In press at *Quarterly Journal of Experimental Psychology*

Individual Differences in Cantonese Chinese word recognition:

Insights from the Chinese Lexicon Project

Rosemarie Y. H. Lim[1], Melvin J. Yap[1], & Chi-Shing Tse[2]

[1]National University Singapore

[2]The Chinese University of Hong Kong

Address correspondence to:

Melvin J. Yap

Department of Psychology, Faculty of Arts and Social Sciences

National University of Singapore

Block AS4, #02-07

Singapore 117570, Republic of Singapore

melvin@nus.edu.sg

Abstract

The Chinese Lexicon Project (Tse et al., 2017) is a repository of lexical decision data for 25,286 Cantonese Chinese two-character compound words. In order to create that repository, 594 participants responded to approximately 1,404 words and 1,404 nonwords over three sessions. Using the data in this repository, the present study examines the variability and reliability of Chinese lexical decision performance, along with the moderating influence of individual differences on lexical processing. We generally found high to very high within- and between-session reliabilities for mean response times, ex-Gaussian parameters, accuracy rates, and a composite proficiency measure tapping lexical processing fluency. Using linear mixed effects models, we also found reliable interactions between fluency and two lexical effects. Specifically, more fluent readers showed *larger* effects of word frequency and semantic transparency. These results attest to the stability of Chinese word recognition performance, and are most consistent with a flexible lexical processing system that adapts optimally to task demands.

(155 words)

The Chinese language stands as one of the few languages today that is logographic, with each character representing both a phonetic syllable and a unit of meaning (morpheme). In contrast to English and many other languages that rely on an alphabetic writing system, each Chinese character comprises strokes, which form radicals that in turn form characters and ultimately compound words. The smallest meaningful unit in written Chinese is a single free-standing morpheme (e.g., 媽 maa1 “mother”); the majority of these characters are comprised of a semantic and phonetic radical (e.g., for 媽, the semantic radical is 女 neoi5 “female” while the phonetic radical is 馬 maa5 “horse”). Chinese is also a tonal language, relying on tonal contrasts to distinguish between words (e.g., 媽 maa1 and 麻 maa4 “sesame” share the same syllable but different tones).

Chinese words are most commonly formed through the compounding of two characters (e.g., 花 faa1 “flower” and 園 jyun4 “park” → 花園 faa1 jyun4 “garden”). Indeed, two-character compound words represent 73.6% of all Chinese words (Institute of Language Teaching and Research, 1986). There exist a number of Chinese varieties or dialects; Mandarin Chinese is used in mainland China, Taiwan, and Singapore, while Cantonese Chinese is used in Hong Kong and Macau. While Mandarin and Cantonese broadly share the same vocabulary and characters, there are a couple of key differences. Specifically, the two varieties are not mutually intelligible, with different pronunciations for the same word. Cantonese Chinese also employs more tones (nine) compared to Mandarin Chinese (four). The present study focuses on the lexical processing of Cantonese Chinese.

The visual word recognition of isolated Chinese characters and compound words has been found to be influenced by orthographic, phonological, and semantic properties (see Sze, Yap, & Rickard Liow, 2015; Tse & Yap, 2018, for reviews). In line with Sze et al. (2015) and

Tse and Yap (2018), orthographic properties[1] include *number of strokes*, *character frequency*, and *word frequency* (i.e., frequency of the compound word). Phonological properties include *homophone density* (i.e., the number of other characters sharing the same pronunciation as the target character) and *phonological consistency* (i.e., whether a character has one or more than one pronunciation; e.g., 曾 is phonologically inconsistent because it is pronounced as cang4 in 曾經 cang4 ging1 "already" and as zang1 when it is a last name). Finally, semantic properties include *semantic transparency*, which reflects the extent to which a compound word is semantically related to its constituent characters. For example, 花園 (faa1 jyun4 flower-park "garden") is transparent whereas 花生 (faa1 saang1 flower-grow "peanut") or 東西 (dung1 sai1 east-west "thing") are opaque. Generally, Chinese characters and words are recognized faster when they have fewer strokes, are more frequent, are more semantically transparent, and are higher in phonological consistency and homophone density (Sze et al., 2015; Tse & Yap, 2018).

To provide insights into the mechanisms underlying Chinese lexical processing, researchers have conducted factorial experiments that systematically examine the influence of various lexical variables (e.g., number of strokes) on *lexical decision* (classifying stimuli as words or nonwords) and *speeded pronunciation* (reading stimuli aloud) performance (e.g., Peng, Liu, & Wang, 1999; Taft, Liu, & Zhu, 1999; Zhou & Marslen-Wilson, 2000). While these factorial experiments have yielded a wealth of important findings that inform and constrain influential models (e.g., Peng et al., 1999; Taft et al., 1999; Tan & Perfetti, 1999; Zhou & Marslen-Wilson, 2000), they are also associated with several drawbacks. For

---

[1] There is as yet no universally agreed upon system for classifying Chinese word properties. For example, Sun, Hendrix, Ma, and Baayen (2018) define character and word frequency as frequency measures, and stroke count as a complexity measure, whereas Tsang et al. (2018) distinguish between word- and character-level variables. To further complicate matters, some researchers (e.g., Baayen, Feldman, & Schreuder, 2006) have proposed that word frequency be treated as a semantic variable. However, for ease of comparing the present study to earlier related work by Sze, Yap, and Rickard Liow (2015) and Tse and Yap (2018), we adopt their convention for classifying variables.

example, factorial experiments require continuous variables to be turned into categorical dimensions, which may reduce statistical power (Cohen, 1983) or increase Type I error (MacCallum, Zhang, Preacher, & Rucker, 2002). It is also increasingly challenging for researchers to select words for the different cells of a factorial design, while controlling for an ever growing list of correlated word properties (see Balota, Yap, Hutchison, & Cortese, 2012, for an extended discussion of these issues). In light of these challenges, the factorial approach has recently been complemented by *megastudies* (see Balota et al., 2012, for a review), wherein participants provide responses for very large sets of words (e.g., all commonly used two-character Chinese words). In that sense, instead of selecting stimuli based on a limited set of criteria, which may be vulnerable to list context effects or the experimenter's implicit knowledge, megastudies allow the language to define the stimuli.

Megastudy data allow the main and interactive effects of predictors to be explored in a continuous, rather than categorical, manner, while statistically controlling for the influence of correlated variables. For example, Sze et al. (2015) used item-level regression analyses to examine the predictive power of various variables on lexical decision performance for 1,560 simplified Chinese characters, while Liu, Shu, and Li (2007) adopted a similar procedure to examine speeded pronunciation performance for 2,423 simplified Chinese characters. Turning to compound words, Tse and Yap (2018) conducted analyses on lexical decision data for 18,983 two-character traditional Chinese compound words. Both Sze et al. and Tse and Yap reported that orthographic variables accounted for the most variance in lexical decision, followed by semantic variables, then by phonological variables. Overall, these results suggest that skilled Chinese readers largely attend to orthographic and semantic information when processing visually presented characters and compound words, and that phonology plays a relatively marginal role in Chinese word recognition (cf. Perfetti & Dunlap, 2008).

Individual Differences in Chinese Word Recognition

Word recognition studies, in Chinese and other languages, have predominantly focused on *group*-level performance, i.e., data are aggregated across participants (Yap, Balota, Sibley, & Ratcliff, 2012). Similarly, extant models of Chinese lexical processing have not yet considered individual differences amongst skilled readers. This emphasis on the characterization of an average or “prototypical” Chinese reader seems at odds with empirical evidence that there are substantial, stable individual differences in reading fluency, which are systematically related to word recognition performance (see Yap et al., 2012). For example, in English word recognition literature, more proficient readers (as reflected by their vocabulary knowledge or degree of print exposure) show *decreased* sensitivity to a number of word dimensions, including number of letters (Butler & Hains, 1979) and word frequency (Chateau & Jared, 2000). This is consistent with the idea that skilled readers, who possess higher quality orthographic and phonological representations (Perfetti & Hart, 2002), are relying to a greater extent on automatized lexical processing mechanisms (Stanovich, 1980).

Indeed, it is conceivable that certain empirical discrepancies in the word recognition literature are at least partly driven by individual differences in participants between studies (for examples in English, see Yap, Balota, Tse, & Besner, 2008; Yap, Tse, & Balota, 2009). For instance, the facilitatory effect of phonological frequency has been reported in some studies (e.g., Ziegler, Tan, Perry, & Montant, 2000), but not in others (e.g., Chen, Vaid, & Wu, 2009). Similarly, some studies (e.g., Shen & Zhu, 1994) showed an interaction between character frequency and number of strokes in speeded pronunciation, wherein the effect of stroke count is stronger for low-frequency characters, while others have failed to replicate the pattern (e.g., Peng & Wang, 1997).

