In press in *Scientific Studies of Reading*

The Impact of Character Complexity on Chinese Literacy: A Generalized Additive Modeling Approach

Shawn Hemelstrand[1], Brian W. L. Wong[1,2], Catherine McBride[3], Urs Maurer[1], and Tomohiro Inoue[1]

[1]The Chinese University of Hong Kong, Hong Kong

[2]Basque Center on Brain, Language and Cognition, Spain

[3]Purdue University, The United States

Author Note

Shawn Hemelstrand https://orcid.org/0000-0002-9627-1571

Brian W. L. Wong https://orcid.org/0000-0002-0519-6116

Catherine McBride https://orcid.org/0000-0002-6212-3511

Urs Maurer https://orcid.org/0000-0002-4156-8597

Tomohiro Inoue https://orcid.org/0000-0001-8163-087X

This research was supported by the Collaborative Research Fund (CUHK8/CRF/13G; C4054-17WF to C. McBride, PI), the Theme-based Research Scheme (T44-410/21-N to U. Maurer, PC, C. McBride, PC, and T. Inoue, Co-PI), and the General Research Fund (Project Number 14617721 to T. Inoue, PI) from the Hong Kong Special Administrative Region Research Grants Council.

Correspondence concerning this article should be addressed to Tomohiro Inoue, Department of Psychology, The Chinese University of Hong Kong, Shatin, NT, Hong Kong. Email: tinoue@cuhk.edu.hk

Word Count: 8879

Abstract

Purpose: We examined the effect of character complexity on early Chinese literacy (word reading and writing). We also investigated whether cognitive skills (phonological awareness, morphological awareness, and rapid automatized naming [RAN]) could moderate the influence of character complexity on literacy outcomes.

Method: Our pre-registered study included a sample of 342 Cantonese–speaking Hong Kong Chinese children (162 males, 180 females, mean age = 7.32, *SD* = 0.87) from Grade 1 to Grade 3. We used a generalized additive mixed model (GAMM) to estimate the main effect of complexity on word recognition and dictation as well as interaction effects with each of the three cognitive skills. Age and nonverbal IQ were also included as controls.

Results: Character complexity had a slightly curvilinear but consistently negative association with word reading, with odds ratios ranging from .62 to .71 per 10% increase in complexity. RAN and morphological awareness buffered the negative influence of character complexity. For writing to dictation, character complexity was more curvilinear, but character complexity and the interactions between cognitive skills and character complexity were not significant.

Conclusion: Complex visual input can be a barrier to early reading development in Chinese, and fostering morphological awareness may help mitigate this effect.

The Impact of Character Complexity on Chinese Literacy: A Generalized Additive Modeling Approach

Literacy is universally dependent on visual processing (Perfetti et al., 2013). Barriers to perceiving objects may consequently alter how we read and write. Graphic complexity, the density in visual input provided by scripts within orthographies, may in turn influence literacy (Chang et al., 2016, 2018; Miton & Morin, 2021; Nag et al., 2014). Despite the fact that complexity varies greatly within and between orthographies (McBride et al., 2022; Miton & Morin, 2021; Verhoeven & Perfetti, 2022), the direct effect of graphic complexity on literacy has been understudied. However, examining what effect complex visual input has on literacy is important. Navigating visually complex words is likely vital to learning how to read and write. Of additional relevance, specific literacy-related skills have been shown to improve reading ability (see e.g., Wu et al., 2009; Zhou et al., 2012, for findings from intervention studies in Chinese), and may in turn help disable barriers caused by this relationship.

What is graphic complexity exactly? The typical alphabetic definition of "graphs" only contains letters that are self-contained graphemes (Chang et al., 2018). However, this hard-line distinction does not exist for scripts like Chinese hanzi or Japanese kanji, whose graphemes contain strokes, radicals, and components (Inoue et al., 2017; McBride, 2016). Traditional visual complexity measures such as perimetric complexity miss graph-by-graph differences as a result of these discrepancies (Verhoeven & Perfetti, 2022). For example, graphs in Latin script often contain varying levels of connected points, whereas more complex scripts such as Cree, Thai or Telugu have more disconnected components. This trade-off between continuity and discontinuity between graphs may affect how they are learned (Chang et al., 2018; Chang & Perfetti, 2018). As typical definitions of graphs seem to be imprecise, we choose to use a definition more generalizable to all orthographies (Chang et

al., 2018). Thus, we define graphic complexity as the level of visual input from a minimal unit of the graph, otherwise a written form that can be combined with other graphs to form graphemes. This defines graphs as alphabetic letters, Chinese characters, Japanese kana, and others which form the "basic writing unit" of each script (Chang et al., 2018). A word's graphic complexity by extension is simply the combination of letters, characters, or other graphs that make up a word.

Thus far, research has shown that orthographies contain both universal and system-specific cognitive skills associated with successful reading, but visual complexity effects, as well as shifts in cognitive skills due to this complexity, are presently unclear (Perfetti et al., 2013; Verhoeven & Perfetti, 2022). One study (McBride-Chang et al., 2011) revealed that Chinese and Korean children outperformed their Israeli and Spanish counterparts on visual skill performance. Given that Chinese and Korean have more visually complex characters than Western scripts, the authors theorized that visual complexity of graphs may influence visual skill acquisition in early literacy development. Another study on Arabic (Tibi et al., 2022) showed that the level of visual similarity in graphs (e.g., ت / ث vs. ب / ش) was not predictive of letter knowledge, indicating that differentiation of visual features may not be extreme enough to exhibit influence on literacy. This seems to run contrary to some Asian orthographies, as visual similarity has been suggested to impact literacy in Chinese (Chang et al., 2016; Wang et al., 2021) and akshara-based (Nag et al., 2014; Padakannaya et al., 2015) writing systems.

Research on alphabets has similarly revealed little information on how graphic complexity influences literacy. A study on the Latin script in English (Treiman & Kessler, 2011) has shown that letters with more typical orientations, such as circular-shaped appendages on the right of letters like 'b' and 'p', were more predictive of copying ability when writing, indicating that irregularity of visual input may influence writing ability.

Multiple studies have also shown that uppercase letters are easier to identify in Latin script than lowercase letters, with their comparative number of connected points argued by some as a form of graphic complexity (Bowles et al., 2014; Chang et al., 2018). Unfortunately, studies in non-Asian contexts have not tested graphic complexity directly as a global construct, so it is unclear what relationship dense visual input has on literacy based on these studies alone. In addition, as the graphic complexity in these writing systems varies little (Chang et al., 2018), conclusions from these studies are likely limited.

Previous research, which has examined graphic complexity effects more directly, has typically involved the contrast of different scripts, though some rare studies have looked at graph-by-graph complexity directly (e.g., Inoue et al., 2017; McBride-Chang et al., 2005b; Nag et al., 2014). One study showed that complexity was associated with a negative effect on akshara recognition among emergent readers in the Kannada script, but it seemed unclear if participant reading skills interacted with this complexity in any way (Nag et al., 2014). However, some studies have indeed shown consequent changes in cognitive predictors of literacy skills. One unique study on Japanese readers (Inoue et al., 2017) compared literacy outcomes between visually simple hiragana (median perimetric complexity: 24.1) and visually complex kanji (median perimetric complexity: 28.9; see Chang et al., 2018). When all paths were entered into the analysis, it seemed that each script had different associations with reading predictors. Phonological awareness was a uniquely significant but weak predictor of hiragana accuracy, whereas kanji accuracy had a stronger association with morphological awareness and rapid automatized naming (RAN). However, as hiragana is a syllabary and kanji is a logography (Tamaoka, 1991), these differences may in part be due to the structural components of each script rather than complexity alone.

Chinese offers a unique solution to this problem, as Chinese contains two structurally similar logographic scripts: simplified Chinese and traditional Chinese. Traditional Chinese is

a logographic script which is commonly used in Hong Kong, Taiwan, and Macau, whereas simplified Chinese is the same script with reductions in strokes and radicals (which is most commonly used in Mainland China and Singapore). Each character in both scripts are self-contained syllables and morphemes, and the majority of Chinese words contain two or more characters (McBride et al., 2022). While traditional Chinese and simplified Chinese share an inventory of thousands of characters, traditional forms are generally more complex (e.g. 这里 in simplified script and 這裏 in traditional script) and some simplified characters have merged multiple traditional characters into one character (發/髮 in traditional is only 发 in simplified). A study that directly compared outcomes for these two showed that simplified script readers had higher visual skill scores (visual closure, visual discrimination, and visual spatial relationships) compared to traditional script readers (McBride-Chang et al., 2005b). Another study with adults seemed to validate this finding, as simplified script readers reported higher visual discrimination rates than traditional script readers (Yang & Wang, 2018). Conversely, another study indicated that traditional script readers seemed to respond faster to visual discrimination tasks than simplified script readers, while the actual difference in accuracy seemed to be negligible between groups (Chang & Perfetti, 2018).

With these differences apparent, caveats must be made for these three studies. First, although these scripts are similar, the history of Chinese character simplification has resulted in the reduction of several phonetic and semantic components of characters (Han et al., 2022). In essence, this does not provide a pure test of within-script complexity effects, as differences may be due in part to orthographic features unrelated to visual complexity (see McBride, 2016, for a relevant discussion). Second, these studies used fairly dissimilar measures of visual skills (e.g. categorical perception, pattern discrimination) with two different age groups (children versus adults), which make comparison between the two scripts difficult. A direct look at some of the most important skills required for reading and writing may help

researchers understand how complexity changes the nature of writing systems.

The Role of Three Cognitive Skills

Before outlining the role of these distinct cognitive skills, we first explain why we concentrated on these three skills in the first place. First, we tried to find three of the most common predictors of literacy that were as script-independent as possible. For example, using a task like radical awareness, which may have had direct interactions with script complexity, would not generalize well across scripts given it is so specific to Chinese characters. Selecting "universal" skills is of course impossible to a degree given how much scripts vary, but we tried to utilize those that were most applicable. Second, because there is a lack of research on interactions between complexity and literacy skills, we tried to at least use skills that had some evidence of effects from complexity in past studies, documented in this section. Third, there were other variables we were originally interested in using for our analysis, but they were unfortunately not available in the original dataset. As such, we were limited to what variables were actually usable. With this in mind, which skills specifically merit inclusion?

