VISUAL ORTHOGRAPHIC VARIATION AND LEARNING TO READ ACROSS WRITING SYSTEMS

by

Li-Yun Chang

B.Ed., National Taiwan Normal University, 2008

M.Ed., National Taiwan Normal University, 2010

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts & Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS & SCIENCES

This dissertation was presented

by

Li-Yun Chang

It was defended on

January 8th, 2015

and approved by

Julie A. Fiez, Professor, Department of Psychology

David C. Plaut, Professor Department of Psychology, Carnegie Mellon University

Natasha Tokowicz, Associate Professor, Department of Psychology

Dissertation Advisor: Charles A. Perfetti, University Professor of Psychology

VISUAL ORTHOGRAPHIC VARIAION AND LEARNING TO READ ACROSS WRITING SYSTEMS

Li-Yun Chang, PhD

University of Pittsburgh, 2015

This research examined the extent to which visual characteristics of orthographies affect learning to read within and across writing systems, with an eye toward the role of mapping principles – the manner in which graphemes map to linguistic units (e.g., phonemes, syllables, and morphemes) in this process. Study 1 explained visual orthographic variation by developing a measurement system to quantify complexity of graphemes in 131 orthographies. The results show that grapheme complexity varies across writing systems and that this variation is driven by grapheme inventory, a consequence of mapping principles. Next, we questioned how visual orthographic variation impacts individuals' perceptual learning of graphemes – one of the initial stages of learning to read. Study 2 tested the degree to which mastering first-language (L1) graphemes with different complexities affects visual perceptual discrimination for individuals using different mapping principles (Online cross-writing-system experiment; eight participant groups: Hebrew, English, Russian, Arabic, Hindi, Telugu, Japanese, and Chinese, n = 60, respectively) and individuals using the same mapping principle (Lab within-writing-system experiment: simplified vs. traditional Chinese, n = 60, respectively). Consistent results from both experiments show that discrimination difficulty is a function of grapheme stimulus complexity itself as well as its relationship to the complexity of participants' L1, regardless of mapping principles. These results were confirmed in Study 3, in which we developed a universal orthographic neural network encoder focus on statistical properties of visual patterns to simulate human behaviors. We trained each of 131 identical encoders to learn the structure of a different

orthography; a strong, positive association was found between grapheme complexity and network learning difficulty. Taken together, our results suggest that visual orthographic variation, encompassing both grapheme complexity and grapheme inventory required for orthographic mastery, affects visual discrimination processing of graphemes; these complexity effects are driven significantly, but not absolutely, by mapping principles across writing systems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

To those who have tread softly,

I offer my sincerest appreciation for your guidance, care, and support throughout my journey in pursuit of self-actualization.

“Had I the heavens’ embroidered cloths,
Enwrought with golden and silver light,
The blue and the dim and the dark cloths
Of night and light and the half-light,
I would spread the cloths under your feet:
But I, being poor, have only my dreams;
I have spread my dreams under your feet;
Tread softly because you tread on my dreams.”

\- *W.B. Yeats (1865–1939)*

"He Wishes for the Cloths of Heaven"

1.0 INTRODUCTION

Our world's writing systems span quite a massive range of variety; not surprisingly, their visual forms vary greatly. Why are some writing systems more visually complex than others? How does variability in the visual characteristics of graphemes, the smallest writing unit with implications for meaning, affect the process of learning to read across writing systems? What is the impact of grapheme mastery in one language on the approach to graphemes in another language – specifically, do individuals using different writing systems perceive graphemes differently? These are key questions that this dissertation research addressed.

We aimed to examine whether – and how – visual orthographic variation (i.e., the visual characteristics of orthography) affects the process of learning to read, among learners both within and comparatively across writing systems. Three studies were designed to achieve this aim: a content analysis of graphemes from across five writing systems, a behavioral study comparing individuals' perceptual processing of graphemes of varying complexity, and a computational modeling study serving as a demonstrative explanation of how visual orthographic variation affects learning. We begin this dissertation work by reviewing cross-writing-systems research, posing broader research questions, sharpening specific aims, introducing each study with relating work, method, results, interim summary, and then summarizing key findings in a general discussion.

1.1 BACKGROUND

1.1.1 Writing system variation

The manner in which writing systems convey meaning, in terms of mapping to spoken language, is highly varied. Many reading studies have involved in-depth discussions on how writing systems vary along several dimensions – phonological grain size (Ziegler & Goswami, 2005; 2006), orthographic depth (Katz & Frost, 1992), semantic transparency (Wydell, 2012), visual symbols set (Nag, 2014), and how such writing system variation can have an impact on learning to read, as compared within and across writing systems (e.g., Frost, 2012; Perfetti & Harris, 2013; Perfetti, Liu, Fiez, Nelson, Bolger, & Tan, 2007; Seidenberg, 2011).

In reading research, writing systems are generally grouped into three categories (Gelb, 1963), each delineated by mapping principles, i.e. the manner of correspondence between graphemes and linguistic units (e.g., phonemes, syllables, or morphemes). Systems in which graphemes map to phonemes, lower-level phonological units (i.e., systems with low mapping level), are known as *alphabetic* writing systems – there exist three subtypes of alphabetic systems, differing in their representations of vowels. In true alphabets (e.g., Finnish, English, and Korean), graphemes each map to independent and equal representations of consonants or vowels. In abjads (or consonantal writing systems; e.g., Hebrew and Arabic), although graphemes also map to consonants and vowels, vowels are marked by secondary diacritics – these are visually less prominent than primary consonant graphemes, and are not independent representations; however, these diacritics are generally left unmarked in practice. In alphasyllabaries (e.g., Hindi and Kannada), whole syllables are generally written with consonant-vowel graphemes combined to form symbol blocks called akshara, using vowel diacritics attached to consonants; however, in contrast to abjads, vowels in alphasyllabaries can be present as independent symbols, quasi-independent symbols, or not at all – for example, in Hindi, the phoneme /a/ is considered inherent when pronouncing any consonant grapheme, and thus is not explicitly written. Of systems using higher-level phonological mapping, where graphemes map to full syllables, there are two types: *syllabaries* (e.g., Japanese hiragana and katakana), in which graphemes only represent whole syllables (higher mapping level), and *morphosyllabaries* (e.g., Japanese Kanji and Chinese), in which graphemes can represent syllables and whole morphemes (highest mapping level) (DeFrancis, 1989). A notable aspect of these categories is that their boundaries, when examined at the level of mapping between graphemes and their corresponding linguistic units in individual orthographies, distinctly overlap. The above example of written vowel omission in Hindi provides a case in which some alphasyllabic graphemes could be categorized as syllabic (e.g. written "k" pronounced /ka/) whereas others could be categorized as alphabetic (in cases of quasi-independent vowel graphemes), and some syllabic graphemes representing only independent vowels could also be categorized as alphabetic (e.g. written "a" in Japanese hiragana). These overlaps illustrate that mapping principle alone may be insufficient for categorizing orthographies in examining reading differences across writing systems.

The differences among these five writing systems are generally captured by the morpho-phonological dimensions of the languages that use them (Frost, 2012). Recently, some reading scholars have proposed that these differences can be equivalently captured by another dimension, visual symbol set (Nag, 2011; 2014; Nag, Caravolas, & Snowling, 2011). These scholars place writing systems on a continuum, split between "contained" and "extensive" sides, in terms of the number of visual symbols that a given system requires. On the contained side are alphabets requiring fewer symbols (24-36); on the extensive side are morphosyllabaries, which require greater numbers of symbols (Chinese: over 2500); alphasyllabaries (200-500) fall between these two extremes.

To capture how five major writing systems vary along these multiple dimensions – morpho-phonology and visual symbol set - we illustrate their relative positions in Figure 1. Generally, as mapping level increases, the number of visual symbols also increases. In interpreting this covariance, we speculate that it is mapping principle that drives number of visual symbols (as opposed to the reverse) because spoken language is generally thought to have existed long before written language.

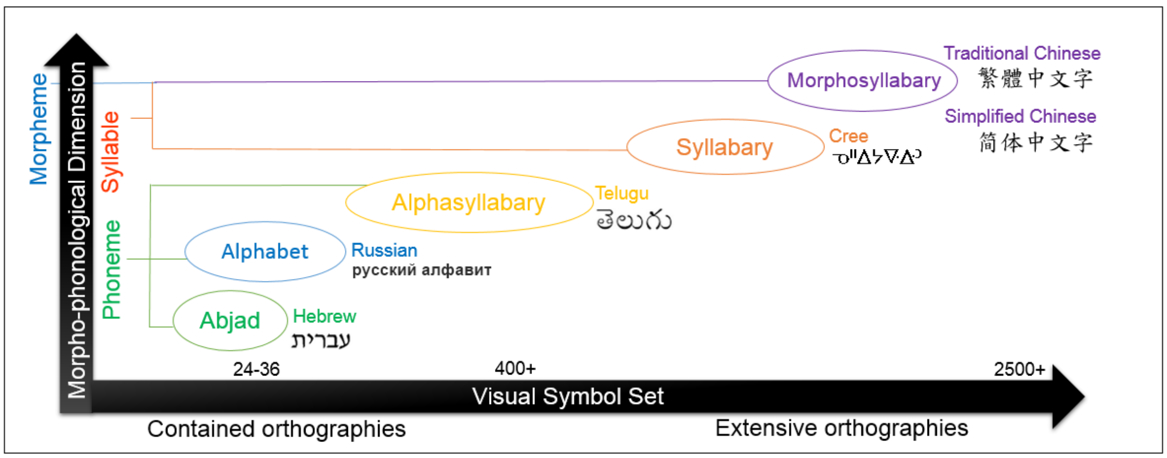

Figure 1. Illustration of writing variety across the five world's major writing systems

1.1.2 Writing system variation and learning to read

Several key terms distilled from thorough review of prior research into the comparative processes of learning to read across writing systems are used in the current research. We introduce these terms here: writing systems, script, and orthographies. The following definitions are provided so as to avoid any confusion on the part of our readers. Writing systems are defined as larger families of written language, delineated by the linguistic units represented by their graphemes (e.g., abjads: consonants; alphabet: phonemes; alpha-syllabaries: consonant-vowel units; syllabaries: syllables; morpho-syllabaries: syllables and morphemes; Cook & Bassetti, 2005). Scripts are visual forms of writing (Perfetti, Liu, Fiez, Nelson, Bolger, & Tan, 2007). A written language can be presented in many scripts; for example, cursive or typeface (e.g., "*font*" and "font" in written English). Orthographies, different from scripts, are the implementations of writing systems used by specific languages (Perfetti et al., 2007). Whereas writing systems are categorized by level of mapping to linguistic units, orthographies are implemented in varying ways depending on their "parent" writing system. The terms of such implementation are determined by the following, somewhat detailed paradigm. Within the morpho-phonological dimension, in abjad, alphabet, and alphasyllabary writing systems, orthographies rely on grapheme-phoneme correspondence (GPC); in syllabaries or morphosyllabaries, orthographies rely on grapheme-syllable correspondence (Scheerer, 1986). This correspondence, i.e., the level at which graphemes map to phonological units, is described in reading literature as a continuum between "transparent" (or shallow; one-to-one mapping) and "opaque" (or deep; one-to-many, many-to-one, or many-to-many mappings). Within the visual symbol set dimension, orthographies are delineated by number of graphemes (or grapheme inventory).

How do writing system variations affect learning to read across writing systems? There is no simple answer, because writing systems vary along non-orthogonal dimensions, limiting ease of comparison, and there are many apparent trade-offs among relevant features of each dimension when relationships between dimensions are closely examined (for reviews, see Frost, 2012; Hirshorn & Fiez, 2014; Perfetti & Harris, 2013; Seidenberg, 2011). Although a growing body of research has investigated how learning to read is influenced by grapheme-linguistic unit mapping correspondence (e.g., Ellis, Natsume et al., 2004; Perfetti, Liu, & Tan, 2005; Perfetti, Zhang, & Berent, 1992), little attention has been given to graphemes *per se*, or visual characteristics of orthography, in terms of their role in learning differences across the world's wide variety of written language.

However, and importantly, accurate, stable orthographic representations are required for associations to be reliably learned between visual forms and aspects of spoken language in order for skilled reading to be achieved (Perfetti & Hart, 2002). Visual complexity of orthographies itself may constrain efficient development of these orthographic representations, thus contributing to difficulty in learning to read. Moreover, orthographies with visually complex graphemes are also likely to contain a larger grapheme set, providing an additional source of difficulty during learning (e.g., Nag, 2011; Nag, Treiman & Snowling, 2010). Therefore, an investigation into differences in learning to read across writing systems that fails to consider visual characteristics of orthography may result in a misleading conclusion that would be difficult to generalize.

1.1.3 Visual orthographic variation and learning to read

We categorize visual orthographic variation among two levels: grapheme and orthography. At the grapheme level, variation concerns the visual characteristics of individual graphemes; at the orthography level, variation deals with the number of graphemes contained within a given orthography. We are particularly interested in how such multi-layered visual orthographic variation impacts the process of learning to read, as compared across writing systems.

The visual demands of grapheme processing can pose a significant challenge to beginning learners. Empirical studies covering a wide range of orthographies have demonstrated that grapheme complexity is negatively correlated with grapheme identification efficiency (Liu, Chen, Liu, & Fu, 2012; Pelli, Burns, Farell, & Moore-Page, 2006). These studies suggest that more complex graphemes impose greater demands on visual perceptual processing as learners attempt to develop robust orthographic representations.

Learners are further challenged in mastering the complete inventory of graphemes in their own orthography, the size of which varies across orthographies. In alphabetic orthographies (average grapheme inventory: 20-30) such as Finnish, children master all graphemes after first grade (Seymour, Aro, & Erskine, 2003; White, Grave, & Slater, 1990); in alphasyllabic orthographies (average grapheme inventory: 400) such as Kannada, children require three to four years of formal instruction to master all graphemes (Nag, 2007); in morphosyllabic orthographies (average grapheme inventory: > 1800) such as Chinese, Japanese Kanji, and Korean Hanja, children continue to learn novel characters after six years of formal education (Shu, Chen, Anderson, Wu, & Xuan, 2003, for Chinese; Tamaoka, Kirsner, Yanase, Miyaoka, & Kawakami, 2002, for Japanese Kanji; Cho & Chen, 1999, for Korean Hanja).

Reading orthographies with large inventories and more complex graphemes may require stronger visual perceptual skills and may, in turn, strengthen such skills. Tan, Spinks et al. (2005) found that early progress in reading Chinese was linked more to copying skills than to phonemic awareness. Moreover, McBridge-Chang, Zhou et al. (2011) reported a link between orthographic learning and general visuospatial skill in typically developing readers from orthographies of varying complexity. Children learning to read traditional Chinese, an orthography with highly complex graphemes (average 10 strokes per character; Huang & Hanley, 1995) outperformed children learning to read Spanish, an orthography with relatively simple graphemes (average 2.5 strokes per letter; Changizi & Shimojo, 2005), on a standardized visuospatial relationship task.

These findings highlight the implications of orthographic visual complexity for learning to read; however, such implications have not been specifically addressed in reading research, leaving our understanding limited.

1.2 RESEARCH QUESTIONS

This research seeks to advance our understanding about how the visual characteristics of orthographies affect learning to read across the world's wide range of writing systems. We pose several key questions on the front of this research:

1. How do writing systems handle variability in visual characteristics of graphemes?

   Specifically, what are the relationships among mapping principle, grapheme inventory (number of graphemes contained within a given orthography), and grapheme complexity?

2. What are the implications of visual orthographic variation on learning to read, from beginning learners to skilled readers, across writing systems?

   (1) Do more complex graphemes impose demands on perceptual processing?

   (2) Does increased grapheme inventory size hinder grapheme learning efficiency?

   (3) Does mastering more complex graphemes or larger grapheme inventories require higher-order visual skills, in turn strengthening such skills?

   Collectively, to what extent does the visual complexity of orthographies, encompassing both grapheme complexity and grapheme inventory size, affect visual perceptual processing of individuals within and across writing systems?

3. Given that orthographies map in different ways to phonemes, syllables, and morphemes, if the visual complexity of orthographies affects individuals' visual perceptual processing, to what extent does an orthography's mapping principle influence this visual processing?

1.3 SPECIFIC AIMS OF THE PRESENT RESEARCH

The overarching goal of this research is to examine the extent to which the visual orthographic variation impacts individuals' perceptual processing of graphemes – one of the initial stage of learning to read. Three specific aims are addressed in one content analysis, one behavioral study, and one computational modeling study.

Aim 1 is to study the relationships among mapping principle, grapheme inventory, and grapheme complexity by proposing a comprehensive measurement system to quantify complexity – both number of graphemes and complexity of individual graphemes, over 131 orthographies. In Study 1 (grapheme complexity quantification), we expected that mapping principle would govern grapheme inventory, which, in turn, would drive grapheme complexity.

Aim 2 is to examine the extent to which the visual orthographic variation affects visual perceptual processing in individuals, as compared within and across writing systems, and to do so by using the measurement system developed in Study 1 to systematically vary the complexities of grapheme stimuli and of participants' L1 orthographies. In Study 2 (behavioral experiment), an identical experimental design is applied to individuals using different mapping principles (Study 2A) and individuals using the same mapping principle (Study 2B), while complexities of L1 orthographies mastered by all individuals varied. Overall, we expected to find a complexity effect – grapheme discrimination efficiency should be subject to an interaction between the complexity of perceived stimuli and of participant L1 orthography, with an eye toward how mapping principle plays differing roles in Study 2A and Study 2B.

Aim 3 is to demonstrate a causal relationship between visual orthographic variation and grapheme perceptual learning across writing systems by developing a universal learning device that solely focuses on visual processing. In Study 3 (computational modeling), the first demonstration is to show that grapheme complexity leads to learning difficulty in mastering all graphemes in a given orthography; all 131 orthographies in Study 1 will be used. The second demonstration is to replicate Study 2A with results that can attributed to experience of orthography complexities only, without any input from mapping principle. We expected that both demonstrations would provide insights to clarify the relationship between the visual complexity of orthographies, mapping principle, and learning to read across writing systems.

Taken together, these three studies form a comprehensive narrative of the impact of visual complexity on the development of reading processes. In Study 1, we characterize the complexity variation over 131 orthographies, serving as a basis for Study 2 and Study 3. In Study 2, we compare individuals using different mapping principles (Study 2A) and the same mapping principle (Study 2B). In Study 3, we conduct an experiment parallel with Study 2A and demonstrate the process of learning to read the 131 orthographies in Study 1. These studies help us understand how visual characteristics of orthographies affect reading development, while taking mapping principle into consideration.

2.0 STUDY 1: GRAPHEME COMPLEXITY QUANTIFICATION

The goal of Study 1 was to develop a tool for studying key visual characteristics of graphemes as they vary over the wide range of orthographies present in the worlds' writing systems – namely, the complexity of these graphemes, the similarities and differences between grapheme complexity patterns within and across writing systems, and the factors underlying these complexity patterns.

In Study 1, our assumption was that a writing system's mapping principle – its manner of correspondence between graphemes and their linguistic units (e.g., phonemes, syllables, or morphemes) – determines the number of graphemes (or grapheme inventory) that the writing system needs. A writing system with lower mapping level (e.g., graphemes map to smaller phonological units such as phonemes) should need fewer graphemes, whereas a writing system with higher mapping level (e.g., graphemes map to larger phonological units such as syllables) should require many graphemes. Taking this assumption, we asked whether or not grapheme inventory, an implementation of mapping principle, is related to grapheme complexity.

To quantify grapheme complexity, we propose a comprehensive measure of four dimensions, each of whose strength has been demonstrated in prior reading research. We applied this measure to quantify grapheme complexity of 131 orthographies, examined the relationships among complexity patterns within and across writing systems, and associated the overall complexity of orthographies with number of graphemes. We expected a strong correlation to be found, supporting the claim that mapping principles govern number of graphemes, which, in turn, drives grapheme complexity.

2.1 OVERVIEW: A VISUAL ORTHOGRAPHY MEASURE

Every grapheme is a basic, two-dimensional visual object whose shape is composed of distinctive features such as lines, curves, intersections, and terminations. It is a natural tendency of grapheme complexity to increase along with number of graphemes, as more intricate combinations of simple features are required to construct larger sets of unique graphemes (cf: Information theory, Shannon & Weaver, 1949). Varying feature combinations give rise to different levels of visual complexity of graphemes which may, in turn, place varying loads on perceptual processing. Indeed, numerous studies of object identification have indicated that stimulus complexity affects recognition efficiency (e.g., Liu, Chen, Liu, & Fu, 2012).

In attempting to compare grapheme complexity across writing systems, we asked which measures are necessary and sufficient for capturing various visual characteristics of individual graphemes. Prior research has proposed several measures of object complexity. For instance, pattern goodness (Checkosky & Whitlock, 1973) is a subjective property of visual configuration, indexed by differences in rotation-reflection equivalence set size; information load (Alvarez & Cavanagh, 2004) is a measure of the visual features of an object that are encoded and stored in memory, indexed by the effect of search object numerosity on visual search speed; perimetric complexity (Pelli et al., 2006) is an objective measure of the complexity of binary images, indexed by the ratio between the square of inside-and-outside perimeter and “ink” area of a shape (for size invariance). Among these measures, the perimetric complexity has been demonstrated to have the most merits. It is objective, size-invariant, commonly used in shape analysis (Grainger et al., 2008), and well-correlated with pattern goodness and information load (Jiang, Shim, & Makovski, 2008). Moreover, this measure has been used in studying letter recognition among different orthographies. Pelli et al., (2006) applied perimetric complexity measures to six orthographies (Arabic, Armenian, Chinese, Devanagari, English, and Hebrew) using various type styles, sizes, and contrasts. Participants, who ranged widely in age and experience, completed a letter identification task and it was found that, across different orthographies, greater complexity of letter form was associated with lower identification efficiency.

Perimetric complexity seemed to be a promising measure in quantifying grapheme complexity; we questioned further whether it is a sufficient measure. Other measures have been used to study visual characteristics of grapheme stimuli in reading research across writing systems. In alphabets, disconnected components in graphemes (e.g. < j >; the dot is not connected to the main body) reportedly increased memory load on young readers (Treiman & Kessler, 2005), whereas line terminations (e.g., connected points in < R >) were reported as the features most critical to college students in letter identification (Fiset et al., 2008). In alphasyllabaries, vowels' featuring of disjointed components (e.g., < โ◌ะ >) was highly associated with vowel placement confusion in early literacy (e.g., Hindi: Gupta, 2004; Thai: Winskel, 2010). In morphosyllabaries, number of simple features (e.g., strokes) was varied to serve as a visual complexity manipulation of character stimuli (e.g., Japanese: Tamaoka & Kiyama, 2013; simplified Chinese: Wu, Zhou, & Shu, 1999; traditional Chinese: Chen, Allport, & Marshall, 1996). Although these studies suggest that visual complexity affects perceptual processing of graphemes, their results are not comparable because their measures captured different characteristics of grapheme complexity.

In Study 1, we aimed to establish a visual orthography measurement system to assess the complexity of any grapheme in the world and to allow fair comparisons of grapheme complexity within and across writing systems. This system comprises four dimensions: perimetric complexity, number of disconnected components, number of connected points, and number of simple features; each of these dimensions has been established in prior reading research. The primary goals were to apply this visual orthography measure to a larger number of orthographies across writing systems and to examine the relationships of grapheme complexity within and across writing systems. The secondary goal was to determine which constituent dimension is better able to properly distinguish writing systems. The ultimate goal was to investigate the degree to which mapping principle, by writing system, plays a role in grapheme complexity by governing grapheme inventory.

2.2 METHOD

2.2.1 Visual orthography measure

Before introducing each dimension of our visual orthography measure, we defined the following three key terms: A *simple feature* is a discrete element of an image that can be discriminated independently from other features (Pelli et al., 2006). For example, < T > has two simple features. A *connected point* (or a junction) is an adjoining of at least two features. For example, < F > has two connected points. A *disconnected component* is a simple feature or a feature that is not linked to other features in a set. For example, < i > and < 云 > have two disconnected components respectively.

Perimetric complexity (PC): PC is defined as the square of the sum of the inside and outside perimeters of a grapheme divided by the product of 4π and the foreground area (Pelli et al., 2006; Watson, 2011). For example, in a 500-pixel × 500-pixel bitmap, 1's represent "ink" and 0's represent "paper"; if upper-case W has a 4,656-pixel perimeter and 136,602-pixel squared area, its perimetric complexity is 12.6287 ($= 4656 \times 4656/ 136602 /4\pi$). This dimension is sensitive to the changes in luminance across space (i.e., spatial frequency) of a grapheme and its value is invariant to the size of the grapheme (Grainger et al., 2008).

Number of disconnected components (DC): DC is defined as a simple feature or a feature that is not linked to other features in a set. This dimension is sensitive to discontinuity information (Gibson, 1969).

Number of connected points (CP): CP is a point of contact between features. This dimension is sensitive to information regarding continuity (Lanthier, Risko, Stolz, & Besner, 2009) and provide clues in the relationships between simple features (Biedeman, 1987), counter to the DC dimension.