The Present Study

To our knowledge, there is currently no published work examining individual differences in skilled Chinese lexical processing. The present study leverages the power of a megastudy, as conducted in the Chinese Lexicon Project (Tse et al., 2017), to address this important gap in the literature. Tse et al. reported a database of lexical variables and lexical decision trial-level data for more than 25,000 traditional Chinese two-character compound words, based on 594 native Cantonese-speaking participants from the Chinese University of Hong Kong, who were tested over three sessions. Importantly, for our purposes, each participant was asked to report their amount of time they read Chinese materials per week, cumulative grade point average, grade on Chinese language university entrance exam, and self-rated knowledge for traditional characters and spoken Cantonese (using 7-point rating scales).

The goal of the present study is to extend the previous work by Tse and Yap (2018) in the following ways. As discussed earlier, Tse and Yap explored the predictive power of orthographic (number of strokes, character and word frequency), phonological (phonological consistency, homophone density), and semantic (semantic transparency) variables on Chinese lexical decision performance, while also examining theoretically important interactions between word-level and character-level properties. They reported that skilled readers seem to rely predominantly on orthographic and semantic information when processing visually presented Chinese words. They also observed that effects of cumulative character frequency (i.e., the summed frequencies of constituent characters) were weaker for compound words that were either semantically opaque or high in frequency. Consider the character frequency × word frequency interaction. Given that the constituent characters are encountered more often in high frequency words, compared to low frequency words, readers tend to process them holistically as a whole word unit (e.g., Caramazza, Laudanna, & Romani, 1988), thereby attenuating the influence of character-level variables (e.g., character frequency). Such findings shed light on the holistic versus analytic modes of processing in Chinese word recognition.

However, it is unclear to what extent these effects were moderated by individual differences such as self-rated language proficiency or academic ability. In our study, the individual differences measures reported by the participants were supplemented by *lexical processing fluency*[2], a composite measure derived by first separately z-score transforming mean response times (RTs) and mean error rates, before averaging the two values (Chignell, Tong, Mizobuchi, Delange, Ho, & Walmsley, 2015). Specifically, we first computed the mean RT and error rate for each participant in the study. Next, for these two measures, the mean and standard deviation for the entire sample were obtained; these were used to generate two z-scores for each participant, one for RT and one for error rate. The two z-scores were then averaged and the composite score reversed in sign for ease of interpretation. In summary, faster, more accurate performance is reflected by more positive scores while slower, less accurate performance is reflected by more negative scores. This measure allows overall word recognition performance to be characterized even when speed-accuracy trade-offs exist. The first objective of the present study is to examine how readers' sensitivity to different word properties is systematically moderated by the foregoing individual differences.

However, before we can properly address the question of individual differences, an important prerequisite is to determine whether individual differences in Chinese lexical processing are reliable, and the extent of variation in these differences. Reliability is highly crucial in psychological research; the reliability of RT measures limits their usefulness and constrains the extent to which they can correlate with other measures (Lowe & Rabbitt, 1998). Without first establishing reliability, we are not able to distinguish genuine individual differences in processing from measurement noise. Despite the intuitive importance of exploring the long-range stability or alternate-form reliability of word recognition measures,

---

[2] We thank an anonymous reviewer for suggesting this metric.

there has been very little work addressing this question, even in English. Hence, in order to make inferences about individual differences in Chinese word recognition processing, one has to first estimate the stability of Chinese lexical decision performance - this is the second objective of the present study. Such stability is reflected by the consistency of performance across different time points (i.e., test-retest reliability) and across odd- and even-numbered trials (i.e., internal consistency).

Individual performance can be quantified in terms of mean RTs and accuracy rates, or more granularly at the level of the underlying RT distribution. RT distributional characteristics have been captured by fitting empirical distributions to theoretical distributions such as the ex-Gaussian distribution (Ratcliff, 1979). This is in line with recent interest (e.g., Yap et al., 2012; Pexman & Yap, 2018) in examining the RT distributions of individual participants via estimating ex-Gaussian model parameters for each participant. The ex-Gaussian distribution (the convolution of an exponential and Gaussian distribution) has been found to provide a good fit for positively skewed empirical RT distributions; it contains three parameters, μ and σ, the mean and standard deviation of the Gaussian distribution respectively, and τ, the mean of the exponential distribution. Estimating ex-Gaussian parameters for each participant yields a more nuanced RT profile that characterizes the *shape* of the RT distribution produced by that participant.

To recapitulate, the present study addresses two broad questions. First, how much stability is there in Chinese lexical decision performance over multiple time points, with respect to between-session test-retest reliability and within-session internal consistency? Performance is characterized across multiple dimensions, including mean RT, mean accuracy, and ex-Gaussian parameters. Second, assuming performance is shown to be reliable, what are the systematic relationships that exist between individual differences (self-rated language proficiency, academic ability, and lexical processing fluency) and sensitivity to various underlying dimensions of Chinese words? For example, consider the hypothesis that skilled readers rely more heavily on automatic lexical processing mechanisms (LaBerge & Samuels, 1974). This yields the prediction that more proficient readers will be influenced to a lesser extent by word properties. The Chinese Lexicon Project (Tse et al., 2017), with its exceptionally large and well-characterized set of words and participants, allows us to address the foregoing interrelated questions in a powerful, unified manner.

Besides addressing questions related to Chinese psycholinguistics, our study serves to address a more general question that relates to the robustness and generalizability of research findings that are largely based on alphabetic writing systems. Although some form of Chinese is used by over a billion people (Li & Thompson, 2009), empirical work and modeling in lexical processing have been overwhelmingly dominated by the study of alphabetic writing systems (Perfetti & Liu, 2006). Relatedly, Share (2008) has highlighted the perils of having a field of reading science that devotes a disproportionate amount of research to English, an "outlier" alphabetic orthography that is characterized by extreme ambiguity with respect to spelling-sound correspondence. For example, the extensive interest in dual-route (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) and connectionist (Plaut, McClelland, Seidenberg, & Patterson, 1996) models of reading was largely motivated by the unusual properties of English spelling-sound relations. Indeed, there is no guarantee that an optimal cognitive system for English and other alphabetic languages will be equally optimized for a logographic language like Chinese. In order for researchers to make more definitive language-general claims, it is important to study effects across multiple languages, so as to better tease apart language-general and language-specific processing.

Method

Dataset

All analyses are based on archival trial-level data from the Chinese Lexicon Project

(Tse et al., 2017), a large-scale repository of psycholinguistic properties and lexical decision data for 25,286 two-character Chinese compound words, collected from 594[3] native Cantonese-speaking students from the Chinese University of Hong Kong (CUHK). The full set of 25,286 words were randomly divided into 18 word sets (about 1,404 words each), with the constraint that the proportion of words with a specific first character was approximately the same across all subsets. For each word set, nonwords were created by recombining characters within the word set randomly, ensuring that the character-level lexical properties for words and nonwords were perfectly matched; this process yielded another 18 nonword sets that were derived from their respective word sets. Word sets were then paired with nonword sets that were created from some other set, e.g., nonwords created from Set 8 words were paired with Set 12 words. Next, 2,808 stimuli (1,404 words and 1,404 nonwords) obtained by aggregating a word and nonword set, were presented over three sessions to a participant, who went through approximately 936 lexical decision trials on each session.

Additional demographic variables collected included the: a) amount of time they read Chinese materials a week, b) grade point average (GPA), c) Chinese language university entrance exam grade, d) self-rated knowledge for traditional characters, and e) self-rated knowledge for spoken Cantonese. To reduce these to a smaller number of orthogonal predictors, principal components analysis was carried out using varimax rotation with Kaiser normalization (Baayen et al., 2006). Two principal components with Eigenvalues greater than 1 were extracted, accounting for 53% of the variance. As indicated in Table 1, the rotated component matrix indicated that self-rated knowledge for traditional characters and self-rated knowledge for spoken Cantonese loaded on the first principal component (PC1), while GPA, Chinese language university entrance score, and Chinese print exposure loaded on the second principal component (PC2). As such, whereas PC1 appears to capture participants' self-rated

---

[3] Five participants who missed at least one of the three sessions were excluded from the analysis.

knowledge in written and spoken forms of Chinese, PC2 appears to capture participants' general academic ability. These principal components, which we respectively term *self-rated language proficiency* and *academic ability*, were used in subsequent analyses of individual differences. As mentioned previously, to complement language proficiency and academic ability, we computed an additional measure of individual difference, lexical processing fluency, derived by averaging z-score transformed participant-level mean RTs and error rates (Chignell et al., 2015).