One of the cognitive skills that may modify the effect of visual complexity is phonological awareness. An attribute of successful readers and writers, this skill involves an understanding of sound rules in written text, and its effects seem consistent across studies (Melby-Lervåg et al., 2012; Verhoeven & Perfetti, 2022). However, some research has shown that the role of phonological awareness may vary between visually simple and visually complex scripts. For example, phonological awareness has been shown to be a stronger predictor of English reading compared to Chinese across several studies (Ruan et al., 2018). This may be in part due to the fact that a limited set of visually simple letters in English allow more direct graph-to-sound mappings than Chinese, which requires navigating an estimated 800 phonetic radicals of varying complexity (McBride, 2016). Contrasts among other scripts

in a small number of studies have shown mixed results. Japanese hiragana, a visually simple script, seems to be more dependent on phonological awareness than Japanese kanji, a visually complex script, but this association appeared fairly weak (Inoue et al., 2017). In contrast, a direct comparison between traditional Chinese and simplified Chinese has shown negligible differences in phonological awareness, which likely yield no major differences in consequent reading ability (McBride-Chang et al., 2005b).

These results in Chinese and Japanese are difficult to define by complexity alone and should be cautiously interpreted. In the Japanese case, it is self-evident that syllabaries like hiragana provide simple consonant-vowel structures that lend themselves to easy detection with heightened phonological awareness skills (Verhoeven & Perfetti, 2022). However, it is also possible that less visual information needs to be dissected in each graph, and so a focus on phonological accuracy is more easily attained given the direct phonology-to-graph correspondence awarded by simple hiragana graphs. Differences between Chinese scripts are less extreme compared to differences between Japanese scripts given they are both logographic scripts (McBride, 2016). However, these are comparisons between scripts rather than between individual graphs in both studies, and so there is no definitive answer on whether complexity is part of this relationship. As an example of how complexity may still matter, the phonological readings of characters in Chinese scripts are somewhat unreliable (McBride, 2016); thus, complexity may additionally add redundant visual information that competes with the attention of early readers. Consequently. an awareness of sound rules, specifically how they map to phonetic radicals, may reduce the amount of visual information required to read a character, as early readers would only have to focus their attention to the semantics of the characters. The role of phonological awareness in terms of visual complexity effects in any case remains unclear; yet its influence may be important to understanding whether it is a useful skill which prevents potential negative influences from dense visual

input.

Another cognitive skill, morphological awareness, may also play a role in how complexity affects literacy rates. Morphological awareness is essentially an understanding of how to combine morphemes, the smallest units of meaning in a language, to derive meaning from words (Ku & Anderson, 2003; Lee et al., 2022). Similar to phonological awareness, the positive association between morphological awareness and literacy is robust across languages, but this may vary to a degree based on which script is used (Lee et al., 2022; Ruan et al., 2018). For example, one meta-analysis showed that English and Chinese exhibited no significant differences in partial correlations between morphological awareness and literacy, but morphological awareness was more closely associated with Chinese literacy than phonological awareness (Ruan et al., 2018). The previously referenced Japanese literacy study (Inoue et al., 2017) also showed that morphological awareness was more predictive of visually complex kanji compared to visually simple hiragana. As Chinese hanzi and Japanese kanji have similar levels of graphic complexity (Chang et al., 2018; Miton & Morin, 2021), this seems to suggest that morphological awareness is potentially more important for visually complex scripts. Finally, a cross-script study between Korean, English, simplified Chinese, and traditional Chinese showed that Latin script, the most visually simple of these four, had the weakest association between morphological awareness and reading ability (McBride-Chang et al., 2005a).

Again, the effects of complexity must be carefully construed. These differences in morphological awareness skills between scripts may arise purely from orthographic features of each writing system and may not reflect complexity influences alone. As an example, alphabets utilize more complex morphology (such as derivation and inflection), and so differences between Chinese and English may be completely due to their respective morphology rules (Verhoeven & Perfetti, 2022). Comparatively, Chinese is derived from an

inventory of thousands of characters (McBride, 2016), and one must construct words among a limited set of morphemes that compound units of meaning, therefore lexical compounding may be of greater importance. However, this may also allow one to overcome difficulties in graphic complexity by limiting the range of morphological units required to guess the reading of complex characters in words (e.g., when reading the word 米飯 'rice,' one may read the simpler character 米 and quickly associate it with highly probable words like 米飯 or 米線 'rice noodle'). Assuming that complex characters are generally more difficult to learn, and given that characters are closely associated with meanings in logographic scripts (Inoue et al., 2023), children's knowledge of morphology and semantics may help their learning of complex characters. Despite this, only a direct investigation of complexity effects can tease apart which differences between complex graphs are due to visual input and which differences are due to orthographic features of a written language.

A final cognitive skill important to literacy is RAN, defined as the speeded naming of highly familiar stimuli, such as objects, letters, colors, and numbers (Denckla & Cutting, 1999). Its importance for literacy has been, like the previously mentioned predictors, well established across studies and languages (Araújo et al., 2015; Song et al., 2016). However, its relative influence in relation to graphic complexity is hardly studied. One cross-cultural investigation of several writing systems (Georgiou et al., 2022) showed that the association of RAN with reading might vary between scripts, but it was unclear from the study if graphic complexity influenced this association. Some cross-script studies have also shown that RAN seems in part graph-specific (e.g., Altani et al., 2017; Georgiou et al., 2016; Inoue et al., 2017). For example, Inoue et al. (2017) found that reading in Japanese kanji (a visually complex script) was more closely associated with RAN than reading in hiragana (a visually simple script). Finally, a meta-analysis conducted on RAN and reading ability showed non-alphabetic orthographies had higher correlations between RAN and reading fluency

(Araújo et al., 2015). Non-alphabetic scripts in general have higher visual complexity than alphabets (Miton & Morin, 2021). Therefore, RAN may serve as an important tool for quick pattern recognition of visually complex words, but past experiments have not made this distinction completely certain. The usual caveats apply here. While complexity may in some part be the cause of these differences, it must be stressed that these differences may be purely due to the orthographic demands of each script, thus only a direct examination of complexity effects can be conducted to attempt to figure out a solution to this question.

To summarize, past research has shown that these three predictors have a consistent influence on reading and writing across languages. Despite this, it is not clear whether these differences are purely because of the orthographic demands of each script or if this effect may be due to their comparative graphic complexity. Given that these predictors have positive influences on literacy, whereas complex visual input may not have this effect, these factors may serve as beneficial components to overcoming visually complex words. With this being the case, investigating the effect of graphic complexity and its influence on these predictors seemed to be a notable research gap worth exploring, as it appears no specific study has directly examined how graphs alter this relationship. Observing this effect may help researchers better understand how the density of visual input, and the implied difficulty it may bring, affects literacy and its associated predictors.

The Present Study

With this background context in place, the purpose of this study was to examine whether graphic complexity may affect literacy acquisition and to identify important literacy skills that may subvert this effect. We define this proposed relationship between graph-level complexity, cognitive skills, and literacy ability here as the *graphic complexity hypothesis* (GCH). In this study, we operationalized graphic complexity using GraphCom (Chang et al., 2018; Chang & Perfetti, 2018), a measure devised to discriminate the complexity of

characters across different languages (see Method section for details). Based on the literature reviewed above showing that variation in graphic complexity may influence cognitive predictors of literacy between writing systems (e.g., Chinese vs. English), we expected that it would also impact their associations within writing systems (e.g., graphs within Chinese scripts). More specifically, given these two presuppositions, we hypothesized that graphic complexity, a form of visual input and potential barrier to reading, would have a negative effect on reading and writing in Chinese. We also speculated that three primary cognitive skills may attenuate the effects of complexity: phonological awareness, morphological awareness, and RAN. Understanding the influence of these predictors would make clear the relationship between complex visual input and literacy ability.

To test this relationship, we looked at how these influences manifested in a single script, traditional Chinese. Of any known single writing system, traditional Chinese has the most variation in visual complexity between its graphs (Chang et al., 2018; Miton & Morin, 2021). Accordingly, one may test the rich variation in graph complexity in this script directly, while simultaneously removing the influence of structural differences between writing systems, such as the mapping principles that result in differences of orthographic depth and breadth (Inoue et al., 2017; Katz & Frost, 1992; Verhoeven & Perfetti, 2022). To achieve this goal, we sought to utilize a logistic generalized additive mixed model (GAMM), using crossed random effects to test outcomes of reading and writing ability (Baayen et al., 2008; Pedersen et al., 2019; see Supplementary Material I for an introduction to GAMMs). A crossed effects design specifically allows one to model in the aggregated effects of graph-level and subject-level predictors as well as their respective random variation per subject and item. By doing so, our analysis would enable us to inspect the overall influence of subject literacy skills and graph complexity while accounting for “random noise” in the model.

In this study, we utilized a dataset originally collected for a longitudinal project on Chinese literacy (Authors, 2020). The analysis plans were pre-registered with the Open Science Framework (OSF). Details about this study and its pre-registration can be found at this link: [https://osf.io/63nzj/?view_only=1d485c53c36e4bed8aaba3b8830ab1d1].

Method

Participants

The data from 342 children (162 males, 180 females, mean age = 7.32, *SD* = 0.87, range = 6–10) was used for this study.[1] Mixed models that utilize crossed random effects expand the data frame by each random effect (Baayen et al., 2008). This in turn gave us 6,840 observations (342 subjects x 20 writing items) for the writing model and 51,300 observations (342 subjects x 150 reading items) for the reading model. As a subject x item model is recommended to have greater than 1,600 observations per condition, our models had adequate power to detect medium effects (Brysbaert & Stevens, 2018). The participants were recruited via local primary schools and social media platforms. The selection criteria required students from grades 1 to 4 who spoke Cantonese as a native language. Participants were also required to have lived in Hong Kong for at least 2.5 years. This criterion was used to select subjects who lived in Hong Kong long enough to accurately reflect normal educational development. None reported experiencing any intellectual, sensory, or behavioral difficulties (based on parents' reports). All children used as participants in this study were included only after obtaining written consent from their parents. Ethics approval was obtained from [university's name anonymized for blind review].