Number of simple features (SF): SF is a discrete element that can be discriminated from others; a typical example is a stroke within a Chinese character (Wu, Zhou, & Shu, 1999). This dimension is sensitive to the degree of combination of simple grapheme building blocks.

Collectively, these four dimensions provide objective, quantitative, and size invariant estimations about complexity of graphemes. Table 1 shows how these four dimensions capturing different characteristics of five example graphemes.

Table 1. Complexity values of five graphemes on four complexity dimensions

Writing System	Abjad	Alphabet	Syllabary	Alphasyllabary	Morphosyllabary
Orthography	Hebrew	Russian	Crree	Telugu	Chinese
Example Grapheme	בּ	З	ᐄᐧ	అః	面
PC	6.02	7.83	12.04	18.06	20.85
DC	2	1	3	3	1
CP	1	1	3	2	14
SF	3	2	6	5	9

Note. PC = Perimetric complexity, DC = number of disconnected components, CP = number of connected points, SF = number of simple features.

2.2.2 Language selection

We selected 131 orthographies from five writing systems (Alphabet: 60; Abjad: 16; Alphasyllabary: 41; Syllabary: 11; Morphosyllabary: 3). These orthographies were selected because they have been specifically examined in previous cross-writing-system (Changizi & Shimojo, 2005), cross-alphabet (Seymour, Aro, Erskine, 2003) and cross-Chinese-orthography (Chen, Chang, Chiou, Sung, Chang, 2011) studies. To retrieve the number of graphemes and writing system categories for these orthographies, we used the same source as Changizi et al.'s (2005) study: Ager's *Omniglot*: a guide to writing systems (Ager, 1998). For the three orthographies for which Omniglot offers no information, we consulted other sources: Chen et al. (2011) for two Chinese orthographies (i.e., traditional and simplified) and Wikipedia for the Japanese Kanji orthography (http://en.wikipedia.org/wiki/Ky%C5%8Diku_kanji).

2.2.3 Grapheme quantification

We generated images of 21,821 graphemes before quantifying their complexity. The *Processing* software (Fry & Reans, 2004) was used to construct a simple image of each grapheme. Graphemes were presented in white Arial font against a 500×500-pixel black background. Of the selected orthographies, 25% were not supported by Arial font; for these, an alternative font similar to Arial was adopted. Appendix A summarizes detailed information for these 131 orthographies.

2.3 RESULTS

To develop a fuller picture of how writing systems handle variability in visual characteristics of graphemes, we analyzed the data in two ways: across writing systems and within writing systems.

2.3.1 Relationships among dimensions within and across writing systems

The critical question here was how grapheme complexity behaves within and across writing systems. To address this question, we correlated complexity values from all four dimensions within writing systems and across five writing systems. Table 2 summarizes the results. Across five writing systems, there were strong, positive correlations among complexity dimensions (all rs > .6; all ps < .001). Intriguingly, separation of data by individual writing systems revealed the relationship between number of disconnected components (DC) and number of connected points (CP) to be strongly positive in alphasyllabaries and morphosyllabaries, yet significantly negative in abjads, while no such relationship was found in alphabets and syllabaries. Figure 2 illustrates the direction and magnitude of correlations among complexity dimensions with six heat maps. These results suggest that some complexity dimensions behave differently in different writing systems.

Table 2. Correlations of grapheme complexity within and across writing systems

	Abjad 443 graphemes				Alphabet 3,232 graphemes				Syllabary 1,021 graphemes			
	PC	DC	CP	SF	PC	DC	CP	SF	PC	DC	CP	SF
PC	1.00				1.00				1.00			
DC	.39***	1.00			.34***	1.00			.42***	1.00		
CP	.44***	-.13***	1.00		.45***	.01	1.00		.43***	.01	1.00	
SF	.57***	.35***	.82***	1.00	.57***	.32***	.92***	1.00	.60***	.52***	.79***	1.00

	Alphasyllabary 2,795 graphemes				Morphosyllabary 14,330 graphemes				All writing systems 21,821 graphemes			
	PC	DC	CP	SF	PC	DC	CP	SF	PC	DC	CP	SF
PC	1.00				1.00				1.00			
DC	.44***	1.00			.63***	1.00			.82***	1.00		
CP	.48***	.13***	1.00		.79***	.24***	1.00		.89***	.65***	1.00	
SF	.59***	.35***	.92***	1.00	.94***	.64***	.83***	1.00	.95***	.83***	.93***	1.00

Note. *** $p < .001$

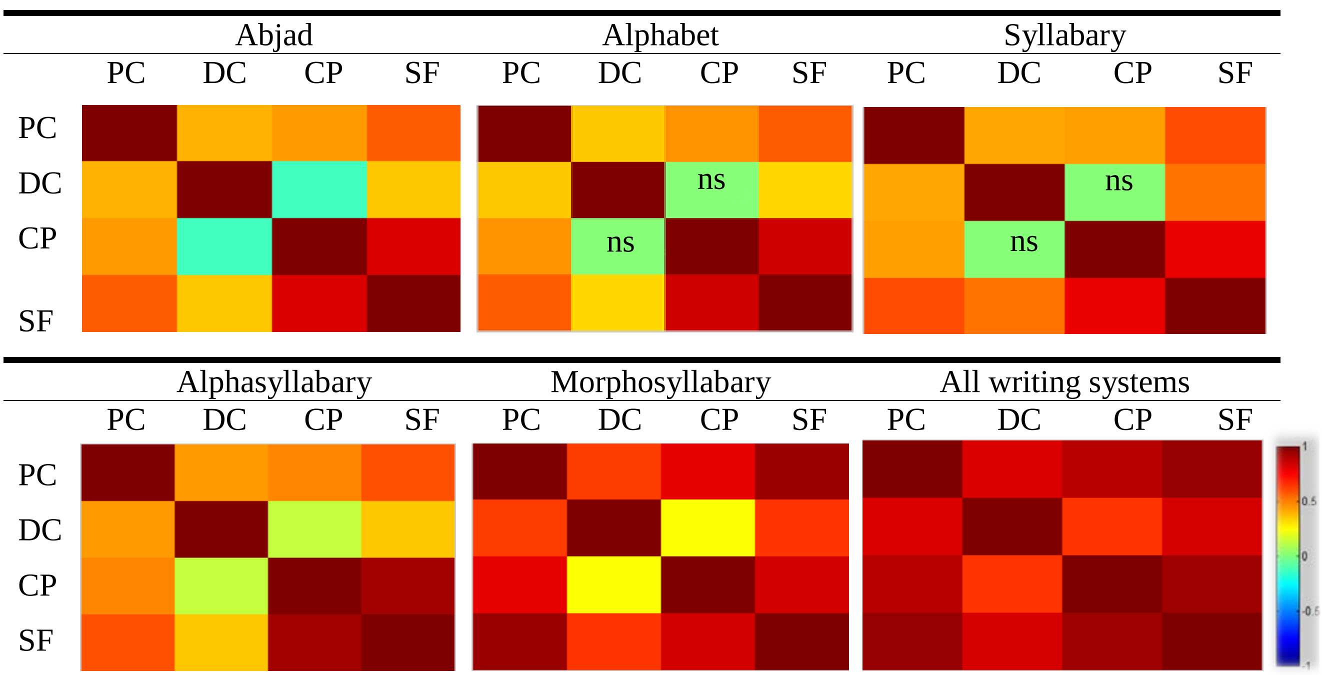

Figure 2. Heat maps of grapheme complexity within and across writing systems

To elucidate the relationships among complexity dimensions across writing systems, a multiple regression with four dimensions as predictors (i.e., mean scores from four dimensions for each orthography) was performed to determine what weighted combination of dimensions can best predict grapheme inventory size across 131 orthographies. The perimetric complexity dimension was entered first given its reported significance in comparing grapheme complexity across orthographies (Pelli et al., 2006). Next, the three other dimensions (i.e., number of disconnected components, number of connected points, and number of simple features) were entered in a stepwise manner to determine whether any of them could account for remaining variance, above and beyond that explained by perimetric complexity. The stepwise model selection method was chosen because it combines the virtues of both forward and backward selection (Hocking, 1976). The resulting, best-fitting model included all four dimensions as significant predictors (R^2 = .82, p < .01), suggesting that the four dimensions collectively can best predict grapheme inventory size. Table 3 provides details about the model summary.

Table 3. Summary of multiple regression for dimensions predicting grapheme inventory (*n* of graphemes = 21,821)

	Model 1			Model 2			Model 3			Model 4		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
PC	94.98	10.79	.61**	3.59	13.48	.02	-24.31	12.42	-.16	.65	10.35	.01
DC				1087.74	124.00	.76**	795.43	116.15	.56**	1185.76	103.48	.83**
CP							244.38	36.99	.48**	796.41	71.12	1.57**
SF										-702.71	82.34	-1.46**
R^2			.38			.61			.71			.82
R^2 change						.23			.10			.11
F for change in R^2		77.49			100.03**			103.45**			139.69**	

Note. PC = Perimetric complexity, DC = number of disconnected components, CP = number of connected points, SF = number of simple features.

**p < .01.

2.3.2 Optimal dimensions in differentiating writing system pairs

In the correlation analysis above, we observed different patterns among complexity dimensions across individual writing systems. In the regression analysis, we discovered that four dimensions together can best predict grapheme inventory size across all writing systems. We then sought to determine whether one dimension more reliably distinguished between graphemes' parent writing systems than others. To implement this, we used the nonparametric Kolmogorov–Smirnov test, because this method is sensitive to difference in the cumulative distribution functions of two samples without assuming normality of the distribution (Stephens, 1974); in our case, the two samples correspond to two writing systems. The difference between two writing systems is represented with Kolmogorov–Smirnov distance (KS-distance). Among the five given writing systems, there were 10 writing system pairs. We calculated 10 KS distances from each dimension; the dimension from which results displayed the greatest KS distances between paired writing systems was taken as the index most sensitive to differences between those two writing systems.

Table 4 shows optimal complexity dimensions in differentiating pairs of writing systems. For instance, for the Alphabet-Abjad writing system pair, the optimal dimension is DC, suggesting that the Alphabet and Abjad writing systems differ the most in terms of their number of disconnected components. Interestingly, perimetric complexity, the only dimension to have been used in comparing grapheme complexity across writing systems in prior research (Pelli et al., 2006), was only found to most reliably differentiate the Alphasyllabary-Alphabet writing system pair. The dimension which may be the most effective in differentiating writing systems pairs overall is number of disconnected components (DC); in Table 4, DC is the optimal complexity dimension (i.e., the dimensions showing greatest KS distance) for 6 out of 10 writing system pairs. However, speaking specifically, the results suggest that the maximally distinctive complexity dimension is different for each pair of writing systems - no single dimension is universally the most effective in distinguishing between any two writing systems.

Table 4. Optimal complexity dimension in differentiating writing system pairs.

	Abjad	Alphabet	Syllabary	Alphasyllabary	Morphosyllabary
Abjad	--				
Alphabet	DC	--			
Syllabary	DC	DC	--		
Alphasyllabary	DC	PC	DC	--	
Morphosyllabary	SF	DC	SF	SF	--

Note. DC = Number of disconnected components, SF = Number of simple features; PC = Perimetric complexity

2.3.3 Mapping principle, number of graphemes, and grapheme complexity

The ultimate goal of Study 1 was to examine the extent to which mapping principles of writing systems govern grapheme inventory and, thus, drive grapheme complexity. We used visualization techniques (e.g., box plots and scatter plot matrices) to approach this goal.

First, we visualized the variation in orthographies between their writing system categories and number of graphemes. We labeled each orthography by color corresponding to the writing system used, and then plotted descriptive statistics (e.g., the mean, median, first and third quantiles, and outliers) for each. Figure 3 shows the variation of grapheme inventories; the x-axis reflects number of graphemes, and the y-axis covers writing system categories, ordered roughly by mapping unit size from low (e.g., phoneme; alphabet) to high (e.g., syllable and morpheme; morphosyllabary). Given the unusually large number of graphemes in morphosyllabic orthographies (i.e., traditional Chinese, simplified Chinese, and Japanese Kanji), we excluded this category from our visualization; Figure 4 displays this new, localized visualization. Generally, as mapping unit granularity increases, the number of graphemes increases; however, there is no fine-grained separation of mapping principles and grapheme inventory.

Second, we examined how grapheme inventory relates to grapheme complexity – as determined by different complexity dimensions. We further consolidated data across dimensions to create a composite score representing overall complexity. For each orthography, we calculated mean score from each dimension; given the scaling difference between dimensions, we transformed the resulting means to within-dimension *z*-scores, averaging these to form a standardized composite score. We gave each dimension equal weight because, in terms of theory, these dimensions were highlighted in different study contexts, and, empirically, our Kolmogorov–Smirnov test suggested that each dimension provides a unique contribution. Figure 5 shows the variation of grapheme overall complexity in 131 orthographies by writing systems.

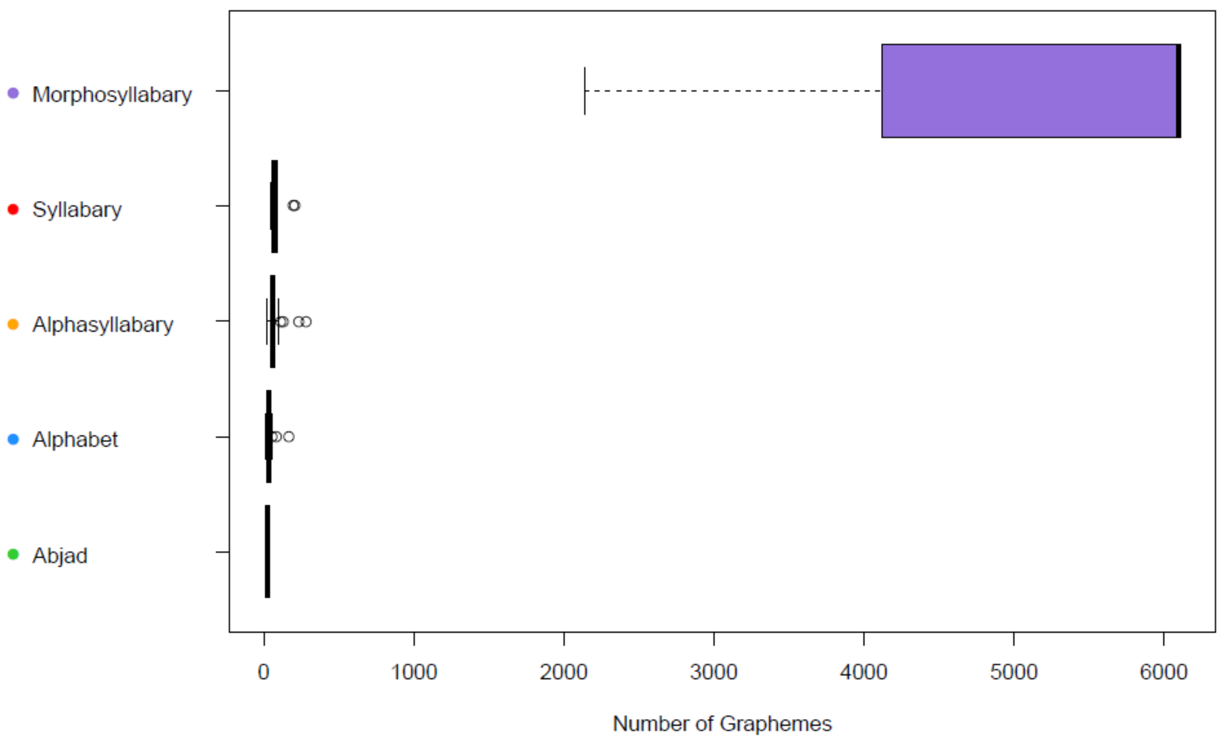

Figure 3. Variation of number of graphemes by writing systems – 131 orthographies

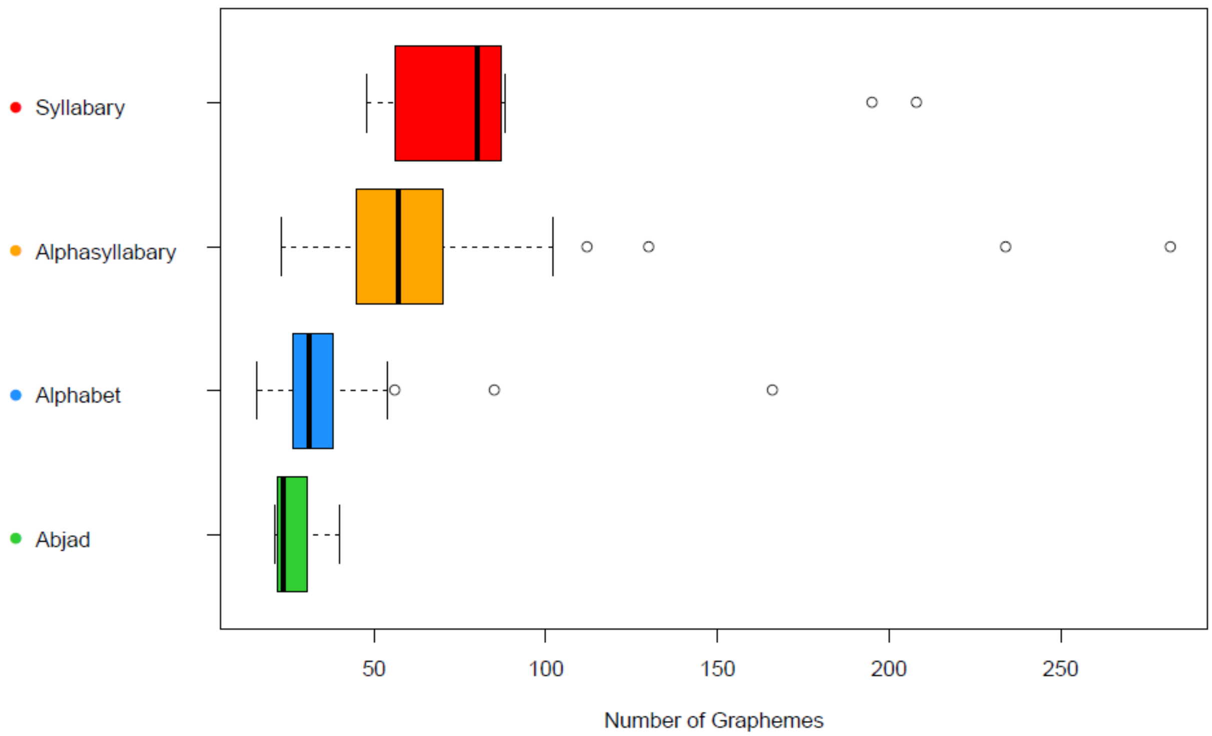

Figure 4. Variation of number of graphemes by writing systems – 129 orthographies

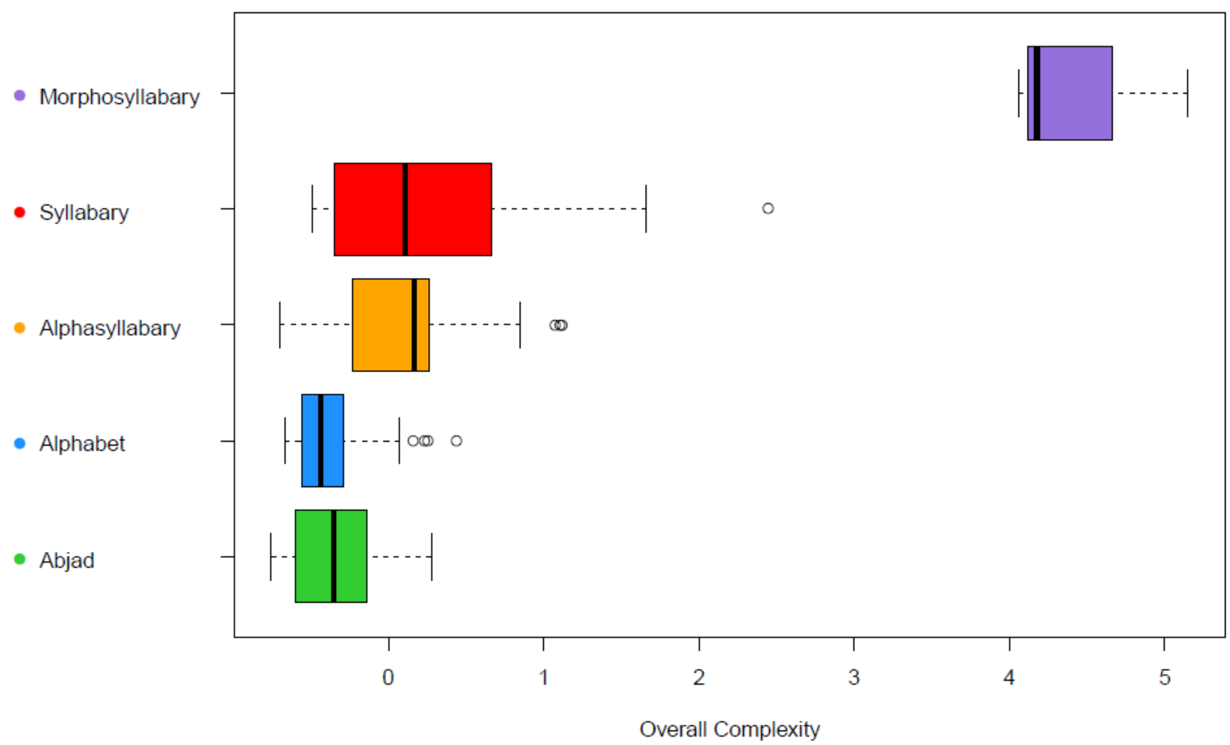

Figure 5. Variation of overall complexity by writing systems – 131 orthographies

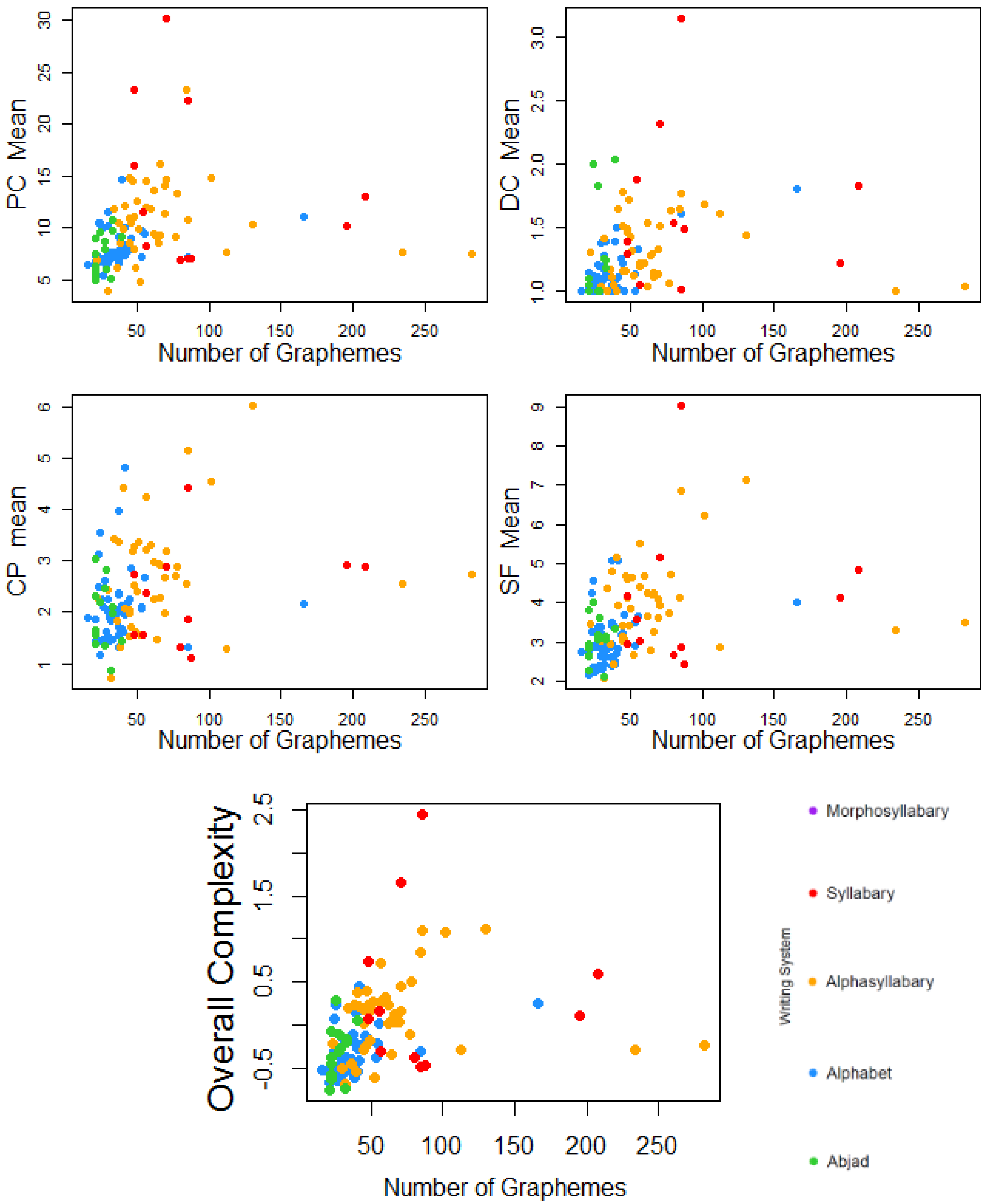

Figure 6. Distribution between number of graphemes and complexity values – 129 orthographies

Figure 6 shows the relationships between number of graphemes and scores from each of the four dimensions as well as the overall standardized score; each observation represents one orthography. These scatter plots show a clear trend in how mapping principles, as indicated by number of graphemes, govern variability in visual characteristics, as measured on different dimensions: orthographies with a larger grapheme inventory (e.g., alphasyllabaries and syllabaries) tend to be visually complex and widely distributed, whereas orthographies with fewer graphemes (e.g., alphabet and abjad) tend to be less complex and show a systematic, linear pattern. Consistent with these results, we found a strong, positive association between number of graphemes and overall complexity ($r = .78$, $p < .001$) across all 131 orthographies.