Results

First, error trials (11%) and trials with response latencies slower than 3000 ms or faster than 200 ms (0.8%) were excluded. Among the remaining correct trials, RTs that were more than 2.5 standard deviations away from each participants' mean (2.7%) were identified as outliers. For ease of exposition, the reliability analyses will be first described, before we consider the relationships between individual differences and sensitivity to different lexical dimensions.

Analysis 1: Reliability Analyses

Reliability was examined in two ways. Each participant's word data were first partitioned such that trials were organized into Day 1 (D1) trials, Day 2 (D2) trials, Day 3 (D3) trials, odd-numbered trials (O), and even-numbered trials (E). Using split-half correlations, comparing D1, D2, and D3 trials allows us to assess between-session reliability, while comparing O and E trials allow us to assess within-session reliability. For each set of trials, the following parameters were computed separately for words and nonwords for each participant: a) mean RT, b) standard deviation, c) mean accuracy, d) *d'*, and e) ex-Gaussian parameters. *d'*, an alternative conceptualization of lexical decision accuracy (Diependaele, Brysbaert, & Neri, 2012), was calculated by subtracting the z-score of the false alarm rate (i.e., proportion of incorrect "word" response to nonwords) from the z-score of the hit rate (i.e., proportion of correct “word” response to words), whereas ex-Gaussian parameters were estimated in R (R Development Core Team, 2015) using Nelder and Mead’s (1965) simplex algorithm. The measures described above are presented on Table 2 as a function of trial type (D1, D2, D3, O, and E). As can be seen, there was evidence for a practice effect; mean RT, standard deviation, μ, σ, and τ decreased over the three sessions.

Interestingly, the speed-up was accompanied by a *decrease* in accuracy, reflected by a monotonic decrease in accuracy and *d’*. Table 3 presents the Pearson correlations for these parameters. Within- and between-session reliabilities were reassuringly high for lexical decision performance for both words and nonwords. Within-session reliability was very high for the mean, standard deviation, ex-Gaussian parameters, accuracy, and *d’* (all rs ≥ .90). Turning to between-session reliability, correlations were moderate to very high for the same parameters (rs from .44 to .83). Interestingly, the D2-D3 correlations, compared to the D1-D2 and D1-D3 correlations, were always highest, suggesting that performance might be relatively noisy during the first session, and that some time is needed for performance to stabilize. Overall, these results suggest that Chinese readers are associated with an RT distributional signature that extends beyond simple mean performance, which is stable across multiple testing sessions with distinct, non-overlapping sets of words and nonwords.

Analysis 2: Individual Differences Analyses

Having established the reliability of our data, we next examined the relationships between individual differences (self-rated language proficiency, academic ability, lexical processing fluency) and different aspects of Chinese lexical decision performance, including mean RT, mean accuracy, and *d’*. Figure 1 presents the intercorrelations between these variables. Self-rated language proficiency and academic ability were unexpectedly poor predictors of lexical decision performance. Language proficiency weakly predicted mean RT (r = -.108, p = .014) but not mean accuracy or *d’*. Academic ability modestly predicted accuracy ($r = .162$, $p < .001$) and *d'* ($r = .187$, $p < .001$), but not mean RT. Interestingly, the moderate positive correlations between RT and accuracy ($r = .309$, $p < .001$) and between RT and *d'* ($r = .314$, $p < .001$) point to a speed-accuracy trade-off, i.e., slower participants were more accurate.

This trade-off implies that the efficiency of an individual's lexical processing system should not be solely captured by RT or accuracy, but is more appropriately tapped by a composite measure like lexical processing fluency that integrates both measures. Such an observation is nicely consistent with the observation that lexical processing fluency, in contrast to mean RT, showed the expected positive relationships with *d'* ($r = .549$, $p < .001$) and mean accuracy ($r = .570$, $p < .001$), and a negative relationship with mean RT ($r = -.581$, $p < .001$); participants associated with higher fluency scores were faster and more accurate in their responses. As mentioned earlier, self-rated language proficiency did not predict mean accuracy or *d'*; more troublingly, its positive correlation with lexical processing fluency was also only borderline significant ($r = .086$, $p = .051$).

Analysis 3: Analyses of Joint Effects of Item and Participant Properties

Using linear mixed effects (LME) models, we next assessed the effects of participant and word properties on lexical decision performance. As stated in the Introduction, our original intent was to assess the moderating effect of participant characteristics such as self-rated language proficiency, academic ability, and lexical processing fluency on word processing. However, the above analyses have already highlighted the limitations of the first two measures, which poorly predicted word recognition performance. In light of that, only the composite lexical processing fluency measure, which balances processing speed and accuracy, is used to quantify the efficiency of a participant's word recognition system. It is also worth noting that the fluency variable is associated with high psychometric reliability (see correlations in Table 3), an important prerequisite for an individual difference measure.

The word properties comprised the full set of variables investigated in Tse and Yap (2018), including word frequency, character frequency, stroke count, phonological consistency, homophone density, and semantic transparency. For character-specific measures (e.g., character frequency), we averaged counts across both characters. Table 4 presents the intercorrelations between predictors.

Using R (R Core Team, 2015), z-score transformed RTs (Faust, Balota, Spieler, & Ferraro, 1999) were fitted using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) and *p*-values were obtained for fixed effects using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2016). Each participant's raw latencies were standardized using a z-score transformation, following the same procedure as the one adopted by Balota et al. (2007). The rationale for standardizing raw RTs using a z-score transformation is to control for spurious effects due to slowing. Specifically, a participant's overall processing speed is positively correlated with the magnitude of effects produced by that participant. To illustrate this point, Faust et al. (1999) described two hypothetical participants, wherein the average RT for the slower participant will be two times that of the faster participant minus 300 ms for any given experimental condition. Suppose the two participants went through an experiment where frequency was manipulated, and the faster participant took, on average, 400 ms and 500 ms respectively to respond to high and low frequency words; the frequency effect is therefore 100 ms. However, because of the systematic linear relationship between the two individuals, the slower participant would take 500 ms (400 × 2 – 300) and 700 ms (500 × 2 – 300) respectively for high and low frequency words, yielding a frequency effect of 200 ms. In other words, the slower participant may produce larger frequency effects *not* because he/she is actually more sensitive to word frequency, but because he/she takes longer to respond. The use of z-score transformed RTs to rule out processing speed as a confound is particularly pertinent when attempting to interpret Group × Treatment interactions when the groups are not matched on overall latency (Faust et al., 1999; see also Salthouse, 1985).

A hierarchical analytical approach was used; main effects of predictors were examined in Step 1, while the interactions between fluency and predictors were subsequently added in Step 2. In Step 1, the main effects of the orthographic, phonological, and semantic variables, as well as fluency, were entered as fixed effects; random intercepts and slopes for participants and random intercepts for items were also included in the model (see Table 5). In Step 2, we tested a second model that included the main effects and the interactions between fluency and the various word properties (see Table 5). A likelihood ratio test confirmed that the Step 2 model accounted for significantly more variance in lexical decision performance than the Step 1 model alone, $\chi^2(6) = 137.28$, $p < .001$.

Consistent with Tse and Yap (2018), the Step 1 main effects analyses revealed that participants recognized two-character compound words faster when they were associated with fewer strokes, higher character frequencies, higher word frequencies, higher homophone density, higher phonological consistency, and higher semantic transparency. Participants who were associated with higher fluency scores also made faster responses. Figure 2 presents the distribution of random slopes or coefficients as a function of word property. As can be seen, there is a great deal of variation in the magnitude of effects associated with participants. For example, consider the random slopes for word frequency. Even though virtually all participants showed a negative slope (i.e., faster RTs for high frequency words), these slopes were normally distributed, with some participants showing a much smaller frequency effect than others. Turning to the Step 2 interaction effects, we examined the extent to which the effects of different word properties were moderated by participant fluency scores. Using the effects package (Fox et al., 2015), the two statistically significant interactions are plotted as simple slopes on Figure 3. Specifically, participants who could discriminate between words and nonwords more efficiently (as reflected by higher fluency scores) showed *larger* (not smaller) effects of word frequency and semantic transparency.

