



Chinese character handwriting: A large-scale behavioral study and a database

Ruiming Wang¹ · Shuting Huang¹ · Yacong Zhou¹ · Zhenguang G. Cai²

Published online: 25 February 2019
© The Psychonomic Society, Inc. 2019

Abstract

We report on a psycholinguistic database of Chinese character handwriting based on a large-scale study that involved 203 participants, each handwriting 200 characters randomly sampled from a cohort of 1,600 characters. Apart from collecting writing latencies, durations, and accuracy, we also compiled 14 lexical variables for each character. Regressions showed that frequency, age of acquisition, and the word context (in which a character appears) are all-around and influential predictors of orthographic access (as reflected in writing latency), motor execution of handwriting (as reflected in writing duration), and accuracy. In addition, phonological factors (phonogram status, spelling regularity, and homophone density) impacted orthographic access but not handwriting execution. Semantic factors (imageability and concreteness) only affected accuracy. These results suggest, among other things, that phonology is consulted in orthographic access while handwriting. As the first of its kind, this database can be used as a source of secondary data analyses and a tool for stimulus construction in handwriting research.

Keywords Handwriting · Chinese · Database · Lexicon · Phonology · Orthography

Humans have long used handwriting to communicate. In transcoding semantic information into orthographic output, handwriting involves not only central processes underlying conceptual preparation, lexical selection, and orthographic access, but also peripheral processes of motor programming and the actual writing execution (Baxter & Warrington, 1986; van Galen, 1991). The cognitive processes underlying handwriting have attracted much attention in recent years, due to technological innovations (e.g., digital tablets and software packages) that allow for detailed observations of the act of

handwriting. The studies on handwriting to date have employed typical lab research in which a small sample of participants handwrite a small set of selected words. Although these lab studies have shed much light on the cognitive mechanism underlying handwriting, the extents to which findings from small participant and stimulus samples generalize to the population and to all the words remain topics of debate. Indeed, conflicting findings have been observed in handwriting (e.g., regarding the role of phonology in orthographic access; Qu, Damian, Zhang, & Zhu, 2011; Q. Zhang & Wang, 2015). Therefore, a large-scale database with responses from a large sample of participants to a large sample of words would, among other functions, help elucidate on and arbitrate recent conflicting results regarding the cognitive processes underlying handwriting.

Many aspects of language use and communication have seen the construction of large-scale psycholinguistic databases. For instance, using a lexical decision task, researchers working on the Lexicon Project series collected responses (accuracy and response times) to 40,481 (American) English words (from a large sample of participants) in the English Lexicon Project (Balota et al., 2007), to 28,730 mono- and bisyllabic words in the British Lexicon Project (Keuleers, Lacey, Rastle, & Brysbaert, 2012), to 38,840 words in the French Lexicon Project (Ferrand et al., 2010), to 14,089 mono- and disyllabic words in the Dutch Lexicon Project (Keuleers, Diependaele, &

Ruiming Wang and Shuting Huang contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13428-019-01206-4>) contains supplementary material, which is available to authorized users.

✉ Ruiming Wang
wangrm@snu.edu.cn

✉ Zhenguang G. Cai
zhenguangcai@cuhk.edu.hk

¹ Guangdong Provincial Key Laboratory of Mental Health and Cognitive Science, and Center for Studies of Psychological Application, School of Psychology, South China Normal University, Guangzhou, China

² Department of Linguistics and Modern Languages/Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong, China

Brysbaert, 2010), and to 2,500 characters in the Chinese Lexicon Project (Sze, Rickard Liow, & Yap, 2014). Other databases have employed tasks that tap into production processes. The English Lexicon Project collected word-naming responses for 40,481 words (Balota et al., 2007). Psycholinguistic norms for Italian nouns were collected from word-naming responses for 626 nouns in Italian (Barca, Burani, & Arduino, 2002). Psycholinguistic norms for Chinese (Liu, Shu, & Li, 2007) were calculated from recorded naming responses for 2,423 characters, and a psycholinguistic database for traditional Chinese character naming (Chang, Hsu, Tsai, Chen, & Lee, 2016) collected responses for 3,314 characters. Q. Zhang and Yang (2003) collected naming responses for 311 pictures in an attempt to analyze name agreement, image agreement, familiarity, and visual complexity in written picture naming.

The availability of responses to a large set of stimuli from a large sample of participants has allowed researchers to investigate questions in a way that lab-based, small-sample studies are less likely to achieve. For instance, a large-scale dataset allows researchers to establish how performance (e.g., lexical processing) is jointly influenced by an array of different variables (Barca et al., 2002; Chang et al., 2016; Liu et al., 2007; Yap, Liow, Jalil, & Faizal, 2010). It also allows researchers to adjudicate previously contradicting lab-based findings (e.g., Chang et al., 2016). Finally, these databases can also be used as sources of norms for selecting experimental materials that can be varied on a variable of interest while being kept constant on others. Despite the availability of many large-scale psycholinguistic databases on different aspects of language use and communication, so far there is no psycholinguistic database on handwriting.

The present article reports a database on handwriting of Chinese characters. Chinese is of particular interest to handwriting research as the language uses a nonalphabetic script with much opaqueness between a word/character's phonology and its orthography. In recent years, Chinese has become a testing bed in quite a few studies examining whether and how phonology is used to access orthographic representations during handwriting in non-alphabetic writing systems (Qu et al., 2011; C. Wang & Zhang, 2015; Q. Zhang & Wang, 2015). According to the phonology mediation hypothesis (Geschwind, 1969; Luria, 1970), phonology mediates access to orthography; that is, people use phonology to guide the retrieval of orthographic codes in handwriting. The alternative account is the orthography autonomy hypothesis, according to which orthographic codes can be directly accessed from semantic input, without the need for phonological mediation (Rapp, Benzing, & Caramazza, 1997). Research on orthographic production in alphabetic languages has marshaled much evidence of phonological effects on orthographic access, giving support to the phonology mediation hypothesis (Afonso & Álvarez, 2011; Bonin, Peereman, & Fayol, 2001; Q. Zhang & Damian, 2010). However, as pointed out by Levelt, Roelofs,

and Meyer (1999), the effects of phonology and orthography are difficult to disentangle in many studies using alphabetic scripts such as English, in which spelling and pronunciation can be inferred through each other to some extent via phoneme–grapheme correspondences. Because of this, recent research on handwriting has shifted its focus to nonalphabetic languages such as Chinese, which can provide a better testing case to examine the role of phonology in orthographic access, thanks to the lack of one-to-one mappings between phonology (e.g., phonemes) and orthography (e.g., graphemes; Qu et al., 2011). Therefore, we believe a psycholinguistic database of Chinese character handwriting will provide a very useful tool for research on handwriting.

Many characteristics at the word or the character level have been shown to be important factors in word production (e.g., handwriting or naming). First, there are frequency-related factors. Many studies have shown that people are quicker at writing high- than writing low-frequency words or characters (Kandel & Perret, 2015; Qu, Zhang, & Damian, 2016; Søvik, Arntzen, Samuelstuen, & Heggberget, 1994; C. Wang & Zhang, 2015). Syllable frequency has also been shown to affect spoken word production, with people being faster at producing higher-frequency syllables (Carreiras & Perea, 2004; Cholin, Levelt, & Schiller, 2006; Q. Zhang & Wang, 2014), though González-Alvarez and Palomar-García (2016) also showed that people are slower to say words when their first syllable is high than low in frequency. It has also been argued that the effect of age of acquisition (AoA) in language is a frequency effect in disguise, with linguistic units acquired earlier to have a higher cumulative frequency (Brysbaert & Ghyselinck, 2006; Lewis, 1999). It has been shown that people are quicker at initiating the handwriting of words that are acquired earlier in life (though AoA does not have an effect on accuracy; Bonin, Fayol, & Chalard, 2001).

Semantic variables are also important predictors of word/character production (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). One semantic variable is the number of meanings that a word or character has: People are quicker at naming words or characters with more meanings (Chang et al., 2016; Woollams, 2005). Another semantic variable is a word or character's imageability, defined as a word/character's capacity to evoke sensory experience (e.g., visual, auditory, haptic images; Barca et al., 2002). Some studies showed that people are faster at naming high- than low-imageable words/characters (Cortese, Simpson, & Woolsey, 1997; Liu et al., 2007), though others failed to observe the effect (Barca et al., 2002) or only observed the effect in poor readers (Coltheart, Laxon, & Keating, 1988). Finally, a word/character can differ in its concreteness in meaning: A word/character is considered to be high in concreteness if it refers to something (e.g., an object or person) that can be experienced by the senses, but low in concreteness if it refers to an abstract concept (e.g., morality, politics; Barca et al., 2002). People are quicker at initiating the naming

of characters/words with high than with low concreteness (Chen & Peng, 1998; Liu et al., 2007).