2.4 INTERIM SUMMARY

To investigate the relationship between grapheme complexity, grapheme inventory, and mapping principle, Study 1 applied a measurement system to quantify visual complexity among 21,821 graphemes in 131 orthographies. Our analysis revealed several interesting results within and across writing systems:

1. Similarities and differences were found among correlation patterns between complexity dimensions for grapheme scores, notably when compared within and across writing systems. Within each writing system, most of the correlations between dimensions were positively associated with correlations ranging from small (e.g., $r = .32$; the correlation between SF and DC in alphabets) to larger (e.g., $r = .94$; the correlation between SF and PC in mophosyllabaries). The correlation between number of disconnected components and number of connected points, however, behaved differently across writing systems: it was

positive in writing systems with larger number of graphemes (e.g., alphasyllabaries and morphosyllablaries), but negative in writing systems with smaller number of graphemes (e.g., abjads).

2. Number of disconnected components seemed to be the most effective dimension on which to distinguish two writing systems (6 out of 10 pairs in the nonparametric Kolmogorov–Smirnov test), whereas number of simple features (3 out of 10 pairs) and perimetric complexity (1 out of 10 pairs) also functioned uniquely in differentiating writing systems. Although the ability of the number of connected points dimension to convey information regarding continuity has been documented in reading (Lanthier, Risko, Stolz, & Besner, 2009) and object recognition research (Biederman, 1987), this dimension did not stand out in differentiating between writing systems.
3. Overall complexity, a composite score over all four complexity dimensions, was significantly, positively, and strongly associated with number of graphemes across 131 orthographies, with correlation of $r = .78$.
4. Number of graphemes, a factor arising from the mapping principles of writing systems, is closely tied to grapheme complexity. Orthographies with a larger number of graphemes (e.g., alphasyllabaries) generally have higher values on each complexity dimension, and the distributions of these orthographies tend to be more dispersed and less systematic than orthographies with smaller number of graphemes (e.g., abjad and alphabets).
5. Mapping principles, as indicated by writing system categories, show a general trend in guiding number of graphemes. In Figure 4 (Distribution of number of graphemes by writing systems – 129 orthographies), higher mapping levels correspond to larger numbers of

graphemes, although there are substantial overlaps across writing systems. These overlaps between number of graphemes and mapping principles may need further investigation.

In summary, in an attempt to compare grapheme complexity within and across writing systems in Study 1, we found a clear, positive association between grapheme complexity and number of graphemes across writing systems. Within writing systems, we also showed that multiple dimensions were weighted differently depends on the characteristics of graphemes. Collectively, these results suggest that our measurement system is sufficient in revealing how visual characteristics of graphemes vary across writing systems.

3.0 STUDY 2: VISUAL DISCRIMINATION PERFORMANCE COMPARISONS WITHIN AND ACROSS WRITING SYSTEMS

The results of Study 1 established the nature of variation across 131 orthographies, both in terms of grapheme complexity and grapheme inventory. We next considered the implications of this variation on individuals' visual perceptual processing of graphemes – an ability critical in early stages of learning to read. Prior research has shown that recognition efficiency is diminished for more complex graphemes (Pelli et al., 2006), and that children whose first-language (L1) orthographies contain larger numbers of graphemes take longer to fully master their L1 grapheme inventory when learning to read. Given that the process of learning to read in a certain orthography entails the development of visual expertise in that orthography, we posit that reading in orthographies containing higher numbers of graphemes, such graphemes tending to be more complex, may require stronger visual perceptual skills and that learning to read such orthographies may, in turn, strengthen such skills.

In Study 2, we tested the extent to which visual orthographic variation, encompassing both grapheme complexity and grapheme inventory size, affects individuals' visual perceptual processing. We systematically manipulated the complexity of grapheme groups, having verified each individual's L1 background. By comparing individuals across writing systems, we examined the effect differences between L1 complexity levels where level differences are driven by mapping principle variation. By comparing individuals across orthographies of differing complexity within one writing system (Chinese: visually complex “traditional” vs. visually simple “simplified”), we examined the effect differences between complexity levels where such levels are without mapping principle difference. Through this attempt to discern the influential balance between L1 visual orthography and mapping principles, we expect to gain a clearer picture of the degree to which mapping principle is involved in perceptual performance; this has especially important implications for the role of visual orthographic complexity across writing systems.

3.1 OVERVIEW: VISUAL PERCEPTUAL LEARNING IN READING DEVELOPMENT ACROSS WRITING SYSTEM VARIATIONS

From the perceptual learning perspective (Fahle & Poggio, 2004), visual perceptual learning involves improvement in visual discrimination through repeated exposure to visual stimuli. Learning to read can be seen as an instantiation of the development of visual perceptual expertise, functioning in the same manner with regard to the role of experience – this would imply that reading employs both domain-general and domain-specific visual cognitive mechanisms (Gauthier & Nelson, 2001). Indeed, numerous neuroimaging studies have suggested that the extent of reading expertise depends directly on readers’ relevant experience levels (e.g., McCandless, Cohen, & Dehaene, 2003; Dehaene, Pegado et al., 2010).

Given that skilled reading entails rapid, effortless, accurate processing of visually perceived words, readers must make an effort to master their orthography’s full set of graphemes. Further, given that reading skill improves with experience and that complexity varies across orthographies, more complex orthographies can require more effort to learn; if learning a more complex orthography is more challenging, the increased effort made by a reader's visual system in overcoming these perceptual challenges should translate to perceptual ability as he or she achieves skilled reading. Because reading involves domain-general visual mechanisms, improvement in reading skill should be accompanied by improvement in these mechanisms, and so increase in general visual perceptual ability over development should be more pronounced for individuals whose first-language orthographies are more complex. In other words, readers of visually complex orthographies should show more advanced visual perceptual skills than readers of visually simpler orthographies because, from the very start of learning, they must memorize larger numbers of graphemes and their constituent features, and must make more fine-grained visual discriminations when distinguishing one grapheme from another.

Such reasoning is supported by a limited, although significant, body of reading research. First, when comparing performance between literate and illiterate adults on perceptual matching of letter strings, literates were sensitive to letter-position changes, whereas illiterates were almost blind to these changes (Duñabeitia, Orihuela, & Carreiras, 2014), reflecting the role of relevant experience in discrimination ability. Second, when comparing Japanese-speaking and English-speaking second graders on ability to memorize abstract visual designs, Japanese speakers (who learn to read Kanji, equivalent to complex Chinese characters) outperformed English speakers (Mann, 1985). Similarly, in a larger cross-orthography study, kindergarteners learning to read a visually complex orthography (traditional Chinese) outperformed age-matched kindergarteners learning to read less complex orthographies (Hebrew and Spanish) in a visuo-spatial processing task (McBride-Chang et al., 2011). These results reflect the influence of L1 orthographic complexity on general visual ability. Third, in a comparison of 8- to 14-year-old readers of Chinese and Greek, controlling for reading experience, Chinese readers of all ages showed greater visual spatial processing efficiency than age-matched Greek readers (Demetriou et al., 2005), suggesting that the influence of complexity on visual process sing ability holds across experience levels. Collectively, these studies support the idea that learning to read more visually complex orthographies may refine young readers' visual processing skills more than learning to read visually simple orthographies.

However, several gaps exist between these studies and our reasoning. First, our reasoning concerns the perceptual experience of visual complexity – mastering complex visual stimuli may strengthen basic visual perceptual skills. The aforementioned research, which highlights differences across writing systems, deals with the experience of both visual complexity and mapping principles. Although it is generally found that mapping principles govern orthographic complexity (i.e., number of graphemes and grapheme complexity), this may not be the case universally. Based on findings from our grapheme complexity quantification (Study 1), there is overlap between alphabets and syllabaries - some alphabets have more graphemes than some syllabaries, despite syllabaries having a generally higher mapping level. It is unclear whether visual performance variation is driven by experience of mapping principles and visual complexity both, or visual complexity only. Second, previous work has been restricted to beginning learners, regardless of age. If our reasoning holds, then the effect of perceptual learning should also develop concurrently with literacy development, such that it strengthens perceptual abilities of skilled readers. Third, the nature of the stimuli is important. Compared to nonlinguistic stimuli such as pictures, graphemes are usually simpler, and are more easily computed and recognized by human vision (Changizi, Zhang, Ye, & Shimojo, 2006). Using authentic grapheme stimuli in the same visual discrimination task for assessing visual perceptual performance would more closely resemble learning to read. These gaps could be addressed by sampling skilled readers across writing systems and assessing their visual perceptual skills with grapheme stimuli. We took such an approach in the present study.

To reiterate, the question at the core of our examination was how visual orthographic variation across writing systems, both at the grapheme and orthography levels (as seen in Study 1) impacts reading development. At the grapheme level, increasing complexity of graphemes adds visual processing load; at the orthography level, larger grapheme inventory begets increased visual skills. Given the multi-layered relationship between graphemes and their orthographies, the effects of grapheme complexity and participant perceptual experience should interact. This interaction would not necessarily be driven by mapping level.

In Study 2, we aimed to examine the extent to which visual orthographic variation, including grapheme complexity and number of graphemes, impacts individuals' perceptual processing. We codified the variation of grapheme complexity by forming groups of grapheme stimuli selected from different writing systems with varied complexities. We codified variation of number of graphemes through the range of inventory sizes of participants' first-language (L1) orthographies. Moreover, we deliberately used the same perceptual discrimination task (e.g., same-different judgments) in two experiments: Study 2A, which compared participants from differing mapping principle *and* L1 orthographic complexity groups; and Study 2B, which compared participants from differing L1 complexity groups *within* the same mapping principle groups; this design allowed us to dissociate between possible influences of individuals' experienced mapping principles and learned L1 complexity on performance. To the best of our knowledge, this study is the first to systematically manipulate stimulus complexity using authentic graphemes over a range of participant L1 backgrounds within and across the world's writing systems. We expect this study to provide a novel opportunity to gain understanding about how visual characteristics of orthography affect reading development.

3.2 STUDY 2A: GRAPHEME DISCRIMINATION PERFORMANCE ACROSS WRITING SYSTEMS (MTURK STUDY)

The goal of Study 2A was to evaluate the complexity hypothesis with individuals experiencing different mapping principles as well as L1 visual orthographic complexities. Several informed decisions were made before conducting the experiment.

1. To study reading phenomena across writing systems, we selected both stimuli and participants by writing system categories.
2. To ensure a fair complexity comparison, we used the visual orthography measure developed in Study 1 to represent complexity values of both of participants' L1 orthography and experimental stimuli.
3. To allow world-wide data collection, we used Amazon Mechanical Turk (MTurk), an online data collection tool that provides a stable pool of participants with various backgrounds; MTurk data have been demonstrated to be indistinguishable from laboratory data in psycholinguistic experiments (e.g., Buhrmester, Kwang, & Gosling, 2011; Horton, Rand, & Zeckhauser, 2011; Sprouse, 2011)
4. To control data quality (especially response time data), we implemented all of our tasks with Adobe flash together with MTurk, this combination has high reliability in collecting response time (Simcox & Fiez, 2013). We also carried out several keyboard-response timing tests comparing low-, medium-, and high-load computer resource usage conditions; the

resulting timing errors were all under 20ms, under the noted 1% threshold of statistical power loss (Brand & Bradley, 2012).

5. To ensure the quality of collecting response time in the online experiment as much as possible, we recorded computer response lag for each trial and used this lag information to filter out questionable data (e.g., a trial with RT ± 3*SD* from the mean, or all trials with a reliable, systematic delay).
6. To ensure our MTurk participants were in fact representative readers of their self-reported L1 orthography, a language history questionnaire (Tokowicz, Michael, & Kroll, 2004) and a demographic background survey (both revised pursuant to advanced psycholinguistic consultation), as well as a 20-word translation task involving critical words in the task instruction were administered. Because all participations via MTurk were anonymous, the resulting information was used only to filter data for quality (see Participants section for exclusion criteria).
7. To ensure our MTurk participants were able to understand our instructions all in English, an English vocabulary size test (Nation & Beglar, 2007) was administered to estimate participants' English word knowledge; this information was subsequently used to filter the data.

3.2.2 Method

In implementing our key manipulations – complexity of graphemes and of participant L1 orthography – we made comparisons within the multidimensional complexity space as determined through the complexity measure developed in Study 1 to ensure the tested orthographies were varied enough and were representative of their writing systems. This was done by identifying a centroid for each writing system within the standardized four-dimensional space (dimensions: perimetric complexity, number of disconnected components, number of connected points, and number of simple features). A centroid is a geometric center; in our case, the centroid of a writing system corresponded to the mean position of all the orthographies within this writing system for all dimensions. Using the default space without weighting any one dimension relative to others allowed us explore target effects without any prior constraints. We termed the resulting orthographies as "centroid orthographies" and ranked them by overall complexity (standardized score) from least to greatest: Hebrew (abjad; -.58), Russian (alphabet; -.32), Cree (syllabary; -.32), Telugu (alphasyllabary; .07), and Chinese (morphosyllabary; 3.79). Note that this order of complexity levels does not necessarily correspond to the phonological mapping unit size of these orthographies; for instance Cree graphemes have a higher mapping level (syllable) but are less complex than Telugu graphemes, which have a lower mapping level (phoneme).

These centroid orthographies were used for manipulating stimulus complexity. As for varying participant L1 orthography, we selected two orthographies from each writing system. Orthographies of languages with the largest speaker populations of any language within each writing system were selected for study, based on the "centroid" assumption that these orthographies would be appropriately representative of their parent writing system as a whole. The syllabary writing system was excluded from study due to limited online population access - although Cree has the largest speaker population of the seven typical syllabaries currently in use, this population numbers only 60,000, and none were found active on MTurk despite extensive search efforts. The following eight orthographies were selected to serve as participants' L1s, here ranked by overall complexity: Hebrew (abjad; -.58), English (alphabet; -.50), Russian (alphabet; -.32), Arabic (abjad; -.26), Hindi (alphasyllabary; -.02), Telugu (alphasyllabary; .07), Japanese (morphosyllabary Kanji; 4.01), and Chinese (morphosyllabary traditional Chinese; 5.49). Again, we noticed that increasing complexity of these orthographies generally, but not consistently, echoes their phonological mapping granularity.

3.2.2.1 Stimuli

The stimuli comprised graphemes from five centroid orthographies. Given that thousands of graphemes with highly variable complexity exist in the Chinese orthography, two groups of graphemes with contrasting complexity (simple or complex) were formed from the overall orthography. Each "simple" character was a radical, the functional "building block" in Chinese orthography (Shen & Ke, 2007), composed of a small number of strokes (average: 4.52); "complex" characters were those containing multiple radicals, composed of a large number of strokes (average: 13.21). Note that these characters shared the same forms between the traditional and simplified Chinese visual orthographies.

Six grapheme groups of increasing complexity (i.e., Hebrew, Russian, Cree, Telugu, simple Chinese, and complex Chinese) were constructed. For the same-different judgments, graphemes were paired within each orthography, matched to either upper or lower case (for Russian), vowel or consonant (except for Chinese), and simple or complex (for Chinese). We included equivalent numbers of "same" and "different" pairs in each list to ensure equal responses. Graphemes paired with themselves comprised "same" pairs; all graphemes in each orthography (except for Chinese) were exhaustively used. Graphemes paired with other graphemes of similar complexity once comprised "different" pairs; not all combinations of different graphemes were used. We created four lists, each consisting of six grapheme groups, to allow us to generalize results to other grapheme combinations. Within each list, complexity varied by grapheme group according to the following ranking (overall complexity of grapheme pairs per orthography across all four lists); Hebrew (-0.58) < Russian (-0.38) < Cree (-0.38) < Telugu (-0.10) < simple Chinese (0.09) < complex Chinese (2.39), $F_{(5, 1439)} = 2339.61$, $p < .001$. Between lists, no complexity differences in grapheme pairs were found for any grapheme group, $F_{(3, 1439)} = 1.64$, $p = .18$. Each list contained 360 pairs – an upper threshold for participant sensitivity to visual similarity (Simpson, Mousikou, Montoya, & Defior, 2013). Appendix B shows grapheme pairs per list; Table 5 provides further information regarding these grapheme pairs per list.

Table 5. Characteristics and number of grapheme pairs for each orthography in same different judgments (per list)

Writing systems	Abjad	Alphabetic	Syllabary	Alphasyllabary	Morphosyllabary	
Orthography	Hebrew	Russian	Cree	Telugu	Traditional Chinese	
Number of L1 speakers	5 million	150 million+	60,000	75 million	23 million +	
Number of graphemes	32	Upper: 33 Lower: 33	80	Vowels:35 Consonant:35	242 simple characters	5600+ Complex characters
Same pairs	32	33	40	35	20	20
Different pairs	32	33	40	35	20	20
Total pairs	64	66	80	70	40	40

Note. Estimates of number of L1 speakers were retrieved from Wikipedia.

3.2.2.2 Task

Same-different judgment task This task tapped individuals' perceptual processing of graphemes; it emphasizes reliance on perceptual processing while minimizing the possibility of linguistic interference from phonology or semantics. In this task, each trial began with a black fixation cross appearing for 300ms, followed by a pair of graphemes appearing for up to 1000ms, followed by a blank for 1000ms. The participants were instructed to judge whether two graphemes were the same or different using their index fingers; response keys were counterbalanced across the four stimulus lists. After instructions, the participants were given 12 example trials with answers, 36 practice trials without feedback, and 360 critical trials with randomized presentation. Responses and response time were recorded. This task took approximately 15 minutes to complete.

Language questionnaire The language questionnaire (Tokowicz, Michael, & Kroll, 2004) was used to study participants' language learning experiences both quantitatively (e.g., rating general language learning skill and proficiency in learned languages) and qualitatively (e.g., comments about language learning experience). Several items were revised to focus more on participants' exposure to graphemes (e.g., degree of use of reading and writing in multiple languages in different contexts) after consulting the first author of this questionnaire. Participants were encouraged to give their best answers to the questions without any time limit. Appendix C provides the questionnaire administrated.

Demographic background questionnaire The demographic background questionnaire was developed to learn more about participants' educational, cultural, and health status (e.g., visual and hearing problems) as well as their surroundings during participation in this study. The responses on visual and hearing questions were used to filter data quality. There was no time limit to complete this survey (Appendix D).

Vocabulary size task The English Vocabulary Size Task (Nation & Beglar, 2007), a multiple-choice test, was used to assess participants' knowledge of the 14,000 most frequent word families of English. This test consisted of 140 items; participant vocabulary size was estimated by multiplying raw score by 100. This task has good reliability and validity for both first- and second language speakers (Beglar, 2010). There was no time limit to complete this task.

Translation task The translation task was developed to filter the data for quality. This task consisted of 20 English words chosen from the instructions of this experiment. Participants saw one word at a time, and were asked to type the first translation that came to mind in their L1 within 12 seconds; timing was determined in a pilot study. Capability of providing translation in an orthography consistent with reported L1 was taken as evidence that the participant was in fact a representative speaker of their reported L1.

3.2.2.3 Procedure

All participants completed this experiment via the Internet. Eight Human Intelligence Tasks (HITs), for recruiting participants from each of eight orthographies, were posted on MTurk's online recruitment interface. Each HIT had a two hour completion limit. Consent was obtained prior to the experiment; after MTurk volunteers agreed to participate, they were directed via web link to any of the four stimuli lists for same-different judgments. The sequence of tasks was the same for each participant: a same-different judgment task, a language history questionnaire, a demographic background task, and a translation task (except for the English HIT). After completing the last task, a unique 13-digit code associated with the participant's responses appeared on the screen automatically, along with debriefing information. The participant was instructed to report the code to MTurk to obtain monetary compensation. Successful generation of the 13-digit code also indicated that all of the participant's responses were successfully sent from his or her local machine to our server.

3.2.2.4 Participants

We recruited 60 participants for each of eight participant groups, for a total of n = 480. All participants read the following criteria via MTurk (the wording was exactly what the MTruk workers saw; each orthography displayed only to target population in recruitment materials):

"(1) Native language: Hebrew, English, Russian, Arabic, Hindi, Telugu, Japanese, or Mandarin Chinese (By native language, this means that you must have learned the language from birth. It is perfectly okay if you can also speak other languages.)

(2) Age: From 18 to 35.

(3) No vision or hearing impairments.

(4) Other: Need a computer that supports input for your first language when participating."

During the recruitment process, those using MTurk were informed that the task would take approximately 1 hour to complete and that they would receive $3.00 after their work quality was approved. To allow fair participation opportunity, we placed no restrictions on their approval rate of Human Intelligence Tasks (HIT). This lack of restriction created the potential for inclusion of dishonest participants (e.g., those repeatedly giving the same response to obtain a unique survey code for payment without following any task instructions). Thus, we used the following criteria to exclude problematic data:

(1) Proportion accurate on the same-different judgment task was below 50%, and thus below chance.

(2) RT on the same-different judgment task, the English vocabulary size task, and the translation task was not within a reasonable range; problematic examples including systematic delay or lag times further than $\pm$ 3*SD* from the mean.

(3) Translation responses were not written in the orthography consistent with L1 reported by the participant in the language history questionnaire or less than 15 out of 20 translation responses were entered.

(4) Self-reports in the language history questionnaire suggested that the participant was not representative of his/her L1 orthography in this study (e.g., reported multiple first languages and both were orthographies of interest, such as English and Russian; reported native-like reading skills in a non-L1 orthography, i.e. Hindi L1 speaker self-rated her reading skills in L2 English as 7 on a 7-point Likert scale.)

(5) Self-reports in the demographic background questionnaire suggested that the participant was ineligible for this study (e.g., had vision or hearing impairments).

(6) Score on the vocabulary size test was lower than 60 (represents knowledge of ~6000 written English words), ensuring participants understood our instructions in English to a reasonable extent.

Given limited participation from Hebrew- and Japanese-speaking individuals through MTurk, we recruited L1 speakers of these orthographies in their home countries and offered e-gift cards as compensation. Several participants volunteered their time in this manner.

We matched the participants by age (M = 26.88, SD = 5.16), F < 1, across all eight orthographies. All 480 individuals (237 males) were included in the data analysis. Table 6 provides background information for these participants by orthography group.

Table 6. Background characteristics for the final set of participants by orthography group

Background Participant group (ordered by their L1 complexity)	Number of participants	Age (years)	Number of learned languages (including first language)	Self-rated general language learning skills	Vocabulary size task (raw scores)
P1 Hebrew (-.58)	60 (27 males)	26.07 (4.38)	2.82 (0.93)	5.38 (1.12)	103.95 (17.42)
P2 English (-.50)	60 (22 males)	26.37 (5.64)	1.67 (0.86)	4.37 (1.30)	114.88 (11.37)
P3 Russian (-.32)	60 (21 males)	27.85 (5.99)	2.60 (0.89)	5.48 (1.13)	108.37 (16.44)
P4 Arabic (-.26)	60 (29 males)	27.42 (4.86)	2.53 (0.93)	5.42 (1.25)	104.98 (16.42)
P5 Hindi (-.02)	60 (42 males)	27.75 (5.67)	2.53 (0.72)	5.25 (1.19)	99.98 (18.98)
P6 Telugu (.07)	60 (42 males)	26.52 (4.00)	2.60 (0.69)	5.62 (1.25)	93.55 (17.79)
P7 Japanese (.54)	60 (30 males)	27.07 (5.92)	2.45 (0.72)	4.95 (1.41)	97.51 (19.52)
P8 Chinese (3.79)	60 (24 males)	25.97 (4.63)	2.62 (0.87)	5.03 (1.07)	99.25 (22.27)

Note. Means (standard deviation) reported. Self-rated general language learning skills is a self-reported measure of language-learning ability that includes listening, speaking reading, and writing on a 7-point scale, with 1 being the lowest.

3.2.3 Results

3.2.3.1 Data analysis plan: Mixed effect models

We used mixed effect models to analyze the data, given that this approach was well suited to examining the characteristics of the complexity of both grapheme groups and participants' L1 orthographies in terms of their influence on performance (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). Mixed effect models comprise fixed effects and random effects. The fixed effects include variables with levels (categories) of interest; in our case, complexity of grapheme groups and of participant L1 orthography. The random effects include variables with levels randomly sampled from a larger population; in our case, items and participants.