To provide an accurate model of early literacy development, only participants from Grades 1 to 3 were selected for our analysis. This subset was selected because some research shows that effects from certain literacy predictors may diminish in importance as children age. As an example, there is evidence that morphological awareness remains an important

predictor of literacy even after primary school, but phonological awareness and RAN may diminish or change direction as early as Grade 1, and phonological awareness may sometimes reverse association with literacy outcomes by Grade 4 (Lee et al., 2022; Li et al., 2012; Yeung et al., 2013). The majority of the participants were in Grade 1 at the time of testing (n = 187), while the rest were in Grade 2 (n = 93) and Grade 3 (n = 62).

Measures

Item-Level Variable

Graphic complexity. There is no agreed upon measure for graphic complexity; however, we chose to use GraphCom, a measure devised to discriminate the complexity of characters in several different languages (Chang et al., 2018; Pelli et al., 2006). We decided on using this specific measure of complexity because it incorporates extra features about graphs that may be overlooked (discussed below). It is important to note that no scoring method is prescribed by the creators of GraphCom, and it was originally devised for single characters. Given that our items were two-character words, we chose to combine sum scores of each component for both characters. This was accomplished by adding together four features of a graph. The main feature of this score, *perimetric complexity*, is derived from dividing the squared perimeter of a graph by its area times the product of 4 times pi ($PC = \frac{P^2}{A4\pi}$). This compares the "ink space" of a graph compared to its white space and has been used in visual perception studies (Pelli et al., 2006). The other features of GraphCom are more novel, as they consider the less studied elements of graph complexity (Chang et al., 2018). The second feature, *disconnected components*, measures spaces between written elements of a graph. The third feature includes *connected points*, referring to points of contact in a graph. The final feature uses *simple features*, which are elements of a graph that are differentiated from each other (Chang et al., 2018). Readers may refer to Supplementary Material III for a more detailed description of GraphCom.

We provide an example with a Chinese word in traditional script composed of two characters: 火車 'train.' The character for 'fire', 火, has a perimetric complexity of 14, two disconnected components, two connected points, and four simple features. In sum, the total GraphCom score for this character would thus be 22. The character for 'car', 車, has a perimetric complexity of 22, one disconnected component, 10 connected points, and seven simple features. This character's sum GraphCom score is thus 40. Together, this word has a combined GraphCom score of 62. As shown, using this sum allows one to examine the complexity of words with various combinations of simple and complex graphs.

GraphCom scores of two-character words are technically multi-dimensional in nature and thus pose problems of orthogonality when considering reliability. Additionally, though two recent literacy studies have used GraphCom as a composite measure (Arachchillage, 2018; Eisenhauer et al., 2022), its reliability as a composite has never been tested. We consequently fit a two-factor confirmatory factor analysis (CFA) to each GraphCom score to obtain reliability coefficients, as unidimensional reliability coefficients would be inappropriate and potentially misleading (Flora, 2020). The rationale and procedure for this estimation method, including our omission of disconnected components from the composite, are detailed in Supplementary Material III. The reliability for Word Reading Character 1 GraphCom (McDonald's omega = .93), Word Reading Character 2 GraphCom (omega = .95), Dictation Character 1 GraphCom (omega = .96), and Dictation Character 2 GraphCom (omega = .98) was within acceptable ranges. The final composite was formulated by summing the features together for each word used for reading and each word used for dictation.

Participant-Level Variables

Nonverbal IQ. The Raven's Standard Progressive Matrices scale was used to assess participants' nonverbal reasoning ability (Raven, 2000). As a potential correlate of Chinese

literacy and other cognitive skills (Ye et al., 2021; Yeung et al., 2013), IQ score was used as a control variable in the analyses of the present study. There were 60 items in the original standardized measure, with five sets (Sets A to E) of 12 items each. Each set involved different principles of matrix transformation, and within each set the items became progressively more difficult. Each item was made up of a target geometric matrix with a missing piece. Participants were asked to pick the best piece out of six to eight choices to complete the target matrix. All correct items were summed up to form a total raw score.

According to the test user manual, participants who were less than 8.5 years old at the time of testing were administered the short form with 36 items (Sets A to C), while older participants were given the full version. As only a limited number of participants met this criterion for our subset of the data, only 55 participants were tested on the long format of the test with Sets A to E, which caused an artificial right plateau effect in the data distribution of scores. As such, we elected to use only Sets A to C for a more realistic distribution of scores. This transformation caused no obvious ceiling or floor effects after re-aggregating these items for the sample. McDonald's omega was .89.

Morphological awareness. The morphological awareness test consisted of two practice items and 48 test items (McBride-Chang et al., 2003). In each item, a scenario was presented orally in one or two sentences. Participants were asked to construct a novel compound word to describe the object or concept based on that scenario from known morphemes. For example, the word sunset means that the sun goes down, or sets. If one were to come up with a word for the concept of the moon going down, according to the rules of this task, the best answer would be "moonset." As morphological awareness and all of the following participant-level measures were unidimensional, only one-factor CFAs were fit for reliability coefficients for these composites. McDonald's omega coefficient was .89.

Rapid automatized naming (RAN). The RAN task (Denckla & Cutting, 1999)

consisted of eight rows of five digits (2, 4, 6, 7, and 9). The RAN Digits task was used in this study since alphanumeric RAN tasks using letters and digits correlate more strongly with reading than non-alphanumeric RAN tasks using objects and colors (see Araújo et al., 2015; Song et al., 2016, for meta-analyses). Specific to Chinese, using characters for RAN would result in a single-character reading task, thus digits were used to avoid this problem. These digits were arranged in different orders for each row. They were presented on a page and children were asked to name the digits in Cantonese as quickly and accurately as possible. Two trials were completed, and the average time in seconds was recorded. The number of errors made were also recorded but they were input for data checking purposes and not used in calculating the composite score. RAN was inverse coded by multiplying the composite score by -1 to facilitate easier interpretation. This provides the useful feature of a positive correlation with literacy rather than the non-intuitive alternative of higher RAN scores equating lower literacy rates. The correlation between both RAN times was strong (r = .86, 95% confidence interval [.83, .88], $p < .001$), indicating consistency in their measurement.

Phonological awareness. The original phonological awareness test was composed of 51 items, which measured syllable deletion and onset deletion (McBride-Chang & Ho, 2000; McBride-Chang & Kail, 2002). In this task, the children were asked to take away one syllable from three-syllable words (e.g., /dai6 mun4 hau2/ without /mun4/ would be /dai6 hau2/; the number indicates the tone). Half of the items were real words and half consisted of nonsense syllables that conformed to the phonological constraints of Cantonese. Preliminary data inspection revealed that 10 items were too easy for the participants and sometimes represented linear constants in our subset of the data, thus they were removed from the analysis. McDonald's omega coefficient was .97.

Word reading accuracy. A two-character word reading task was used. In this study, we used the two-character measure instead of a single-character reading measure because it

was directly comparable to our writing measure, which only used two-character words, and had much more substantial bandwidth-fidelity due to a larger number of items (150; for more details of the task selection, see the pre-registration). The 150 test items were adapted from The Hong Kong Test of Specific Learning Difficulties in Reading and Writing (Ho et al., 2000). The items were arranged in order of increasing difficulty. The task was stopped when children failed to read 15 consecutive items. One point was given for correct recognition of both characters whereas zero points were given when participants could not recognize both characters. Additionally, once the 15-item stopping rule was applied, all subsequent items were scored as zero. This coding was adopted because subjects were assumed to no longer be able to correctly identify words and including this missing data would result in either substantial data loss with their removal or inaccurate model estimation with their inclusion. A participant's score was the total number of words read correctly. The omega reliability coefficient for this measure was .99.[2]

Dictation. In the original version of this task (Tong et al., 2009), the children were asked to write 20 two-character Chinese words.[3] The words were arranged by character complexity as well as age of acquisition. In Tong et al.'s (2009) study, a two-point scale was developed to code the children's writing: One point was given if the child could write down at least one radical of the character or he/she only made a mistake in one of the radicals (e.g., missing a stroke, adding additional components); two points were given if he/she could write the whole character correctly. However, as our analysis required a logistic regression model, and we desired a more parsimonious definition of correct scores, we dummy-coded the completely correct answers as "1" and all other answers as "0." Therefore, the total score for this composite was a maximum of 20 points. McDonald's omega coefficient was .88.

Statistical Analysis

As highlighted previously, our analysis utilized a combination of initial data

exploration as well as a logistic generalized additive mixed model (GAMM; Baayen et al., 2017; Simpson, 2018). GAMMs are recently developed regression models that combine generalized additive models (GAMs) and generalized linear mixed models (GLMMs; Pedersen et al., 2019; Wood, 2017). GAMMs measure fixed and random effects, as do GLMMs, but they additionally relax the assumption of linearity by including smoothing terms called splines. These splines are formulated by estimating a set of basis functions that approximate the localized curves of the data distribution (Pedersen et al., 2019; Wood, 2017). The regression line is thus not fit in a "straight line" as typically done in regression, but rather fit to where the data is spatially distributed, which provides unique insights. This allows users of GAMMs, for example, to provide more accurate descriptions of the data from regressions, as one can approximate associations between variables when their functional relationship shifts (see Pedersen et al., 2019 and Simpson, 2018, for examples).

Our model also used a crossed random effects design that has gained recent prominence in psycholinguistic mixed models (Baayen et al., 2008). To summarize, this type of model expands a dataset into a subject-by-item data frame which can be used to estimate the amount of random variance, or random "noise," attributed to differences in subjects and stimuli. For example, poor readers may score worse on average than good readers and hard items may be difficult to guess compared to easy items. Estimating these random effects separate from the predicted outcome provides a more accurate representation of the overall influence of typical regression predictors (Baayen et al., 2017). This estimation can also potentially increase the power of the test, consequently increasing the chances of replication (Baayen et al., 2008; Brysbaert & Stevens, 2018).