As discussed earlier, the motivation for using z-score transformed RTs in place of raw RTs is to guard against spurious overadditive interactions between fluency and properties. Essentially, because less fluent participants are associated with slower, more variable RTs, the larger effects produced by these participants may therefore be mediated by slowing rather than by greater sensitivity to a particular word dimension (Faust et al., 1999). As can be seen in Figure 1, there is a strong negative correlation ($r = -.58$) between fluency and mean RT. To verify the effects of slowing in the present dataset, we conducted a supplementary analysis to examine the interactions between fluency and word properties when raw RTs were used as the dependent variable (see Table 6). Interestingly, this analysis revealed three reliable interactions; more fluent participants showed smaller effects of stroke count, character frequency, and word frequency (see Figure 4). However, the story becomes more complicated when one examines Table 7, which presents the descriptive statistics behind the simple slopes. As can be seen, faster raw RTs are associated with smaller effects, making it difficult to unequivocally disentangle the influence of slowing from the influence of genuine enhanced sensitivity. The present data further underscore the interpretative ambiguity of Group × Treatment interactions based on raw RTs when groups vary on processing speed.

Discussion

To our knowledge, the present study is the first large-scale investigation of individual differences in Cantonese Chinese lexical processing, using the well-characterized behavioral dataset from the Chinese Lexicon Project (Tse et al., 2017). There were a few noteworthy findings. First, across non-overlapping sets of stimuli, between- and with-session reliability were high to very high across a range of performance indicators. Second, measures based on self-reported language proficiency and academic ability were unexpectedly poor predictors of word recognition performance. Finally, the LME models not only yielded main effects that were consistent with previous work by Tse and Yap (2018), but importantly revealed novel and theoretically interesting interactions between word properties and individual differences.

Variability and Reliability of Word Recognition Performance

Our data revealed unequivocal and substantial variability between participants in word recognition performance across virtually all dimensions of interest, including mean RTs, accuracy rates, RT distributional characteristics, lexical processing fluency, and sensitivity to lexical dimensions (see Figures 1 and 2). Characterizing the scale of these individual difference using a very large sample of participants serves as a useful benchmark for future individual differences work in Chinese. More crucially, our analyses revealed reassuringly high within- and between-session reliability for these measures. Evaluating the reliability of word recognition measures is an important first step in the study of individual differences in word recognition. Without first doing that, it is unclear if variability amongst readers reflects systematic individual differences or simply measurement noise.

Interestingly, reliability was just as high for responses to both words *and* nonwords (Yap, Sibley, Balota, Ratcliff, & Rueckl, 2015). While not entirely unexpected, these findings collectively attest to the reliability of Chinese word recognition performance, which goes beyond simple average processing speed and which applies to both words and nonwords. More fundamentally, they demonstrate that stable RT distributional profiles in lexical processing are not idiosyncratic to English but also generalize to other languages with markedly different properties.

Reconsidering Self-Rated Language Proficiency

In the Chinese Lexicon Project (Tse et al., 2017), participants provided information on self-rated language proficiency (written and spoken) and academic ability (cumulative grade-point average and grades on the Chinese language university entrance exam). With the benefit of hindsight, these may not have been the optimal individual differences measures to quantify reader proficiency. As reported in the Results, both measures were relatively poor predictors of word recognition performance.

The low correlations between academic ability and word recognition measures can probably be explained by the fact that most courses in the university do not tap proficiency in the Chinese language that strongly. However, our *a priori* expectation was that self-rated language proficiency would tap the efficiency of the lexical processing system. For example, one might predict that higher proficiency participants would be associated with faster and more accurate lexical decision performance. However, to our surprise, there was essentially no correlation between self-rated language proficiency and lexical decision accuracy (see Figure 1). Although there was a weak negative correlation (r = -.108) between proficiency and mean RT (i.e., faster responses for more proficient participants), this relationship was qualified by the moderate positive correlation (r = .309, p < .001) between mean RT and mean accuracy, indicative of a speed-accuracy tradeoff.

The low predictive power of self-rated proficiency is possibly driven by a range of factors, including range restriction. Specifically, a selective admission process prevents applicants with lower Chinese fluency from securing a place in the university, and this restriction of range is reflected in the data. For the 7-point self-rated knowledge for traditional characters, nearly 90% of the sample produced responses between 5 and 7. Likewise, for the 7-point self-rated knowledge for spoken Cantonese, over 95% of the sample produced responses between 5 and 7. In light of the limitations of the self-rated language fluency measure, Chinese proficiency in our study was operationally defined using a composite lexical processing fluency measure that combines z-score transformed mean RTs and accuracy rates (Chignell et al., 2015). Lexical processing fluency was associated with high within-session and between-session reliabilities (see Table 3), and also showed strong correlations (in the correct directions) with RT and accuracy.

Of course, one could ask if there is any independent evidence that fluency actually taps Chinese proficiency, given the very weak correlation (r = .086) between fluency and self-rated language proficiency. Fortunately, we have some indirect evidence that speaks to this question. Specifically, the English Lexicon Project (Balota et al., 2007) contains lexical decision data for approximately 800 participants, along with their Shipley (1940) vocabulary scores; vocabulary knowledge serves as a proxy for language proficiency. We computed fluency scores for these participants, and found a strong positive correlation of .542 between fluency and vocabulary scores, yielding converging evidence that lexical processing fluency can be validly used to tap the efficiency of an individual's word recognition processes.

Individual Differences in Chinese Lexical Processing

After having established the reliability of word recognition measures, and equipped with a reliable and valid index of lexical processing fluency, we next examined how sensitivity to different word properties in Chinese lexical processing was moderated by fluency. Interestingly, more fluent participants produced larger effects of word frequency and semantic transparency. This pattern is inconsistent with the intuition that highly proficient Chinese readers are simply less sensitive to all word dimensions. Instead, our results point to a more complex situation where participants with higher quality representations are more sensitive to certain dimensions.

Specifically, participants who are more fluent lexical processors can better capitalize on their word knowledge to emphasize the processing of lexical properties that are most task-relevant. In the lexical decision task, the objective is to discriminate familiar/meaningful words from unfamiliar/meaningless nonwords (Balota & Chumbley, 1984). To effectively carry out this task, participants should attend to word properties that are diagnostic of a letter string's familiarity or meaningfulness. In the present study, nonwords were generated through the random recombination of characters within the word set, ensuring that the character-level properties (e.g., character frequency, stroke count, phonological consistency) for words and nonwords are perfectly matched. However, compound words are higher in word frequency than compound nonwords; nonwords, which do not occur in the language, by definition have a word frequency of 0. Similarly, semantic transparency reflects the degree of relatedness between a compound word and its constituent characters. Nonwords are necessarily more semantically opaque than words, since they, by definition, do not possess meaning and therefore can never be related to their constituent characters. It is worth noting that the moderating influence of individual differences was observed for the *only* two properties that are systematically different between words and nonwords.

In sum, our data support the hypothesis that more proficient Chinese readers show enhanced sensitivity to task-relevant word dimensions (i.e., word frequency and semantic transparency) that allow them to discriminate more efficiently between words and nonwords. Yap et al. (2012) found a similar pattern in English wherein vocabulary knowledge was positively correlated with participant-level frequency effects in lexical decision, but not speeded pronunciation. In addition, a more recent study by Pexman and Yap (2018) reported that readers with more vocabulary knowledge were more sensitive to concreteness but less sensitive to word frequency in a semantic categorization task (is a word concrete or abstract?). Broadly speaking, these findings are consistent with an adaptive and flexible lexical processing framework in which attention can be strategically deployed towards word features that allow performance to be optimized on a given task (Balota, Paul, & Spieler, 1999; Balota & Yap, 2006).

That being said, we were also expecting more fluent readers to show smaller effects of task-*irrelevant* dimensions such as character frequency or number of strokes. This is predicted by the perspective that as readers become more skilled, they rely to a greater extent on automatized lexical processing mechanisms, and consequently show attenuated sensitivity to lexical dimensions (Stanovich, 1980). One might also expect readers with high-quality representations to process compound words as a whole word unit, rather than as discrete characters (e.g., Caramazza et al., 1988, Augmented Addressed Morphology model). However, it is quite clear from our results (see Table 5) that the task-irrelevant word properties did *not* interact with fluency. The reasons for this pattern are not entirely clear, but it is possible that the lexical processing fluency measure may not be sensitive or nuanced enough to detect the effects of interest.