Moreover, mixed effect models can include random slopes by items or by participants to capture variability in a given effect across items or participants. Although all possible random slopes could be combined for analysis in one model, this may result in the model failing to converge, and these slopes are often not of theoretical interest (Freeman, Heathcote, Chalmers, & Hockley, 2010); thus, we tested random slopes one at a time. We always tested the random slopes by items first given the limitation (by design) of cross-list item variability; if inclusion of the slopes by items did not result in significant improvement to the model's account of the data, we continued data analysis without these slopes. In the model comparison process, the Likelihood Ratio Test (Lehmann, 1986) was used to determine a best-fit model for predicting performance.

Independent variable (i.e., participant L1 orthography and grapheme group) levels were ordered progressively by complexity, thus we used the Helmert contrast to code predictors. Helmert contrast coding allowed us to evaluate several key contrasts in a single model by testing each level of the variable against the mean of all previous levels. We designed the Helmert code such that the resulting estimates would correspond to actual differences between conditions, making the estimates easier to interpret. For the dependent variables, we ran a mixed effect logit model using the *glmer()* function to analyze response accuracy, given that accuracy is a binary variable (responses are accurate or inaccurate); we ran a linear mixed effect model using the *lmer()* function to analyze the response time (i.e., how quickly participant gave correct response) given that this variable is continuous. All models were fit using the *lme4* package (Bates, Maechler, & Bolker, 2011) in the R software.

We followed the same hypothesis-testing plan to process the accuracy and RT data. The analysis aimed to examine how complexity of both grapheme groups and participants' L1 orthographies would influence perceptual performance. First, we created an intercept-only model with the following parameters: (1) all main effect and interaction terms between levels of grapheme groups and participants' L1 orthographies (for the fixed effects) and (2) random intercepts for participant and item (for the random effects). Second, we created an item slope model by adding random slopes of participant L1 orthography by items to account for variability in the L1 orthography effect across items; random slopes of grapheme groups by items were not added, because such slopes would not have been appropriate for a between-items manipulation. Next, we used the Likelihood Ratio Test to compare the two models. We took the best-fit of the two models and then tested whether it was further improved by adding random slopes of grapheme groups by participants to account for variability in the grapheme complexity effect across participants. The best-fit of these remaining two models was taken as the best account of the data. Once we had identified the best-fitting random effect structure, we included the main effect of L1 orthography to control for potential L1 effects (i.e., participant response bias toward graphemes from their L1). This method not only facilitated clear examination of grapheme group and participant L1 orthographic complexity effects, but also provided a means of observation of L1 effects.

Below, we present the results of response accuracy followed by those of response time (RT). For each type of data, we summarize the comparisons of models and the parameter estimates from the final model with the L1 effect controlled.

3.2.3.2 Response accuracy (MTurk study)

We tested each model's strength in accounting for response accuracy variation. After adding all random slopes for the effect of participant L1 orthography by item to the intercept-only model, the model failed to converge. Upon examining the partially-converged model, we found variance explained by item slopes to be extremely small (all variances < .01; Freeman et al., 2010), and thus we did not consider the item slopes further. The Likelihood Ratio Test revealed that adding all random slopes for the effect of grapheme group over participants significantly improved the intercept-only model, $\chi^2_{(5)} = 1161.90$, $p < .001$. The results indicated some variability across participants in how their accuracies were predicted by grapheme group. We added an L1 main-effect term to this better-fitting model to form a final model. The results revealed that both grapheme group and participant L1 orthography significantly affected accuracy even when L1 was controlled.

Table 7 displays fixed effect parameter estimates for the final model. Crossing the seven participant L1 orthography contrasts and five grapheme group contrasts resulted in 35 total interaction terms; although we included these interactions in the model, not all were included in the reports because many were not of primary theoretical interest. For simplicity, we used increasing numbers to denote the ordering of grapheme groups and participants' L1 orthographies by increasing complexity; note that numbers used for the grapheme groups were not used for the L1 orthographies. Results of Helmert contrasts, which involved comparing each level of a variable with the mean of the previous levels, indicated major differences in performance between ordered complexity levels. Take the contrast of grapheme groups 6 vs. 1,2,3,4,5, for instance, its estimate indicated how the odds of being accurate were affected by a visually complex condition (i.e., complexity of grapheme group 6) relative to a visually simple condition (i.e., an averaged complexity of grapheme group 1, 2, 3, 4, and 5). Similarly, in one contrast of participant L1 orthography, the estimate indicated how the odds of responding accurately for participants in a more visually complex L1 condition differ from their simple relative counterpart.

The critical question was how accuracy would be affected by the complexity of grapheme groups and the participants' L1 orthographies respectively. Table 8 shows the descriptive statistics, and Figure 7 displays proportions of accurate responses as a function of grapheme groups and participant L1 orthography.

Among grapheme groups, as we expected, more visually complex graphemes yielded lower accuracy. We found that the OBA (i.e., odds of being accurate) were reliably lower for complex Chinese, simple Chinese, and Telugu graphemes than those for the combination of all groups of lesser complexity than each (all *p*s $< .001$; odds = 0.73, 0.88, 0.80, respectively). The OBA were approximately the same for Cree graphemes and the combined simpler groups ($p = .985$). There was an unexpected pattern – the OBA for Russian graphemes were 1.06 times greater than Hebrew graphemes ($p < .001$); however, the effect size of .06 log odds was small.

For participant L1 orthography, we expected that the odds of participants with more visually complex L1 would be greater than those with visually simple L1. This pattern held for Chinese participants (odds of responding accurately were 1.06 times greater than the other seven groups, p = .005) and Japanese participants (odds of responding accurately were 1.17 times greater than the other six groups, $p < .001$), but not for Hindi participants (odds = 0.93, p = .046) or English participants (odds = 0.79, p = .029), all as compared with the combination of all other participant groups with less-complex L1. We did not find any differences in the other contrasts of participant-L1 groups. The L1 effect was significant ($p < .001$); the OBA were two times greater when participants responded to their L1 graphemes.

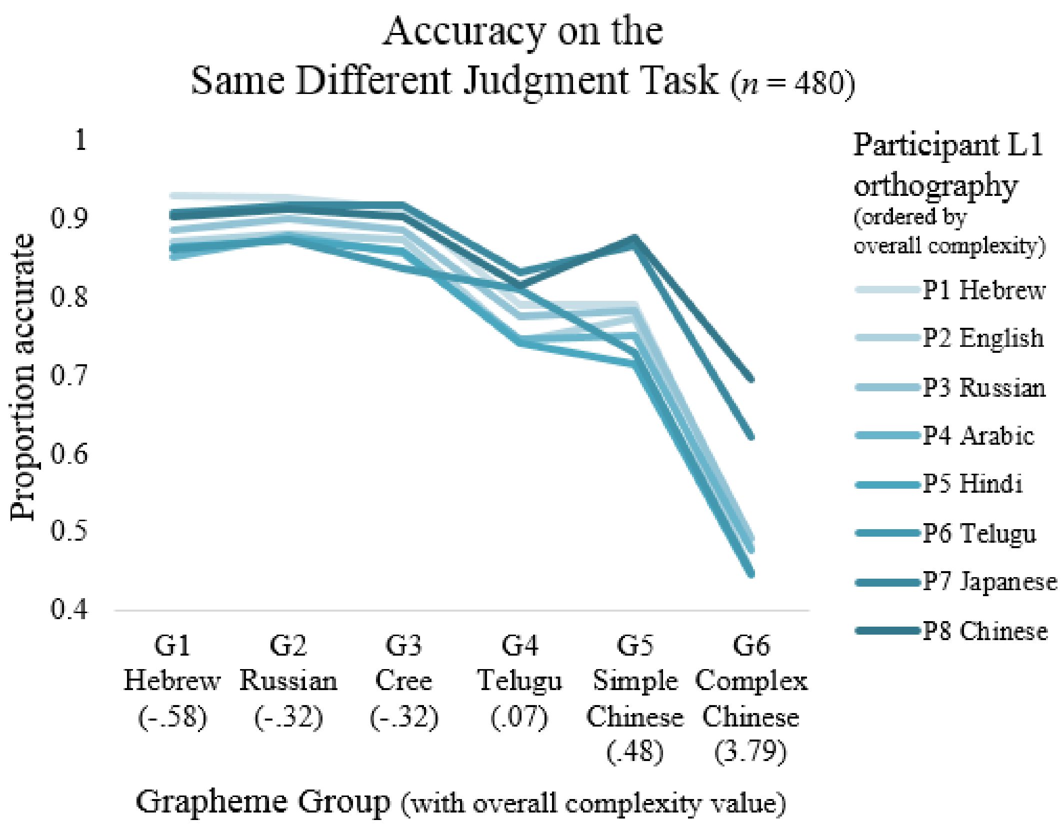

Figure 7. Proportion accurate (n = 60 for each participant L1; MTurk data)

Table 7. Fixed effect parameter estimates for the final logit model for accuracy data in the same-different judgment task (n = 60 for each participant group; MTurk data)

Fixed effect	Estimate (Odds)	Estimate (Log odds)	Standard error	z-value	p-value
(intercept)	4.60	1.53	0.04	34.91	<.001***
Grapheme group 6 vs. 1,2,3,4,5	0.73	-0.31	<0.01	-69.57	<.001***
Grapheme group 5 vs. 1,2,3,4	0.88	-0.12	<0.01	-25.54	<.001***
Grapheme group 4 vs. 1,2,3	0.80	-0.22	<0.01	-38.61	<.001***
Grapheme group 3 vs. 1,2	1.00	-0.01	<0.01	-0.02	.985
Grapheme group 2 vs. 1	1.06	0.06	0.01	4.44	<.001***
Participant group 8 vs.1,2,3,4,5,6,7	1.06	0.06	0.02	2.83	.005**
Participant group 7 vs.1,2,3,4,5,6	1.17	0.16	0.02	6.46	<.001***
Participant group 6 vs.1,2,3,4,5	0.95	-0.05	0.03	-1.88	.060
Participant group 5 vs.1,2,3,4	0.93	-0.07	0.03	-1.99	.046*
Participant group 4 vs.1,2,3	0.93	-0.07	0.04	-1.66	.098
Participant group 3 vs.1,2	0.97	-0.03	0.06	-0.41	.680
Participant group 2 vs.1	0.79	-0.24	0.11	-2.19	.029*
First language (L1) graphemes	2.02	0.70	0.09	7.62	<.001***

Note. For grapheme groups order by complexity, 1 = Hebrew, 2 = Russian, 3 = Cree, 4 = Telugu, 5 = simple Chinese, 6 = complex Chinese. For participant groups ordered by their L1 complexity, 1 = Hebrew, 2 = English, 3 = Russian, 4 = Arabic, 5 = Hindi, 6 = Telugu, 7 = Japanese, 8 = Chinese.

Table 8. Means and standard deviations (in parentheses) of proportion accurate of same-different judgments (n = 60 for each participant group; MTurk data)

Grapheme group (G) / Participant group (P)	G1 Hebrew	G2 Russian	G3 Cree	G4 Telugu	G5 Simple Chinese	G6 Complex Chinese	Marginal means/*SD* of P
P1 Hebrew	.93	0.93	0.91	0.79	0.79	0.48	0.83
(-.58)	(0.26)	(0.26)	(0.28)	(0.41)	(0.41)	(0.50)	(0.37)
P2 English	.87	0.88	0.87	0.74	0.77	0.45	0.79
(-.50)	(0.34)	(0.32)	(0.33)	(0.44)	(0.42)	(0.50)	(0.41)
P3 Russian	.89	0.90	0.89	0.78	0.78	0.49	0.81
(-.32)	(0.32)	(0.30)	(0.32)	(042.)	(0.41)	(0.50)	(0.39)
P4 Arabic	0.85	0.88	0.86	0.75	0.75	0.48	0.78
(-.26)	(0.36)	(0.33)	(0.35)	(0.44)	(0.43)	(0.50)	(0.41)
P5 Hindi	0.86	0.87	0.86	0.74	0.71	0.45	0.78
(-.02)	(0.34)	(0.33)	(0.35)	(0.44)	(0.45)	(0.50)	(0.42)
P6 Telugu	0.86	0.87	0.84	0.81	0.73	0.45	0.79
(.07)	(0.34)	(0.33)	(0.37)	(0.39)	(0.44)	(0.50)	(0.41)
P7 Japanese	0.91	0.92	0.92	0.83	0.87	0.62	0.86
(.54)	(0.29)	(0.27)	(0.28)	(0.37)	(0.34)	(0.49)	(0.35)
P8 Chinese	0.90	0.91	0.90	0.81	0.88	0.69	0.86
(3.79)	(0.30)	(0.28)	(0.30)	(0.39)	(0.33)	(0.46)	(35.)
Marginal means/*SD* of G	0.88	0.90	0.88	0.78	0.79	0.51	
	(0.32)	(0.31)	(0.32)	(0.41)	(0.41)	(0.50)	

3.2.3.3 Response time (MTurk study)

Following the same data analysis plan, we used the Likelihood Ratio Test to examine which model can best account for the RT data. As with the accuracy data, adding random slopes for the effect of participants L1 orthographies by items did not improve the intercept-only model, $\chi^2_{(7)} = 0.73$, $p = .99$, but adding random slopes for the effect of grapheme groups by participants did, $\chi^2_{(5)} = 1161.90$, $p < .001$. Again, the results suggest some variability across participants in how their reaction times differed across grapheme groups. Furthermore, after adding the L1 main effect term, the final model confirmed that both grapheme group and participant L1 orthography significantly affected reaction time, controlling for the L1 effect.

Fixed effect parameter estimates for the final model are summarized in Table 9. The estimates here reflect actual RT differences between conditions with the L1 effect controlled. Table 10 shows the descriptive statistics and Figure 8 displays reaction times on accurately responded items as a function of grapheme group and participant L1 orthography. Among grapheme groups, increased visual complexity clearly led to longer RT. Responses to complex Chinese graphemes were 110 ms slower than to the other five grapheme groups (t = 46.63); simple Chinese graphemes were 44 ms slower than the other four groups (t = 33.55); Telugu graphemes were 79 ms slower than the other three groups (t = 56.51); Cree graphemes were 4 ms slower than the other two groups (t = 2.80). The only exception was that responses to Russian graphemes were 5 ms faster than to the visually simpler Hebrew graphemes (t = -5.69); this pattern was unexpected but it echoed the accuracy results of the Russian-Hebrew contrast. Conversely, it was expected that increased visual complexity of participant L1 orthography would result in faster response time. This expectation was confirmed for Chinese participants (responses about 18 ms faster than the other seven groups, t = -2.09) and Japanese participants

(25 ms faster than the other six groups, t = -3.02) but not Hindi participants (23 ms slower than the other four groups). Again, these specific, reversed trends for Chinese, Japanese, and Hindi participants echoed the patterns in the accuracy results. Furthermore, unlike the accuracy results, the L1 effect in the RTs data was not significant (t = -1.39).

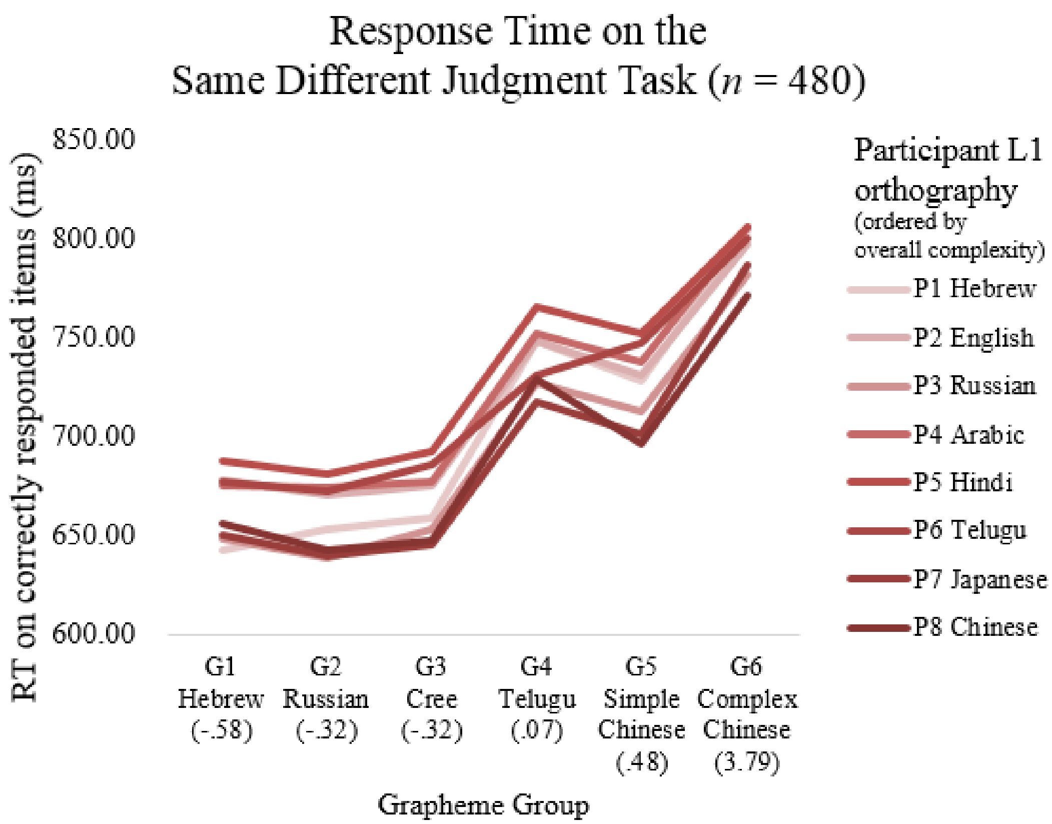

Figure 8. Response time on accurately responded items (MTurk data)

Table 9. Fixed effect parameter estimates for the final model for response time (RT) in the same-different judgment task (n = 60 for each participant group; MTurk data)

Fixed effect	Estimate (ms)	Standard error	*t*-value	Significance
(intercept)	712.14	2.86	248.59	*
Grapheme group 6 vs. 1,2,3,4,5	109.89	2.36	46.63	*
Grapheme group 5 vs. 1,2,3,4	44.38	1.32	33.55	*
Grapheme group 4 vs. 1,2,3	78.57	1.39	56.51	*
Grapheme group 3 vs. 1,2	3.82	1.36	2.80	*
Grapheme group 2 vs. 1	-5.49	0.95	-5.69	*
Participant group 8 vs.1,2,3,4,5,6,7	-17.74	8.50	-2.09	*
Participant group 7 vs.1,2,3,4,5,6	-25.55	8.46	-3.02	*
Participant group 6 vs.1,2,3,4,5	6.61	8.54	0.77	ns.
Participant group 5 vs.1,2,3,4	23.09	8.70	2.65	*
Participant group 4 vs.1,2,3	6.39	9.01	0.71	ns.
Participant group 3 vs.1,2	-16.89	9.52	-1.77	ns.
Participant group 2 vs.1	11.80	11.05	1.07	ns.
First language (L1) graphemes	-10.73	7.70	-1.39	ns.

Note. * indicates significance, the absolute value of $t > 2$ (Baayen, 2008); ns = not significant. For grapheme groups order by complexity, 1 = Hebrew, 2 = Russian, 3 = Cree, 4 = Telugu, 5 = simple Chinese, 6 = complex Chinese. For participant groups ordered by their L1 complexity, 1 = Hebrew, 2 = English, 3 = Russian, 4 = Arabic, 5 = Hindi, 6 = Telugu, 7 = Japanese, 8 = Chinese.

Table 10. Means and standard deviations (in parentheses) of response time (RT) on accurately responded items in same-different judgments (n = 60 for each participant group; MTurk data)

Grapheme group (G) / Participant group (P)	G1 Hebrew	G2 Russian	G3 Cree	G4 Telugu	G5 Simple Chinese	G6 Complex Chinese	Marginal means/*SD* of P
P1 Hebrew	643.13	653.43	659.24	748.51	728.84	804.01	687.97
(-.58)	(120.28)	(120.39)	(126.49)	(126.77)	(127.79)	(132.49)	(134.51)
P2 English	678.67	670.46	675.43	748.39	731.40	797.69	702.16
(-.50)	(126.62)	(126.49)	(123.05)	(130.28)	(126.91)	(131.12)	(132.67)
P3 Russian	648.37	639.50	653.62	727.89	713.34	781.95	678.56
(-.32)	(125.85)	(129.89)	(128.72)	(141.01)	(139.47)	(153.93)	(140.81)
P4 Arabic	675.19	674.70	677.91	752.85	738.54	806.07	705.70
(-.26)	(118.30)	(118.37)	(121.54)	(128.73)	(130.19)	(134.52)	(130.47)
P5 Hindi	688.03	681.38	693.19	765.73	752.52	806.23	716.41
(-.02)	(135.79)	(118.77)	(137.15)	(149.91)	(125.83)	(129.80)	(140.19)
P6 Telugu	677.84	672.70	686.00	731.70	747.94	800.08	704.42
(.07)	(114.94)	(116.19)	(118.69)	(121.79)	(124.30)	(135.11)	(125.25)
P7 Japanese	650.93	640.20	645.61	717.63	701.38	786.62	676.65
(.54)	(125.00)	(119.02)	(116.70)	(126.23)	(122.33)	(123.10)	(129.51)
P8 Chinese	655.98	643.27	647.27	729.13	696.39	771.68	679.86
(3.79)	(118.30)	(112.40)	(108.89)	(125.29)	(115.33)	(119.99)	(123.88)
Marginal means/*SD* of G	664.41	659.18	666.88	739.73	725.03	792.65	
	(124.27)	(121.35)	(124.02)	(132.15)	(127.97)	(132.55)	

3.2.4 Interim summary of Study 2A

To examine how the complexity of grapheme group and participant L1 orthography affected individuals' same-different judgments, we analyzed accuracy and RT data and determined the model that could best account for performance. These best-fitting models (including the intercept-only model - improved by adding random slopes over participants) suggested that perceptual judgment was mainly influenced by stimulus complexity and to a lesser extent by individuals' L1 background. Moreover, controlling for the L1 effect, these models provided several significant results:

(1) The complexity effect of grapheme groups.

More visually complex stimuli (e.g., complex Chinese, simple Chinese, and Telugu) tended to reliably yield lower accuracies and slower RT when compared to their simple relative counterparts, as expected. However, visually simpler stimuli (e.g., Cree, Russian, and Hebrew) did not have a consistent effect of complexity. Russian graphemes showed higher accuracy and faster RT than the simpler Hebrew graphemes, while no difference in accuracy or RT was found between the Cree graphemes and their relative counterparts.

(2) The complexity effect of participant L1 orthography

Participants with more visually complex L1 (e.g., Chinese, Japanese participants) responded more accurately and faster than their counterparts respectively. However, no complexity effect was observed for other participants with visually simpler L1s relative to Chinese and Japanese.

(3) The effect of testing within L1.

Regardless of L1 background, participants responded more accurately on L1 graphemes than non-L1 graphemes, but this L1 advantage was not statistically significant in RT.

In short, individuals' perceptual performance on same-different judgments was strongly affected by grapheme complexity such that greater complexity hindered performance, and was further affected, although to a lesser extent, by complexity of individuals' L1 orthographies, when the effect of responding to L1 graphemes was controlled.

3.3 STUDY 2B: VISUAL DISCRIMINATION PERFORMANCE WITHIN A WRITING SYSTEM (LAB STUDY)

The results of Study 2A generally suggest the complexity effect: more visually complex graphemes yield lower discrimination efficiency and participants who mastered more visually complex L1 (especially those who learned to read Chinese characters) outperformed those who mastered less visually complex L1. However, results also revealed several unexpected patterns that may need to be replicated. Moreover, Study 2A does not speak to whether the complexity of participants' L1 orthographies, independent from mapping principles, directly drives individuals' perceptual difference. One possible way to answer this question is to examine perceptual performance between two groups of individuals who speak the very same language but use orthographies with varied complexities. The best test-bed may be the Chinese language, because it employs two visual orthographies – the more visually complex "traditional" Chinese orthography used in Taiwan (and Hong Kong), and the more visually simple "simplified" Chinese orthography used in China. Given that groups in both Taiwan and China use Mandarin

Chinese as their official language, by reducing possible interference from other linguistic factors to minimum, the observed group differences can be attributed to the visual complexity of their orthographies.

The goals of Study 2B were threefold. First, we replicated the findings of Study 2A by recruiting age- and gender-matched participants and tested contrasting visual L1 orthographies with the same mapping principle and language (e.g., traditional: Taiwan; simplified: China). Second, we examined the extent to which the complexity effect can be generalized from linguistic stimuli to non-linguistic stimuli by testing the same participants on a pattern discrimination task. Third, we sought to enhance the internal validity of the findings by switching the experimental setting from the Internet to the lab. We expected the Taiwan group to outperform the China group on discriminating both the grapheme and non-grapheme stimuli.