The entire statistical analysis was conducted using R Version 4.2.1 (R Core Team 2022). GAMM fitting and plotting was performed with the mgcv package (Wood, 2017; Wood & Wood, 2015). As many functions were used for the analysis, additional packages are

listed in the Supplementary Materials as well as the R script available on our OSF project page (https://osf.io/gv2qb/?view_only=23caf1c254814e83a9254734dce64106). The GAMM models were fit using the mgcv package's "bam" function using fast restricted maximum likelihood (fREML), which improves estimation time for large data sets (Pedersen et al., 2019; Wood, 2017; Wood & Wood, 2015). All significance cutoffs were set at $p < .05$. Syntax for both GAMMs can be viewed within the R script on our OSF project page. The model was fit with three types of splines, which are the smoothing functions approximated across the data distribution (Simpson, 2018; Wood, 2017). Main effects without interaction terms were fit with cubic regression (CR) splines. Interaction effects and their respective independent main effects were fit with tensor product splines, and random effects were fit with random effect splines (Baayen et al., 2017; Pedersen et al., 2019; Simpson, 2018; Wood, 2017). Testing of regression assumptions and model performance can be found in the Supplementary Materials under "Model Performance".

Results

Preliminary Data Analysis

Descriptive statistics for the pre-split dataset are shown in Table 1 and descriptive plots from this data can be found in the Supplementary Materials. Correlations of each variable pair are shown in Table 2. Both GAMM models used nonverbal IQ and age as control variables, while phonological awareness, morphological awareness, and RAN were entered as main effects. Graphic complexity was modeled as a main effect and an interaction for these three main effects. Results from the GAMM models can be seen in Table 3 (see Supplementary Material I for explanation of terminology). To summarize, GAMM *p* values are approximations of the "zero-effect" of each curve's smoothing parameters (Wood, 2012). If flat horizontal lines are shown within the majority of the confidence interval bands, this will result in non-significant results. Effective degrees of freedom (*edf*) represent the degree

of curvilinear structure, with values of 1 equating a linear fit. As our model is not Gaussian, explained deviance is used to ascertain the amount of variance estimated by each model.

Plots of each predictor's association with literacy outcomes, however, are more illustrative than the summary output alone. These plots are shown in Figures 1 and 2. Percentages on the y-axis indicate the predicted probability of reading or writing successfully after holding all other predictors at their average value. The x-axis represents the raw value of the predictor shown (e.g., the probability of reading correctly when RAN is -50 is 0%). Filtering for the lowest RAN score in our data, the associated child with this score only correctly answered three out of 150 reading items. However, at the opposite end of the range, the probability of reading correctly was more than 60% after holding all other predictors at their average value. The child with the highest RAN score answered 137 out of 150 reading items correctly, indicating a much higher rate of reading accuracy (91.3%). This particular child was also older than others (9 years old) and scored higher than average on many of the measures, highlighting why average ability across measures must be controlled when looking at the partial effects produced by each GAMM. As the intercept is included and a probability transformation is utilized, the plot curves are sometimes slightly different from the effective degrees of freedom from the original model estimation.

Reading GAMM

After fitting the reading GAMM, graphic complexity had an overall negative linear effect on reading accuracy, with slight curvature in the upper and lower extremes of the distribution (*edf* = 1.04, χ^2 = 6.31, *p* = .014). Odds ratios for each 10% increase in graphic complexity ranged from .63 to .71 across the regression curve.[4] Confidence intervals (CIs) suggested that this effect was most accurate in the middle range of values, which was very likely due to the paucity of characters at both extremes of the distribution. Morphological awareness had a significant linear association with reading outcomes (*edf* = 1.00, χ^2 = 36.99,

$p < .001$), with a consistent odds ratio of 1.65 per 10% increase in morphological awareness. RAN had a similar effect, with odds ratios ranging from 1.46 to 1.65 per 10% increase due to slight curves (*edf* = 1.24, $\chi^2 = 41.13$, $p < .001$), leading to less precision in CI at the lower range of values. In contrast, phonological awareness had a non-significant effect on reading (*edf* = 2.50, $\chi^2 = 3.49$, $p = .33$), which was originally posited to be due to suppression effects from morphological awareness in the model (Friedman & Wall, 2005). However, additional exploratory data analysis revealed that omission of morphological awareness in the model did not substantially alter this effect (see Supplementary Material IV for more details). This resulted in a semi-flat curve with only minor shifts in reading ability.

Interactions for graphic complexity and cognitive skills partially confirmed our pre-specified hypothesis, as the complexity x phonological awareness interaction was not significant (*edf* = 3.01, $\chi^2 = 5.17$, $p = .27$). Significant graphic complexity interactions were found for RAN (*edf* =11.89, $\chi^2 = 17.11$, $p = .03$) and morphological awareness (*edf* = 1.00, $\chi^2 = 5.79$, $p = .02$). A heatmap showing significant interaction effects can be seen in Figure 3. As shown in the heatmaps, graphic complexity had a stronger effect on reading outcomes. However, RAN and morphological awareness seemed to be associated with diminished consequences of this effect, as higher values of each skill increased the chances of reading accuracy, even when graphic complexity was quite high. For example, if a subject saw a word with a graphic complexity score of 110 (i.e., 抽屜) and the same subject had a morphological awareness score of 15, the predicted probability of a correct reading score was around 20%. However, if this subject's morphological awareness score was instead 35, they had an 80% chance of getting the same word correct, providing a dramatically improved outcome. An inspection of random effects showed that subject reading scores varied by 1.97 *SD,* while reading score by item varied by 3.57 *SD.* This indicated substantial variability in item difficulty and subject ability.

Writing GAMM

Table 3 includes results from the writing GAMM. The writing model had some considerably contrary effects compared to the reading model. Many variables shared by the reading model were still significant, such as age (*edf* = 1.30, χ^2 = 33.50, $p < .001$), nonverbal IQ (*edf* = 2.59, χ^2 = 18.62, $p < .001$), RAN (*edf* = 2.60, χ^2 = 34.34, $p < .001$), and morphological awareness (*edf* = 1.00, χ^2 = 5.05, $p = .02$). However, all other main effects and interactions were not significant. In general, subject variation (*SD* = 1.02) and item variation (*SD* = 0.95) in scores varied much less than the reading model. This may indicate that developmental level in writing skill was more similar across subjects, with difficulty in each writing task having a narrower range than the reading model.

The relationship between graphic complexity and writing, contrary to our hypothesis, was curvilinear yet non-significant compared to the reading GAMM. An inspection of the partial effects plot in Figure 2 shows an initial curve of this relationship that was negative and linear similar to the reading GAMM, with 10% increases in complexity yielding odds ratios ranging from .71 to .84. However, increases following this point led to a plateau effect with brief fluctuations. The lack of statistical significance in this effect thereafter was apparent, as 10% increases in complexity within this distribution ranged from 0.94 to 1.08, exhibiting meager effects on writing outcomes. This indicates that complexity only had a limited influence until a “complexity threshold” was met. However, results from this model must be interpreted with caution, as explained below.

Discussion

The purpose of the current study was to examine the effects of graphic complexity on early childhood literacy (word reading and writing). Additionally, this study sought to investigate whether specific cognitive predictors (phonological awareness, RAN, and morphological awareness) could change this relationship. Using a generalized additive mixed

model, we estimated the linear and curvilinear trends in fixed and random effects to test the associations of graphic complexity with literacy. The analysis provided new insights on literacy outcomes based on graphic complexity influences, though the conclusions differed between reading and writing outcomes. Based on the results, a number of important conclusions can be made.

The findings from the reading model confirmed previous associations with predictors of literacy, explaining 66.1% of the deviance in reading outcomes (Lee et al., 2022; Ruan et al., 2018; Song et al., 2016). However, our model provided additional unique insights on how graphic complexity affects this relationship. In line with our proposed graphic complexity hypothesis, character complexity in general had a negative effect on reading accuracy rates (see Chang et al., 2016, for a similar finding from a simulation study). This suggests that as visual input becomes more complex, readers typically struggle with learning the readings of characters (Nag et al., 2014). Interestingly, a previous study showed that readers of simplified Chinese had higher reading accuracy rates than readers of traditional Chinese (McBride-Chang et al., 2005b). This seems to support the idea posited earlier that differences between these scripts could be related to orthographic structure rather than complexity alone. This could also be related to the linguistic context of each sample, as Mainland Chinese readers use a different spoken language (Mandarin) and even use different phonological coding systems like pinyin (Chang & Perfetti, 2018; McBride, 2016). As such, our study provides a purer analysis of graphic complexity, as it elucidates this relationship within a singular script using a singular spoken language.

Our results further showed that, despite the negative influence of graphic complexity, morphological awareness and RAN served as a “protective barrier” to this influence. This relationship is sensible given the nature of each predictor. As morphological awareness in Chinese typically involves combining the meaning of characters, readers who face

graphically complex words may derive meaning of one unknown character by using the meaning of another known character via context (Ku & Anderson, 2003; McBride et al., 2022; Ruan et al., 2018). Similarly, because RAN involves integrating visual processing skills (Kail & Hall, 1994), being able to read quickly may bypass the effects of visual complexity via pattern recognition of the global shapes of characters rather than deliberate reading of character features (Ruan et al., 2018; Song et al., 2016). Another possible explanation is that RAN may help children develop high-quality orthographic representations (Bowers et al., 1999) in learning visually complex characters. While these associations from the analysis may shed light on early childhood literacy, it is important to state that these associations may be limited to this age group. For example, the associations between reading and RAN shift over time even in primary school for early Chinese readers (Pan et al., 2020). Consequently, this relationship likely looks different for mature readers, who have the benefit of a lifetime of exposure to complex print.

However, the influence of phonological awareness apparently differs from these two predictors, showing no significant association with either reading or writing. This seems to reflect past research showing weak influences between scripts of varying complexity, as well as the comparative influence of phonological awareness in visually complex scripts compared to visually simple scripts (Inoue et al., 2023; McBride-Chang et al., 2005b; Ruan et al., 2018). However, this feature may not be constant across all visually complex scripts. Phonological awareness may in fact be relevant for visually complex scripts if they are phonology-focused, as in Indic akshara scripts, but less so if they are morpho-syllabic, as in Chinese scripts (McBride, 2016; Nag et al., 2014; Wijaythilake et al., 2018, 2019). It is also possible that the combination of dense visual input with a large number of homophones and phonetic radicals in Chinese (McBride, 2016; Verhoeven & Perfetti, 2022) makes it impossible to bypass the strength of graphic complexity effects. Because Hong Kong children do not use coded

phonological systems such as pinyin or zhuyin like other regions that use Chinese script, this may have also been a contextual factor that influenced the results (McBride, 2016). However, the impact of phonological awareness has been shown to have no major differences between Mandarin and Cantonese in at least one meta-analysis which examined the two (Song et al., 2016), so it is likely not a major point of contention.