Andrews (2012) has pointed out that even though different aspects of skilled reading can be tapped by reading comprehension, vocabulary knowledge, and spelling ability, spelling ability remains the optimal index of the quality of an individual's lexical representations (Perfetti, 1992). While other reading tasks (e.g., reading comprehension) can be supported by partial lexical information and context, accurate performance on a spelling task depends on highly precise lexical representations (Andrews, 2012). Future individual differences studies in Chinese should collect a wider range of performance-based measures, in particular spelling ability, vocabulary knowledge, and reading comprehension; these are likely to provide deeper insights into the interplay between lexical processing and individual differences. Another intriguing possibility is that, unlike English (Yap et al., 2012), more fluent Chinese lexical processors do not show attenuated sensitivity to word properties (e.g., number of strokes). This surprising pattern of results should be followed up in future research.

Conclusion, Limitations, and Future Directions

The Chinese word recognition literature has been almost exclusively dominated by the study of group-level data. Using well-characterized trial-level lexical decision megastudy data for 594 Cantonese-speaking participants responding to almost 1.7 million trials, we established the reliability of word and nonword recognition, and lexical processing fluency measures in Chinese. Our findings also reinforce the value of finer-grained insights afforded by analyses of individual differences. In particular, we found support for flexible lexical processing, whereby highly skilled lexical processors are better able at tuning their attentional systems to lexical dimensions that are most diagnostic of the required decision.

Of course, there are some limitations associated with the present work. Due to range restriction and possibly its self-report nature, self-rated language proficiency did not seem to be a good predictor of lexical decision performance, and future investigations ought to supplement self-rated proficiency with alternative measures such as a combination of reading comprehension, vocabulary knowledge, and spelling ability. College students are selected for their vocabulary knowledge and language fluency; this restriction of range makes it more than likely that we are *underestimating* the strength of the relationship between fluency and word recognition performance.

Finally, our analyses focused on the processing of two-character Cantonese Chinese words in the lexical decision task. Moving forward, future work should consider the interplay between individual differences and word recognition performance, across a wider range of word lengths, lexical properties (http://www.chineselexicaldatabase.com; Sun, Hendrix, Ma, & Baayen, 2018), writing systems (simplified vs. traditional characters), dialect (Mandarin vs. Cantonese), paradigms (speeded pronunciation vs. lexical decision), and individual differences.

References

Andrews, S. (2012). Individual differences in skilled visual word recognition and reading: The role of lexical quality. In J. S. Adelman (Ed.), *Visual word recognition: Meaning and context, individuals and development* (pp. 151-172). New York, NY: Psychology Press.

Baayen, R. H., Feldman, L. F., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory & Language, 53*, 496-512.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance*, *10*, 340-357.

Balota, D. A., & Yap, M. J. (2006). Attentional control and flexible lexical processing: Explorations of the magic moment of word recognition. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 229-258). New York: Psychology Press.

Balota, D. A., Paul, S., & Spieler, D. H. (1999). Attentional control of lexical processing pathways during word recognition and reading. In S. Garrod & M. Pickering (Eds.), Language processing (pp. 15-57). Hove, UK: Psychology Press.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

Balota, D. A., Yap, M. J., Hutchison, K.A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.). *Visual word recognition* (pp. 90-115). Hove: Psychology Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48.

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition, 7*, 68-76.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28*, 297-332.

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition, 28*, 143-153.

Chen, H.-C., Vaid, J., & Wu, J.-T. (2009). Homophone density and phonological frequency in Chinese word recognition. *Language & Cognitive Processes, 24*, 967–982.

Chignell, M., Tong, T., Mizobuchi, S., Delange, T., Ho, W., & Walmsley, W. (2015). Combining multiple measures into a single figure of merit. *Procedia Computer Science, 69*, 36-43.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249–253.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychological Review, 108, 204–256.

Diependaele, K., Brysbaert, M., & Neri, P. (2012). How noisy is lexical decision? *Frontiers in Psychology*, 3:348.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125*, 777-799.

Fox, J., Weisberg, S., Friendly, M., Hong, J., Andersen, R., Firth, D., & Taylor, S. (2015). effects: Effect displays for linear, generalized linear, and other models. *R package version 3.0-0*. http://cran.r-project.org/ package=effects

Institute of Language Teaching and Research. (1986). *A frequency dictionary of Modern*

Chinese. Beijing: Beijing Language Institute Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R package version 2.0-29*. https://cran.r-project.org/web/packages/lmerTest/

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323.

Li, C. N., & Thompson, S. A. (2009). Chinese. In B. Comrie (Ed.), The world's major languages 2nd edition (pp. 703–723). Abingdon: Routledge.

Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods, 39*, 192-198.

Lowe, C., & Rabbitt, P. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. *Neuropsychologia, 36*, 915-923.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19-40.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, *7*, 308-313.

Peng, D. L., & Wang, C. (1997). 汉字加工的基本单元:来自笔画数效应和部件数效应的证据[Basic processing unit of Chinese character recognition: Evidence from stroke number effect and radical number effect]. *Acta Psychologica Sinica, 29*, 8-16.

Peng, D. L., Liu, Y., & Wang, C. (1999). How is access representation organized? The relation of polymorphemic words and their morphemes in Chinese. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 65-89). NJ: Lawrence Erlbaum Associates.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L.

C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145-174). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C. A., & Dunlap, S. (2008). Learning to read: General principles and writing systems variations. In K. Koda and A. M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first- and second-language literacy development*. New York: Routledge.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam: John Benjamins.

Perfetti, C. A., & Liu, Y. (2006). Reading Chinese characters: Orthography, phonology, meaning, and the Lexical Constituency Model. In P. Li (Series and Vol. Ed.), L. H. Tan, E. Bates, & O. J. L. Tzeng (Vol. Eds.), Handbook of East Asian psycholinguistics: Vol. 1. Chinese (pp. 225–236). Cambridge, UK: Cambridge University Press.

Pexman, P. M., & Yap, M. J. (2018). Individual differences in semantic processing: Insights from the Calgary Semantic Decision Project. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *44*, 1091-1112.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56–115.

R Development Core Team. (2015). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*, 446-461.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition. In J. E. Birren

& K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 400-426). New York: Van Nostrand Reinhold.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an “outlier” orthography. *Psychological Bulletin, 134*, 584-615.

Shen, L. M., & Zhu, X. (1994). 汉字识别中笔画数与字频效应的研究[Investigating stroke count and character frequency effects in the visual recognition of characters]. *Psychological Science, 17*, 245–247.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9,* 371-377.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*, 32-71.

Sun, C. C., Hendrix, P., Ma, J. Q., & Baayen, R. H. (2018). Chinese Lexical Database (CLD): A large-scale lexical database for simplified Mandarin Chinese. *Behavior Research Methods, 50,* 2606-2629.

Sze, W. P., Yap, M. J., & Rickard Liow, S. J. (2015). The role of lexical variables in the visual recognition of Chinese characters: A megastudy analysis. *Quarterly Journal of Experimental Psychology, 68*, 1541-1570.

Taft, M., Liu, Y., & Zhu, X. (1999). Morphemic processing in reading Chinese. In J. Wang, A. Inhoff, & H. C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 91-113). NJ: Lawrence Erlbaum Associates.

Tan, L. H., & Perfetti, C. A. (1999). Phonological activation in visual identification of Chinese two-character words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 382-393.

Tsang, Y. K., Huang, J., Lui, M., Xue, M., Chan, Y. F., Wang, S., & Chen H. C. (2018).

MELD-SCH: A megastudy of lexical decision in simplified Chinese. *Behavior Research Methods, 50*, 1763-1777.

Tse, C.-S., & Yap, M. J. (2018). The role of lexical variables in the visual recognition of two-character Chinese compound words: A megastudy analysis. *Quarterly Journal of Experimental Psychology, 71,* 2022-2038.

Tse, C.-S., Yap, M. J., Chan, Y. L., Sze, W. P., Shaoul, C., & Lin, D. (2017). The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods*, *49*, 1503-1519.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception & Performance*, *38*, 53-79.

Yap, M. J., Balota, D. A., Tse, C.-S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34*, 495-513.

Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 41*, 597-613.

Yap, M. J., Tse, C.-S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency: The role of lexical integrity. *Journal of Memory & Language, 61*, 303-325.