3.3.1 Method

3.3.1.1 Participants

Sixty adults were recruited for the Taiwan and China groups respectively. All participants met the following recruiting criteria: (1) Mandarin Chinese as their first spoken language, (2) completed nine-year formal education in Taiwan or China, (3) age from 18 to 35 years old, (4) right-handed, (5) college or graduate students, and (6) no reading difficulty. The Taiwan group participated at the National Taiwan Normal University in Taiwan, and the China group participated at the University of Pittsburgh in the US. They received monetary compensation for their participation.

Table 11 provides background information for the Taiwan and China groups. These two groups were matched on age, $F_{(1, 119)} = 1.39$, $p = .24$ (Taiwan: $M = 24.15$, $SD = 3.64$; China: $M =$

24.87, *SD* = 2.98) and gender, $F_{(1, 119)} = 0.14$, $p = .71$ (Taiwan: 24 males; China: 22 males). No group differences were found for the number of learned languages, $F < 1$ (Taiwan: 2.30; China: 2.24) nor the self-rated general language learning skills, $F < 1$ (Taiwan: 4.54; China: 4.78 on a 7-point scale where 1 indicated the lowest level of skills of learning new languages). However, analysis of variance on the vocabulary size task showed that the Taiwan group scored worse than the China group, $F_{(1, 119)} = 25.09$, $p < .01$; the China group reported that they have studied abroad in the US for 1.34 years on average (*SD* = .55 years).

Table 11. Background characteristics for the China and Taiwan participants

Background / Participant group	Number of participants	Age (years)	Number of learned languages (including first language)	Self-rated general language learning skills	Vocabulary size task (raw scores)
China (Simplified Chinese)	60 (22 males)	24.87 (2.98)	2.24 (0.52)	4.78 (1.10)	94.18 (13.27)
Taiwan (Traditional Chinese)	60 (24 males)	24.15 (3.64)	2.30 (0.61)	4.54 (1.11)	82.08 (10.21)

Note. Means (standard deviation) reported. Self-rated general language learning skills is a self-reported measure of language-learning ability that includes listening, speaking reading, and writing on a 7-point scale, with 1 being the lowest.

3.3.1.2 Design

A between-participants design (Taiwan vs. China) was used to examine the influence of participants' L1 orthographies on visual perceptual performance of grapheme and non-grapheme stimuli. Response accuracy and time served as dependent measures.

3.3.1.3 Task

Both the Taiwan and China groups received identical tasks: a pattern discrimination task and four tasks identical to Study 2A (see Task in Study 2A for details). The same-different judgment task was designed with four lists and each list was used for equal numbers of participants from both participant groups.

Pattern discrimination task The pattern discrimination task was adapted from a complex working memory span task (Chein & Morrison, 2010), which has been shown to cover a broad range of visual processing difficulty for adults (Morrison & Chein, 2011). We revised this task to tap individuals' capacity for visual form discrimination. In this task, participants were required to discriminate between two complex checkerboard patterns while making a categorical decision. Each checkerboard pattern measured 1.5 inches square, yielding a visual angle of approximately 4.8°. The task consisted of 100 trials over 5 minutes with breaks between blocks of 20 trials (four total). For each trial, two checkerboard patterns were presented simultaneously side by side (left-and-right) and participants were encouraged to respond as accurately and quickly as possible, within a limit of 2.5 seconds. They were asked to press "1" if the patterns were both symmetrical or both asymmetrical and "2" if only one was symmetrical.

3.3.1.4 Procedure

Consent was obtained prior to the experiment. The orders of the tasks were the same for both groups: a pattern discrimination task, a same-different judgment task, a language history questionnaire, a vocabulary size test, and then a Chinese-English translation task. All of the tasks were individually administered in a quiet lab space by trained psychology-major students in a one-hour experiment session.

3.3.2 Results

We analyzed the same-different judgments with mixed effect models. For consistency, we used the most appropriate random effects structure as identified in Study 2A. The models, however, were slightly different here due to consideration of L1 which, when combined with all the partially redundant interaction terms, introduced confounds and required us to drop one variable from the model. We decided to drop the contrast of grapheme groups 6 vs. 1, 2, 3, 4, 5 in the analysis and ignored the contrast of grapheme groups 5 vs. 1, 2, 3, 4 in the report, as these effects were confounded by the L1 effect (Chinese participants responding to Chinese graphemes).

We are interested in examining the different performance between Taiwan and Chinese groups when controlling for the L1 effect, with a focus on the effect of grapheme groups.

3.3.2.1 Response accuracy (Lab study)

Results of the same-different judgments showed that the Taiwan group had higher accuracy than the China group. Table 12 lists the descriptive statistics. Figure 9 illustrates accuracy as a function of grapheme group and Chinese group; the MTurk Chinese data were presented for reference and was not included in the current analysis. Table 13 displays fixed effect parameter estimates for the accuracy data. We found that the odds of responding accurately for the Taiwan group were 1.19 times greater than the China group (p = .04). For the grapheme groups, we expected that more visually complex graphemes would lead to lower accuracy. This pattern held for Telugu graphemes (their OBA were 0.49 times lower than all simpler grapheme groups combined: Hebrew, Russian, and Cree graphemes, p < .001) and Cree graphemes (their OBA were 0.90 times lower than all simpler combined: Hebrew and Russian graphemes, p = .027) but not for Russian graphemes (their OBA were 1.17 times greater than Hebrew graphemes, p

= .006). This pattern of complexity effect for grapheme group was consistent with the MTurk results.

Table 12. Means and standard deviations (in parentheses) of proportion accurate of same-different judgments (n = 60 for each Chinese group; Lab data)

Grapheme group (G) / Chinese group(C)	G1 Hebrew	G2 Russian	G3 Cree	G4 Telugu	G5 Simple Chinese	G6 Complex Chinese	Marginal means/*SD* of C
China (Simplified Chinese)	0.90 (0.31)	0.91 (0.28)	0.90 (0.31)	0.83 (0.38)	0.89 (0.31)	0.77 (0.42)	0.87 (0.34)
Taiwan (Traditional Chinese)	0.92 (0.27)	0.93 (0.26)	0.92 (0.27)	0.85 (0.36)	0.91 (0.28)	0.76 (0.43)	0.89 (0.31)
Marginal means/*SD* of G	0.91 (0.29)	0.92 (0.27)	0.91 (0.29)	0.84 (0.37)	0.84 (0.30)	0.76 (0.42)	

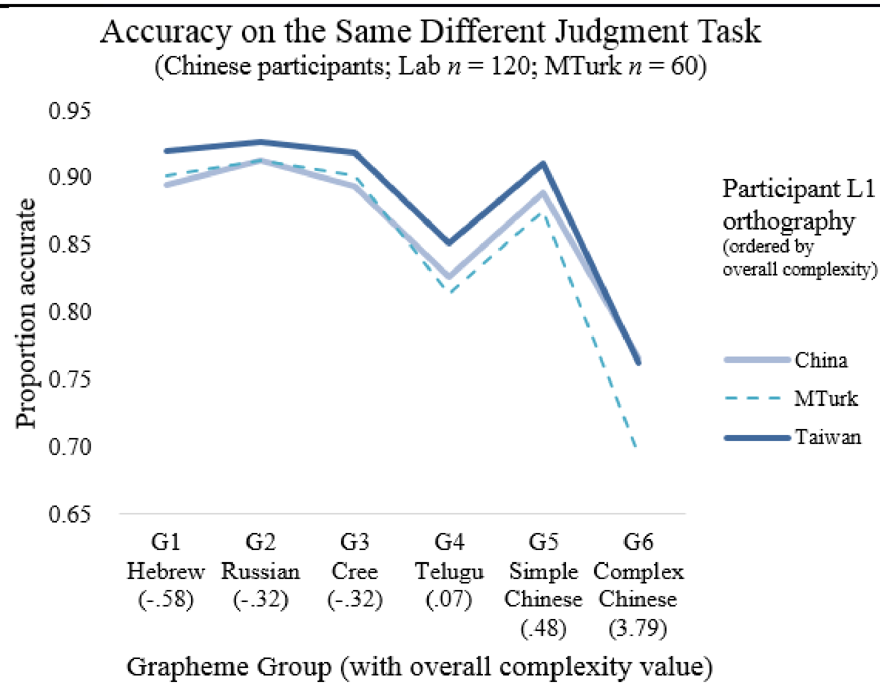

Figure 9. Proportion accurate (Lab data; MTurk data for reference)

Table 13. Fixed effect parameter estimates for the final logit model for accuracy data in the same-different judgment task (n = 60 for each Chinese group; Lab data)

Fixed effect	Estimate (Odds)	Estimate (Log odds)	Standard error	z-value	p-value
(intercept)	12.26	2.51	0.05	53.94	< .001***
Taiwan vs. China participants	1.19	0.17	0.09	2.02	.04*
Grapheme group 5 vs. 1,2,3,4	3.75	1.32	0.07	17.67	< .001***
Grapheme group 4 vs. 1,2,3	0.49	-0.71	0.04	-18.73	< .001***
Grapheme group 3 vs. 1,2	0.90	-0.10	0.05	-2.21	.027*
Grapheme group 2 vs. 1	1.17	0.16	0.06	2.77	.006**
First language (L1) graphemes	0.28	-1.28	0.05	-26.33	< .001***

Note. For grapheme groups, 1 = Hebrew, 2 = Russian, 3 = Cree, 4 = Telugu, 5 = simple Chinese

3.3.2.2 Response time (Lab study)

Consistently, results on response time for accurate responses showed that the Taiwan group outperformed the China group. Table 14 summarizes the descriptive statistics. Figure 10 illustrates response time as a function of grapheme groups and Chinese groups. Table 15 displays fixed effect parameter estimates; these estimates reflect actual RT differences between conditions with the L1 effect controlled. For the two Chinese groups, we found that the Taiwan group responded about 17ms faster than the China group (t = -2.17). For the grapheme groups, we expected that more visually complex graphemes would yield longer response time. We did find that responses to the Telugu graphemes were about 82 ms slower than to their less complex counterparts (t = 53.99). Meanwhile, there was no difference between the Cree graphemes and their less complex counterpart (t = 0.41). Responses to the Russian graphemes were about 11ms faster than to their less complex counterpart, Hebrew graphemes (t = 5.39), a pattern which was unexpected under the hypothesis, yet was consistent with the accuracy results.

Table 14. Means and standard deviations (in parentheses) of response time (RT) of same-different judgments (n = 60 for each Chinese group; Lab data)

Grapheme group (G) / Chinese group(C)	G1 Hebrew	G2 Russian	G3 Cree	G4 Telugu	G5 Simple Chinese	G6 Complex Chinese	Marginal means/*SD* of C
China (Simplified Chinese)	662.55 (125.86)	648.09 (124.58)	657.21 (124.04)	735.61 (128.07)	679.22 (123.70)	761.21 (125.49)	683.59 (131.45)
Taiwan (Traditional Chinese)	639.04 (111.71)	631.92 (112.16)	635.24 (108.61)	717.59 (118.63)	670.44 (111.39)	748.97 (116.06)	665.48 (120.20)
Marginal means/*SD* of G	650.63 (119.46)	639.95 (118.75)	646.08 (116.98)	726.46 (123.69)	674.78 (117.80)	755.11 (121.01)	

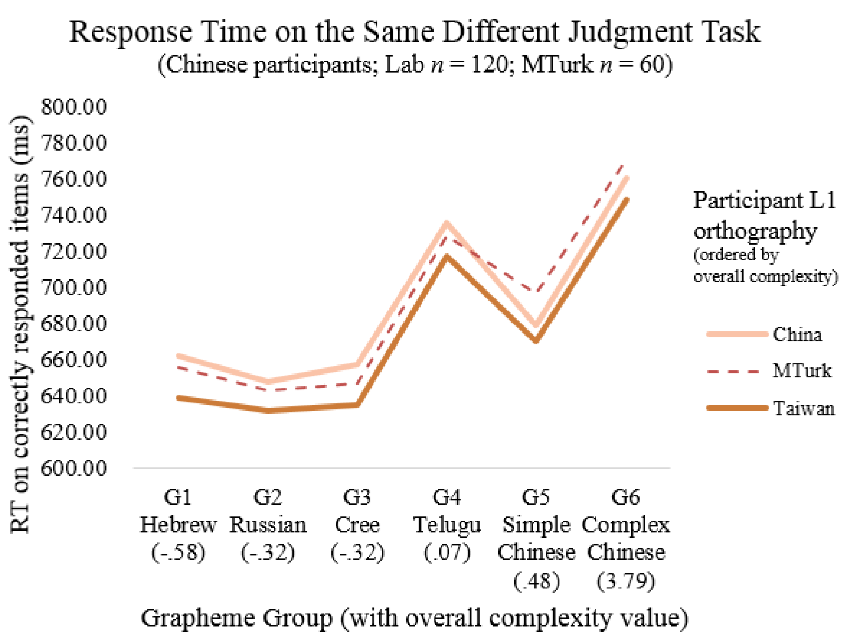

Figure 10. Response time on accurately responded items (Lab data; MTurk data for reference)

Table 15. Fixed effect parameter estimates for the final model for response time (RT) in the same-different judgment task (*n* = 60 for each Chinese group; Lab data)

Fixed effect	Estimate (ms)	Standard error	*t*-value	Significance
(intercept)	645.38	8.10	79.72	*
Taiwan vs. China participants	-17.21	7.94	-2.17	*
Grapheme group 5 vs. 1,2,3,4	-101.86	3.11	-32.71	*
Grapheme group 4 vs. 1,2,3	81.72	1.51	53.99	*
Grapheme group 3 vs. 1,2	0.62	1.61	0.41	ns
Grapheme group 2 vs. 1	-11.01	2.05	-5.39	*
First language (L1) graphemes	110.30	2.42	45.64	*

Note. * indicates significance, the absolute value of $t > 2$ (Baayen, 2008); ns = not significant. For grapheme groups, 1 = Hebrew, 2 = Russian, 3 = Cree, 4 = Telugu, 5 = simple Chinese, 6 = complex Chinese. For participant groups, 1 = Hebrew, 2 = English, 3 = Russian, 4 = Arabic, 5 = Hindi, 6 = Telugu, 7 = Japanese, 8 = Chinese.

3.3.2.3 Pattern discrimination task

The question we wanted to answer with this pattern discrimination task was whether the observed complexity effect (Taiwan vs. China groups) from grapheme stimuli would hold for non-grapheme stimuli. Analysis of variance confirmed that the Taiwan group responded faster when responding accurately than the China group ($F_{(1, 119)} = 5.83$, $p = .01$), whereas no difference was found in response accuracy itself ($F_{(1, 119)} = 0.32$, $p = .57$). Table 16 shows the descriptive statistics.

Table 16. Means and standard deviations (*SD*) of proportion accurate and response time (RT) of the pattern discrimination task (n = 60 for each Chinese group; Lab data)

	Proportion accurate (%)		RT (ms)	
Chinese group	*M*	*SD*	*M*	*SD*
China (Simplified Chinese)	80.41	10.29	1672.45	160.78
Taiwan (Traditional Chinese)	79.33	10.64	1582.45	241.70

3.3.3 Interim summary of Study 2B

Study 2B examined the extent to which the complexity of grapheme group and of participant L1 orthography affects individuals' visual perceptual performance by replicating Study 2A in the lab, by comparing performance of two groups of Chinese speakers using varied visual orthographies (complex: Taiwan vs. simple: China), and by adding a pattern discrimination task involving checkerboard stimuli comparable in visual complexity to the complex graphemes. The results were that the Taiwan group was more accurate and faster in discriminating among grapheme stimuli and faster when accurately judging complex, non-grapheme stimuli, although there was no advantage for accuracy alone in the pattern discrimination task.

To sum up, the results of Study 2B resonated the complexity patterns of grapheme group of Study 2A, and importantly, Study 2B confirmed with the complexity effect of participant L1 orthography as seen in Study 2A: individuals mastering a visually complex orthography may develop stronger visual perceptual skills than those mastering a more simple orthography, such skills serving to enhance relative discrimination performance.

4.0 STUDY 3: MODELING VISUAL ORTHOGRAPHIC LEARNING ACROSS WRITING SYSTEMS

In Study 2A, we did find the complexity effect of participant L1 orthography – when comparing skilled readers from different writing systems on their perceptual judgments, those readers who mastered more visually complex orthographies (mainly Japanese and Chinese) outperformed readers of simpler orthographies. However, across writing systems, the observed effects of grapheme complexity on perception cannot be separated fully from the effects of mapping principles between graphemes and their phonological units in particular writing systems.

In Study 3, we addressed this issue by developing a computational model with no access to phonology, focused solely on visual properties of graphemes. We trained each of 131 identical models to learn the structure of a different orthography, and tested eight trained models that represented skilled L1 readers on stimuli taken from six grapheme groups to replicate Study 2A. Thus, this model provides a test of pure orthographic learning. We used this model as a tool to test three hypotheses repeatedly posed in this research:

(1) Grapheme complexity leads to learning difficulty (as discussed in the Introduction)

(2) Grapheme complexity imposes perceptual demands in processing (as shown in Study 2).

(3) Learners of more visually complex orthographies develop stronger visual perceptual skills (as revealed in Study 2).

We expected the models to provide direct support for the following concepts: the requirement to master more visually complex graphemes in one orthography relative to another orthography would impose more visual perceptual demands on the viewers of that orthography and thus they would develop stronger visual skills; these visual skills are not necessarily driven by mapping principle.

4.1 OVERVIEW: A MODEL WITH A DISTRIBUTED CODING SCHEME SERVES AS A UNIVERSAL ORTHOGRAPHIC LEARNING DEVICE [1]

There are a growing number of computational models addressing orthographic representations (e.g., the Spatial Coding model, Davis, 2010; the Overlap model, Gomez, Ratcliff, & Perea, 2008; the Bayesian Reader model, Norris, Kinoshita, & van Casteren, 2010; the sequential encoding regulated by inputs to oscillating letter [SERIOL] model, Whitney, 2001). These models, however, were developed to code alphabetic orthographies, and are not applicable to more visually complex orthographies such as Chinese. Although some models have been developed to code Chinese (e.g., Perfetti, Liu, & Tan, 2005; Taft, 2006; Yang, McCandliss, Shu, & Zevin, 2009), the orthographic coding schemes used were slot-based, requiring independent coding specific to graphemic forms of Chinese such as radicals or strokes, and thus had no natural generalization to other orthographies.

---

[1] This modeling work has been submitted for a journal review with the following title and all authors' contributions: Chang, L. Y., Plaut, D., & Perfetti, C. A. (2014) Visual-orthographic complexity in learning to read: Modeling learning across writing system variations. *Scientific Studies of Reading*.

To capture the various visual forms of writing systems, what is needed is a way to encode the full range of graphemes in terms of basic, universal elements that apply to any orthography; a model with a universal coding scheme would serve this purpose. To simulate orthographic learning, what is essential is to represent knowledge acquisition; the Parallel Distributed Processing (PDP) framework provides learning outcomes (e.g., changes in the model's output over time as a function of the input it receives). In PDP models (Plaut, 2005; Seidenberg, 2006), processing takes the form of cooperative and competitive interactions over many simple processing units, instead of activation of single units. Knowledge is encoded by weights on the interconnections among these units; learning involves iteratively adjusting these weight values based on performance feedback. After learning, these models can generalize their knowledge to novel input, and performance is determined by the similarity between the novel and learned representations. In short, PDP models instantiate learning as an incremental increase in knowledge. Such models have been used to simulate reading processes in English (e.g., Zevin & Seidenberg, 2006) and in Chinese (e.g., Yang, McCandliss, Shu, & Zevin, 2009); in skilled and less-skilled readers (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996); and in normal and dyslexic readers (e.g., Harm & Seidenberg, 1999; Plaut, 1999; Woollams, Lambon Ralph, Plaut, & Patterson, 2007, among others).

In Study 3, we aimed to develop a PDP model with a distributed coding scheme to serve as a universal orthographic learning device. With this model, our first goal was to demonstrate how visual complexity of orthographies can drive difficulty in orthographic learning across writing systems; we applied the same basic functional architecture to simulate learning in 131 orthographies. Our second goal was to show how difficulty of perceptual processing can be influenced by both complexity of presented stimuli themselves and their relationships to the L1 orthography of the viewer; we tested models representing learners from eight orthographies on stimuli with varied complexities to replicate the perceptual experiment in Study 2A.

4.2 METHOD

4.2.1 Model architecture

The model is a specific form of three-layer neural network known as an *encoder* network. A standard encoder network learns to copy patterns of activity over a group of input units onto an identically-sized group of output units via a smaller number of intermediate or "hidden" units. Because there are fewer hidden units than input (or output) units, the network must learn to re-represent the inputs in a more concise form. In this way, the hidden representations come to emphasize the underlying structure shared by the ensemble of inputs at the expense of more idiosyncratic aspects of only one or a few patterns (Hinton & Salakhutdinov, 2006).

Figure 11 illustrates the architecture of the specific network used in the current work. The input patterns are images of graphemes over the 38×38 array of units at the bottom of the figure—note that, because the input and output groups have exactly the same structure, only a single group of units is shown. In Figure 11, each small square corresponds to a unit. Input is presented as activity values (shown in grayscale, with black = 0.0 and white = 1.0) over the 38×38 array at the bottom; four groups of hidden units, varying in number and in receptive field size and spacing, are shown at the top. Input-to-hidden and hidden-to-output connections were restricted to topographically constrained circular "receptive fields"—the red lines depict the scale of these receptive fields for four representative hidden units (no actual connections are shown). The output units have exactly the same 38×38 form as the input units and are not depicted separately; rather, their activations (for an example complex Chinese character after training) are shown in the central region of each input unit, with the actual input value shown in the surrounding ring. Thus, units for which the center and surround match one another are fully accurate in their reconstructed activations.

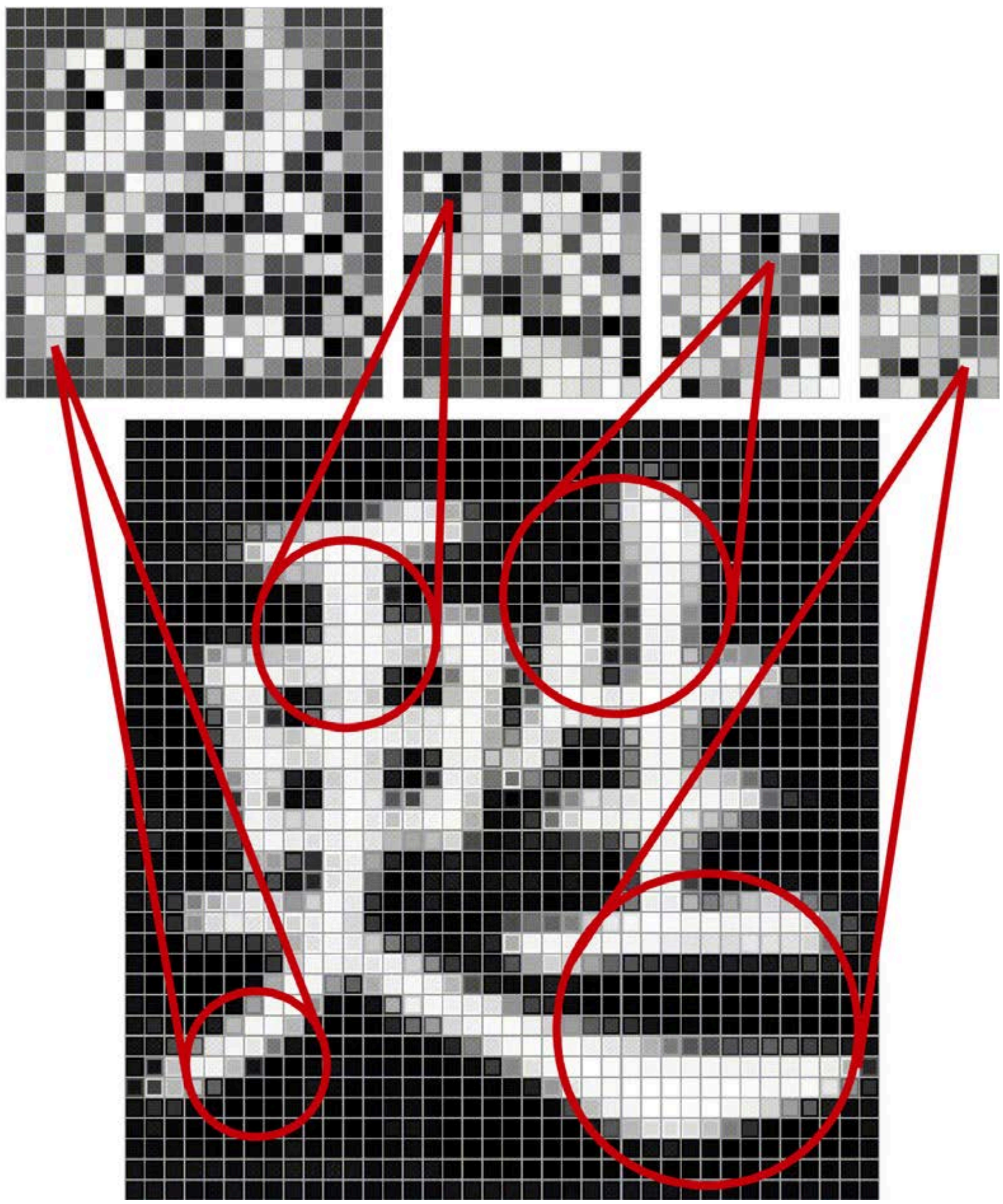

Figure 11. The architecture of the model used in the simulation

The hidden layer is divided into four groups of units (shown at the top of Figure 11) that differ in number of units and in the sizes of their "receptive fields" (RFs). In particular, each hidden unit receives input only from a restricted circular region of the input, and projects to the corresponding circular region of the output (these are depicted in red for four representative units). To allow the network to learn to be sensitive to features of varying scales and positions, different groups of units had different RF sizes, with centers spaced evenly across the input (and output) arrays: a 19×19 group with a RF diameter of 5 units and centers spaced every 2 units horizontally and vertically; a 12×12 group with diameter = 7 and spacing = 3; a 9×9 group with diameter = 11 and spacing = 4; and a 7×7 group with diameter = 15 and spacing = 5. Including "bias" connections (which determine the activation of units in the absence of other inputs), the network had a total of 83,607 connections. As a point of comparison, if all 635 hidden units were fully connected to both the input and output, the network would have required 1,835,959 connections. Using topographically restricted connectivity not only drastically reduces the required number of connections, and is broadly compatible with patterns of connectivity in visual cortex, but also encourages the network to discover largely local features of varying scales.