Statistical reasons for these differences are also apparent. As the phonological awareness measure for this specific subset of the data had clear bimodal distributions, this caused a flattening effect on the relationship between reading and writing. It also appeared that morphological awareness partially suppressed the effect of phonological awareness, as both measures were strongly correlated. Our initial GLMM models (see Supplementary Materials) showed similar suppression effects, but variance inflation factor (VIF) was well below the 10-point cutoff (max VIF: 1.67 [1.62, 1.73]), thus this indicated no major concerns of multi-collinearity (Miles, 2014). Exploratory data analysis also indicated that removing morphological awareness did not substantially alter phonological awareness effects on literacy outcomes. Controlling for the effect of morphological awareness, which had a clear positive linear effect on literacy, it seems morphology performs better as a predictor of reading and writing (see Ruan et al., 2022; Yang et al., 2022, for similar results). Thus, with this in mind, phonological awareness may be less important for dense visual input than we hypothesized.

In contrast to the reading model, evidence for the writing model of the graphic complexity hypothesis was limited, explaining only 33.4% of model deviance. While it appeared that complexity had a negative linear effect at the lower range of values, the effect thereafter became plateaued with only minor shifts in this association. This seems to indicate that for writing, graphic complexity has an initial disabling effect, but after reaching a certain threshold (around 80 GraphCom score; see Figure 2), this effect may become redundant. It is

possible that because reading can be considered an input (i.e., recognition; readers passively obtain information from material), while writing can be considered an output (i.e., production; writers must actively produce information), this may influence differences between reading and writing. Writing visually complex characters like 繁 'many' (85 GraphCom score) and 體 'body' (109 GraphCom score), for example, may be equally difficult given the similar amount of time required for writing them. Features like the number of strokes in a graph, which may be highly correlated with complexity, may also play some part in this relationship. In contrast, reading characters in general is a more rapid process, sometimes performed within milliseconds (see e.g., Reichle & Yu, 2017, for evidence from Chinese). With this being the case, it may be beneficial for researchers to further investigate this input-output modality of complexity.

Some other explanations may be present for differences in the writing model. One cause may be the predictors chosen for the GAMMs. We purposely included the same predictors in each model to provide a more direct comparison between literacy outcomes, as this would allow us to ascertain if reading and writing had shared features in terms of complexity, while eliminating the confounding influence of incomparable skills. Given that writing requires fine motor movements by producing graphs stroke-by-stroke, it could be that other skills in fact better define this relationship, such as visual memory, visual-motor skills, or copying ability (Ye et al., 2021). Statistically, there are also three additional explanations for these differences. First, as writing item random effects varied less in difficulty compared to reading items, this may have led to less extremity in effects. Hence including much easier or more difficult words may be helpful to more accurately capture outcomes. Second, for the same reason, participants varied less in writing skill compared to reading, leading to less random effect variation. Theoretically, it is possible that our specific sample of participants were generally more alike in ability and thus may not be comparable to the entire population

of early developing writers of Chinese. Third, and most important, the sheer number of items in the reading model eclipse those in the writing model, which naturally leads to more precision in estimation.

There are several limitations to our analysis that should be taken into account. First, our study was cross-sectional, and thus may not be indicative of long-term developmental trends in literacy. Second, as our writing model had substantially fewer items than the reading model, it is difficult to ascertain the full breadth of the complexity-literacy relationship. Relatedly, there is no denying that the reading model had considerably more power to detect effects than the writing model. As such, they cannot be easily compared. Third, we did not include a grade-level analysis into our pre-registration. This was in part due to the unequal sample sizes between grades, as first graders far outnumbered second and third graders in the sample. However, if early curricula in Hong Kong begins to structure exposure towards a visually simple character set (see Hong Kong Education Bureau, 2009), then the models may yield different outcomes based on these grade-level changes. Fourth, due to a stopping rule in the original study design, the phonological awareness measure exhibited strong bimodal behavior, which may have affected its association with literacy outcomes. A statistical solution exercised in some psychological experiments is to transform this continuous variable into a binary factorial variable; however, this practice was avoided since it has problematic features (Royston et al., 2006). As such, we left the variable in the analysis as-is. Lastly, our pre-registration did not include word frequency or orthographic awareness as part of our analysis. Word frequency seems to have a mediating influence on phonological and morphological awareness, and visual-orthographic analysis of characters seems to be an important skill in reading Chinese (e.g., Lin et al., 2019; Wang et al., 2021). Therefore, their relation to complexity may necessitate their inclusion into a future study.

This analysis also presents some interesting future directions. It is self-evident that complexity on average is higher in Chinese compared to other orthographies. It is clear from our analysis that our GAMM models also over-represent the middle range of complexity scores, which may be specific to Chinese. For scripts that have lower complexity scores on average, this may change the nature of main effects and interactions. Thus, this should be studied further to test the robustness of our hypothesis across scripts rather than the just the complexity within scripts. A second important direction is related to our main measure of complexity. While perimetric complexity has a long history of use and is the main feature of GraphCom, further studies will need to validate the entire GraphCom composite as a reliable measure of graphic complexity before its reliability and validity can be considered as concrete. This study at least presents the first attempt at validating this composite as a measure of complexity.

To summarize, the present study examined the influences of graphic complexity on early childhood literacy skills (word reading and writing) in Chinese, as well as complexity effects on typical cognitive predictors of literacy (phonological awareness, RAN, and morphological awareness). The results indicate that as visual complexity of characters increases, reading accuracy for children decreases. However, morphological awareness and RAN may mitigate this effect. Additionally, the influence of complexity on writing may be weak, though evidence of this effect is fairly limited. Theoretically, these findings extend our understanding of how cognitive skills may help early readers overcome difficulties in learning to read. As the complexity of visual input increases, the ability for child readers to meaningfully interpret graphs and memorize them diminishes, as it serves as a barrier to successful reading outcomes. Practically speaking, the results indicate that educators and parents should consider fostering skills such as morphological awareness in order to attenuate the effects of difficult-to-read words in Chinese.

Declaration of Interest Statement

We have no known conflict of interest to disclose.

References

Altani, A., Georgiou, G. K., Deng, C., Cho, J.-R., Katopodi, K., Wei, W., & Protopapas, A. (2017). Is processing of symbols and words influenced by writing system? Evidence from Chinese, Korean, English, and Greek. *Journal of Experimental Child Psychology*, *164*, 117–135. https://doi.org/10.1016/j.jecp.2017.07.006

Arachchillage, D. K. M. (2018). *The role of akshara knowledge and phonological processing skills in reading development among Sri Lankan children* [Doctoral dissertation, University of Alberta]. Education and Research Archive. https://era.library.ualberta.ca/items/2ae946b0-38c1-499e-a4c8-415381abdc2e/download/a53d9a9b-295c-4772-99b2-13368f70827c

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, *107*(3), 868–883. https://doi.org/10.1037/edu0000006

Authors (2020).

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234. https://doi.org/10.1016/j.jml.2016.11.006

Bowers, P. G., Sunseth, K., & Golden, J. (1999). The route between rapid naming and reading progress. *Scientific Studies of Reading, 3,* 31–53. https://doi.org/10.1207/s1532799xssr0301_2

Bowles, R. P., Pentimonti, J. M., Gerde, H. K., & Montroy, J. J. (2014). Item response analysis of uppercase and lowercase letter name knowledge. *Journal of*

Psychoeducational Assessment, *32*(2), 146–156. https://doi.org/10.1177/0734282913490266

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Chang, L.-Y., Chen, Y.-C., & Perfetti, C. A. (2018). GraphCom: A multidimensional measure of graphic complexity applied to 131 written languages. *Behavior Research Methods*, *50*, 427–449. https://doi.org/10.3758/s13428-017-0881-y

Chang, L.-Y., Plaut, D. C., & Perfetti, C. A. (2016). Visual complexity in orthographic learning: Modeling learning across writing system variations. *Scientific Studies of Reading*, *20*(1), 64–85. https://doi.org/10.1080/10888438.2015.1104688

Chang, L.-Y. (Wendy), & Perfetti, C. A. (2018). Visual factors in writing system variation: Measurement and implications for reading. In H. K. Pae (Eds.), *Writing systems, reading processes, and cross-linguistic influences: Reflections from the Chinese, Japanese and Korean languages* (Vol. 7, pp. 49–72). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.7.03cha

Denckla, M. B., & Cutting, L. E. (1999). History and significance of rapid automatized naming. *Annals of Dyslexia*, *49*, 29–42. https://doi.org/10.1007/s11881-999-0018-9

Eisenhauer, S., Gagl, B., & Fiebach, C. J. (2022). Predictive pre-activation of orthographic and lexical-semantic representations facilitates visual word recognition. *Psychophysiology*, *59*(3), e13970. https://doi.org/10.1111/psyp.13970

Flora, D. B. (2020). Your coefficient alpha Is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, *3*(4), 484–501. https://doi.org/10.1177/2515245920951747

Friedman, L., & Wall, M. (2005). Graphical views of suppression and multicollinearity in multiple linear regression. *The American Statistician*, *59*(2), 127–136. https://doi.org/10.1198/000313005X41337

Georgiou, G. K., Aro, M., Liao, C.-H., & Parrila, R. (2016). Modeling the relationship between rapid automatized naming and literacy skills across languages varying in orthographic consistency. *Journal of Experimental Child Psychology*, *143*(C), 48–64. https://doi.org/10.1016/j.jecp.2015.10.017

Georgiou, G. K., Cardoso-Martins, C., Das, J. P., Falcón, A., Hosokawa, M., Inoue, T., Li, Y., Martinez, D., Padakannaya, P., Parrila, R., Pollo, T. C., Salha, S. S., Samantaray, S., Shu, H., Tanji, T., Tibi, S., & Vieira, A. P. A. (2022). Cross-language contributions of rapid automatized naming to reading accuracy and fluency in young adults: Evidence from eight languages representing different writing systems. *Journal of Cultural Cognitive Science, 6*, 151–168. https://doi.org/10.1007/s41809-021-00092-7

Han, S. J., Kelly, P., Winters, J., & Kemp, C. (2022). Simplification is not dominant in the evolution of Chinese characters. *Open Mind, 6*, 264–279. https://doi.org/10.1162/opmi_a_00064

Ho, C. S. H., Chan, D. W. O., Tsang, S. M., & Lee, L. H. (2000). *The Hong Kong test of specific learning disabilities in reading and writing (HKT-SpLD)*. The Chinese University of Hong Kong and Education Department, HKSAR Government.