Zhou, X., & Marslen-Wilson, W. (1999). Phonology, orthography, and semantic activation in reading Chinese. *Journal of Memory & Language*, *41*, 579-606.

Ziegler, J. C., Tan, L. H., Perry, C., & Montant, M. (2000). Phonology matters: The phonological frequency effect in written Chinese. *Psychological Science, 11*, 234-238.

Acknowledgements

This research was carried out as an independent research project by R. L. under the direction of M. J. Y. Portions of this research were presented at the 58th Annual Meeting of the Psychonomic Society (November 2017, Vancouver, Canada). This research was based on Tse et al. (2017), and supported by the General Research Fund (#451113) awarded by Hong Kong Research Grants Council. We are grateful for the comments provided by Steve Lupker and anonymous reviewers on an earlier version of this paper. The dataset for this project can be accessed at https://osf.io/k98z7/.

Table 1. *Rotated Component Matrix with Component Loadings.*

	Component	
	1	2
Knowledge of Chinese Characters	.847	.085
Knowledge of Spoken Cantonese	.828	.041
Entrance Score for Chinese Language	.147	.721
Cumulative Grade Point Average	-.250	.702
Time spent reading Chinese Materials	.181	.359

Table 2. *Means, Standard Deviations, Accuracy, Ex-Gaussian Parameters, and d' as a Function of Trial Type*.

	Overall	Day 1	Day 2	Day 3	Odd	Even
M_{word}	657	683	650	635	656	657
SD_{word}	165	174	154	150	164	165
μ_{word}	500	514	502	492	500	500
σ_{word}	48	48	45	45	48	48
τ_{word}	157	169	148	143	156	157
$\text{Accuracy}_{\text{word}}$	0.88	0.90	0.88	0.87	0.88	0.88
M_{nonword}	727	769	716	693	727	726
SD_{nonword}	189	200	175	167	190	189
μ_{nonword}	548	577	550	535	548	548
σ_{nonword}	56	59	52	50	56	55
τ_{nonword}	179	192	166	158	179	178
$\text{Accuracy}_{\text{nonword}}$	0.90	0.90	0.90	0.89	0.90	0.90
d'	2.56	2.63	2.59	2.55	2.57	2.56

Table 3. *Correlations Between Day 1, Day 2, Day 3, Odd-, and Even-Numbered Trial Parameters.*

	D1-D2	D2-D3	D1-D3	O-E
M_{word}	.790	.818	.722	.991
SD_{word}	.754	.757	.670	.973
μ_{word}	.784	.788	.672	.975
σ_{word}	.517	.662	.522	.896
τ_{word}	.752	.775	.682	.966
$\text{Accuracy}_{\text{word}}$	.761	.781	.683	.941
M_{nonword}	.805	.827	.714	.994
SD_{nonword}	.761	.786	.677	.983
μ_{nonword}	.813	.795	.704	.981
σ_{nonword}	.542	.608	.439	.911
τ_{nonword}	.754	.792	.671	.978
$\text{Accuracy}_{\text{nonword}}$	.683	.805	.645	.969
d’	.814	.814	.739	.967
Fluency	.794	.818	.735	.974

All correlations are statistically significant, $p < .001$

Table 4. *Intercorrelations Between Predictor Variables (N = 18,983).*

	RT	zRT	Stroke Count	Character Frequency	Word Frequency	Homophone Density	Semantic Transparency	Phonological Consistency (C1)	Phonological Consistency (C2)
RT	—								
zRT	.411 ***	—							
Stroke Count	.058 ***	.119 ***	—						
Character Frequency	-.106 ***	-.278 ***	-.261 ***	—					
Word Frequency	-.236 ***	-.557 ***	-.093 ***	.328 ***	—				
Homophone Density	-.005	-.007	-.002	-.080 ***	-.017 *	—			
Semantic Transparency	.013	.022 **	.202 ***	-.264 ***	-.116 ***	.004	—		
Phonological Consistency (C1)	.008	.000	.057 ***	-.080 ***	-.020 **	.046 ***	.074 ***	—	
Phonological Consistency (C2)	-.003	.007	.063 ***	-.075 ***	-.043 ***	.046 ***	.063 ***	.019 *	—

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

Table 5. *LME Model Estimates (based on zRT) for the* ***Main*** *and* ***Joint*** *Effects of Word Properties and Fluency for Chinese Compound Words. C1 = Character 1; C2 = Character 2.*

Effect	Variance	SD
Random effects		
Items		
Intercept	0.0320	0.1788
Participants		
Intercept	0.0003	0.0168
Stroke Count	0.0000	0.0053
Character Frequency	0.0052	0.0720
Word Frequency	0.0018	0.0420
Homophone Density	0.0000	0.0005
Semantic Transparency	0.0007	0.0256

	Coefficient	Standard error	*p* value
Fixed effects			
Intercept	-0.1597	0.0031	<.001
Stroke Count	0.0043	0.0005	<.001
Character Frequency	-0.1066	0.0055	<.001
Word Frequency	-0.1803	0.0027	<.001
Homophone Density	-0.0006	0.0001	<.001
Semantic Transparency	-0.0666	0.0045	<.001
Phonological Consistency (C1)	-0.0128	0.0031	<.001
Phonological Consistency (C2)	-0.0130	0.0031	<.001
Fluency	-0.0114	0.0019	<.001

	Coefficient	Standard error	*p* value
Fixed effects			
Stroke Count × Fluency	-0.0009	0.0006	*ns*
Character Frequency × Fluency	0.0049	0.0069	*ns*
Word Frequency × Fluency	-0.0378	0.0033	<.001
Homophone Density × Fluency	0.0000	0.0001	*ns*
Semantic Transparency × Fluency	-0.0179	0.0042	<.001
Phon Consistency (C1) × Fluency	-0.0014	0.0030	*ns*
Phon Consistency (C1) × Fluency	-0.0036	0.0030	*ns*

Table 6. *LME Model Estimates (based on raw RT) for the* ***Main*** *and* ***Joint*** *Effects of Word Properties and Fluency for Chinese Compound Words. C1 = Character 1; C2 = Character 2.*

Effect	Variance	SD
Random effects		
Items		
Intercept	1652.00	40.64
Participants		
Intercept	2617.00	51.16
Stroke Count	1.90	1.38
Character Frequency	213.00	14.59
Word Frequency	127.80	11.30
Homophone Density	0.02	0.15
Semantic Transparency	13.28	3.64

	Coefficient	Standard error	*p* value
Fixed effects			
Intercept	645.93	2.24	<.001
Stroke Count	1.07	0.13	<.001
Character Frequency	-22.64	1.23	<.001
Word Frequency	-41.20	0.67	<.001
Homophone Density	-0.14	0.03	<.001
Semantic Transparency	-15.11	1.04	<.001
Phonological Consistency (C1)	-2.93	0.74	<.001
Phonological Consistency (C2)	-2.83	0.73	<.001
Fluency	-70.36	3.75	<.001

	Coefficient	Standard error	*p* value
Fixed effects			
Stroke Count × Fluency	-0.42	0.16	.007
Character Frequency × Fluency	8.44	1.50	<.001
Word Frequency × Fluency	4.23	0.94	<.001
Homophone Density × Fluency	0.03	0.04	*ns*
Semantic Transparency × Fluency	0.13	1.12	*ns*
Phon Consistency (C1) × Fluency	0.83	0.76	*ns*
Phon Consistency (C1) × Fluency	1.00	0.76	*ns*

Table 7. *Effects of Stroke Count, Character Frequency, and Word Frequency as a Function of Lexical Processing Fluency. Raw Response Times Reported in Milliseconds.*

	Fluency (Centered)				
Variables	-1	-.5	0	.5	1
Low Stroke Count	699	666	633	600	566
High Stroke Count	728	691	653	616	579
Stroke Count Effect	28	24	20	16	12
Low Character Frequency	805	758	710	662	614
High Character Frequency	701	667	634	600	567
Character Frequency Effect	105	90	76	62	47
Low Word Frequency	805	765	725	685	645
High Word Frequency	621	590	560	529	499
Word Frequency Effect	184	175	165	156	147

Figure Captions

Figure 1. Frequency distributions for measures of individual differences and word recognition performance, and scatterplots reflecting relationships between these variables.

Figure 2. Distributions of random slopes across participants as function of lexical property.

Figure 3. Statistically significant interactions of lexical processing fluency with character and word properties (based on z-scored RTs).

Figure 4. Statistically significant interactions of lexical processing fluency with character and word properties (based on raw RTs).