Phonological variables may also play important roles, though this is currently under debate. One such variable is whether a character was historically coined using a sound radical, which represents the pronunciation of the character (and if so the character is a phonogram). Phonograms consists of a sound radical and/or a meaning radical that is semantically related to the character's meaning. In Chinese, most phonograms have the sound radical on the right (and the meaning radical on the left). H. Cai, Qi, Chen, and Zhong (2012) proposed that phonograms can be categorized according to the position of the sound radical (i.e., left, right, top or bottom). Although the exact position of the sound radical may be important for recognition, it may suffice in handwriting simply to categorize a Chinese character as whether or not having its sound radical written first, a measure we adopted in the present study.

Relatedly, though the sound radical in a phonogram is supposed to represent the pronunciation of the character, the resemblance in pronunciation between the sound radical and the character varies due to language evolution—that is, a character can change in pronunciation over time, and may end up with a pronunciation that is different from that of its sound radical.

For instance, the sound radical 青 (*qing*₁) can be found in the character 清 (*qing*₁, with a pronunciation identical to that of the sound radical), 请 (*qing*₃, with a pronunciation identical to the sound radical, except for the tone), and 情 (*qian*₄, with only partial overlap in pronunciation with the sound radical). Because of the varying degrees of resemblance between a phonogram and its sound radical, some researchers have attempted to categorize phonograms as regular ones (with total resemblance) and irregular ones (with partial resemblance; Lee, Tsai, Su, Tzeng, & Hung, 2005; Lien, 1985). One problem with such a classification is that it is based on linguists' expert intuition and may not capture the relationship between a phonogram and its sound radical as perceived by ordinary speakers. For instance, the character 霸 (*ba*₄, containing the same-sounding radical 霸) is a regular phonogram, but people would rarely make use of the sound radical, since it is a rare character in itself. To more directly measure the phonology–orthography consistency or spelling regularity of Chinese characters, in the present study we obtained subjective ratings of the extent to which a character contains a sound radical that provides cues to the character's pronunciation.

In Chinese, a character is a syllable in pronunciation; hence, a character will often have different degrees of homophone density (i.e., the number of characters that are homophones). Homophone density has been shown to impact visual character recognition (Pexman, Lupker, & Jared, 2001) and handwriting (C. Wang & Zhang, 2015), with people being slower at recognizing or handwriting a character/word with more homophones. Importantly, it has also been suggested that if a syllable has more corresponding characters/radicals,

there will be competition among these possible candidates for selection when orthographic codes are being retrieved (C. Wang & Zhang, 2015).

Finally, there are also factors related to orthographic complexity. Chinese characters consist of one or more radicals, and these radicals are further made up of strokes. The number of strokes and the number of radicals have also been shown to be important determinants in Chinese character naming (Leck, Weekes, & Chen, 1995; Leong, Cheng, & Mulcahy, 1987; Peng & Wang, 1997; Zhou, Shu, Bi, & Shi, 1999) and to affect handwriting (Q. Zhang & Feng, 2017). In addition, to write a Chinese character, one needs to arrange the radicals in a particular spatial composition (and in a particular temporal order). The most common composition is the left–right composition, followed by the top–down composition. Left–right composition involves two radicals arranged horizontally (e.g., 好 is composed of 女 on the left and 子 on the right), whereas top–down composition involves two radicals arranged vertically (e.g., 岩 is composed of 山 on top of 石). Whether character composition impacts handwriting (e.g., orthographic access) remains an issue to be explored.

Below we report a study in which we collected handwritten forms of 1,600 characters from 203 participants (each contributing 200 handwritten characters). In the experiment, we opted to use the spelling-to-dictation task, which has been used to investigate the cognitive processes underlying handwriting (see Bonin, Méot, Lagarrigue, & Roux, 2015; Bonin, Peereman, & Fayol, 2001). The choice of such a task was also due to practical considerations: Note that not all the characters (1,600 of them) have a corresponding depictable object/event (as would be required for a task of written picture naming; e.g., C. Wang & Zhang, 2015); in addition, a task such as character copying (e.g., Kandel & Perret, 2015) would not allow for the investigation of the potential role of phonology in handwriting.

In the spelling-to-dictation task, apart from writing accuracy, we also collected writing latencies (the time between dictation and handwriting onset) and writing durations (from handwriting onset to offset). Latencies in speaking and handwriting have been used in prior studies to disentangle the central processes of accessing and transcoding linguistic information and the peripheral processes of executing encoded linguistic information into speech sounds and written symbols (Damian, 2003; Damian & Dumay, 2007, 2009; Schriefers, de Ruiter, & Steigerwald, 1999; Schriefers & Teruel, 1999); we will return to this point later in the Discussion section.

Method

Participants

A total of 205 participants took part in the study for monetary reward. They were undergraduate or postgraduate students

from South China Normal University and neighboring universities. Self-reports indicated that 203 of the participants used pinyin (alphabetic spelling) as the input method in typing, and the remaining two used wubi (a radical-based input method that is now rarely used). Considering that input methods in typing can have an effect on the association among semantics, phonology, and orthography (J. Zhang & Li, 2010), we excluded the two wubi-method participants, leaving 203 participants (151 females, 52 males; age range = 17–26 years, with a mean of 20.02). All participants reported being right-handed, having normal hearing, and having normal or corrected-to-normal vision.

Materials

We selected a total of 1,600 simplified Chinese characters as the experimental materials, according to the following criteria. First, on the basis of SUBTLEX-CH, a corpus of Chinese character/word frequencies derived from film subtitles (Q. Cai & Brysbaert, 2010), we selected only characters with a log frequency between 1.5 and 5.0 (i.e., between 32 and 100,000 counts in the whole corpus) and a context frequency higher than .20 (i.e., a character should appear in 20% of the films in the corpus). This led to the exclusion of characters that were too rare (e.g., 寮, *liao*₂) or too frequent (e.g., 人, *ren*₂). Second, the experimental characters should have more than four strokes; this criterion was adopted to exclude characters that were too simple (e.g., 二, *er*₂). These two criteria resulted in the selection of 2,095 characters from the SUBTLEX-CH corpus. We then selected the 2,095 most frequent bicharacter word in SUBTLEX-CH that contains one of the 2,095 characters, and had another 15 participants (from the same population as the participants for the main experiment) rate the familiarity of these bicharacter context words (one on the scale of 1–7, with 7 meaning *most familiar*). We then excluded any character whose hosting bicharacter word was rated less than 4 on average in terms of familiarity. Each of the remaining characters was composed into a phrase that clearly indicated the target character in the context of the bicharacter word (e.g., 灶台的灶, meaning the character 灶 as in the word 灶台). These phrases were then recorded using the text-to-speech reader Langdunv (version 7.6; <http://www.443w.com/tts>). We then selected the 1,600 phrases (i.e., 1,600 experimental characters) that had the best recording quality.

Apparatus and procedure

The experiment was carried out in E-Prime 2.0 run on a desktop computer. The writing apparatus included an IntuosA4 graphic tablet (linked to E-Prime), an inking digitizer pen (WACOM, Japan), and writing sheets (each with 50 squares in a 5×10 grid). Participants were tested individually in a quiet cubicle and were seated about 60 cm from the computer

screen. Participants wrote ten characters in a practice session and then, in the main experiment, wrote 200 dictated characters, selected randomly from the 1,600-character cohort. Each trial, as is illustrated in Fig. 1, began with a cue sound “ding” for 500 ms, followed by a blank interval of 500 ms and then a spoken phrase specifying the to-be-written character (e.g., 灶台的灶, meaning the character 灶 in the word 灶台). At the offset of the spoken phrase, participants could begin to write down the character on one of the 50 squares on the writing sheet (in a left-to-right and top-to-bottom order) placed on top of the pen tablet. Once the inking pen touched the tablet, an instruction appeared onscreen telling participants to press the ENTER key upon finishing writing.¹ When the ENTER key was pressed, the target character was shown onscreen for 1,500 ms, followed by a screen asking participants to press a number on the number pad to indicate their writing accuracy (0 = *correct writing*, 1 = *did not know what character was supposed to write*, 2 = *knew what character was supposed to write but forgot how to write it*, 3 = *accidentally wrote the character incorrectly*, and 4 = *had always had the wrong orthography for that character*). The response was followed by a 1-s intertrial interval. If participants did not initiate writing within 30 s after the offset of the spoken phrase, the experimental program automatically jumped to the “Write” screen (see Fig. 1). If participants did not finish writing within 30 s after the pen touched the tablet, the “Target character” screen would be shown. If participants did not choose a number to indicate their accuracy within 3 s on the “Report” screen, the intertrial interval would follow, and then the next trial would commence. We collected three behavioral measures. The first was writing latency, measured from the offset of the spoken phrase to the time that the pen touched the writing sheet. The second measure was writing duration, measured from the pen touch to the press of the ENTER key (which signaled the completion of handwriting). The third measure was writing accuracy, as self-reported by the participants.² The whole experiment lasted for about 35 min.