Hong Kong Education Bureau. (2009). *Zhongying duizhao xianggang xuexiao zhongwen xuexi jichu zici* [Lexical items with English explanations for fundamental Chinese learning in Hong Kong schools]. Chinese Language Education Section, Curriculum Development Institute, Education Bureau, The Government of Hong Kong Special Administrative Region. https://www.edbchinese.hk/lexlist_en/

Inoue, T., Georgiou, G. K., Muroya, N., Maekawa, H., & Parrila, R. (2017). Cognitive predictors of literacy acquisition in syllabic hiragana and morphographic kanji. *Reading and Writing*, *30*, 1335–1360. https://doi.org/10.1007/s11145-017-9726-4

Inoue, T., Zheng, M., Lui, K. F.-H., McBride, C., & Ho, C. S.-H. (2023). Early literacy acquisition across logographic orthographies: Evidence from a cross-linguistic study. *Early Childhood Research Quarterly, 63*, 73–84. https://doi.org/10.1016/j.ecresq.2022.11.009

Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology*, *30*(6), 949–954. https://doi.org/10.1037/0012-1649.30.6.949

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. *Advances in Psychology*, *94*, 67–84. https://doi.org/10.1016/S0166-4115(08)62789-2

Ku, Y.-M., & Anderson, R. C. (2003). Development of morphological awareness in Chinese and English. *Reading and Writing, 16*, 399–422. https://doi.org/10.1023/a:1024227231216

Lee, J. won, Wolters, A., & Grace Kim, Y.-S. (2022). The relations of morphological awareness with language and literacy skills vary depending on orthographic depth and nature of morphological awareness. *Review of Educational Research 0*(0), 1-31. https://doi.org/10.3102/00346543221123816

Li, H., Shu, H., McBride-Chang, C., Liu, H., & Peng, H. (2012). Chinese children's character recognition: Visuo-orthographic, phonological processing and morphological skills. *Journal of Research in Reading*, *35*(3), 287–307. https://doi.org/10.1111/j.1467-9817.2010.01460.x

Lin, D., Mo, J., Liu, Y., & Li, H. (2019). Developmental changes in the relationship between character reading ability and orthographic awareness in Chinese. *Frontiers in Psychology*, *10*(2397), 1-10. https://doi.org/10.3389/fpsyg.2019.02397

McBride, C. A. (2016). Is Chinese special? Four aspects of Chinese literacy acquisition that might distinguish learning Chinese from learning alphabetic orthographies. *Educational Psychology Review*, *28*, 523–549. https://doi.org/10.1007/s10648-015-9318-2

McBride, C., Pan, D. J., & Mohseni, F. (2022). Reading and writing words: A cross-linguistic perspective. *Scientific Studies of Reading*, *26*(2), 125–138. https://doi.org/10.1080/10888438.2021.1920595

McBride-Chang, C., Cho, J.-R., Liu, H., Wagner, R. K., Shu, H., Zhou, A., Cheuk, C. S.-M., & Muse, A. (2005a). Changing models across cultures: Associations of phonological awareness and morphological structure awareness with vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea, and the United States. *Journal of Experimental Child Psychology*, *92*(2), 140–160. https://doi.org/10.1016/j.jecp.2005.03.009

McBride-Chang, C., Chow, B. W. Y., Zhong, Y., Burgess, S., & Hayward, W. G. (2005b). Chinese character acquisition and visual skills in two Chinese scripts. *Reading and Writing*, *18*(2), 99–128. https://doi.org/10.1007/s11145-004-7343-5

McBride-Chang, C., & Ho, C. S.-H. (2000). Developmental issues in Chinese children's character acquisition. *Journal of Educational Psychology*, *92*(1), 50–55. https://doi.org/10.1037//0022-0663.92.1.50

McBride–Chang, C., & Kail, R. V. (2002). Cross–cultural similarities in the predictors of reading acquisition. *Child Development*, *73*(5), 1392–1407. https://doi.org/10.1111/1467-8624.00479

McBride-Chang, C., Shu, H., Zhou, A., Wat, C. P., & Wagner, R. K. (2003). Morphological awareness uniquely predicts young children's Chinese character recognition. *Journal of Educational Psychology*, *95*(4), 743–751. https://doi.org/10.1037/0022-0663.95.4.743

McBride-Chang, C., Zhou, Y., Cho, J.-R., Aram, D., Levin, I., & Tolchinsky, L. (2011). Visual spatial skill: A consequence of learning to read? *Journal of Experimental Child Psychology*, *109*(2), 256–262. https://doi.org/10.1016/j.jecp.2010.12.003

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138*, 322–352. https://doi.org/10.1037/a0026744

Miles, J. (2014). Tolerance and variance inflation factor. In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri & J. L. Teugels (Eds.), *StatsRef: Statistics reference online*. Wiley. https://doi.org/10.1002/9781118445112.stat06593

Miton, H., & Morin, O. (2021). Graphic complexity in writing systems. *Cognition*, *214*, 104771. https://doi.org/10.1016/j.cognition.2021.104771

Nag, S., Snowling, M., Quinlan, P., & Hulme, C. (2014). Child and symbol factors in learning to read a visually complex writing system. *Scientific Studies of Reading*, *18*(5), 309–324. https://doi.org/10.1080/10888438.2014.892489

Padakannaya, P., Pandey, A., Saligram, D., & Rao, S. R. (2015). Visual-orthographic complexity of Akshara and eye movements in reading: A study in Kannada alphasyllabary. *Writing Systems Research*, *8*(1), 1–11. https://doi.org/10.1080/17586801.2015.1071235

Pan, J., Cui, X., McBride, C., & Shu, H. (2020). An investigation of the bidirectional relations of word reading to timed visual tasks involving different levels of

phonological processing in Chinese. *Scientific Studies of Reading*, *24*(4), 275–291. https://doi.org/10.1080/10888438.2019.1663857

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, *7*, e6876. https://doi.org/10.7717/peerj.6876

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, *46*, 4646–4674. https://doi.org/10.1016/j.visres.2006.04.023

Perfetti, C., Cao, F., & Booth, J. (2013). Specialization and universals in the development of reading skill: How Chinese research informs a universal science of reading. *Scientific Studies of Reading*, *17*, 5–21. https://doi.org/10.1080/10888438.2012.689786

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Reichle, E. D., & Yu, L. (2017). Models of Chinese reading: Review and analysis. *Cognitive Science, 42*, 1154–1165. https://doi.org/10.1111/cogs.12564

Revelle, W., & Revelle, M. W. (2015). *Package 'psych'.* The comprehensive R archive network.

Rosseel, Y. (2014). *The lavaan tutorial*. Department of Data Analysis: Ghent University. https://lavaan.ugent.be/tutorial/tutorial.pdf

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, *25*, 127–141. https://doi.org/10.1002/sim.2331

Ruan, Y., Georgiou, G. K., Song, S., Li, Y., & Shu, H. (2018). Does writing system influence the associations between phonological awareness, morphological awareness, and reading? A meta-analysis. *Journal of Educational Psychology*, *110*(2), 180–202. https://doi.org/10.1037/edu0000216

Ruan, Y., Ye, Y., Lui, K. F. H., McBride, C., & Ho, C. S. H. (2022). How do word reading and word spelling develop over time? A three-year longitudinal study of Hong Kong Chinese–English bilingual children. *Reading Research Quarterly*. https://doi.org/10.1002/rrq.478

Simpson, G. L. (2018). Modelling palaeoecological time series using generalised additive models. *Frontiers in Ecology and Evolution*, *6*(149), 1–21. https://doi.org/10.3389/fevo.2018.00149

Song, S., Georgiou, G. K., Su, M., & Hua, S. (2016). How well do phonological awareness and rapid automatized naming correlate with Chinese reading accuracy and fluency? A meta-analysis. *Scientific Studies of Reading*, *20*(2), 99–123. https://doi.org/10.1080/10888438.2015.1088543

Tamaoka, K. (1991). Psycholinguistic nature of the Japanese orthography. *Studies in Language and Literature*, *11*(1), 49–82.

Tibi, S., Edwards, A. A., Kim, Y.-S. G., Schatschneider, C., & Boudelaa, S. (2022). The contributions of letter features to Arabic letter knowledge for Arabic-speaking kindergartners. *Scientific Studies of Reading*, *26*(5), 357–372. https://doi.org/10.1080/10888438.2021.2016769

Tong, X., McBride-Chang, C., Shu, H., & Wong, A. M.-Y. (2009). Morphological awareness, orthographic knowledge, and spelling errors: Keys to understanding early Chinese literacy acquisition. *Scientific Studies of Reading*, *13*(5), 426–452. https://doi.org/10.1080/10888430903162910

Treiman, R., & Kessler, B. (2011). Similarities among the shapes of writing and their effects on learning. *Written Language & Literacy*, *14*(1), 39–57. https://doi.org/10.1075/wll.14.1.03tre

Verhoeven, L., & Perfetti, C. (2022). Universals in learning to read across languages and writing systems. *Scientific Studies of Reading, 26*, 150–164. https://doi.org/10.1080/10888438.2021.1938575

Wang, L., Wang, J., Liu, D., & Lin, D. (2021). The role of cognitive awareness and character properties in early Chinese reading. *Journal of Experimental Child Psychology*, *210*(105185), 1-10. https://doi.org/10.1016/j.jecp.2021.105185

Wijaythilake, M. A. D. K., Parrila, R., Inoue, T., & Nag, S. (2018). Instruction matters to the development of phoneme awareness and its relationship to akshara knowledge and word reading: Evidence from Sinhala. *Scientific Studies of Reading*, *22*, 420–433. https://doi.org/10.1080/10888438.2018.1466890

Wijaythilake, M. A. D. K., Parrila, R., Inoue, T., & Nag, S. (2019). Cognitive predictors of word reading in Sinhala. *Reading and Writing*, *32*(7), 1881–1907. https://doi.org/10.1007/s11145-018-9927-5

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). CRC Press, Taylor and Francis Group.