Figure 1.

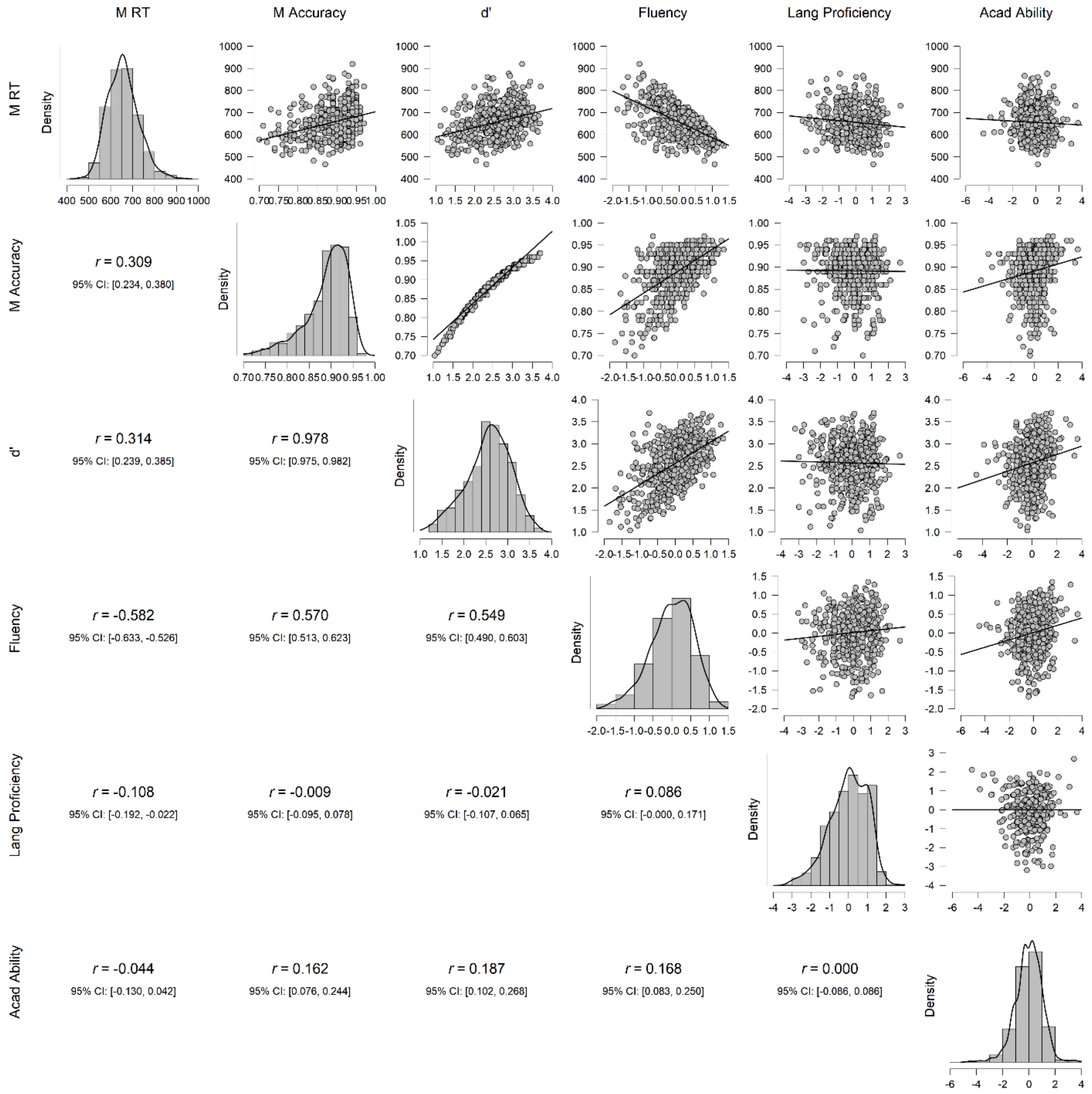

M RT
M Accuracy
d'
Fluency
Lang Proficiency
Acad Ability
Density
r = 0.309
95% CI: [0.234, 0.380]
r = 0.314
95% CI: [0.239, 0.385]
r = 0.978
95% CI: [0.975, 0.982]
r = -0.582
95% CI: [-0.633, -0.526]
r = 0.570
95% CI: [0.513, 0.623]
r = 0.549
95% CI: [0.490, 0.603]
r = -0.108
95% CI: [-0.192, -0.022]
r = -0.009
95% CI: [-0.095, 0.078]
r = -0.021
95% CI: [-0.107, 0.065]
r = 0.086
95% CI: [-0.000, 0.171]
r = -0.044
95% CI: [-0.130, 0.042]
r = 0.162
95% CI: [0.076, 0.244]
r = 0.187
95% CI: [0.102, 0.268]
r = 0.168
95% CI: [0.083, 0.250]
r = 0.000
95% CI: [-0.086, 0.086]
400 500 600 700 800 900 1000
0.70 0.75 0.80 0.85 0.90 0.95 1.00
1.0 1.5 2.0 2.5 3.0 3.5 4.0
-2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
-4 -3 -2 -1 0 1 2 3
-6 -4 -2 0 2 4

Figure 2.

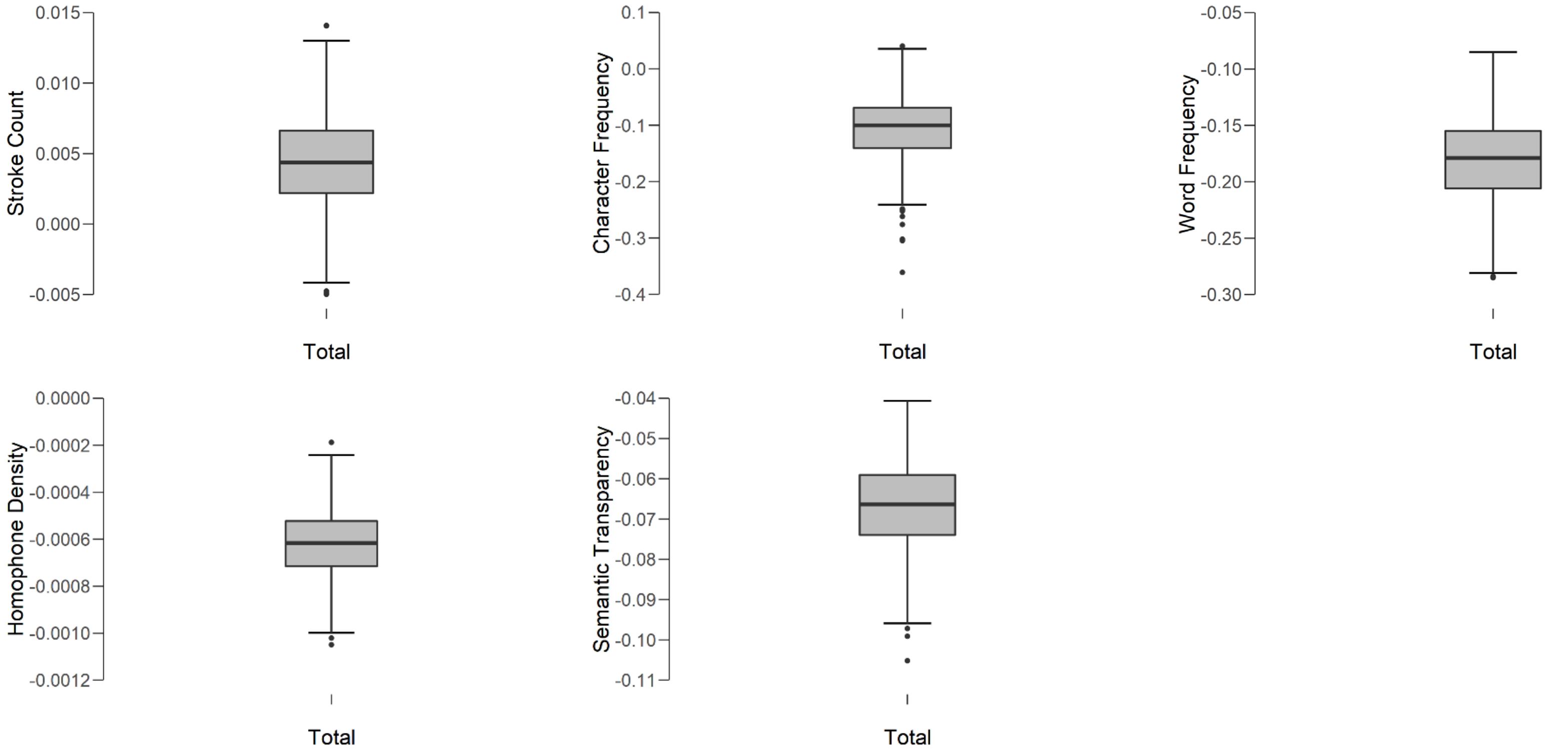

Stroke Count
0.015
0.010
0.005
0.000
-0.005
Total
Character Frequency
0.1
0.0
-0.1
-0.2
-0.3
-0.4
Total
Word Frequency
-0.05
-0.10
-0.15
-0.20
-0.25
-0.30
Total
Homophone Density
0.0000
-0.0002
-0.0004
-0.0006
-0.0008
-0.0010
-0.0012
Total
Semantic Transparency
-0.04
-0.05
-0.06
-0.07
-0.08
-0.09
-0.10
-0.11
Total

Figure 3.