Lexical variables

Character frequency Two frequency measures were obtained for each of the 1,600 characters using the SUBTLEX-CH corpus, which contains about 46.8 million character counts

¹ Although it was desirable to automatically measure the offset of handwriting, pilot work showed that the device was not reliable in measuring the offset when used together with a writing sheet on top; hence, we chose to have participants press the ENTER key to signal the end of writing. Although this might add some additional time to the writing duration, this added time should be constant and independent of the lexical effects that we were looking at.

² In the present article, we only report analyses on accuracy in general (i.e., whether a character was reported to be correctly written). It should be noted that our database also contains information about why a character was incorrectly written (e.g., accidentally miswriting, incorrect orthographic representation, or tip of the pen; see the [Apparatus and Procedure](#) section).

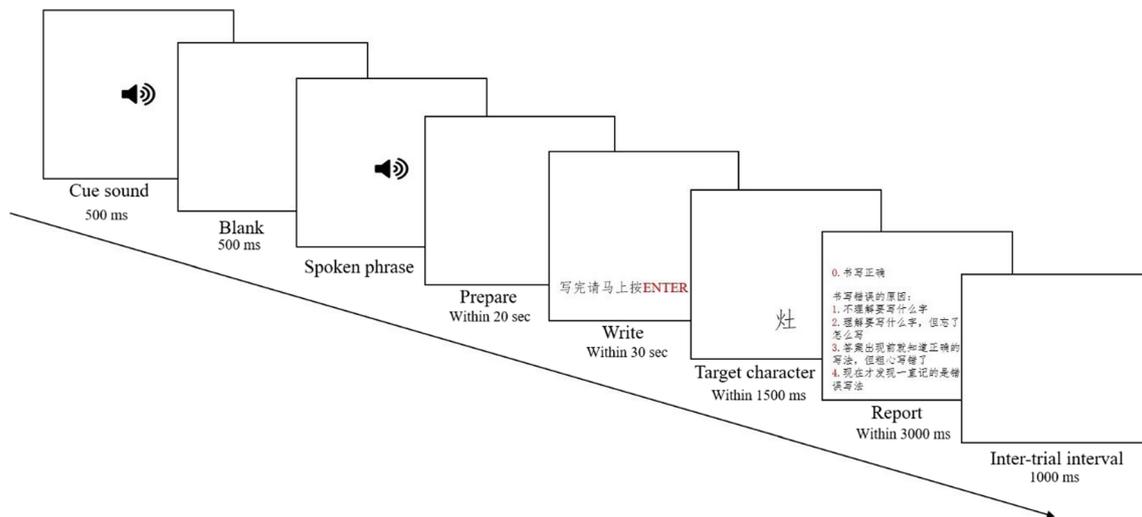


Fig. 1 Schematic illustration of the trial structure. On the “Write” screen, the instruction “Please press ENTER upon finishing writing” was presented; on the “Target character” screen, the target character was shown; on the “Report” screen, participants were asked to press a

number on the number pad indicating whether their handwriting was accurate (i.e., 0) or not (1, 2, 3, or 4, depending on the reason for the inaccuracy; see the text for more details)

based on film subtitles (Q. Cai & Brysbaert, 2010). As we described in the **Materials** section, the *count frequency* refers to the log of a character’s counts in the corpus, whereas the *context frequency* refers to the number of films in which a character appears. Note that these two measures are highly correlated, and to avoid collinearity, we only used the former frequency measure in our regression analyses (see the **Results** section).

AoA Characters in Chinese are formally taught and learned in school, and Chinese textbooks provide lists of characters that are learned in each of the 12 semesters in primary school in China. Following Shu, Chen, Anderson, Wu, and Xuan (2003), we used these lists as a measure of character AoA. Because Chinese children start school at age 6, a character learned in the first semester would mean that it is acquired at age 6–6.5, and a character learned in semester 12 would be formally acquired at age 11.5–12 (in our study, we used the upper limit of these estimates as the AoA; e.g., 6.5 and 12, in the examples). Shu et al. compiled the AoA information for 2,570 characters based on six primary school Chinese textbooks used in the period around 2003 in areas such as Beijing, Jiangsu, and Guangdong (Hua Shu, personal communication). Note that the majority of our participants (undergraduate students) started primary school around the year 2003 so that the AoA corpus in Shu et al. (2003) provided accurate AoA information for our participants. In Shu et al.’s corpus we found AoA information for 1,255 of the 1,600 target characters. We then found two more online sources of textbook lists of characters, compiled by primary school teachers (<https://wenku.baidu.com/view/7c0abf1ce53a580216fcfeb0.html?from=search>; http://www.sohu.com/a/62481121_101008), which helped us find AoAs

for another 223 characters, leaving 122 characters without AoAs from primary school textbooks. Since these are mostly rare characters that are not taught in primary schools, we assumed that these characters are acquired after primary school (and hence assigned an AoA of 12.5 to these characters).

Number of meanings The *Dictionary of Chinese Character Information* (Chinese Character Encoding Group of Shanghai Jiao Tong University & Chinese Pinyin Characters Research Group in Shanghai, 1988) used the *Xinhua Dictionary* (新华字典) to calculate the number of meanings a character has. We followed this practice by resorting to the newest *Xinhua Dictionary* (11th edition, Linguistics Institute of the Chinese Academy of Social Sciences, 2011).

Imageability and concreteness Another 60 participants (from the same population as those in the main experiment) each rated the imageability of 400 characters randomly sampled from the cohort of 1,600, and the concreteness of another 400 randomly sampled characters (the order of the two tasks was counterbalanced across participants). In each trial, they were presented a character and provided a rating on the imageability/concreteness of its meaning(s) on a 7-point scale (with 7 indicating *most imageable/concrete*). Ratings were individually normalized (i.e. the ratings for each participant were transformed into *z* scores).

Phonograms and sound radical order We used the *Dictionary of Modern Chinese Phonograms* (现代汉字形声字字汇; Ni, 1982) to determine whether or not a character is (historically) a phonogram. In addition, we categorized a character as sound-first or not in accordance with the writing order

of the composing radicals; the nonphonograms, in which there is no sound radical, were all categorized as not-sound-first.

Spelling regularity As we reviewed above, a phonogram does not necessarily have a sound radical that is identical or similar to the character in pronunciation; therefore, we needed a separate measure to indicate a character's spelling regularity (or phonology–orthography consistency). Another 53 participants from the same population as the main experiment participants each rated 400 characters randomly selected from the 1,600 target characters. A trial started with a written character; participants could press a key to listen to its pronunciation, if needed. Then they rated the extent to which the character contained a sound radical that indicated the pronunciation of the character (0 = *not at all*, 7 = *containing an identically sounding radical*). Again, ratings were individually normalized.

Homophone density This variable reflects the number of characters in the SUBTLEX-CH corpus that are homophonous with a target character. We first determined the pronunciations of all the characters in the SUBTLEX-CH corpus using the *Xinhua Dictionary* (Linguistics Institute of the Chinese Academy of Social Sciences, 2011). If a character has multiple pronunciations, homophonous characters of all the pronunciations were all included. Then we calculated the number of characters that had the same pronunciation (i.e., identical syllable and identical tone) with a target character. We then took the log of the number of homophonous characters as the measure of homophone density.

Number of strokes and number of radicals Simplified Chinese characters are composed of five basic strokes: 一, |, 丿, 丶, and 冫, according to the *Dictionary of Common Chinese Characters in Print* (印刷通用汉字字形表; Chinese Ministry of Culture & State Language Affairs Commission, 1986) and the *Modern Dictionary of Common Characters in Chinese* (现代汉语通用字表; Chinese Ministry of Culture & State Language Affairs Commission, 1988). The number of strokes was taken from the above two dictionaries. The number of radicals is a bit hard to operationalize, as it would depend on what counts as a radical. For instance, the character 横 is said to have two radicals (i.e., 木 and 黄) in the *Dictionary of Chinese Character Information* (汉字信息字典; Chinese Character Encoding Group of Shanghai Jiao Tong University & Chinese Pinyin Characters Research Group in Shanghai, 1988), but to have four (i.e., 木, 艹, 由, and 八) in the *Dictionary of Chinese Character Properties* (汉字属性字典; Fu, 1989). Since there seems to be inconsistency in the *Dictionary of Chinese Character Information* (e.g., while it defines 横 as having two radicals, it defines 黄, which is part of 横, as having three radicals), we decided to use the radical counts given by *Dictionary of Chinese Character Properties*.