Wood, S. N. (2012). On p-values for smooth components of an extended generalized additive model. *Biometrika*, *100*, 221–228. https://doi.org/10.1093/biomet/ass048

Wood, S., & Wood, M. S. (2015). Package ‘mgcv’. *R package version*, *1*(29), 729.

Wu, X., Anderson, R. C., Li, W., Wu, X., Li, H., Zhang, J., Zheng, Q., Zhu, J., Shu, H., Jiang, W., Chen, X., Wang, Q., Yin, L., He, Y., Packard, J., & Gaffney, J. S. (2009). Morphological awareness and Chinese children’s literacy development: An

intervention study. *Scientific Studies of Reading*, *13*(1), 26–52. https://doi.org/10.1080/10888430802631734

Yang, X., Pan, D. J., Lo, C. M., & McBride, C. (2022). Same or different: Chinese character reading and word reading of young readers with development. *Reading and Writing*. https://doi.org/10.1007/s11145-022-10255-9

Yang, R., & Wang, W. S. Y. (2018). Categorical perception of Chinese characters by simplified and traditional Chinese readers. *Reading and Writing*, *31*(5), 1133–1154. https://doi.org/10.1007/s11145-018-9832-y

Ye, Y., Yan, M., Ruan, Y., McBride, C., & Yeung, C. F. (2021). Literacy learning in early Chinese-English bilinguals: The role of pure copying skill. *Early Childhood Research Quarterly*, *55*, 263–274. https://doi.org/10.1016/j.ecresq.2020.12.004

Yeung, P., Ho, C., Wong, Y., Chan, D., Chung, K., & Lo, L. (2013). Longitudinal predictors of Chinese word reading and spelling among elementary grade students. *Applied Psycholinguistics, 34*(6), 1245–1277. https://doi.org/10.1017/S0142716412000239

Zhou, Y.-L., McBride-Chang, C., Fong, C. Y. C., Wong, T. T. Y., & Cheung, S. K. (2012). A comparison of phonological awareness, lexical compounding, and homophone training for Chinese word reading in Hong Kong kindergartners. *Early Education and Development*, *23*(4), 475–492. https://doi.org/10.1080/10409289.2010.530478

Footnotes

[1] The original sample we used was a smaller sample of a much larger longitudinal dataset. Details about the selection of this subset are documented in the Supplementary Material II.

[2] Estimation of omega in R requires a confirmatory factor loading structure using lavaan (Rosseel, 2014) and the psych package (Revelle & Revelle, 2015), but the initial model variance-covariance matrix was not positive-definite, which may be due to the number of binary items that were estimated. To alleviate this issue, we used the "cor.smooth" function from the psych package, which is designed for transforming matrices that are not positive-definite. Results from the follow-up lavaan summary showed no problems with model estimation thereafter.

[3] The writing task (Tong et al., 2009) originally included 20 items, but this test later utilized a 25-item version which was made to be more difficult for older participants. We only used the 20-item version of the test, as those who were only tested on the 20-item test would naturally hit a different ceiling of scores compared to those who were tested on all items.

[4] The odds ratios and confidence intervals for GAMs and GAMMs vary based on the location of the fitted line. The odds ratios listed are based on 10% incremental increases in raw predictor units and thus fluctuate based on where the percentage "slice" occurs. This results in several odds ratio estimations and their respective confidence intervals for each GAMM model. A detailed list of each odds ratio and its respective confidence interval per percentage increase can be seen in the Supplementary Material V.

Table 1

Descriptive Statistics for GAMM Models

	M	*SD*	Skew	Kurtosis
Nonverbal IQ	24.36	6.15	0.27	2.69
Age (in years)	7.33	0.87	-0.02	1.67
Morphological awareness	19.35	6.44	-1.11	4.66
Rapid automatized naming (RAN)	-22.78	6.78	-0.66	2.81
Phonological awareness	25.16	10.37	0.67	2.71
GraphCom reading items	115.50	23.79	2.98	0.23
GraphCom writing items	97.45	32.73	2.14	0.28
Word reading accuracy	63.49	33.93	-0.03	1.96
Dictation	24.04	8.96	-0.25	2.30

Note. Only composite literacy scores are listed, as binary response data does not have meaningful skew/kurtosis values. As our final sample used complete data, all measures contain $n = 342$.

Table 2

Participant-Level Correlation Matrix

	1.	2.	3.	4.	5.	6.
1. IQ						
2. Age	.38 [.28, .46]					
3. MA	.44 [.35, .53]	.40 [.31, .49]				
4. RAN	.25 [.15, .34]	.35 [.26, .44]	.39 [.29, .47]			
5. PA	.38 [.28, .47]	.24 [.14, .34]	.55 [.47, .62]	.39 [.29, .47]		
6. WR	.44 [.35, .53]	.58 [.50, .64]	.58 [.50, .65]	.51 [.42, .58]	.35 [.25, .44]	
7. DICT	.41 [.32, .50]	.52 [.44, .59]	.44 [.35, .52]	.48 [.39, .56]	.32 [.22, .41]	.77 [.72, .81]

Note. PA = phonological awareness; MA = morphological awareness; RAN = rapid automatized naming; WR = word reading; DICT = word dictation. Numbers in brackets represent 95% confidence intervals (CI). Numbers left of brackets indicate Pearson's *r* coefficient of each variable pair. All correlations were significant ($p < .001$). As our final sample used complete data, all measures contain $n = 342$.

Table 3

GAMM Model Summaries

	Reading GAMM				Writing GAMM			
Parametric Coefficients	Est.	*SE*	*z*	*p*	Est.	*SE*	*z*	*p*
(Intercept)	-1.07	0.31	-3.44	.00	-0.40	0.23	-1.76	.08
Smooth Terms	*edf*	ref. *edf*	χ^2	*p*	*edf*	ref. *edf*	χ^2	*p*
Random effects								
Subjects	322.88	336.00	24,304.01	.00	251.59	336.00	1,491.93	.00
Items	147.23	148.00	178,256.52	.00	16.67	18.00	739.62	.00
Main effects								
IQ	1.99	2.00	12.75	.00	2.59	2.75	18.62	.00
Age	2.92	2.94	51.56	.00	1.30	1.36	33.50	.00
MA	1.00	1.00	36.99	.00	1.00	1.00	5.05	.02
RAN	1.24	1.25	41.13	.00	2.60	2.76	34.34	.00
PA	2.50	2.51	3.49	.33	1.00	1.00	0.60	.44
GC	1.04	1.04	6.31	.01	2.01	2.02	2.36	.34
Interactions								
GC x MA	1.01	1.02	5.79	.02	5.91	8.55	8.51	.43
GC x PA	3.01	3.79	5.17	.27	2.03	2.55	3.14	.41
GC x RAN	11.89	17.11	29.69	.03	1.00	1.00	0.04	.85
Model Metrics	Adj. R^2	Dev. Exp.	fREML	Obs.	Adj. R^2	Dev. Exp.	fREML	Obs.
(Value)	.71	66.1%	124,114.60	51,300	.37	33.4%	9,564.20	6,840

Note. Est. = estimate, *SE* = standard error, *edf* = effective degrees of freedom, ref. *edf* = reference effective degrees of freedom, χ^2 = chi-square value, Adj. R^2 = adjusted R^2, Dev. Exp = deviance explained, fREML = fast restricted maximum likelihood value, obs. = observations, MA = morphological awareness, PA = phonological awareness, GC = graphic complexity. For a more detailed description of GAMM summaries and their terminology, see Supplementary Material I.