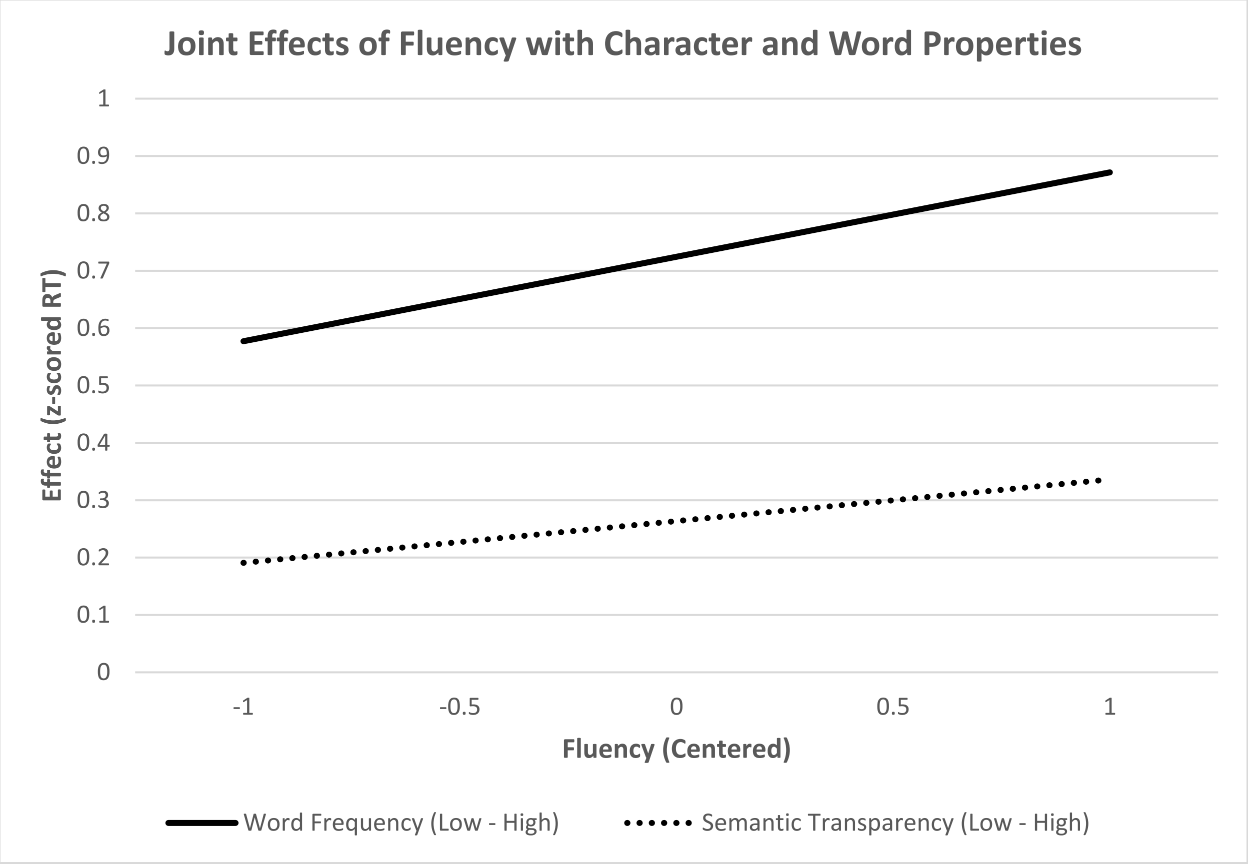
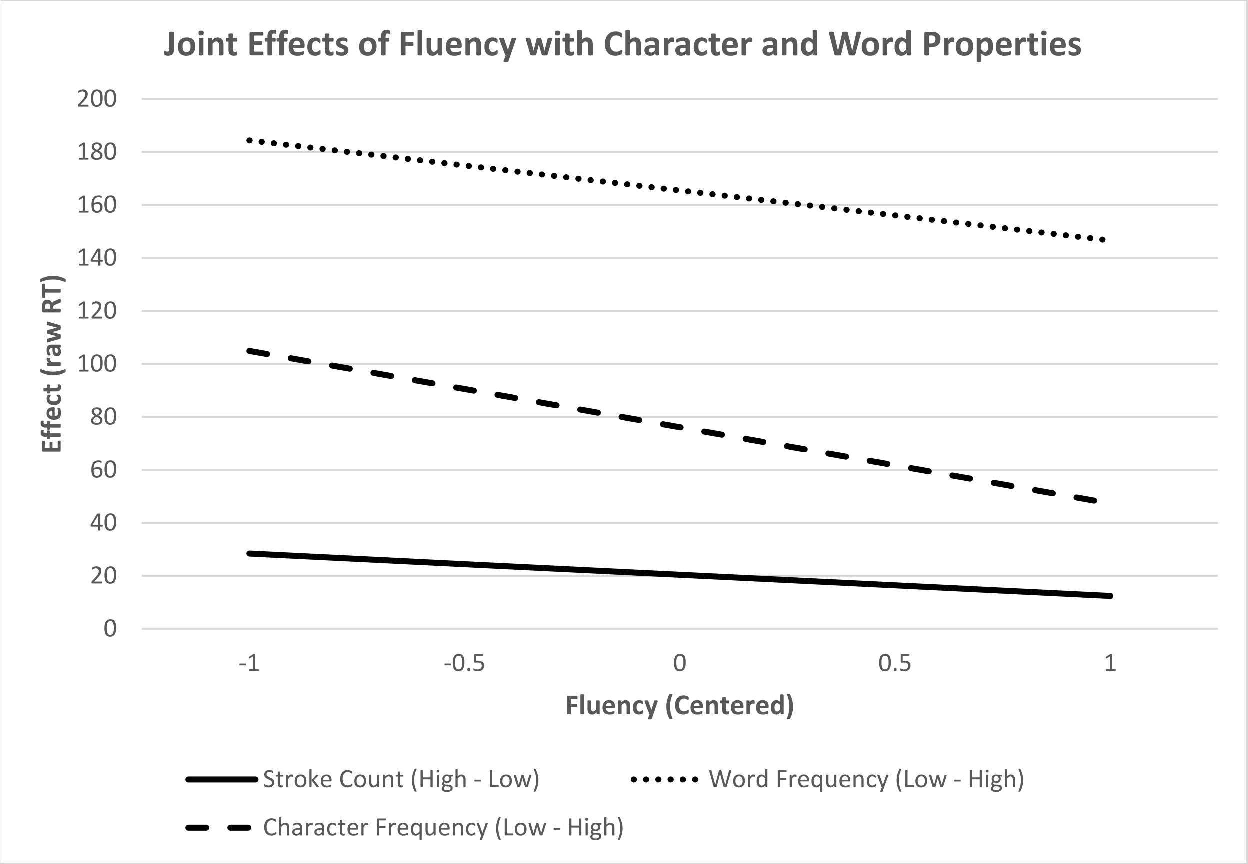

Joint Effects of Fluency with Character and Word Properties
Effect (z-scored RT)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-1
-0.5
0
0.5
1
Fluency (Centered)
Word Frequency (Low - High)
Semantic Transparency (Low - High)

Figure 4.


Joint Effects of Fluency with Character and Word Properties
Effect (raw RT)
200
180
160
140
120
100
80
60
40
20
0
-1
-0.5
0
0.5
1
Fluency (Centered)
Stroke Count (High - Low)
Word Frequency (Low - High)
Character Frequency (Low - High)