Character composition We used the *Dictionary of Chinese Character Properties* to determine the character composition of the selected characters, though we made some simplifications. In particular, a character was coded as having a left–right composition if its radicals are horizontally arranged (there were 943 such characters in our 1,600-character cohort), or a top–down composition if the radicals are vertically arranged (429 such characters in our cohort); otherwise, a character was coded as having an “other” composition (228 such characters in our cohort). Because character composition is a categorical variable, in our analyses below, it was dummy-coded with two variables: “Left–right” refers to whether a character has a left–right composition (coded as 1) or not (coded as 0), and “top–down” refers to whether a character has a top–down composition (coded as 0) or not (coded as 1).

Context word familiarity Because each character was embedded in a two-character spoken word (see the [Materials](#) section), we also measured participants' familiarity with these words by having 15 participants from the same population as those in the main experiment rate their familiarity with the word that a character was embedded in on 7-point scale, with 7 meaning *most familiar*.

Results

An experimenter manually examined the accuracy of the reports (i.e., whether the writing was accurate, and if not, what was the cause of the inaccuracy) by 12 participants (6% of our 203 participants). Comparison between experimenter-reported accuracy and self-reported accuracy showed that six of the participants were 100% reliable in their self-reports; all of the remaining six participants differed from the experimenter in their accuracy reports in fewer than 1% of the trials. Such a result suggests that the participant self-reports were highly reliable (which was not surprising, given that they reported the accuracy of the handwriting upon seeing the target character).

We compiled the handwriting responses together with the lexical characteristics into a database, which (together with the analysis script) is available at Open Science Framework (<https://osf.io/7s9kq/>). Writing latencies and durations were based on trials in which the character was reported to be correctly written. Because one character (孽) was never written correctly, it did not have a latency or duration, and was hence excluded from the latency and duration analyses below (but included in the accuracy analysis). Any writing latencies longer than 10 s were removed as outliers; any writing durations shorter than 1 s or longer than 10 s were also removed as outliers. These trimming steps led to the exclusion of 0.32% of the writing latency data and 0.93% of the writing duration data. To minimize individual differences in timing, we then transformed each participant's writing

latency and duration into z scores. Then we calculated the average z score for the latency and duration of each character. A handwritten character was categorized as accurate if the participant reported it as accurate after seeing the target character; otherwise, it was categorized as inaccurate. The accuracy for each character was the proportion of correct trials out of all trials for that character.

Table 1 presents descriptive results for the behavioral measures and the lexical variables, and Table 2 presents the correlations among these variables. From Table 2, we can see that (1) accuracy is correlated with all the character-level variables except concreteness and sound radical order (SRO); (2) writing latency is correlated with all the lexical variables except SRO and whether or not a character has a top-down composition (top-down); (3) writing duration is correlated with all the lexical variables except homophone density (HomoDen) and whether or not a character has a top-down composition (top-down).

Predictive power of the lexical variables

Since there were correlations among the predictors (Table 2), before conducting regression analyses, we first looked for

Table 1 Descriptive statistics for all variables

	<i>N</i>	Mean	Min	Max	<i>SD</i>
ACC	1,600	.874	.000	1.000	.185
Latency	1,599	0.093	− 0.810	4.010	0.603
Duration	1,599	0.058	− 1.360	2.840	0.732
FreqCHR	1,600	3.358	1.510	5.000	0.727
FreqCTX	1,600	31.359	0.210	99.840	29.114
AoA	1,600	9.075	6.500	12.500	1.724
Meanings	1,600	3.096	1.000	16.000	2.047
Imageability	1,600	− 0.002	− 1.910	1.370	0.593
Concreteness	1,600	0.001	− 1.690	1.560	0.687
Phonogram	1,600	.753	.000	1.000	.431
SRO	1,600	.184	.000	1.000	.387
Regularity	1,600	− 0.001	− 1.300	1.720	0.850
HomoDen	1,600	0.877	0.000	1.720	0.347
Strokes	1,600	9.843	5.000	23.000	3.072
Radicals	1,600	2.995	1.000	7.000	1.027
Left–right	1,600	.589	.000	1.000	.492
Top-down	1,600	.732	.000	1.000	.443
Familiarity	1,600	0.273	− 1.520	1.210	0.388

Latency and duration respectively refer to averages of the individually normalized latencies and durations across participants. FreqCHR (character frequency) refers to the log of character counts, and FreqCTX (contextual frequency) refers to the proportion of contexts (films) that a character appears in in SUBTLEX-CH. Imageability, concreteness, regularity (spelling regularity), and familiarity (context word familiarity) were based on ratings that were individually normalized. HomoDen (homophone density) refers to the log of the number of homophonous characters in SUBTLEX-CH

collinearity issues using a stepwise variance inflation factor (VIF) selection procedure with a wrapper of the vif function in the fmsb R package (for the stepwise VIF selection procedure, see <https://www.r-bloggers.com/collinearity-and-stepwise-vif-selection/>; see also our analysis script). Setting the VIF threshold at 5, we found that having both count frequency and context frequency (which were highly correlated) led to collinearity; to address this, we removed context frequency. Further stepwise VIF selection revealed no collinearity among the remaining predictors. Hence, in the subsequent regression analyses, we included the following predictors: count frequency, AoA, number of meanings, imageability, concreteness, phonogram, sound radical order, spelling regularity, homophone density, number of strokes, number of radicals, left–right composition, top-down composition (note that these two composition dummy variables were enough to capture the three-level variable of radical composition), and context word familiarity. These variables were simultaneously entered into regression models to examine their predictive power for the handwriting measures.

Writing latency The results of the regression model ($R^2 = .524$, adjusted $R^2 = .520$) are presented in Table 3. Assuming that writing latency mainly reflects access to orthographic codes, the results show that people spend a longer time accessing the orthography of a character if the character is less frequent or learned later, as expected. Phonology impacts the time needed for orthographic access: Writing latencies are shorter for characters with higher spelling regularity or with fewer homophones, and also for nonphonograms than for phonograms. Orthographic complexity predictors also have significant effects: Writing latencies are shorter for characters with fewer strokes, and also shorter for left–right characters than for other composition types; there is a marginally significant effect of top-down composition, suggesting that people seem to spend longer times accessing orthographic codes for top-down characters than for characters of other types. Finally, people spend less time preparing writing when the context word (in which the character is embedded) has higher familiarity. Interestingly, semantic predictors did not seem to impact writing latencies.

Writing duration The results of the regression model ($R^2 = .837$, adjusted $R^2 = .835$) are presented in Table 3. People are faster writing a character if it is more frequent, is acquired earlier, or has fewer strokes. In addition, they are also faster at writing a character if it has the left–right composition (but slower if a character has the top-down composition) or if the character is embedded in a more familiar word. There are no semantic or phonological effects on writing duration.

Accuracy The results of the regression model ($R^2 = .431$, adjusted $R^2 = .426$) are again presented in Table 3. People are

Table 2 Correlations of all variables ($N = 1,600$ for all but latency and duration, where $N = 1,599$)

	ACC	Latency	Duration	FreqCHR	FreqCTX	AoA	Meanings	Image-ability	Con-creteness	Phono-gram
ACC	1									
Latency	-.687**	1								
Duration	-.391**	.419**	1							
FreqCHR	.556**	-.618**	-.396**	1						
FreqCTX	.432**	-.520**	-.377**	.921**	1					
AoA	-.448**	.512**	.333**	-.534**	-.492**	1				
Meanings	.272**	-.301**	-.281**	.494**	.510**	-.319**	1			
Imageability	.102**	-.137**	.092**	-.039**	-.114**	-.177**	-.088**	1		
Concreteness	.011	-.077**	.112**	-.125**	-.188**	-.143**	-.124**	.804**	1	
Phonogram	-.163**	.185**	.331**	-.253**	-.248**	.209**	-.234**	.039**	.061**	1
SRO	.003	.036	.106**	.039	.014	.016	.016	-.051*	-.096**	.272**
Regularity	-.123**	.115**	.228**	-.257**	-.252**	.166**	-.271**	.052*	.102**	.509**
HomoDen	-.103**	.110**	-.007	-.073**	-.077**	.075**	-.097**	-.196**	-.135**	.014
Strokes	-.258**	.270**	.899**	-.291**	-.280**	.236**	-.223**	.104**	.116**	.337**
Radicals	-.178**	.211**	.649**	-.248**	-.245**	.183**	-.179**	.082**	.077**	.323**
Left-right	-.085**	.069**	.099**	-.158**	-.149**	.116**	-.135**	.019	.027	.394**
Top-down	-.060**	.045**	.026	-.076**	-.067**	.064**	-.081**	-.025	-.012	.269**
Familiarity	.560**	-.606**	-.269**	.627**	.533**	-.419**	.275**	.059*	-.022	-.135**
SRO		Regu-larity	Homo-Den	Strokes	Radicals	Top-Down	Left-Right	Familiarity		
Latency										
Duration										
FreqCHR										
FreqCTX										
AoA										
Meanings										
Imageability										
Concreteness										
Phonogram										
SRO	1									
Regularity	-.020	1								
HomoDen	.032	.099**	1							
Strokes	.110**	.220**	-.023	1						
Radicals	.125**	.217**	.004	.728**	1					
Left-right	-.231**	.305**	-.018	.110**	.128**	1				
Top-down	-.205**	.221**	-.014	.007	-.021	.725**	1			
Familiarity	.054*	-.170**	-.019	-.167**	-.142**	-.118**	-.067**	1		