Figure 1

Predicted Probability Plots for Reading GAMM

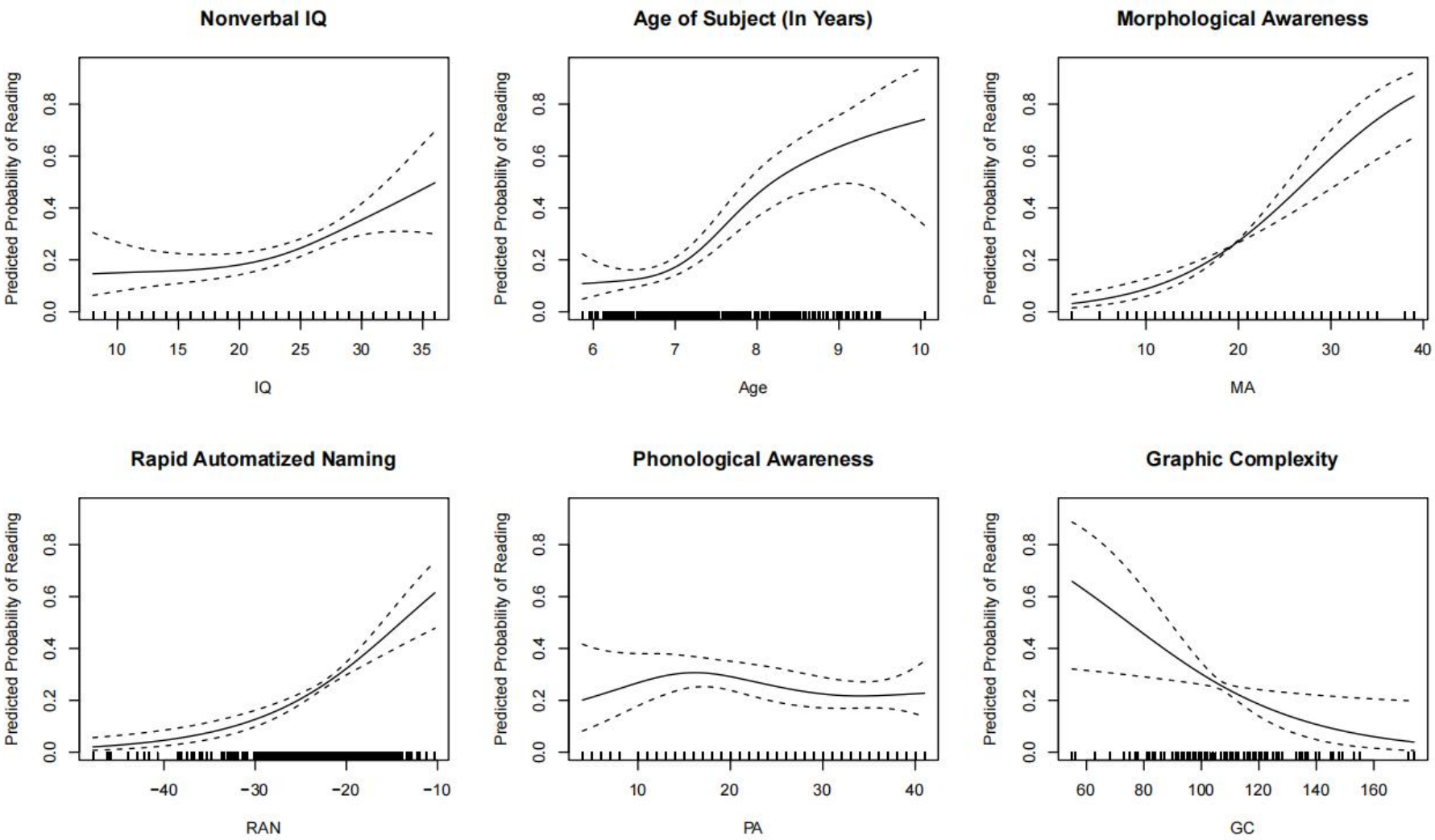


Figure 2

Predicted Probability Plots for Writing GAMM

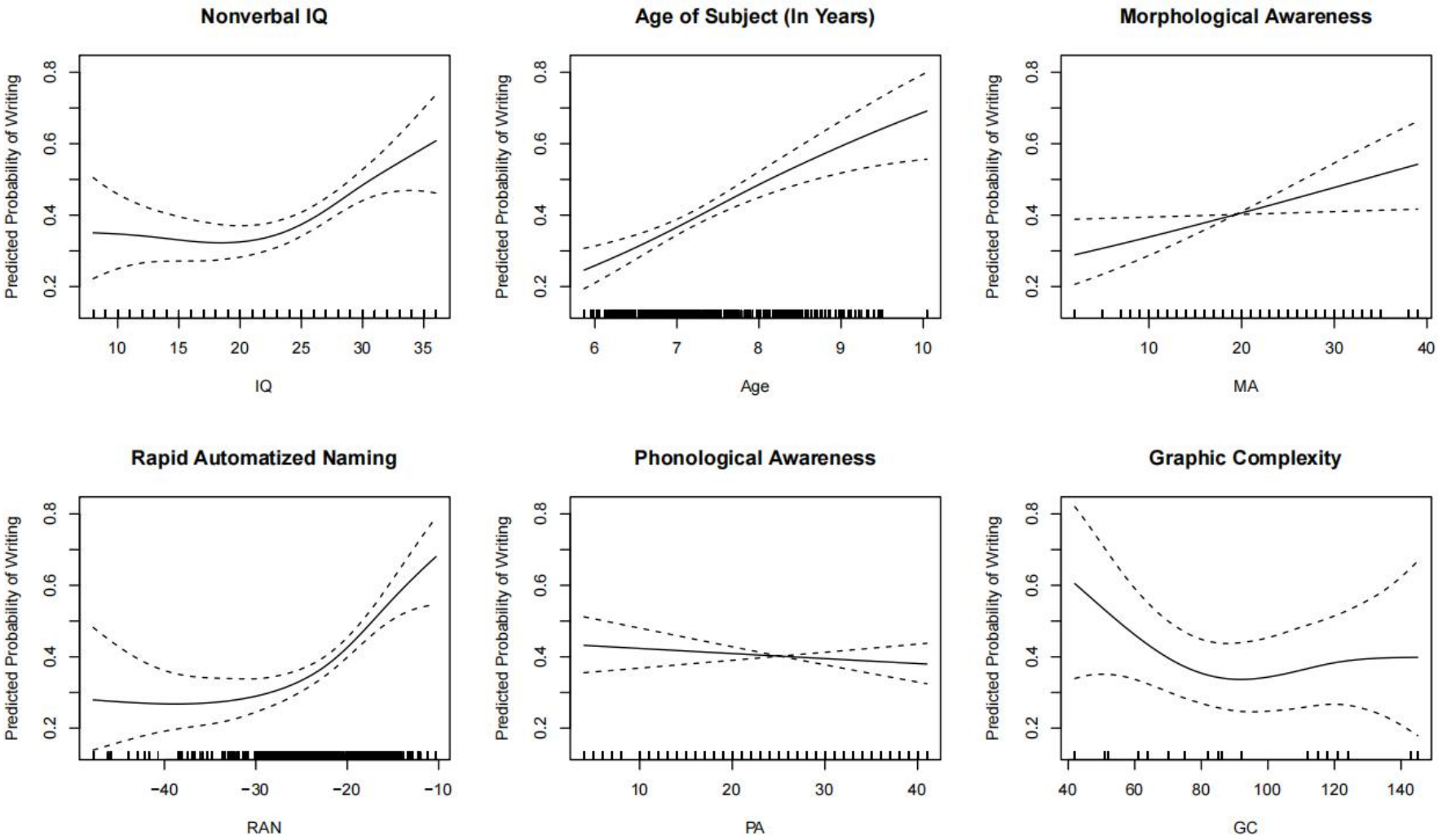


Figure 3

Reading GAMM Interaction Contour Plots

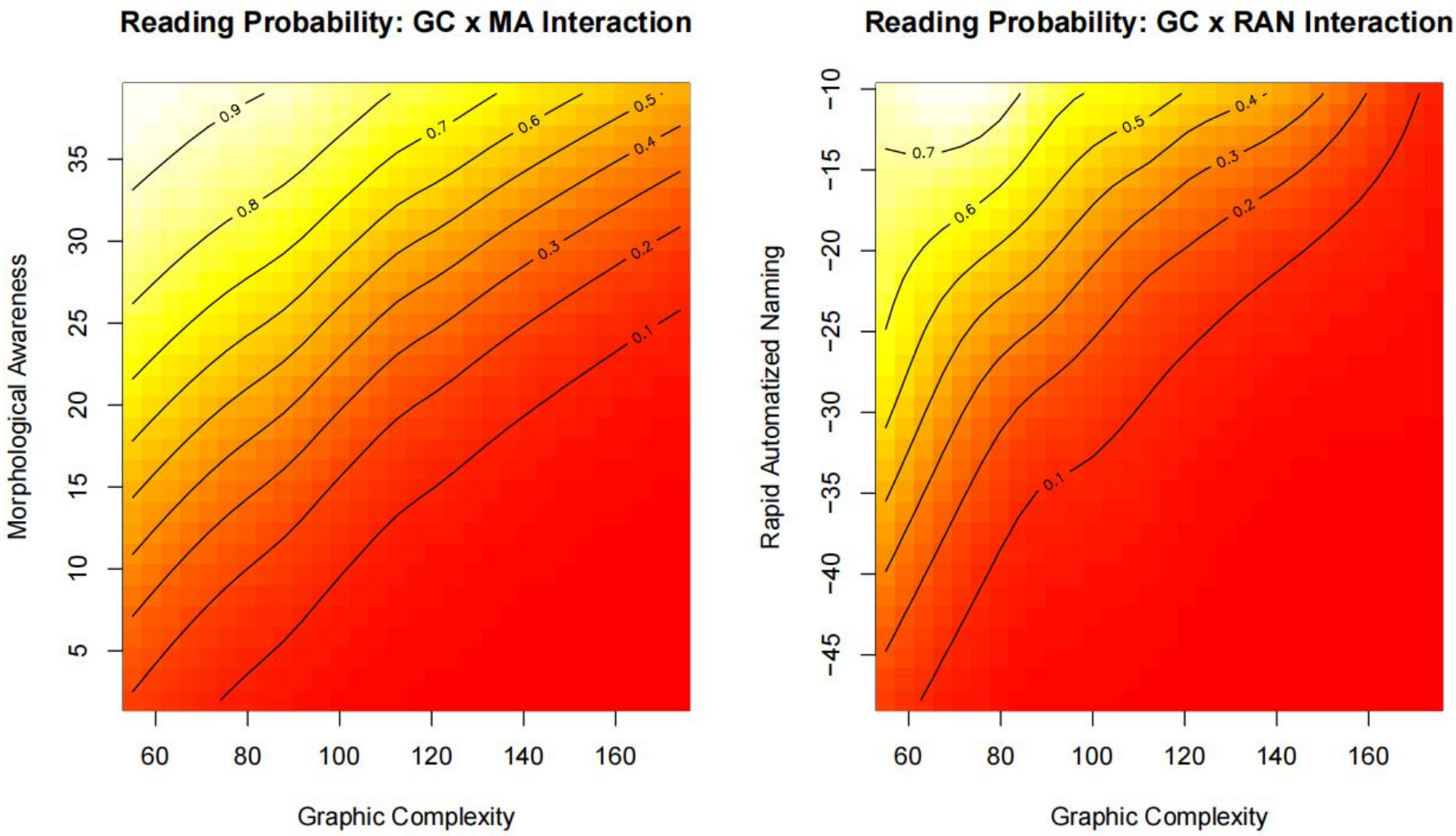


Figure Captions

1) *Note.* Reading GAMM predictors transformed to predicted probabilities. X-axis represents raw units of measures. Y-axis represents probability of reading success (for example .4 = 40% chance of reading correctly). Solid lines show GAMM regression fit. Dashed lines represent confidence intervals that include uncertainty about the overall mean. Each plot has the intercept of the regression added into the transformation, which allows each plot to be interpreted as the probability of success after controlling for the average value of other predictors.

2) *Note.* Writing GAMM predictors transformed to predicted probabilities. X-axis represents raw units of measures. Y-axis represents probability of writing success (for example .4 = 40% chance of writing correctly). Solid lines show GAMM regression fit. Dashed lines represent confidence intervals that include uncertainty about the overall mean. Each plot has the intercept of the regression added into the transformation, which allows each plot to be interpreted as the probability of success after controlling for the average value of other predictors.

3) *Note.* Contour plot heatmaps depicting attenuating effects of cognitive skills on reading complexity. Numeric values on contour lines represent predicted probability (for example .7 = 70% chance of reading accurately) dependent on the given values of GraphCom and the paired reading predictor. Rapid automatized naming is inverse coded for ease of interpretation. Yellow regions show increases in reading accuracy while red regions depict decreases in reading accuracy. Axis values are the raw scores for each measure. More details of graph interpretation can be found in Supplementary Material I.