	5014.30
Validation data 5 (*N* = 115)											
Model 1: TR/TW on WA	48.07	49	.511	0.99		0.00 (0.00, 0.06)	.895	1.00	1.00	0.05	7060.70
Model 2: TW on WA	48.86	51	.559	1.00	1.08	0.00 (0.00, 0.06)	.918	1.00	1.00	0.05	7058.13
Validation data 6 (*N* = 94)											
Model 1: TR/TW on WA	61.60	49	.107	1.01		0.05 (0.00, 0.09)	.439	0.97	0.96	0.06	5714.59
Model 2: TW on WA	75.46	51	.015	0.99	22.19^{***}	0.07 (0.03, 0.10)	.153	0.94	0.92	0.07	5723.23

* $p < .05$, ** $p < .01$, *** $p < .001$

Note. χ^2 = chi-square statistic, *df* = degrees of freedom, SCF = scaling correction factor for maximum likelihood estimation with robust standard errors, TRd = Satorra-Bentler scaled chi-square difference test statistic, RMSEA = root mean square error of approximation, CI = confidence interval of RMSEA, CFI = comparative fit index, TLI = Tucker-Lewis index, SRMR = standardized root mean square residual, AIC = Akaike's information criterion, TR = text-level reading, TW = text-level writing, WA = word-level writing accuracy.

Supplement Table 1. Correlation matrix of variables included in the structural equation modeling with the main data

	1	2	3	4	5	6	7	8	9	10	11	12
1. Reading												
2. Homophones	0.71											
3. Compounds Completion	0.56	0.48										
4. Compounds Structure	0.65	0.61	0.49									
5. Radicals	0.49	0.48	0.35	0.49								
6. Antonyms/Synonyms	0.62	0.51	0.48	0.58	0.43							
7. Kana Suffixes	0.63	0.54	0.38	0.55	0.49	0.62						
8. Four-character Idioms	0.66	0.54	0.44	0.65	0.47	0.72	0.69					
9. Error Correction	0.59	0.52	0.40	0.50	0.42	0.64	0.53	0.65				
10. Writing	0.65	0.52	0.42	0.49	0.42	0.73	0.71	0.73	0.59			
11. Text reading	0.50	0.43	0.36	0.44	0.31	0.36	0.37	0.43	0.34	0.38		
12. Text writing	0.45	0.39	0.27	0.39	0.45	0.39	0.46	0.44	0.38	0.37	0.44	