** Correlation is significant at the .01 level (two-tailed); * Correlation is significant at the .05 level (two-tailed)

Table 3 Results of regressions on writing latency, writing duration, and accuracy

	Writing latency			Writing duration			Accuracy		
	<i>B</i>	<i>t</i>	<i>p</i>	β	<i>t</i>	<i>p</i>	β	<i>t</i>	<i>p</i>
Intercept	.093	8.945	< .001	.059	7.980	< .001	.874	248.805	< .001
Frequency	-.186	-11.503	< .001	-.054	-4.678	< .001	.046	8.398	< .001
AoA	.103	7.849	< .001	.053	5.701	< .001	-.027	-6.076	< .001
Meanings	-.001	-0.089	.929	-.013	-1.517	.129	.000	-0.084	.933
Imageability	-.030	-1.674	.094	.014	1.069	.285	.021	3.412	.001
Concreteness	-.032	-1.788	.074	-.002	-0.136	.892	-.012	-1.912	.056
Phonogram	.029	1.976	.048	.006	0.574	.566	-.005	-1.052	.293
SRO	.014	1.169	.243	.012	1.385	.166	.000	-0.040	.968
Regularity	-.052	-4.059	< .001	-.002	-0.205	.838	.011	2.553	.011
HomoDen	.036	3.369	.001	.002	0.292	.770	-.012	-3.203	.001
Strokes	.066	4.190	.000	.631	56.272	< .001	-.027	-5.078	< .001
Radicals	-.002	-0.160	.873	-.011	-1.015	.310	.009	1.684	.092
Left-right	-.041	-2.470	.014	-.041	-3.458	.001	.008	1.366	.172
Top-down	.027	1.733	.083	.032	2.948	.003	-.008	-1.516	.130
Familiarity	-.202	-14.827	< .001	-.039	-4.007	< .001	.060	13.172	< .001

more accurate at writing a character if it is more frequent, is acquired earlier, or has fewer strokes. There are also semantic effects: Accuracy increases as a function of imageability but, interestingly, decreases as a function of concreteness. Phonological effects can also be seen: A character is more often correctly written if it has a higher spelling regularity or lower homophone density. A character is also correctly written more often if it is embedded in a more familiar word.

Relative importance of the predictors Although the regression results reveal whether and how each predictor contributes to handwriting, they do not allow us to assess the relative contribution of a predictor as compared to other predictors. Although one can compute the variance explained (R^2) by each predictor, such a metric is not straightforward to calculate when the predictors in the model are correlated (as in the present study); for instance, a predictor's R^2 can change depending on whether the predictor is added prior to or after a correlated predictor. To address this issue, researchers have proposed the *lmg* metric, which can be seen as the relative importance of a predictor, based on its R^2 when its ordering is taken into consideration (Grömping, 2006; Johnson & Lebreton, 2004). We used the R package *relaimpo* to calculate *lmg*. As is shown in Fig. 2, for writing latency (as a measure of orthographic access) and accuracy, the most important character variable is the context in which a character is used (i.e., context word familiarity), followed by its frequency, then by its AoA, and then by its number of strokes and number of meanings. Interestingly, the above predictors did not contribute substantially to writing duration, except for the number of strokes, which explained more 60% of the variance; the

number of radicals is also a key determinant of writing duration (though we did not observe a significant effect for this predictor in the regression analysis, probably because of its high correlation with the number of strokes). These results suggest that writing duration mainly reflects the execution of motor programs to carry out handwriting. Phonological predictors also play some role (albeit with relatively small contributions); in particular, homophone density and spelling regularity impact handwriting latency and accuracy.

Discussion

The effects of character frequency and context word familiarity

We observed that two variables, as expected, have large effects on all measures of handwriting: character frequency and the familiarity of the context word in which a character appears. People spend less time preparing to write and actually writing a more frequent character, and they are also more accurate at writing it. These frequency effects are consistent with previous findings in handwriting (e.g., Kandel & Perret, 2015; Qu et al., 2016; C. Wang & Zhang, 2015), and indeed also in spoken word/character production (e.g., Balota & Ferraro, 1993; Chang et al., 2016; Jescheniak & Levelt, 1994; Lee et al., 2005). A similar advantage in writing latency, writing duration, and accuracy is also observed when a character is embedded in a more familiar word. This finding is consistent with previous observations that word processing

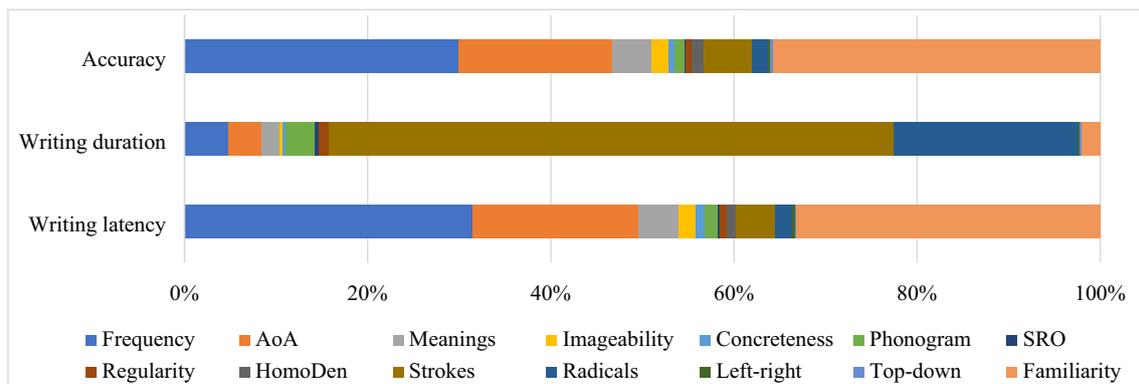


Fig. 2 Relative contributions (lmg) of the different predictors to the handwriting measures

is constrained by the context in which it appears (see Z. G. Cai & Vigliocco, 2018, for a review).

The effect of age of acquisition (AoA)

Our results show that AoA has an effect on all measures of handwriting (especially on latency and accuracy). These results are consistent with the finding in Bonin, Fayol, & Chalard (2001), who showed that people are quicker at accessing orthographic codes for early-acquired characters; in addition, the results further indicate that people are also quicker and more accurate at writing them. It is worth asking what caused the AoA effects, even after we controlled for correlated variables such as character frequency and stroke number (early-acquired characters are more frequent and contain fewer strokes; see Table 2). One possibility is that the AoA effects are cumulative-frequency effects in disguise (Lewis, 1999; Lewis, Gerhand, & Ellis, 2001). That is, early-acquired characters, compared to late-acquired ones, have higher cumulative frequencies in both reading and writing over one's life. Note that these cumulative frequencies are unlikely to be fully captured by corpus counts, which are proxies for a character's synchronic frequency of use rather than its cumulative frequency over time. An alternative account is that AoA effects reflect plasticity in the language-learning system, such that early-acquired characters are more entrenched in their neural representations (Brysbaert & Ghyselinck, 2006; Ellis & Lambon Ralph, 2000; for a review, see Ghyselinck, Lewis, & Brysbaert, 2004); in other words, the language system is more plastic in earlier than in later years, and thus early-acquired characters, as compared to late-acquired ones, have the benefit of more stable or acute representations, due to the brain being more plastic at an early age. Further research will be needed to test this hypothesis by controlling for the cumulative frequency of words/characters (e.g., Perez, 2007).

The effects of semantic predictors

We did not observe any impact of a character's number of meanings on any measure, which is inconsistent with the observation that speakers are quicker at naming words with more meanings (Chang et al., 2016; Woollams, 2005). Imageability and concreteness (differentially) affected writing accuracy but not writing latency and duration in our study, which is again not entirely consistent with prior findings that these semantic variables affect the speed with which lexical information is accessed (Chen & Peng, 1998; Cortese et al., 1997; Liu et al., 2007; but see Barca et al., 2002, for a lack of imageability and concreteness effects on speaking latencies). We believe these inconsistencies may be due to task idiosyncrasies. In particular, in our study we used a phrase to specify the target character (e.g., 辣椒的辣: *la₄ in la₄jiao₁*). Therefore, it is very likely that the lack of a meaning quantity effect in our task could have been due to the strong constraint that the context word placed on the semantic interpretation of a character; that is, even if a character may have different meanings in isolation, only the consistent meaning would be accessed, due to the word context (e.g., Sereno, Brewer, & O'Donnell, 2003). A similar explanation can also apply to the lack of latency and duration effects for imageability and concreteness.