4.2.2 Stimuli

We developed two sets of stimuli: one training set and one testing set. The training patterns were used to simulate L1 orthographic learning; they consisted of all 131 orthographies in Study 1 (alphabetic: 60; alphasyllabary: 41; abjad: 16; syllabary: 11, and morphosyllabary: 3). To generate these training patterns, the 21,821 grapheme images in Study 1 were resized from 500×500 to 38×38 pixel dimension for computational convenience, then converted to 8-bit integer values, further inverted and normalized to real values between 0.0 and 1.0 (in gray scales, with black = 0.0 and white = 1.0). The testing patterns were used to simulate human behavior in the same-different judgment; they were comprised of all stimuli in Study 2 (six grapheme groups ranked form least to greatest in complexity: Hebrew, Russian, Cree, Telugu, simple Chinese characters, and complex Chinese characters). The patterns were generated in the same manner as the training patterns.

4.2.3 Training

In the real world, successful orthographic learning occurs when learners are able to correctly identify a grapheme, recognizing it as one they have previously seen. In computational modeling, learning occurs when hidden units can detect feature differences in the input layer and reconstruct the representations onto the output layer with minimal difference between the target activations and the actual activations in the hidden layer—that is, minimal reconstruction error. To reduce reconstruction error, we used the back-propagation algorithm (Rumelhart, Hinton, & Williams, 1986) in the present model, with online learning, a learning rate of 0.01 and momentum of 0.8.

To simulate orthographic learning across writing systems, we created 131 encoder networks and trained them on grapheme patterns from each of 131 orthographies. Training was halted when the average reconstruction error across the entire set of graphemes in that orthography fell below 10. The number of learning epochs that the model required in reaching the average error of 10 was taken as the primary measure of the difficulty of learning a given orthography. This learning epoch measure is important because it allows us to reliably compare different models given that number of graphemes varied across orthographies in training.

4.2.4 Testing

To model how individuals with different L1 experiences approach graphemes with various complexities, we first selected eight trained encoders (average error < 10) to represent skilled L1 readers, and presented these encoders (i.e., Hebrew, English, Russian, Arabic, Hindi, Telugu, Japanese, and Chinese) with testing patterns consisting of pairs of both identical and differing graphemes, taken from six grapheme groups (i.e., Hebrew, Russian, Cree, Telugu, simple Chinese characters, and complex Chinese characters). Each grapheme in a pair was presented separately to the network, and the activation values over all 635 hidden units were recorded.

We assume that same-different judgments are made on the basis of the similarity of these representations. To measure this similarity, we repeatedly added noise to each element of the two hidden patterns, computed the correlations between each pair of patterns, and then averaged the results (n.b. noise was added to both "same" and "different" trials because the hidden representations of the "same" trials were identical). We characterized each model's performance in terms of average correlation within each grapheme group, and tied these correlations to human performance. Higher correlations indicated worse performance (e.g., longer reaction time or lower accuracy) because the two given graphemes in "different" pairs were more similar and thus more difficult to discriminate; lower correlations indicated better performance. Finally, to approximate the accuracy data in the behavioral experiment, these correlations were inverted for data analysis.

4.3 RESULTS

4.3.1 Grapheme complexity is strongly associated with learning difficulty

In training encoder networks to learn internal representations of graphemes, we were particularly interested in the relationship between learning performance and grapheme complexity. We correlated the number of learning epochs (to reach average error < 10) from 131 encoders with the overall complexity measure of the 131 corresponding orthographies in Study 1. This resulted in a significant correlation of .68 ($p < .001$; two-tailed), suggesting that grapheme complexity and learning difficulty were strongly, positively associated.

4.3.2 Encoder accuracy is a function of stimulus complexity and L1 orthography

In simulating readers from eight L1 orthographies perceiving graphemes pairs from six grapheme groups, we asked whether L1 background differentially affects perceptual processing across grapheme complexity levels in viewers. We conducted an 8 × 6 (L1 background × grapheme group) analysis of variance with encoder accuracy as the dependent measure; we did not use mixed effect modeling as in Study 2 because the structures of human and simulation data were not identical (e.g., no random slopes by participants in the simulation data).

Results revealed that although simple stimuli were equally difficult for all encoders, complex stimuli were more difficult for encoders trained on less complex orthographies, but not for encoders trained on more complex orthographies. Figure 12 illustrates how encoder accuracy for each tested stimuli set is presented as a function of trained L1 orthography. The main effect of L1 orthography was significant, $F(7, 12960) = 39.70$, $p < .001$, $\eta_p^2 = .021$; the main effect of stimulus complexity was also significant, $F_{(5, 12960)} = 277.94$, $p < .001$, $\eta_p^2 = .097$. There was a significant interaction between L1 background and complexity, $F_{(35, 12960)} = 6.58$, $p < .001$, $\eta_p^2 = .018$. To better understand this interaction, pair-wise comparisons using Bonferroni adjustments were conducted to control for the overall Type I error. Table 17 provides a summary of the comparisons along with the means and standard deviations of encoder accuracy.

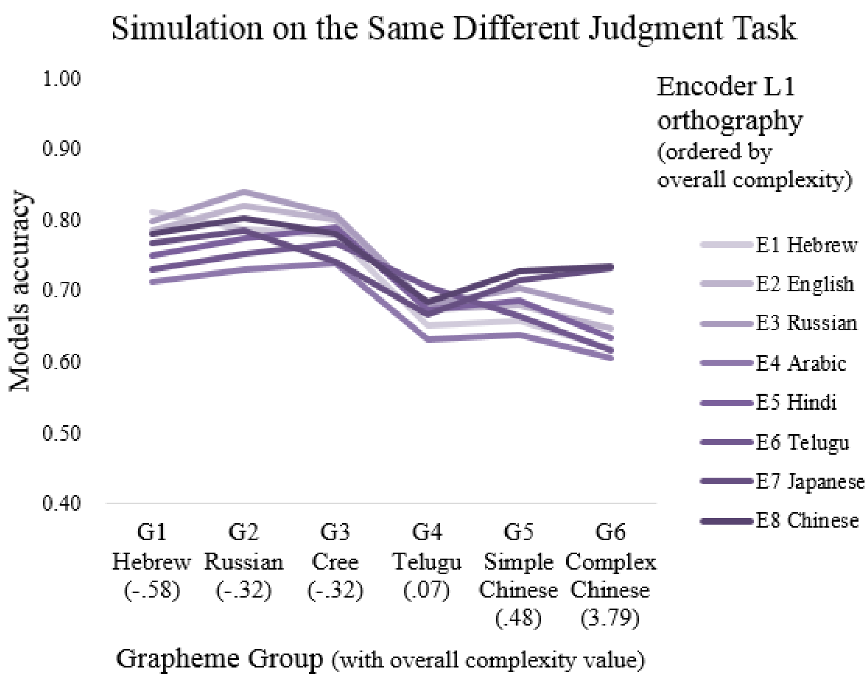

Figure 12. Modeling results of same-different judgments grapheme pairs drawn from different grapheme groups, made by encoders trained with different orthographies

As a general trend, all encoders performed worse on complex stimuli (e.g., complex Chinese character, simple Chinese characters, and Telugu than simple stimuli (e.g., Cree, Russian, and Hebrew), regardless of L1 background. This complex-simple distinction is consistent with that we found in Study 2: complex Chinese characters, simple Chinese characters, and Telugu yielded lower accuracies and slower RT when compared to the combination of all groups of lesser complexity than each.

Table 17. Means and standard deviations (in parentheses) of encoder accuracy in examining whether L1 orthographic complexity differentially affects perceptual variability across stimulus complexity levels

Encoder (ordered by trained orthographies from simple to complex)	Stimulus group (ranked complexity from simple to complex)						F	η_p^2	Pairwise comparison (with Bonferroni adjustments)
	1 Hebrew	2 Russian	3 Cree	4 Telugu	5 Simple Chinese	6 Complex Chinese			
Hebrew (-.58)	.81 (0.19)	.79 (0.15)	.78 (0.12)	.65 (0.18)	.66 (0.18)	0.62 (0.17)	58.81**	.170	1,2,3>4,5,6
English (-.50)	0.79 (0.18)	0.82 (0.15)	.80 (0.11)	.67 (0.18)	.68 (0.18)	0.65 (0.17)	51.32**	.152	1,2,3>4,5,6
Russian (-.32)	.80 (0.19)	.84 (0.14)	0.81 (0.11)	.68 (0.20)	.70 (0.18)	0.67 (0.17)	46.24**	.139	2>1,3>4,5,6
Arabic (-.26)	.71 (0.17)	0.73 (0.14)	.74 (0.10)	.63 (0.19)	.64 (0.18)	0.60 (0.21)	29.55**	.093	1,2,3>4,5,6
Hindi (-.02)	.75 (0.19)	0.77 (0.14)	.79 (0.10)	.67 (0.18)	.69 (0.17)	0.63 (0.21)	33.33**	.104	1,2,3>4,5,6
Telugu (.07)	.73 (0.19)	0.75 (0.15)	.77 (0.11)	.71 (0.20)	.66 (0.17)	0.62 (0.20)	23.43**	.076	1,2,3,4>5,6
Japanese (.54)	.77 (0.16)	0.78 (0.13)	.74 (0.09)	.67 (0.15)	.71 (0.14)	0.73 (0.13)	26.79**	.085	1,2,3>5,6>4
Chinese (3.79)	0.78 (0.18)	.80 (0.14)	.78 (0.10)	.69 (0.17)	.73 (0.16)	0.73 (0.15)	46.19**	.074	1,2,3>5,6>4

** $p < .001$; For the pairwise comparison with Bonferroni adjustments, all ps <.001.

Interestingly, encoders trained on different orthographies showed varying difficulty when tested on stimuli from their original training orthography, analogous to testing within L1. For the Hebrew encoder, the Hebrew stimuli were just as difficult as other simple stimuli (i.e., Russian, and Cree) and less difficult than complex stimuli (e.g., Telugu, simple Chinese characters, and complex Chinese characters). For the Russian encoder, however, the Russian stimuli were the least difficult, whereas the decreasing accuracy gradient from simple to complex was maintained for other stimuli. Notably, although Telugu stimuli were complex, the Telugu encoder deviated from the complex-simple distinction and showed less difficulty with Telugu stimuli than with some of the simple stimuli, although not the least difficulty. Finally, for the Chinese L1 encoder, gradient directions switched between Chinese and Telugu stimuli. Both Chinese stimulus sets became less difficult than Telugu, even though both Chinese sets were more complex than Telugu; Chinese stimuli remained more difficult than the simple stimuli. These results suggested that, although the L1 effect (e.g., higher accuracy in processing L1 graphemes) may play a role in encoders' perceptual performance, stimulus complexity seemed to have a greater effect.

To further elucidate the stimulus complexity effect, the patterns from L1 encoders tested on non-L1 stimuli were investigated. For the L1 English, Arabic, and Hindi encoders, accuracy on complex stimuli (e.g., Telugu, simple Chinese and complex Chinese characters) were clearly lower than for simple stimuli (e.g., Hebrew, Russian, and Cree). The Japanese encoder showed nearly the same trend, although accuracy was higher for Chinese stimuli than for Telugu, both of which were complex sets. Given that the Japanese encoder was trained on all three character types in the Japanese orthography, namely Katakana, Hiragana, and Kanji, and that 56% of the trained Kanji characters overlapped with the Chinese training set, this gradient reversal between Chinese and Telugu stimuli was not surprising. Collectively, results from the four encoders tested on non-L1 stimuli confirmed the complexity effect as witnessed in the other four encoders – overall, stimulus processing performance decreased as stimulus complexity increased, whereas complexity of L1 orthography interacted with stimulus complexity to produce patterns of performance that differed by L1.

4.4 INTERIM SUMMARY

In attempting to focus on visual complexity for the effects of visual orthographic variation on learning to read across writing systems, Study 3 developed a universal orthographic learning model simulating learning to read, as well as simulating directed reading behavior in the form of a grapheme discrimination task.

The results of Study 3 demonstrated pure visual orthographic learning across writing systems and resonated with the findings of Study 2A:

(1) Visual orthographic learning across writing systems:

After training 131 encoders to reach the same average error level, roughly equivalent to reaching the same level of grapheme mastery, difficulty of learning a given orthography (as represented by number of training epochs required to reach the aforementioned error level) was found to be positively, strongly associated with overall grapheme complexity, in all 131 orthographies (r = .68). The results suggest that grapheme complexity, shown to be governed by grapheme inventory in Study 1, drives learning difficulty in terms of mastering the full grapheme set of a particular orthography.

(2) Discrimination judgments across writing systems:

When testing eight trained encoders (average error < 10) on identical and differing stimulus pairs from six grapheme groups, a significant interaction was found between encoder L1 background and grapheme group, and both factors displayed significant main effects. The results show the general trend that, for all encoders, more visually complex stimuli (e.g., complex Chinese characters, simple Chinese characters, and Telugu) are more difficult to process than less complex stimuli (e.g., Cree, Russian, and Hebrew), whereas L1 graphemes were encountered with varying response difficulty for the four encoders trained on corresponding L1 orthographies.

5.0 GENERAL DISCUSSION

5.1 KEY FINDINGS

The overarching goal of this research was to advance our understanding about how visual orthographic variation, at both the orthography and grapheme levels, affects learning to read within and across writing systems. The research questions included:

1. How do writing systems handle variability in visual characteristics of graphemes?
2. To what extent does the visual complexity of orthographies, encompassing both grapheme complexity and grapheme inventory size, affect visual perceptual processing in individuals both within and across writing systems?
3. Although the visual complexity of orthographies affects individuals' visual perceptual processing, to what extent is mapping principle involved in this processing?

In what follows, we link key results from each study to answer these questions.

1. How do writing systems handle variability in visual characteristics of graphemes?

After applying our measurement system to quantify the complexity of 21,821 graphemes, correlations among four dimensions revealed similarities and differences among complex patterns across writing systems. Generally, most complexity dimensions are positively correlated; however, the correlation between two dimensions in particular, number of connected points and number of disconnected components, showed different directions and magnitudes in different writing systems. The correlation between the two is positive in morphosyllabaries and alphasyllabaries, negative in abjad, and inconclusive in alphabets and syllabaries. These results suggest that graphemes with different mapping levels weigh differently on different complexity dimensions (Figure 2), echoing the results of the Kolmogorov–Smirnov test which showed that different dimensions have their unique contributions in differentiating graphemes in different writing systems.

Among 131 orthographies, grapheme inventory is strongly, positively correlated with grapheme complexity (r = .78), the averaged complexity of graphemes in a given orthography. When categorizing orthographies by mapping principle and examining the relationships between grapheme inventory and grapheme complexity (Figure 6), orthographies associated with higher mapping levels such as syllables (e.g., syllabaries) are more dispersed and less structured than those with lower mapping levels such as phonemes (e.g., alphabets and abjads). Meanwhile, orthographies associated with both phonemes and syllables (e.g., alphasyllabaries) behave similarly to both syllabaries and alphabets, with substantial overlaps in a certain range of grapheme inventory (approximately 30 to 60).

Figures 4 and 5 provide further insights into the relationship among mapping principle, grapheme inventory, and grapheme complexity. Generally, as mapping level increases, grapheme inventory increases and overall complexity of graphemes increases. Strikingly, the substantial overlaps between orthographies across mapping levels are noticeable in both number of graphemes and overall complexity of graphemes. The results suggest that, although grapheme inventory and grapheme complexity behave highly similarly under mapping principle classification, the association between grapheme inventory and grapheme complexity is much stronger than the association of each (grapheme inventory and grapheme complexity) with mapping principle. In short, although mapping principle is certainly a factor that governs orthographic complexity (i.e., grapheme inventory and grapheme complexity), no clear-cut distinction exists between mapping principle and orthographic complexity.

2. To what extent does the visual complexity of orthographies, encompassing both grapheme complexity and grapheme inventory size, affect visual perceptual processing of individuals within and across writing systems?

We examined the effect of complexities by systematically manipulating stimulus complexity and varying participant L1 orthography to compare individuals' discrimination of grapheme pairs. Results of comparisons between eight participant groups using different mapping principles (Study 2A and its parallel simulation in Study 3) and comparing Chinese groups using the same mapping principle (Study 2B) show consistent patterns: discrimination efficiency decreases as the complexity of grapheme groups increases and the complexity of participant L1 orthography decreases. We focus on the main effects of grapheme group and participant L1 orthography as these are of primary theoretical interest to our study aims.

The effect of grapheme group is particularly robust. In the behavioral experiments, the structure of best-fit models (i.e., including variability of grapheme groups across participant results in best accounts for both accuracy and RT data) confirmed that discrimination efficiency is mainly influenced by grapheme group, and to a lesser extent by participant L1 orthography. More visually complex graphemes impose perceptual processing demands. The complexity increase is so demanding that the effect of grapheme group surpasses the effect of L1 bias when responding to L1 graphemes. When controlling for the L1 effect, all individuals performed reliably worse (for both response accuracy and RT) on complex Chinese characters, simple Chinese characters, and Telugu graphemes than on these graphemes' simple relative counterparts (Study 2A and 2B). Consistently, simulation results show a clear distinction of all encoders' discrimination efficiency between visually complex (complex Chinese, simple Chinese, and Telugu) and simple grapheme groups (Cree, Russian, and Hebrew graphemes), with varying difficulty in responding to graphemes from encoders' trained "L1" orthographies.

The effect of participant L1 orthography is also significant. In the cross-writing-system behavioral experiment (Study 2A), participants with more visually complex L1 orthographies (i.e., Chinese and Japanese participants) responded more accurately and faster than did all participants with less complex L1 orthographies. This advantage of mastering more visually complex L1 orthographies holds for the within-writing-system experiment (Study 2B) with participants using the same mapping principle of the Chinese language. The Taiwan group, who had learned to read more visually complex graphemes, outperformed the China group. This superiority effect was also observed in a complex pattern discrimination task with nonlinguistic stimuli. The Taiwan group was quicker to accurately discriminate patterns of checkerboard pairs than the China group, whereas no difference was found in response accuracy. These results afford several insights. First, the visual perceptual skills developed by mastering visually complex graphemes transfer to novel visual stimuli. Because the effect of visual expertise is found in grapheme and non-grapheme stimuli, the developed skills may be orthography-independent, or domain-general. Second, the rather similar patterns found among the three groups of participating Chinese L1 readers (Lab: Taiwan and China; MTurk: a half-half mixture of Taiwan and China participants) confirmed that MTurk data collection methods has good internal validity in this setting. More importantly, the similar effects of participant L1 orthography found in the MTurk groups using different mapping principles and the Lab groups using the same mapping principle suggest that the observed effect can be attributed to visual perceptual experience, regardless of experience of mapping principle.

Although most of the complexity effects of grapheme groups and participants L1 orthographies were significant and directionally consistent with our predictions, several unexpected patterns were found. In particular, the reverse pattern showing that visually complex Russian stimuli are easier to discriminate (both response accuracy and time) than visually simpler Hebrew stimuli was observed consistently across the MTurk and the Lab studies. In a follow-up analysis that specifically compared complexity of Russian and Hebrew stimuli over individual dimensions, we verified that Russian stimuli were more visually complex than Hebrew in all dimensions (all ps $< .01$), excluding number of disconnected components ($p = .16$). Thus, we speculate that the reverse pattern between Russian and Hebrew stimuli was not likely to have been a result of the visual properties of these stimuli; such reversal was more likely due to participants' perceptual experience. Given that our international participants were able to read task instructions in English, an orthography sharing similar visual forms with Russian (e.g., А, В, Е, Н, or Т) relative to Hebrew (e.g. ו, א, ם, נ, or ל), the similarity between English and Russian graphemes may have played a role in processing Russian stimuli. However, this speculation demands further investigation. In addition to participants' multiple language exposure, the influence of additional factors such as schooling practices (Nag, 2011), instructional methods (Landerl, 2000) or even culture differences cannot be reliably extricated from the results of any cross-national comparisons; thus, our behavioral results should be interpreted with caution.

3. Although the visual complexity of orthographies affects individuals' visual perceptual processing, to what extent is mapping principle involved in this processing?

Given that we posited that mapping principle governs grapheme inventory which, in turn, drives grapheme complexity, we examined the effects of orthographic complexity with an eye toward the role of mapping principle. We observed its role in several ways. First, we selected grapheme groups and participants' L1 orthographies not only by complexity but also by writing systems. Interestingly, increasing complexity level generally, but not absolutely, corresponded to increasing mapping levels for both selected stimulus and participant groups. For example, Cree graphemes, which map to higher phonological levels (i.e., syllable), are visually less complex than Telugu graphemes, which map to lower phonological levels (i.e., phoneme). Another example from participant L1 orthography is Arabic and Hebrew; they use the same mapping principle, but they are not equivalent to each other by their L1 complexity ranking. These results suggest that it is not necessarily the case that graphemes with lower mapping levels (e.g., alphabets) are absolutely visually simpler than graphemes with higher mapping levels (e.g., syllabaries); this finding also resonates with the substantial overlaps across orthographies at different mapping levels as revealed in Figure 4 (variation of number of graphemes) and Figure 5 (variation of overall grapheme complexity).

Moreover, when attributing difference to visual experience in the PDP modeling, we witnessed that the simulation of encoder accuracy (Figure 12) showed a similar pattern to that underlying the cross-writing-system behavioral phenomena (Study 2 A; Figure 7). Note that the model's orthography-focused design affords inference of a closer causal link between visual complexity of both stimuli and readers' L1 orthography and discrimination efficiency. Thus, the effects found in the behavioral experiments can be attributed to individuals' experiences of orthographic complexity, without much consideration for mapping principles.

Furthermore, findings from purely visual orthographic learning over 131 orthographies demonstrated that more visually complex orthographies are linked to slower pace of orthographic learning as discussed in literacy literature. For the given task of mastering all graphemes in an orthography, encoders learning to master larger number of graphemes take longer than encoders learning to master smaller number of graphemes. The efforts they made (i.e., learning epochs for reaching the averaged error < 10), during this orthographic learning, are positively, strongly correlated with the complexity of these learned graphemes (r = .68). This significant correlation allows us to attribute the difficulty of mastering graphemes to their visual characteristics, again, without regarding to mapping principles.

The key finding is that visual processing efficiency is determined by complexity of the perceived stimuli themselves as well as their relationship to the viewers' L1 orthographies. Although such complexity, encompassing both visual characteristics and number of graphemes, is theoretically dictated by mapping principles, empirically, its effect on viewers' perceptual processing surpasses the boundaries of mapping principles. In other words, it is sufficient for viewers using higher mapping levels, which happen to be coded by more visually complex graphemes (e.g., Chinese readers), to exhibit perceptual performance superior to those using lower mapping levels (e.g., Chinese readers' simple counterparts); however, a higher mapping level is not necessary for stronger perceptual ability (e.g., the Taiwan group outperformed the China group). In short, experiences in overcoming orthographic complexity count the most for visual perceptual performance.

5.2 DISCUSSION

In linking our key findings with prior research, we focus on the aspects of such work most relevant to learning to read across writing systems. We begin by discussing the importance of our modeling demonstration of pure orthographic learning, followed by linking this discussion to empirical studies investigating learning to read across orthographies, concluding with a reflection on how orthographies implement multiple mapping principles to yield variety in patterns of learning to read across writing systems.

5.2.1 Visual orthographic learning across writing systems

It is worth highlighting the distributed-coding scheme that we developed for the model to simulate orthographic learning universally, restricting task performance to visual aspects of reading processing only (e.g., identification and discrimination). Each grapheme was represented by a particular pattern of activity over many units. Thus, the distributed coding scheme was sensitive to the similarities and differences among patterns representing graphemes, and was able to authentically capture the statistical properties shared across many graphemes, including those from significantly differing orthographies and of disparate complexity levels. Another important property of the model was its use of hidden units with varying receptive (and projective) field sizes; this design feature assured the model would not be biased to any particular orthography and thus allowed fair cross-orthographic comparisons of resulting performance data.