Another finding worth noting is that, whereas concreteness and accuracy had a (nonsignificant) positive correlation (see Table 2), concreteness actually negatively predicted accuracy when entered in conjunction with other predictors. We note that more recent studies have cast doubt on the role of concreteness in word processing (e.g., Bonin, Méot, & Bugajska, 2018; Guasch, Ferré, & Fraga, 2016; Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011); thus, further research will be needed to investigate whether and how concreteness might affect handwriting.

The role of phonology in handwriting

We observed a homophone density effect in orthographic access, with people having longer latencies to start writing if a

character had more homophones. This finding sheds light on the debate on whether homophone density affects access to orthographic codes. An earlier finding by Bonin, Peereman, and Fayol (2001) showed no difference between the handwriting of words with one or two homophones. However, a more recent study by C. Wang and Zhang (2015) showed that people were slower at accessing orthographic codes when asked to write Chinese characters with more homophones. Our results agree with the latter finding and suggest that different orthographic representations can be co-activated by a pronunciation, leading to competition in orthographic access. Such a conclusion is also in line with the inhibitory effect of homophone density in word recognition (e.g., Pexman et al., 2001).

Access to orthographic codes was also influenced by spelling regularity in our study: People are quicker at initiating writing if a character has a more indicative sound radical. This finding is consistent with some previous studies. For instance, Qu and colleagues (Qu et al., 2011; Qu et al., 2016) showed that a masked prime that is phonologically related to a subsequent target character can facilitate handwriting of the target character; Kandel and Perret (2015) showed longer handwriting latencies (for accessing orthographic codes) for words that are irregular rather than regular in spelling. These findings suggest that people can make use of phonological cues to facilitate access to the target orthographic representations. It should be noted that C. Wang and Zhang (2015) found an inhibitory rather than facilitative effect of spelling regularity; however, as these authors argued, it is possible that such inhibition was due to homophone density rather than spelling regularity. Further research will be needed to establish how spelling regularity facilitation and homophone interference may interact in orthographic access.

It is interesting to note that although homophone density and spelling regularity impacted orthographic access (as reflected in writing latencies) and accuracy, they did not have any effect on writing duration. Writing duration is likely to mainly reflect motor execution and not orthographic access in writing—hence, the lack of phonology effects. Alternatively, phonology comes into play at the initial stage of handwriting; therefore, its effects are reflected in writing latency but not in writing duration. Note that, consistent with the second account, it has been found that phonology mediates orthographic access only at an early stage (e.g., Q. Zhang & Damian, 2010).

The effects of orthographic complexity

Another all-around effect is that of a character's orthographic complexity (especially its stroke number): People are slower at accessing the orthographic codes and in the actual handwriting, and are also more error-prone, when a character has more strokes. Such a finding speaks to conflicting results from small-sample lab-based studies concerning the effect of a

character's number of strokes on orthographic access. Whereas some studies have provided positive evidence that handwriting latencies increase as a function of stroke number (Leong et al., 1987; Peng & Wang, 1997), others have failed to find such an effect (Su & Samuels, 2010). Using much larger participant and stimulus samples than previous studies, the present database allows for better control of other lexical variables when assessing the possible effect of stroke number. These results revealed longer latencies for characters with more strokes. Such a result is also consistent with the ERP finding that characters with more strokes elicited a larger P200 and a smaller N200 in a delayed character-matching task (Yang, Zhang, & Wang, 2016).

We did not find a significant effect of radical number when stroke number was added as a covariate. However, we caution against using such a null finding to argue that the number of radicals does not impact orthographic access. It should be noted that in our database there is (unavoidably) a high correlation ($r = .728$; see Table 2) between a character's stroke and radical numbers; thus, our finding only suggests that a character's radical number does not seem to explain variance in the handwriting data beyond what is explained by stroke number. Please also note that it has been shown that, at least in character recognition, characters with more incomplete radicals elicited a larger N200 and a larger N400, suggesting the possibility that radicals serve as an intermediate or subcharacter representation (Q. Wang & Dong, 2013).

Finally, the results also showed an effect of a character's composition type. People need less time to access orthographic codes and to actually handwrite left–right characters (as compared to characters of other compositions); in contrast, people need more time to access orthographic codes and to handwrite top–down characters. We suspect that the composition effects on latency and duration may reflect the habit of writing, in that people are more used to composing radicals in a left–right than in a top–down fashion.

Central and peripheral processes in handwriting

As we briefly mentioned above, prior work on language production (speaking or handwriting) has made use of speaking/writing latencies and durations in an attempt to disentangle the central processes of lexical access and the peripheral processes of executing speaking and handwriting. There is much evidence from both speaking and writing that latencies (the time it takes to prepare speaking/writing) reflect central processes of accessing linguistic information, and hence should be sensitive to manipulations that modulate linguistic access. Such a hypothesis has been confirmed in quite a few studies (Damian, 2003; Damian & Dumay, 2007, 2009; Schriefers et al., 1999; Schriefers & Teruel, 1999). For instance, Damian (2003) showed that, across three different experimental

paradigms, speaking latencies (but not speaking durations) in single-word production were sensitive to semantic and form relatedness factors that arguably play a role in the central process (e.g., planning), but not in the peripheral processes of motor programming and execution, during speaking. Damian and Dumay (2007) also showed that the effect of phonological priming in a picture–word interference task was reflected in speaking latencies, further suggesting that speaking latencies reflect the central processes of speaking (e.g., planning; see also Schriefers et al., 1999). In handwriting, Damian and Stadthagen-Gonzalez (2009) found that phonological relatedness impacted writing latencies but not writing durations in the handwriting of single words and short phrases.

Whether speaking/writing durations also reflect the central processes of speaking/writing remains an issue of debate. In contrast to Damian (2003), some studies have shown that effects on linguistic access can show up in speaking duration, especially when there was a variable deadline to speak (Ferreira & Swets, 2002) or time pressure (Kello, Plaut, & MacWhinney, 2000). More critically, in handwriting, orthographic regularity and word frequency impacted writing latencies, but such effects also cascade into movement production as a result of increasing the writing durations (Bloemsaat, Van Galen, & Meulenbroek, 2003; Delattre, Bonin, & Barry, 2006; Kandel & Perret, 2015; Roux, McKeef, Grosjacques, Afonso, & Kandel, 2013). More recently, Q. Zhang and Feng (2017) investigated whether effects on the central processes of handwriting spill over to the actual handwriting. They found that both lexicality (whether or not an experimental stimulus was a character) and radical complexity (number of strokes) affected the actual execution of handwriting.

In line with previous studies (e.g., Damian & Dumay, 2007; Damian & Stadthagen-Gonzalez, 2009), our results revealed that handwriting latencies are very sensitive to lexical variables (e.g., frequency, semantic factors, phonological factors, and word context) that arguably only impact the central processes of lexical/orthographic access and planning but not the motor processes of handwriting execution. However, it should be also pointed out that some of these effects, especially those of dominant variables such as frequency and AoA, may extend from writing preparation (as reflected in writing latencies) to writing execution (as reflected in writing durations). This is especially likely considering that, compared to speaking, handwriting is time-consuming and not as automatic. For instance, after accessing the orthographic codes of the first radical of a character, people may initiate handwriting while continuing to access the orthographic codes of subsequent radicals. Such an account would then predict that, whereas some characteristics (e.g., complexity, frequency, AoA) of a temporally early (e.g., the first) radical's characteristics would mainly impact writing latencies, those of temporally later radicals (e.g., the second or third one) would be more likely to manifest their effects in

writing durations. It is also interesting to consider the account we discussed above concerning the lack of phonological effects on writing durations; that is, phonological information is used early in orthographic access (e.g., to determine the sound radical) but is not consulted at later stages (see also Q. Zhang & Damian, 2010, for a similar account). However, it is important to note that, in our present task, the dichotomy between writing latencies and writing durations does not strictly equate to the distinction between central and peripheral processes in handwriting, and further research will be needed to investigate the time course of these processes.

Possible use of the database and possible further extension

Although our findings have shed light on the current issues of handwriting, we believe the database can be used for further analyses of handwriting. It can be used to conduct secondary analyses that may shed light, for instance, on the interaction of lexical variables in handwriting. For instance, frequency has been shown to interact with spelling regularity in Chinese word/character naming (H. Cai et al., 2012; Chang et al., 2016; Lee et al., 2005). The database can further be used to test whether this finding generalizes to a much larger sample after other variables are taken into account. It should also be noted that, although we only reported accuracy in our present analyses, the database also contains information about tip-of-the-pen (TOP) reports; that is, participants sometimes have partial access to the word form of a character but fail to produce it. TOPs in Chinese have been worrying in recent years, due to the decreasing use of handwriting as a result of the availability of computers and smartphones; the phenomenon has been reported in nonacademic outlets such as newspapers (e.g., www.cctv.com, 2006), but there has been no scientific research into it. Finally, the database can also be used as a resource for stimulus construction for lab-based studies on handwriting (in Chinese, at least). For instance, we have shown that dictionary-defined phonograms may not be valid as materials for testing the role of phonology in orthographic access; instead, the phonological cues embedded in a character are better measured by the spelling regularity rating we collected in the database.

The present database can be developed further. For instance, more characters could be added to the database. It may also be possible to determine how a character is written incorrectly (e.g., within which radical) and to provide more information about the time course of handwriting (e.g., how much time is spent on each stroke and how long people pause between strokes). In addition, more character-level characteristics could be collected and added to the database. Future work might also consider a comparison of writing to dictation and character copying, in an attempt to delineate possible cognitive differences among different handwriting tasks.

Conclusion

In the present large-scale behavioral study, we investigated the effects of different character-level variables on writing latencies (as a measure of orthographic access), writing durations (as a measure of motor execution), and accuracy. We showed that linguistic context, frequency, stroke number, and AoA are the most important determinants of handwriting. In addition, the findings clearly reveal that phonological information in the character impacts orthographic access. Finally, the database resulting from the study can be used as a resource for further lab-based research into handwriting.

Author note This work was supported by the Guangdong Province Universities and Colleges Pearl River Younger Scholar Funded Scheme (2016), a Project of the Key Institute of Humanities and Social Sciences, MOE (17JJD190001), and the Economic and Social Research Council (L010224/2). We thank Gabriella Vigliocco for discussion of the work, and Hua Shu for communication on character age of acquisition.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Afonso, O., & Álvarez, C. J. (2011). Phonological effects in handwriting production: Evidence from the implicit priming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1474–1483. <https://doi.org/10.1037/a0024515>
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316. <https://doi.org/10.1037/0096-3445.133.2.283>
- Balota, D. A., & Ferraro, F. R. (1993). A dissociation of frequency and regularity effects in pronunciation performance across young-adults, older adults, and individuals with senile dementia of the Alzheimer-type. *Journal of Memory and Language*, *32*, 573–592. <https://doi.org/10.1006/jmla.1993.1029>
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459. <https://doi.org/10.3758/BF03193014>
- Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers*, *34*, 424–434. <https://doi.org/10.3758/BF03195471>
- Baxter, D. M., & Warrington, E. K. (1986). Ideational agraphia: A single case study. *Journal of Neurology, Neurosurgery & Psychiatry*, *49*, 369–374. <https://doi.org/10.1136/jnnp.49.4.369>
- Bloemsaat, G., Van Galen, G. P., & Meulenbroek, R. G. (2003). Lateralized effects of orthographical irregularity and auditory memory load on the kinematics of transcription typewriting. *Psychological Research*, *67*, 123–133. <https://doi.org/10.1007/s00426-002-0112-7>
- Bonin, P., Fayol, M., & Chalard, M. (2001). Age of acquisition and word frequency in written picture naming. *Quarterly Journal of Experimental Psychology*, *54A*, 469–489. <https://doi.org/10.1080/171375596>
- Bonin, P., Méot, A., & Bugaiska, A. (2018). Concreteness norms for 1,659 French words: Relationships with other psycholinguistic variables and word recognition times. *Behavior Research Methods*, *50*, 2366–2387. <https://doi.org/10.3758/s13428-018-1014-y>
- Bonin, P., Méot, A., Lagarrigue, A., & Roux, S. (2015). Written object naming, spelling to dictation, and immediate copying: Different tasks, different pathways? *Quarterly Journal of Experimental Psychology*, *68*, 1268–1294. <https://doi.org/10.1080/17470218.2014.978877>
- Bonin, P., Peereman, R., & Fayol, M. (2001). Do phonological codes constrain the selection of orthographic codes in written picture naming? *Journal of Memory and Language*, *45*, 688–720. <https://doi.org/10.1006/jmla.2000.2786>
- Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition*, *13*, 992–1011. <https://doi.org/10.1080/13506280544000165>
- Cai, H., Qi, X., Chen, Q., & Zhong, Y. (2012). Effects of phonetic radical position on the regularity effect for naming pictophonetic characters (in Chinese). *Acta Psychologica Sinica*, *44*, 868–881. <https://doi.org/10.3724/sp.j.1041.2012.00868>
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*, e10729:1–8. <https://doi.org/10.1371/journal.pone.0010729>
- Cai, Z. G., & Vigliocco, V. (2018). Word processing. In J. T. Wixted & S. Thompson-Schill (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience, Volume 3: Language and thought* (4th ed., pp. 75–110). New York, NY: Wiley.
- Carreiras, M., & Perea, M. (2004). Naming pseudowords in Spanish: Effects of syllable frequency. *Brain and Language*, *90*, 393–400. <https://doi.org/10.1016/j.bandl.2003.12.003>
- Chang, Y. N., Hsu, C. H., Tsai, J. L., Chen, C. L., & Lee, C. Y. (2016). A psycholinguistic database for traditional Chinese character naming. *Behavior Research Methods*, *48*, 112–122. <https://doi.org/10.3758/s13428-014-0559-7>
- Chen, B. G., & Peng, D. L. (1998). The influence of concreteness on word recognition (in Chinese). *Acta Psychologica Sinica*, *30*, 387–393.
- Chinese Character Encoding Group of Shanghai Jiao Tong University, & Chinese Pinyin Characters Research Group in Shanghai. (1988). *Dictionary of Chinese Character Information (汉字信息字典)*. Beijing, China: Science Press.
- Chinese Ministry of Culture & State Language Affairs Commission. (1986). *Dictionary of Common Chinese Characters in Print (印刷通用汉字字形表)*. Beijing, China: Language Reform Publishing House
- Chinese Ministry of Culture & State Language Affairs Commission. (1988). *Modern Dictionary of Common Characters in Chinese (现代汉语通用字表)*. Beijing, China: The Press and Publications Administration of the P.R.C.
- Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, *99*, 205–235. <https://doi.org/10.1016/j.cognition.2005.01.009>
- Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology*, *79*, 1–12. <https://doi.org/10.1111/j.2044-8295.1988.tb02270.x>
- Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review*, *4*, 226–231. <https://doi.org/10.3758/BF03209397>
- Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 416–431. <https://doi.org/10.1037/0278-7393.29.3.416>
- Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, *57*, 195–209. <https://doi.org/10.1016/j.jml.2006.11.001>