Given the well-established capacity of PDP learning models to reveal, through simulation, patterns similar to those found in the underlying behavioral phenomena (Plaut, 2005; Seidenberg, 2006), our simulation results are informative about how visual orthographic characteristics contribute to learning to read in human learners. In simulating L1 learning across 131 orthographies, the strong, positive correlation found between grapheme complexity and learning difficulty was consistent with prior research that reported that perceptual load from letters themselves can hinder individuals' recognition efficiency (Pelli et al, 2006; Vogel, Woodman, & Luck, 2001; Xu & Chun, 2006). This interpretation is also consistent with reading studies that implicate perceptual load of orthography as a source of processing difficulty (e.g., Nag, Snowling, Quinlan, & Hulme, 2014, for Kannada; Rao, Vaid, Srinivasan, & Chen, 2011, for Urdu).

Moreover, in replicating the cross-writing-system behavioral experiment, the consistencies discovered between modeling and behavioral results strengthen the account in which behavioral difference is attributed to varied complexities of stimulus and participant L1 orthography. These findings also echo those of cross-orthography studies that suggest that visual orthographic variation plays a role in learning to read (e.g., Abdelhadi, Ibrahim, & Eviatar, 2011; McBride-Chang, Zhou, et al., 2011). Concerning the effect of stimulus complexity, Abdelhadi et al. (2011) compared visual vowel detection among Arabic-Hebrew bilingual children and reported that the same individuals had higher accuracy in Hebrew, a visually simple orthography, than in Arabic, a visually complex orthography. Concerning the effect of complexity of individuals' L1 orthography, McBride-Chang, Zhou, et al. (2011) compared performance on a visuospatial task among age-matched children and found that Chinese children outperformed peers who were learning the visually less-complex Hebrew and Spanish. Although interpretation of the response differences found in these studies cannot exclude potential input from various linguistic units such as phonology or other internationally relevant issues such as methods of instruction, the account offered by our simulation underscores the importance of the role of visual complexity in learning to read.

5.2.2 Learning to read across orthographies

In the broader context of learning to read, the effects of visual orthographic variation cannot be fully established based solely on grapheme- or orthography-level observation. Learning to read and write is fundamentally a process of learning to associate orthography with phonology and semantics (Perfetti, Liu, & Tan, 2005).

Indeed, visual complexity of orthography is related to the transparency of the correspondence between graphemes and phonological units, i.e., orthographic transparency (or orthographic depth). Note we now emphasize the term orthography on both visual aspects of graphemes (G) and their corresponding transparency in relation to phonological units (P) such as phonemes or syllables. Within the alphabet family, those that are more opaque (e.g., in English, every grapheme but < r > and < v > corresponds to at least two phonemes) require learners to associate the same letter with multiple pronunciations, increasing memory load and thus slowing learning (Gough, 1996) compared to transparent orthographies, which have one-to-one GP mapping. In a series of vocabulary learning experiments, English preschoolers who learned to sight-read words were reported to learn quickly and accurately in the initial stages of learning by merely memorizing the rough visual forms of words (e.g., dissimilar letter-strings or different word lengths). However, learners became overwhelmed as learning set size increased (Gough, 1993; Gough & Juel, 1991), indicating that such rough strategies were not useful in generalizing to unfamiliar words. These findings inspired stage theories of reading development and discussions of instructional methods in reading in English (for a review, see Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001).

In orthographies with higher-level mappings and larger grapheme inventory, visual complexity of orthography and GP transparency also go hand in hand with one another. In alphasyllabaries such as Kannada, there exist conditional GP rules such as vowel suppression (e.g., inherent vowels are pronounced in speaking but are unmarked in writing), making the GP representation opaque (Nag, 2014). Nag and Snowling (2011) reported that poor visual processing skill may be a risk factor for poor reading given the visual complexity of the Kannada orthography. Similarly, in morphosyllabaries with very opaque GPCs and extremely large grapheme inventories such as Chinese, a meta-analysis summarizing 64 L1-Chinese reading studies reported that visual perceptual skill is significantly, positively correlated with Chinese character recognition ability (Yang et al., 2013). When visual processing is highlighted as an important component of learning to read, the cognitive profiles found for more visually complex orthographies are not seen for alphabetic orthographies.

The differences across orthographies between cognitive profiles relevant to learning to read suggest that visual perceptual skills can be weighted differently; skill effects are more prominent in learning orthographies with greater visual complexity and more opaque GPC (Kannada: Nag & Snowling, 2011; Chinese: Ho et al., 2012; Yang et al., 2013), whereas they are less likely to affect reading in orthographies that are visually less complex, such as alphabets (Goswami, 2004; Vellutino Steger, Moyer, Harding, & Niles, 1977). Furthermore, this effect of variety of cognitive profiles in learning to read across orthographies also resonates with prior in-depth discussions on how readers adapt themselves to the ways in which different writing systems represent languages (e.g., Perfetti & Harris, 2013).

5.2.3 Orthographic variation as implementation of multiple mapping principles

As writing systems interact with the structure of the spoken language they are trying to capture, they adapt themselves through a variety of implementations, namely orthographies. From the perspective of evolution, orthographies evolve under selective pressures to be efficient to record and easy to recognize (Changizi & Shimojo, 2005). From the perspective of semiotics, ideal visual characteristics of orthographies should be similar (e.g., have a degree of homogeneity), contrasting (e.g., be distinguishable from one to another), economical (e.g., be easy to perceive and produce), redundant, attractive, and expressive (Watt, 1983; see Treiman & Kessler, 2011 for a discussion). More careful and broader considerations of orthographies are given by the perspective of reading science: because the world's languages vary greatly along multiple non-orthogonal dimensions (e.g., phonology, morphology, and semantics), their orthographies do so as well, and, in a sense, "every language gets the writing system it deserves" (Halliday & Webster, 2003, p. 103, reprinted from Halliday, 1977; for discussion, see Frost, 2012; Perfetti & Harris, 2013; Seidenberg, 2011). Taking into account these considerations, in investigating how orthographies develop their variation to enable their parent writing systems to adapt to specific languages, we believe the key is mapping principle, the manner of correspondence between grapheme and linguistic units.

"Mapping principle of a language dictates grapheme inventory that an orthography needs" and, in turn, "grapheme inventory drives visual complexity of graphemes in any particular orthography" – these two "propositions" are the central assumptions of this study.

Indeed, our data support these two logic chains. The scheme of grapheme complexity quantification encompassing 131 orthographies demonstrates that grapheme complexity is strongly, positively associated with grapheme inventory size, which, in turn, is positively associated with mapping level. Interestingly, we found no clear-cut distinction between mapping levels of orthographic complexity (for grapheme inventory, see Figure 4; for grapheme complexity, see Figure 5). The noted substantial overlaps (as well as, in some instances, notable lack of overlap) of orthographies across writing system categories suggest that some orthographies (e.g., outliers in alphabets, alphasyllabaries, and syllabaries) may employ greater numbers of graphemes, or more visual information within graphemes, than do other orthographies in the same writing system category.

An alternative interpretation for the overlaps of orthographies is that some orthographies implement multiple mapping principles. For instance, all orthographies that map their graphemes to phonemes (i.e., alphasyllabaries, alphabets, and abjads) generally implement the alphabetic principle (Gelb, 1963), whereas, specifically, alphasyllabaries implement both alphabetic and syllabic principles (Nag, 2011), and abjads implement both alphabetic and morphemic principles (Frost, 2012). Using the same rationale of multiple mappings, morphosyllabaric orthographies can be seen as implementing both morphemic and syllabic principles. This multiple-mapping-principle interpretation is in line with reading research that discussed how writing systems vary along multiple dimensions (e.g., Frost, 2012; Hirshorn & Fiez, 2014; Perfetti & Harris, 2013; Seidenberg, 2011) and writing research suggesting that further classifications of writing systems could be discovered or be developed (Daniels & Bright, 1996).

Future attempts to establish understanding of how orthographies implement multiple mapping principles in different writing systems must be done in a broader linguistic environment, because studies focusing on learning to read cannot treat orthography as an isolated domain. Although the current research focuses on visual characteristics of orthographies, it does so to gain more attention from the reading field, to test the hypothesis that visual complexity affects learning to read, and to generalize the results to various writing systems. By demonstrating the effect of visual orthographic variation in highlighting how complexity can play a role in learning to read across orthographies, we hope that visual complexity will continue to garner more investigative attention. At the same time, by acknowledging the theoretical role of mapping principles in writing system implementation, we appreciate how orthographies develop variation in enabling writing to adapt to spoken language.

5.3 CONCLUSION

In closing, we have attempted to examine the extent to which the visual characteristics of orthographies affect learning to read across writing systems. We did this particularly by revealing visual orthographic variation across 131 orthographies, testing how orthographic complexity affects visual perceptual processing in individuals within and across writing systems, and by implementing a PDP model that represents orthographic knowledge in a distributed fashion and using this model to simulate individual task performance, comparing the results of such performance along dimensions relevant to our research aims.

The broader contribution of this research is threefold. Theoretically, it shines a light on the visual perceptual processes that are important but often ignored in reading research. Methodologically, we used the PDP framework to demonstrate the value of a distributed-coding scheme with the ability to accommodate graphemes from any orthography, encouraging comparative examination of orthographic learning across writing systems. Practically, our visual orthography measure, as demonstrated when applied over a wide range of orthographies, provides a means of comparison of grapheme complexity between any two graphemes from any and every orthography. Importantly, this surpasses the issues of limitation that hinder the generalizability of other methods of comparison, giving those in the field greater opportunities to examine universal reading and writing.

APPENDIX A

STUDY 1 CHARACTERISTICS OF 131 ORTHOGRAPHIES

Table 18. Orthographies (by alphabetic order) in five writing systems with complexity values on different dimensions: Grapheme inventory (GI), perimetric complexity (PC), number of disconnected components (DC), number of connected points (CP), number of simple features (SF), and overall complexity (in standardized composite scores)

Orthographies by writing system	Font for creating grapheme images	GI	PC mean	DC mean	CP mean	SF mean	Overall complexity
Alphabet							
Albanian (Elbasan)	Arial	40	7.73	1.08	1.45	2.53	-0.54
Albanian (Todhri)	Arial	53	7.18	1.13	2.08	2.92	-0.39
Armenian (Eastern)	Arial	38	7.43	1.03	1.55	2.63	-0.55
Asomtavruli	Arial	38	7.53	1.00	2.00	2.97	-0.43
Avestan	Ahuramazda	54	9.83	1.00	2.11	3.52	-0.23
Bassa	Arial	30	7.66	1.02	1.43	2.68	-0.55
Belarusian	Arial	32	7.35	1.17	1.83	2.74	-0.43
Bosnian	Arial	30	7.59	1.37	1.52	2.63	-0.40
Bulgaria	Arial	30	7.40	1.03	2.05	2.85	-0.44
Celtiberian	Arial	28	6.59	1.11	2.61	3.39	-0.28
Cyrillic (Abkhaz)	Arial	56	9.41	1.33	2.68	3.68	0.01
Danish	Arial	29	7.25	1.05	1.62	2.38	-0.57
Deseret	Code 2000	38	6.65	1.00	1.71	2.61	-0.58
Dutch	Arial	26	6.85	1.04	1.44	2.25	-0.64

Orthographies by writing system	Font for creating grapheme images	GI	PC mean	DC mean	CP mean	SF mean	Overall complexity
English	Arial	26	6.85	1.04	1.44	2.25	-0.64
Enochian	Enochian	22	7.41	1.14	1.59	2.68	-0.49
Finnish	Arial	28	7.00	1.20	1.41	2.34	-0.56
Fraser	Arial	40	7.82	1.00	1.68	2.43	-0.55
French	Arial	26	6.85	1.04	1.44	2.23	-0.65
Glagolitic	Arial	42	10.13	1.02	4.81	5.07	0.44
Gothic (Wulfila)	Alphabetum Unicode	25	7.11	1.08	1.16	2.36	-0.64
Greek	Arial	24	7.09	1.06	1.43	2.27	-0.62
Greman	Arial	26	6.85	1.04	1.44	2.25	-0.64
Hungarian Runes	Arial	46	9.09	1.00	2.85	3.70	-0.12
Icelandic	Arial	32	7.11	1.28	1.47	2.45	-0.49
Italian	Arial	21	6.74	1.02	1.45	2.17	-0.66
Kazakh	Arial	42	7.31	1.11	1.94	2.85	-0.43
Korean (Hangeul)	MS Mincho	40	14.71	1.38	2.15	3.40	0.15
Kyrgyz	Arial	36	7.48	1.11	2.04	2.89	-0.40
Latin (ancient)	Alphabetum Unicode	21	6.12	1.00	1.86	2.71	-0.56
Latin (modern)	Arial	41	8.41	1.49	1.61	2.71	-0.28
Macedonian	Arial	31	7.27	1.08	1.89	2.81	-0.45
Marsiliana	Arial	26	9.97	1.00	2.27	2.88	-0.29
Mkhedruli	BPG Glaho	38	7.80	1.00	1.32	2.39	-0.61
Mongolian	Arial	35	7.49	1.11	2.01	2.87	-0.40
Montenegrin	Arial	33	7.65	1.39	1.47	2.64	-0.39
N’Ko	JG Nko	27	5.50	1.00	2.11	2.96	-0.52
Norwegian	Arial	29	7.25	1.05	1.62	2.38	-0.57
Nuskhuri	BPG Nino Khutsuri U	38	7.12	1.00	3.97	5.08	0.16
Old Church Slavonic	Arial	45	8.42	1.21	2.26	3.24	-0.22
Old Permic (Abur)	Arial	38	9.16	1.05	2.39	3.42	-0.21
Pahawh Hmong	Naadaa	166	11.05	1.80	2.16	4.02	0.26
Pollard Miao	Ahmao	85	7.19	1.61	1.31	2.87	-0.31
Portuguese	Arial	26	6.85	1.04	1.44	2.25	-0.64
Romanian	Arial	31	6.96	1.19	1.45	2.34	-0.56

Orthographies by writing system	Font for creating grapheme images	GI	PC mean	DC mean	CP mean	SF mean	Overall complexity
Runic (Danish Futhark)	Code 2000	16	6.55	1.00	1.88	2.75	-0.53
Runic (Elder Futhark)	Code 2000	24	7.20	1.04	2.50	3.25	-0.31
Russian	Arial	33	7.51	1.12	2.05	2.89	-0.39
Santali (OlCemet')	Arial	30	10.27	1.07	2.43	3.40	-0.15
Serbian	Arial	30	7.34	1.02	2.02	2.83	-0.45
Somali (Osmanya)	MPH 2B Damase	30	11.52	1.00	1.47	2.63	-0.38
Sorang Sompeng	Arial	24	10.55	1.00	3.13	4.25	0.07
Spanish	Arial	27	6.93	1.07	1.48	2.31	-0.61
Swedish	Arial	29	7.14	1.19	1.47	2.40	-0.54
Tajik	Arial	35	7.48	1.14	1.94	2.90	-0.40
Theban	Theban	25	10.49	1.12	3.56	4.56	0.23
Ukrainian	Arial	33	7.16	1.11	1.89	2.79	-0.45
Varang Kshiti	Arial	30	6.62	1.00	2.27	3.23	-0.40
Yupik	Arial	44	7.84	1.25	2.19	3.18	-0.25
Zhuyin Fuhao	DFKai-SB	37	10.51	1.11	2.35	3.51	-0.11
Abjad							
Ancient Berber (Vertical)	Tamalout Standard Unicode	25	9.64	2.00	2.20	4.00	0.28
Arabic	Arial	28	8.78	1.82	1.36	3.07	-0.11
Aramaic (Early)	Aramaic Early Br Rkb	22	6.31	1.00	2.32	2.91	-0.45
Hebrew	Arial	32	5.21	1.25	0.88	2.13	-0.74
MiddlePersian (Pahlavi)	Arial	22	4.95	1.00	1.64	2.82	-0.64
Nabataean	Arial	22	5.99	1.09	1.59	2.68	-0.57
Neo Tifinagh	Hapax Berbère	33	9.82	1.18	2.12	3.15	-0.20
Parthian	Arial	22	5.37	1.05	1.64	2.82	-0.60
Pashto	Arial	40	9.14	2.03	1.43	3.35	0.05
Phoenician	MPH 2B Damase	22	7.55	1.00	2.32	2.95	-0.39
Psalter	Arial	21	4.94	1.00	1.38	2.29	-0.75

Orthographies by writing system	Font for creating grapheme images	GI	PC mean	DC mean	CP mean	SF mean	Overall complexity
Sabaean Minean	Sabaen44	29	6.09	1.00	2.83	3.62	-0.28
Samaritan	Samaritan	22	9.02	1.00	3.05	3.82	-0.08
South Arabian	Arial	28	7.95	1.00	2.46	3.18	-0.31
Syriac	Estrangelo Edessa	22	5.74	1.09	1.55	2.64	-0.60
Tifinagh	MPH 2B Damase	33	10.81	1.24	1.97	3.09	-0.16
Alphasyllabary							
‘Phags-pa	BabelStone Phags-pa Book	41	9.87	1.00	4.44	5.17	0.37
Ahom	Ahom	45	11.00	1.51	2.04	3.42	0.02
Amharic	GF Zemen Unicode	282	7.47	1.03	2.74	3.50	-0.23
Balinese	JG Aksara Bali	84	23.32	1.64	2.56	4.13	0.85
Batak (KaraBatak)	Arial	32	5.19	1.41	0.72	2.09	-0.70
Bengali	Akaash Normal	57	14.60	1.21	4.26	5.51	0.71
Brahmi	Brahmi TTF	52	4.89	1.12	1.56	2.67	-0.62
Buhid (Mangyan)	Arial	48	8.03	1.46	3.29	4.60	0.22
Burmese	Myanmar1	62	13.72	1.53	2.27	3.68	0.23
Dehong	Arial	30	4.00	1.03	2.43	3.03	-0.51
Devanagari	Sanskrit 2003	62	9.41	1.03	2.98	4.27	0.01
Dives Akuru	Arial	46	10.45	1.15	1.70	3.09	-0.26
Ethiopic (Ge’ez)	Code 2000	234	7.63	1.00	2.56	3.32	-0.29
Gujarati	Shruti	64	9.23	1.28	1.47	2.81	-0.34
Gurmukhi	Anmol Uni	60	11.81	1.22	3.32	4.68	0.32
Hanuno’o (Mangyan)	Arial	48	11.05	1.48	2.52	4.13	0.19
Hindi	Sanskrit 2003	66	9.25	1.14	2.91	4.27	0.04
Inuktitut	Aboriginal Serif Regular	112	7.65	1.61	1.28	2.88	-0.29
Kannada	Tunga	50	12.55	1.42	2.40	3.84	0.17
Kharosthi	MPH 2B Damase	39	8.57	1.05	1.33	2.44	-0.54
Khmer	Khmer OS	130	10.42	1.44	6.02	7.12	1.12
Lao	Saysettha Web	78	13.40	1.63	2.90	4.71	0.51
Lepcha_Rong	JG Lepcha	77	9.13	1.06	2.71	3.74	-0.11

Orthographies by writing system	Font for creating grapheme images	GI	PC mean	DC mean	CP mean	SF mean	Overall complexity
Limbu	MPH 2B Damase	45	8.60	1.16	1.98	3.16	-0.29
Malayalam	ML-NILA01	69	14.13	1.13	1.97	3.64	0.03
Manipuri	Akaash Normal	57	11.98	1.19	3.21	4.42	0.26
Marithi	Sanskrit 2003	65	8.53	1.29	2.94	4.22	0.06
Meroitic (non-hieroglyphic)	Arial	23	6.89	1.30	2.30	3.48	-0.21
Oriya	Raghu Oriya	66	16.25	1.11	2.30	3.27	0.12
Redjang (Kaganga)	Arial	36	6.12	1.17	1.83	2.97	-0.46
Sindhi	Bahij Nassim-Regular	51	9.86	1.31	3.37	4.63	0.26
Sinhala	Potha	71	14.71	1.51	3.20	3.93	0.45
Soyombo	JG Soyombo	86	10.79	1.76	5.16	6.87	1.10
Syloti-Nagri	Arial	38	10.49	1.11	3.37	4.79	0.23
Tagalog	Tagalog Stylized	45	14.87	1.78	1.53	3.02	0.18
Tagbanwa	Arial	42	12.18	1.64	2.07	3.93	0.21
Tamil	Code 2000	47	14.58	1.15	3.19	4.68	0.40
Telugu	NATS	70	11.41	1.33	2.68	4.10	0.16
Thaana	Free Serif	49	6.20	1.71	1.63	3.41	-0.18
Thai	Angsana New	102	14.88	1.68	4.54	6.24	1.07
Tibetan	Arial	34	11.79	1.00	3.44	4.38	0.20
Syllabary							
Carrier Dene	Code 2000	195	10.27	1.22	2.93	4.14	0.10
Cherokee	Aboriginal Sans	85	7.07	1.01	1.87	2.86	-0.49
Cree (Woodland)	Aboriginal Serif	80	7.00	1.53	1.33	2.68	-0.38
Cypriot	Alphabetum Unicode	55	11.60	1.87	1.55	3.58	0.15
Japanese (Hiragana)	MS Mincho	48	23.32	1.29	2.75	4.19	0.73
Japanese (Katakana)	MS Mincho	48	16.06	1.38	1.56	2.96	0.06
Kpelle	JG Kpelle A	86	22.26	3.14	4.43	9.01	2.44
LinearB	Penuturesu	71	30.20	2.31	2.89	5.17	1.66
Ndjuka’	Arial	57	8.34	1.04	2.39	3.05	-0.31
Ojibwe	Aboriginal Serif	88	7.04	1.48	1.11	2.43	-0.47
Vai	Dukor	208	13.07	1.82	2.88	4.86	0.59

Orthographies by writing system	Font for creating grapheme images	GI	PC mean	DC mean	CP mean	SF mean	Overall complexity
Morphosyllabary							
Chinese (Simpified)	DFKai-SB	6097	29.47	4.01	9.54	10.60	4.18
Chinese(Traditional)	DFKai-SB	6097	32.47	4.55	11.64	12.50	5.15
Japanese (Kanji)	DFKai-SB	2136	28.62	3.84	9.65	10.43	4.06

Note. Source for grapheme inventory (GI) includes: Chen et al. (2011) for the simplified and traditional Chinese orthographies, Wikipedia for the Japanese Kanji orthography (http://en.wikipedia.org/wiki/Ky%C5%8Diku_kanji), and Omniglot for other 128 orthographies (http://www.omniglot.com/)

APPENDIX B

STUDY 2 EXPERIMENTAL STIMULI

Table 19. Grapheme pairs in Study 2 (Each list contains 180 “same” pairs and 180 “different” pairs; 360 pairs in total per list)

	List 1				List 2				List 3				List 4			
	Same		Different		Same		Different		Same		Different		Same		Different	
Hebrew	א	א	א	צ	א	א	ב	א	א	א	ס	א	א	א	א	ץ
	בּ	בּ	תּ	בּ	בּ	בּ	בּ	ג	בּ	בּ	בּ	ה	בּ	בּ	ר	בּ
	ב	ב	ב	ד	ב	ב	בּ	ב	ב	ב	ת	ב	ב	ב	ב	ך
	ג	ג	פּ	ג	ג	ג	ג	ן	ג	ג	ג	ז	ג	ג	ק	ג
	ד	ד	ד	ך	ד	ד	שׁ	ד	ד	ד	בּ	ד	ד	ד	ד	ח
	ה	ה	ם	ה	ה	ה	ה	פ	ה	ה	ה	שׁ	ה	ה	כּ	ה
	ו	ו	ו	נ	ו	ו	ת	ו	ו	ו	ל	ו	ו	ו	ו	א
	ז	ז	ו	ז	ז	ז	ז	י	ז	ז	ז	ג	ז	ז	ן	ז
	ח	ח	ח	מ	ח	ח	ה	ח	ח	ח	ב	ח	ח	ח	ח	ת