- Damian, M. F., & Dumay, N. (2009). Exploring phonological encoding through repeated segments. *Language & Cognitive Processes*, 24, 685–712. <https://doi.org/10.1080/01690960802351260>
- Damian, M. F., & Stadthagen-Gonzalez, H. (2009). Advance planning of form properties in the written production of single and multiple words. *Language & Cognitive Processes*, 24, 555–579. <https://doi.org/10.1080/01690960802346500>
- Delattre, M., Bonin, P., & Barry, C. (2006). Written spelling to dictation: Sound-to-spelling regularity affects both writing latencies and durations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1330–1340. <https://doi.org/10.1037/0278-7393.32.6.1330>
- Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1103–1123. <https://doi.org/10.1037/0278-7393.26.5.1103>
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Meot, A., ... Pallier, C. (2010). The French Lexicon Project: lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 42, 488–496. <https://doi.org/10.3758/BRM.42.2.488>
- Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46, 57–84. <https://doi.org/10.1006/jmla.2001.2797>
- Fu, Y. (1989). *Dictionary of Chinese character properties* (汉字属性字典). Beijing, China: Language and Culture Press.
- Geschwind, N. (1969). Problems in the anatomical understanding of the aphasias. In A. L. Benton (Ed.), *Contributions to clinical neuropsychology*. Chicago, IL: Aldine.
- Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta Psychologica*, 115, 43–67. <https://doi.org/10.1016/j.actpsy.2003.11.002>
- González-Alvarez, J., & Palomar-García, M.-A. (2016). Syllable frequency and spoken word recognition: An inhibitory effect. *Psychological Reports*, 119, 263–275. <https://doi.org/10.1177/0033294116654449>
- Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software*, 17, 1–27.
- Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods*, 48, 1358–1369. <https://doi.org/10.3758/s13428-015-0684-y>
- Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic in formation and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 824–843. <https://doi.org/10.1037/0278-7393.20.4.824>
- Johnson, J. W., & Lebreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7, 238–257. <https://doi.org/10.1177/1094428104266510>
- Kandel, S., & Perret, C. (2015). How does the interaction between spelling and motor processes build up during writing acquisition? *Cognition*, 136, 325–336. <https://doi.org/10.1016/j.cognition.2014.11.014>
- Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech perception. *Journal of Experimental Psychology: General*, 129, 340–360. <https://doi.org/10.1037/0096-3445.129.3.340>
- Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, 1, 174:1–15. <https://doi.org/10.3389/fpsyg.2010.00174>
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, 44, 287–304. <https://doi.org/10.3758/s13428-011-0118-4>
- Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, 140, 14–34. <https://doi.org/10.1037/a0021446>
- Leck, K. J., Weekes, B. S., & Chen, M. J. (1995). Visual and phonological pathways to the lexicon: Evidence from Chinese readers. *Memory & Cognition*, 23, 468–476. <https://doi.org/10.3758/BF03197248>
- Lee, C.-Y., Tsai, J.-L., Su, E. C.-I., Tzeng, O. J. L., & Hung, D. L. (2005). Consistency, regularity, and frequency effects in naming Chinese. *Language and Linguistics*, 6, 75–107.
- Leong, C. K., Cheng, P.-W., & Mulcahy, R. (1987). Automatic processing of morphemic orthography by mature readers. *Language and Speech*, 30, 181–196. <https://doi.org/10.1177/002383098703000207>
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38, disc. 38–75. <https://doi.org/10.1017/S0140525X99001776>
- Lewis, M. B. (1999). Age of acquisition in face categorisation: Is there an instance-based account? *Cognition*, 71, B23–B39. [https://doi.org/10.1016/S0010-0277\(99\)00020-7](https://doi.org/10.1016/S0010-0277(99)00020-7)
- Lewis, M. B., Gerhand, S., & Ellis, H. D. (2001). Re-evaluating age-of-acquisition effects: Are they simply cumulative-frequency effects? *Cognition*, 78, 189–205. [https://doi.org/10.1016/S0010-0277\(00\)00117-7](https://doi.org/10.1016/S0010-0277(00)00117-7)
- Lien, Y.-W. (1985). *Consistency of the phonetic clues in the Chinese phonograms and their naming latencies*. Psychological Department. National Taiwan University, Taipei.
- Linguistics Institute of Chinese Academy of Social Sciences. (2011). *Xinhua dictionary* (新华字典, version 11). Beijing, China: Commercial Press.
- Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods*, 39, 192–198. <https://doi.org/10.3758/BRF03193147>
- Luria, A. R. (1970). *Traumatic aphasia*. The Hague, The Netherlands: Mouton.
- Ni, H. (1982). *Dictionary of modern Chinese phonograms* (现代汉字形声字字汇). Beijing, China: Language and Culture Press.
- Peng, D., & Wang, C. (1997). Basic processing unit of Chinese character recognition: Evidence from stroke number effect and radical number effect (in Chinese). *Acta Psychologica Sinica*, 29.
- Perez, M. A. (2007). Age of acquisition persists as the main factor in picture naming when cumulative word frequency and frequency trajectory are controlled. *Quarterly Journal of Experimental Psychology*, 60, 32–42. <https://doi.org/10.1080/17470210600577423>
- Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 139–156. <https://doi.org/10.1037/0278-7393.27.1.139>
- Qu, Q., Damian, M. F., Zhang, Q., & Zhu, X. (2011). Phonology contributes to writing: Evidence from written word production in a nonalphabetic script. *Psychological Science*, 22, 1107–1112. <https://doi.org/10.1177/0956797611417001>
- Qu, Q., Zhang, Q., & Damian, M. F. (2016). Tracking the time course of lexical access in orthographic production: An event-related potential study of word frequency effects in written picture naming. *Brain and Language*, 159, 118–126. <https://doi.org/10.1016/j.bandl.2016.06.008>

- Rapp, B., Benzing, L., & Caramazza, A. (1997). The autonomy of lexical orthography. *Cognitive Neuropsychology*, *14*, 71–104. <https://doi.org/10.1080/026432997381628>
- Roux, S., McKeeff, T. J., Grosjacques, G., Afonso, O., & Kandel, S. (2013). The interaction between central and peripheral processes in handwriting production. *Cognition*, *127*, 235–241. <https://doi.org/10.1016/j.cognition.2012.12.009>
- Schriefers, H., de Ruitter, J. P., & Steigerwald, M. (1999). Parallelism in the production of noun phrases: Experiments and reaction time models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 702–720. <https://doi.org/10.1037/0278-7393.25.3.702>
- Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology*, *11*, 17–50. <https://doi.org/10.1080/713752301>
- Sereno, S. C., Brewer, C. C., & O'Donnell, P. J. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science*, *14*, 328–333. <https://doi.org/10.1111/1467-9280.14471>
- Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of school Chinese: Implications for learning to read. *Child Development*, *74*, 27–47. <https://doi.org/10.1111/1467-8624.00519>
- Søvik, N., Arntzen, O., Samuelstuen, M., & Heggberget, M. (1994). Relations between linguistic word groups and writing. In C. Faure, G. Lorette, & A. Vinter (Eds.), *Advances in handwriting and drawing: A multidisciplinary approach* (pp. 231–246). Paris, France: Europa.
- Su, Y.-F., & Samuels, S. J. (2010). Developmental changes in character-complexity and word-length effects when reading Chinese script. *Reading and Writing*, *23*, 1085–1108. <https://doi.org/10.1007/s11145-009-9197-3>
- Sze, W. P., Rickard Liow, S. J., & Yap, M. J. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods*, *46*, 263–273. <https://doi.org/10.3758/s13428-013-0355-9>
- van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science*, *10*, 165–191. [https://doi.org/10.1016/0167-9457\(91\)90003-G](https://doi.org/10.1016/0167-9457(91)90003-G)
- Wang, C., & Zhang, Q. (2015). Phonological codes constrain output of orthographic codes via sublexical and lexical routes in Chinese written production. *PLoS ONE*, *10*, e0124470. <https://doi.org/10.1371/journal.pone.0124470>
- Wang, Q., & Dong, Y. (2013). The N2- and N400-like effects of radicals on complex Chinese characters. *Neuroscience Letters*, *548*, 301–305. <https://doi.org/10.1016/j.neulet.2013.05.074>
- Woollams, A. M. (2005). Imageability and ambiguity effects in speeded naming: Convergence and divergence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 878–890. <https://doi.org/10.1037/0278-7393.31.5.878>
- www.cctv.com (2006). 你提笔忘字吗? <http://news.cctv.com/education/20061011/104804.shtml> [Retrieved 14/05/2018].
- Yang, S., Zhang, S., & Wang, Q. (2016). P2 and behavioral effects of stroke count in Chinese characters: Evidence for an analytic and attentional view. *Neuroscience Letters*, *628*, 123–127. <https://doi.org/10.1016/j.neulet.2016.06.006>
- Yap, M. J., Liow, S. J., Jalil, S. B., & Faizal, S. S. (2010). The Malay Lexicon Project: A database of lexical statistics for 9,592 words. *Behavior Research Methods*, *42*, 992–1003. <https://doi.org/10.3758/BRM.42.4.992>
- Zhang, J., & Li, M. (2010). The effects of Chinese input methods on the links among the orthography, phonology and meaning of Chinese characters (in Chinese). *Journal of Psychological Science*, *33*, 835–838.
- Zhang, Q., & Damian, M. F. (2010). Impact of phonology on the generation of handwritten responses: Evidence from picture-word interference tasks. *Memory & Cognition*, *38*, 519–528. <https://doi.org/10.3758/MC.38.4.519>
- Zhang, Q., & Feng, C. (2017). The interaction between central and peripheral processing in Chinese handwritten production: Evidence from the effect of lexicality and radical complexity. *Frontiers in Psychology*, *8*, 334. <https://doi.org/10.3389/fpsyg.2017.00334>
- Zhang, Q., & Wang, C. (2014). Syllable frequency and word frequency effects in spoken and written word production in a non-alphabetic script. *Frontiers in Psychology*, *5*, 120. <https://doi.org/10.3389/fpsyg.2014.00120>
- Zhang, Q., & Wang, C. (2015). Phonology is not accessed earlier than orthography in Chinese written production: Evidence for the orthography autonomy hypothesis. *Frontiers in Psychology*, *6*, 448. <https://doi.org/10.3389/fpsyg.2015.00448>
- Zhang, Q., & Yang, Y. (2003). The determiners of picture-naming latency (in Chinese). *Acta Psychologica Sinica*, *35*, 447–454.
- Zhou, X., Shu, H., Bi, Y., & Shi, D. (1999). Is there phonologically mediated access to lexical semantics in reading Chinese. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 135–171). Mahwah, NJ: Erlbaum.