Hebrew	ט	ט	כּ	ט	ט	ט	ט	ס	ט	ט	ט	ע	ט	ט	ם	ט
(cont.)	י	י	י	שׁ	י	י	ז	י	י	י	ץ	י	י	י	י	נ
	כּ	כּ	כ	כּ	כּ	כּ	כּ	ץ	כּ	כּ	כּ	פ	כּ	כּ	שׁ	כּ
	כ	כ	כ	ט	כ	כ	ל	כ	כ	כ	א	כ	כ	כ	כ	מ
	ך	ך	ף	ך	ך	ך	ך	ץ	ך	ך	ך	ן	ך	ך	ד	ך
	ל	ל	ל	ר	ל	ל	נ	ל	ל	ל	שׂ	ל	ל	ל	ל	צ
	מ	מ	שׁ	מ	מ	מ	מ	ע	מ	מ	מ	כּ	מ	מ	כ	מ
	ם	ם	ם	בּ	ם	ם	מ	ם	ם	ם	ח	ם	ם	ם	ם	ט
	נ	נ	ג	נ	נ	נ	נ	ר	נ	נ	נ	י	נ	נ	פּ	נ
	ן	ן	ן	ז	ן	ן	ך	ן	ן	ן	נ	ן	ן	ן	ן	ו
	ס	ס	ק	ס	ס	ס	ס	כּ	ס	ס	ס	פּ	ס	ס	תּ	ס
	ע	ע	ע	ל	ע	ע	צ	ע	ע	ע	ר	ע	ע	ע	ע	שׂ
	פּ	פּ	פ	פּ	פּ	פּ	פּ	כ	פּ	פּ	פּ	ף	פּ	פּ	ע	פּ
	פ	פ	פ	ב	פ	פ	ו	פ	פ	פ	ט	פ	פ	פ	פ	בּ
	ף	ף	ה	ף	ף	ף	ף	פּ	ף	ף	ף	ד	ף	ף	ג	ף
	צ	צ	צ	א	צ	צ	ט	צ	צ	צ	ם	צ	צ	צ	צ	ל
	ץ	ץ	ן	ץ	ץ	ץ	ץ	תּ	ץ	ץ	ץ	ק	ץ	ץ	י	ץ
	ק	ק	ק	ח	ק	ק	ד	ק	ק	ק	צ	ק	ק	ק	ק	ה
	ר	ר	י	ר	ר	ר	ר	ף	ר	ר	ר	ו	ר	ר	ז	ר
	שׁ	שׁ	שׁ	שׂ	שׁ	שׁ	ח	שׁ	שׁ	שׁ	מ	שׁ	שׁ	שׁ	שׁ	ס
	שׂ	שׂ	ע	שׂ	שׂ	שׂ	שׂ	ק	שׂ	שׂ	שׂ	תּ	שׂ	שׂ	פ	שׂ
	תּ	תּ	תּ	ת	תּ	תּ	א	תּ	תּ	תּ	ך	תּ	תּ	תּ	תּ	ב
	ת	ת	ס	ת	ת	ת	ת	ם	ת	ת	ת	כ	ת	ת	ף	ת

Russian	А	А	А	В	а	а	н	а	А	А	Б	А	а	а	а	ж
	Б	Б	Т	Б	б	б	б	ъ	Б	Б	Б	Ф	б	б	о	б
	В	В	В	З	в	в	ю	в	В	В	Д	В	в	в	в	ш
	Г	Г	С	Г	г	г	г	р	Г	Г	Г	Ь	г	г	р	г
	Д	Д	Д	Л	д	д	ц	д	Д	Д	Н	Д	д	д	д	п
	Е	Е	Ё	Е	е	е	е	г	Е	Е	Е	И	е	е	с	е
	Ё	Ё	Ё	Ш	ё	ё	в	ё	Ё	Ё	В	Ё	ё	ё	ё	ю
	Ж	Ж	К	Ж	ж	ж	ж	щ	Ж	Ж	Ж	М	ж	ж	э	ж
	З	З	З	У	з	з	ш	з	З	З	М	З	з	з	з	х
	И	И	М	И	и	и	и	о	И	И	И	Э	и	и	щ	и
	Й	Й	Й	Ы	й	й	ч	й	Й	Й	Ю	Й	й	й	й	л
	К	К	И	К	к	к	к	м	К	К	К	Й	к	к	г	к
	Л	Л	Л	А	л	л	х	л	Л	Л	Ш	Л	л	л	л	ч
	М	М	Е	М	м	м	м	э	М	М	Ц	Щ	м	м	й	м
	Н	Н	Н	П	н	н	ы	н	Н	Н	З	Н	н	н	н	б
	О	О	Ф	О	о	о	о	и	О	О	О	С	о	о	е	о
	П	П	П	Д	п	п	л	п	П	П	У	П	п	п	п	н
	Р	Р	Ь	Р	р	р	р	е	Р	Р	Р	О	р	р	ф	р
	С	С	С	Ч	с	с	з	с	С	С	А	С	с	с	с	я
	Т	Т	Г	Т	т	т	т	с	Т	Т	Т	К	т	т	ь	т
	У	У	У	Х	у	у	б	у	У	У	П	У	у	у	у	в
	Ф	Ф	Э	Ф	ф	ф	ф	ь	Ф	Ф	Ф	Р	ф	ф	ё	ф

Russian (cont.)	Х	Х	Х	Б	х	х	я	х	Х	Х	Я	Х	х	х	х	а
	Ц	Ц	Й	Ц	ц	ц	ц	ё	Ц	Ц	Ц	Ъ	ц	ц	к	ц
	Ч	Ч	Ч	Н	ч	ч	п	ч	Ч	Ч	Х	Ч	ч	ч	ч	у
	Ш	Ш	Щ	Ш	ш	ш	ш	й	Ш	Ш	Ш	Е	ш	ш	м	ш
	Щ	Щ	Щ	Ц	щ	щ	ж	щ	Щ	Щ	Л	Щ	щ	щ	щ	ы
	Ъ	Ъ	Р	Ъ	ъ	ъ	ъ	к	Ъ	Ъ	Ъ	Г	ъ	ъ	т	ъ
	Ы	Ы	Ы	Ю	ы	ы	д	ы	Ы	Ы	Ж	Ы	ы	ы	ы	ц
	Ь	Ь	Ъ	Ь	ь	ь	ь	т	Ь	Ь	Ь	Т	ь	ь	и	ь
	Э	Э	Э	Я	э	э	у	э	Э	Э	Ч	Э	э	э	э	з
	Ю	Ю	О	Ю	ю	ю	ю	ф	Ю	Ю	Ю	Ё	ю	ю	ъ	ю
	Я	Я	Я	Ж	я	я	а	я	Я	Я	Ы	Я	я	я	я	д
Cree	ᐃ	ᐃ	ᐃ	ᐅ	ᐋ	ᐋ	ᐋ	ᐕ	ᐋ	ᐋ	ᐄ	ᐋ	ᐃ	ᐃ	ᐸ	ᐃ
	ᐏ	ᐏ	ᐏ	ᐃ	ᐚ	ᐚ	ᐚ	ᐑ	ᐚ	ᐚ	ᐆ	ᐚ	ᐏ	ᐏ	ᐓ	ᐏ
	ᐱ	ᐱ	ᐱ	ᐳ	ᐹ	ᐹ	ᐹ	ᖭ	ᐹ	ᐹ	ᑖ	ᐹ	ᐱ	ᐱ	ᐃ	ᐱ
	ᑎ	ᑎ	ᑎ	ᒍ	ᑖ	ᑖ	ᑖ	ᒎ	ᑖ	ᑖ	ᑏ	ᑖ	ᑎ	ᑎ	ᑐ	ᑎ
	ᑭ	ᑭ	ᑭ	ᓇ	ᑳ	ᑳ	ᑳ	ᑰ	ᑳ	ᑳ	ᓃ	ᑳ	ᑭ	ᑭ	ᑯ	ᑭ
	ᒥ	ᒥ	ᒥ	ᒧ	ᒫ	ᒫ	ᒫ	ᐴ	ᒫ	ᒫ	ᖩ	ᒫ	ᒥ	ᒥ	ᑎ	ᒥ
	ᓂ	ᓂ	ᓂ	ᑭ	ᓈ	ᓈ	ᓈ	ᓅ	ᓈ	ᓈ	ᑮ	ᓈ	ᓂ	ᓂ	ᒍ	ᓂ
	ᓯ	ᓯ	ᓯ	ᓱ	ᓵ	ᓵ	ᓵ	ᒨ	ᓵ	ᓵ	ᓲ	ᓵ	ᓯ	ᓯ	ᑲ	ᓯ
	ᔨ	ᔨ	ᔨ	ᔭ	ᔫ	ᔫ	ᔫ	ᓲ	ᔫ	ᔫ	ᒨ	ᔫ	ᔨ	ᔨ	ᓴ	ᔨ
	ᒋ	ᒋ	ᒋ	ᓯ	ᒑ	ᒑ	ᒑ	ᒌ	ᒑ	ᒑ	ᑰ	ᒑ	ᒋ	ᒋ	ᓱ	ᒋ
	ᖨ	ᖨ	ᖨ	ᖬ	ᖫ	ᖫ	ᖫ	ᓵ	ᖫ	ᖫ	ᔩ	ᖫ	ᖨ	ᖨ	ᔭ	ᖨ
	ᐅ	ᐅ	ᐅ	ᐸ	ᐄ	ᐄ	ᐄ	ᐋ	ᐄ	ᐄ	ᐑ	ᐄ	ᐅ	ᐅ	ᒐ	ᐅ
	ᐓ	ᐓ	ᐓ	ᐏ	ᐑ	ᐑ	ᐑ	ᐚ	ᐑ	ᐑ	ᐕ	ᐑ	ᐓ	ᐓ	ᐘ	ᐓ

Cree	ᐳ	ᐳ	ᐳ	ᐱ	ᐲ	ᐲ	ᐲ	ᖩ	ᐲ	ᐲ	ᐴ	ᐲ	ᐳ	ᐳ	ᐅ	ᐳ
(cont.)	ᑐ	ᑐ	ᑐ	ᑎ	ᑏ	ᑏ	ᑏ	ᒫ	ᑏ	ᑏ	ᒦ	ᑏ	ᑐ	ᑐ	ᓇ	ᑐ
	ᑯ	ᑯ	ᑯ	ᓂ	ᑮ	ᑮ	ᑮ	ᓃ	ᑮ	ᑮ	ᑑ	ᑮ	ᑯ	ᑯ	ᑭ	ᑯ
	ᒧ	ᒧ	ᒧ	ᒥ	ᒦ	ᒦ	ᒦ	ᐹ	ᒦ	ᒦ	ᐲ	ᒦ	ᒧ	ᒧ	ᐳ	ᒧ
	ᓄ	ᓄ	ᓄ	ᑯ	ᓃ	ᓃ	ᓃ	ᐆ	ᓃ	ᓃ	ᑰ	ᓃ	ᓄ	ᓄ	ᓂ	ᓄ
	ᓱ	ᓱ	ᓱ	ᓴ	ᓰ	ᓰ	ᓰ	ᒦ	ᓰ	ᓰ	ᒌ	ᓰ	ᓱ	ᓱ	ᓯ	ᓱ
	ᔪ	ᔪ	ᔪ	ᐊ	ᔩ	ᔩ	ᔩ	ᓰ	ᔩ	ᔩ	ᐹ	ᔩ	ᔪ	ᔪ	ᔨ	ᔪ
	ᒍ	ᒍ	ᒍ	ᒋ	ᒌ	ᒌ	ᒌ	ᑑ	ᒌ	ᒌ	ᓲ	ᒌ	ᒍ	ᒍ	ᑕ	ᒍ
	ᖪ	ᖪ	ᖪ	ᖨ	ᖩ	ᖩ	ᖩ	ᔮ	ᖩ	ᖩ	ᐚ	ᖩ	ᖪ	ᖪ	ᔭ	ᖪ
	ᐊ	ᐊ	ᐊ	ᐘ	ᐆ	ᐆ	ᐆ	ᐄ	ᐆ	ᐆ	ᐋ	ᐆ	ᐊ	ᐊ	ᒪ	ᐊ
	ᐘ	ᐘ	ᐘ	ᐔ	ᐖ	ᐖ	ᐖ	ᔫ	ᐖ	ᐖ	ᒎ	ᐖ	ᐘ	ᐘ	ᐏ	ᐘ
	ᐸ	ᐸ	ᐸ	ᒪ	ᐴ	ᐴ	ᐴ	ᐲ	ᐴ	ᐴ	ᔫ	ᐴ	ᐸ	ᐸ	ᐱ	ᐸ
	ᑕ	ᑕ	ᑕ	ᑐ	ᑑ	ᑑ	ᑑ	ᑖ	ᑑ	ᑑ	ᒫ	ᑑ	ᑕ	ᑕ	ᒧ	ᑕ
	ᑲ	ᑲ	ᑲ	ᓄ	ᑰ	ᑰ	ᑰ	ᑳ	ᑰ	ᑰ	ᓈ	ᑰ	ᑲ	ᑲ	ᒋ	ᑲ
	ᒪ	ᒪ	ᒪ	ᒐ	ᒨ	ᒨ	ᒨ	ᑏ	ᒨ	ᒨ	ᒑ	ᒨ	ᒪ	ᒪ	ᒥ	ᒪ
	ᓇ	ᓇ	ᓇ	ᑲ	ᓅ	ᓅ	ᓅ	ᓈ	ᓅ	ᓅ	ᑳ	ᓅ	ᓇ	ᓇ	ᑯ	ᓇ
	ᓴ	ᓴ	ᓴ	ᔭ	ᓲ	ᓲ	ᓲ	ᔩ	ᓲ	ᓲ	ᓰ	ᓲ	ᓴ	ᓴ	ᐊ	ᓴ
	ᔭ	ᔭ	ᔭ	ᔨ	ᔫ	ᔫ	ᔫ	ᑮ	ᔫ	ᔫ	ᖭ	ᔫ	ᔭ	ᔭ	ᖪ	ᕙ
	ᒐ	ᒐ	ᒐ	ᕗ	ᒎ	ᒎ	ᒎ	ᒑ	ᒎ	ᒎ	ᖫ	ᒎ	ᒐ	ᒐ	ᖨ	ᒐ
	ᖬ	ᖬ	ᖬ	ᑕ	ᖫ	ᖫ	ᖫ	ᖭ	ᖫ	ᖫ	ᔮ	ᖫ	ᖬ	ᖬ	ᕙ	ᖬ
	ᕗ	ᕗ	ᕗ	ᕙ	ᐣ	ᐣ	ᐣ	‡	ᐣ	ᐣ	ᕽ	ᐣ	ᕗ	ᕗ	ᐤ	ᕗ
	ᕙ	ᕙ	ᕙ	ᖬ	ᐢ	ᐢ	ᐢ	ᐦ	ᐢ	ᐢ	ᐣ	ᐢ	ᕙ	ᕙ	ᕗ	ᔭ
	ᐤ	ᐤ	ᐤ	ᑊ	ᐩ	ᐩ	ᐩ	ᐨ	ᐩ	ᐩ	‡	ᐩ	ᐤ	ᐤ	ᖬ	ᐤ
	ᑊ	ᑊ	ᑊ	ᐟ	ᐨ	ᐨ	ᐨ	ᕽ	ᐨ	ᐨ	ᐦ	ᐨ	ᑊ	ᑊ	ᐠ	ᑊ
	ᐟ	ᐟ	ᐟ	ᐠ	‡	‡	‡	ᐣ	‡	‡	ᐢ	‡	ᐟ	ᐟ	ᑦ	ᐟ

Cree (cont.)	ᐠ ᑦ	ᐠ ᑦ	ᐠ ᑦ	ᑦ ᐤ	ᐦ ᕁ	ᐦ ᕁ	ᐦ ᕁ	ᐩ ᐢ	ᐦ ᕁ	ᐦ ᕁ	ᐨ ᐩ	ᐦ ᕁ	ᐠ ᑦ	ᐠ ᑦ	ᐟ ᑊ	ᐠ ᑦ
Telugu	అ	అ	అ	అం	క	క	క	శ	క	క	డ	క	అ	అ	ఓ	అ
	ఆ	ఆ	అ	ఆ	ఖ	ఖ	ఛ	ఖ	ఖ	ఖ	ఖ	భ	ఆ	ఆ	ఆ	ఇ
	ఇ	ఇ	ఇ	ఓ	గ	గ	గ	ర	గ	గ	బ	గ	ఇ	ఇ	ఐ	ఇ
	ఈ	ఈ	ఉ	ఈ	ఘ	ఘ	ఝ	ఘ	ఘ	ఘ	ఘ	ఫ	ఈ	ఈ	ఈ	ఋ
	ఉ	ఉ	ఉ	ఊ	ఙ	ఙ	ఙ	న	ఙ	ఙ	ష	ఙ	ఉ	ఉ	అః	ఉ
	ఊ	ఊ	ఈ	ఊ	చ	చ	గ	చ	చ	చ	చ	హ	ఊ	ఊ	ఊ	ౡ
	ఋ	ఋ	ఋ	ఔ	ఛ	ఛ	ఛ	ఢ	ఛ	ఛ	చ	ఛ	ఋ	ఋ	ఊ	ఋ
	ౠ	ౠ	ఋ	ౠ	జ	జ	అ	జ	జ	జ	జ	ళ	ౠ	ౠ	ౠ	ఉ
	ఌ	ఌ	ఌ	ఐ	ఝ	ఝ	ఝ	డ	ఝ	ఝ	య	ఝ	ఌ	ఌ	ఏ	ఌ
	ౡ	ౡ	ఌ	ౡ	ఞ	ఞ	ధ	ఞ	ఞ	ఞ	ఞ	గ	ౡ	ౡ	ౡ	ఈ
	ఎ	ఎ	ఎ	ఏ	ట	ట	ట	ల	ట	ట	ద	ట	ఎ	ఎ	అం	ఎ
	ఏ	ఏ	ఇ	ఏ	ఠ	ఠ	భ	ఠ	ఠ	ఠ	ఠ	ర	ఏ	ఏ	ఏ	ఌ
	ఐ	ఐ	ఐ	ఎ	డ	డ	డ	త	డ	డ	శ	డ	ఐ	ఐ	ఒ	ఐ
	ఒ	ఒ	ఆ	ఒ	ఢ	ఢ	ఙ	ఢ	ఢ	ఢ	ఢ	ఝ	ఒ	ఒ	ఒ	అ
	ఓ	ఓ	ఓ	ఒ	ణ	ణ	ణ	బ	ణ	ణ	స	ణ	ఓ	ఓ	ఎ	ఓ

Telugu (cont.)	ఔ	ఔ	ౠ	ఔ	త	త	ళ	త	త	త	త	అ	ఔ	ఔ	ఔ	ఆ
	అం	అం	అం	అః	థ	థ	థ	ప	థ	థ	ఠ	థ	అం	అం	ఔ	అం
	అః	అః	ౡ	అః	ద	ద	ర	ద	ద	ద	ద	జ	అః	అః	అః	ౠ
	కా	కా	కా	కౌ	ధ	ధ	ధ	ధ	ధ	ధ	ఢ	ధ	కా	కా	కె	కా
	కి	కి	కీ	కి	న	న	క	న	న	న	న	వ	కి	కి	కి	కా
	కీ	కీ	కీ	కో	ప	ప	ప	చ	ప	ప	స	ప	కీ	కీ	కొ	కీ
	కు	కు	కూ	కు	ఫ	ఫ	ఘ	ఫ	ఫ	ఫ	ఫ	ఖ	కు	కు	కు	కం
	కూ	కూ	కూ	కౄ	బ	బ	బ	ట	బ	బ	ల	బ	కూ	కూ	కు	కూ
	కృ	కృ	కం	కృ	భ	భ	ఖ	భ	భ	భ	భ	ఛ	కృ	కృ	కృ	కః
	కౄ	కౄ	కౄ	కృ	మ	మ	మ	ష	మ	మ	ట	మ	కౄ	కౄ	కై	కౄ
	కౢ	కౢ	కౣ	కౢ	య	య	హ	య	య	య	య	ఘ	కౢ	కౢ	కౢ	కూ
	కౣ	కౣ	కౣ	కొ	ర	ర	ర	ద	ర	ర	ప	ర	కౣ	కౣ	కౄ	కౣ
	కె	కె	కి	కె	ల	ల	వ	ల	ల	ల	ల	ధ	కె	కె	కె	కీ
	కే	కే	కే	కె	వ	వ	వ	య	వ	వ	ణ	వ	కే	కే	కొ	కే
	కై	కై	కౢ	కై	శ	శ	జ	శ	శ	శ	శ	ఝ	కై	కై	కై	కౣ
	కొ	కొ	కొ	కే	ష	ష	ష	స	ష	ష	మ	ష	కొ	కొ	కో	కొ

Telgugu (cont.)	కో	కో	కా	కో	స	స	ఫ	స	స	స	స	క	కో	కో	కో	కి
	కా	కా	కా	కై	హ	హ	హ	మ	హ	హ	థ	హ	కా	కా	కే	కా
	కం	కం	కః	కం	ళ	ళ	ఝ	ళ	ళ	ళ	ళ	జ	కం	కం	కం	క్ష
	కః	కః	కః	కు	ఱ	ఱ	ఱ	ణ	ఱ	ఱ	త	ఱ	కః	కః	కృ	కః
Simple Chinese	人	人	七	乜	日	日	夫	矢	孓	孓	了	十	千	千	大	夫
	夕	夕	本	禾	入	入	矢	天	丈	丈	九	几	刀	刀	丁	十
	了	了	口	日	七	七	中	史	口	口	于	干	天	天	人	入
	大	大	刀	力	力	力	夭	大	丁	丁	己	已	子	子	土	工
	十	十	尸	尺	孑	孑	廾	卅	土	土	木	本	中	中	巳	巴
	工	工	子	孑	九	九	丁	了	廾	廾	壬	士	尸	尸	千	牛
	天	天	目	早	木	木	孓	丈	乜	乜	夭	大	士	士	夫	失
	史	史	失	矢	王	王	曰	白	几	几	矢	夭	曰	曰	夕	歹
	止	止	目	自	由	由	于	手	白	白	中	申	古	古	曰	田
	卅	卅	甲	曰	歹	歹	九	子	牛	牛	王	工	于	于	日	目
	尺	尺	土	王	失	失	舌	古	目	目	日	有	矢	矢	斤	斥
	夫	夫	丰	井	未	未	田	由	皿	皿	廿	世	禾	禾	止	丘
	吏	吏	且	日	正	正	甲	中	廿	廿	另	吊	生	生	友	反
	己	己	白	百	已	已	甘	耳	本	本	禾	木	丰	丰	未	朱
	斤	斤	王	生	甘	甘	早	卓	巳	巳	止	正	壬	壬	吏	串
	丘	丘	占	舌	巴	巴	史	吏	反	反	生	圭	手	手	皿	血
	另	另	杏	杳	友	友	李	季	占	占	甲	申	世	世	申	曰
	早	早	用	甬	申	申	叓	申	斥	斥	申	里	甲	甲	由	曲
	田	田	里	甲	冉	冉	冉	再	用	用	果	更	且	且	甲	更
	舌	舌	吏	更	弗	弗	里	重	叓	叓	弗	夷	里	里	且	早

Complex	漫	漫	僮	撞	僮	僮	僧	憎	僧	僧	懊	燠	懊	懊	漫	慢
Chinese	塘	塘	憾	撼	憾	憾	揪	楸	揪	揪	搞	塙	搞	搞	塘	瑭
	晴	晴	濡	燸	濡	濡	檄	激	檄	檄	濯	曜	濯	濯	晴	睛
	喋	喋	瞭	瞭	瞭	瞭	暗	喑	暗	暗	蟠	皤	蟠	蟠	喋	牒
	褫	褫	徼	檄	徼	徼	梢	稍	梢	梢	睦	眭	睦	睦	褫	褫
	填	填	褐	裼	褐	褐	情	惜	情	情	植	楣	植	植	填	埧
	梧	梧	漕	潭	漕	漕	棒	椿	棒	棒	棹	楮	棹	棹	梧	楮
	跌	跌	酩	酪	酩	酩	嗑	磕	嗑	嗑	蜊	蜥	蜊	蜊	跌	跦
	擒	擒	跛	跂	跛	跛	跟	踉	跟	跟	鞘	鞠	鞘	鞘	擒	摛
	部	部	梡	棕	梡	梡	媒	媟	媒	媒	勤	鄞	勤	勤	部	敲
	置	置	黑	熏	黑	黑	覃	罩	覃	覃	署	覃	署	署	覃	置
	寥	寥	察	蔡	察	察	熏	墨	熏	熏	窘	宭	窘	窘	寥	廖
	惠	惠	愚	蕙	愚	愚	窠	裹	窠	窠	寞	暮	寞	寞	惠	愚
	眷	眷	紊	索	紊	紊	菟	冤	菟	菟	奢	替	奢	奢	眷	智
	耆	耆	娶	最	娶	娶	煦	照	煦	煦	梨	焚	梨	梨	耆	屠
	管	管	幕	暮	幕	幕	篡	纂	篡	篡	笙	筐	笙	笙	管	篙
	菅	菅	寒	塞	寒	寒	莠	萎	莠	莠	莫	募	莫	莫	菅	萱
	壁	壁	斐	裴	斐	斐	寅	富	寅	寅	髡	髦	髡	髡	壁	擘
	罩	罩	票	粟	票	票	亳	毫	亳	亳	奠	尊	奠	奠	罩	署
	恙	恙	喜	嘉	喜	喜	堂	掌	堂	堂	竟	章	竟	竟	恙	羔

APPENDIX C

STUDY 2 LANGUAGE HISTORY QUESTIONNAIRE

1. What is your first language (i.e., language first spoken)? _______________

2. If you have more than one first language, please specify which language you consider to be your second language? (write n/a if you have only one first language) _______________

3. What languages were spoken in your home when you were a child and by whom? (e.g., English, father; Chinese, grandmother) ____

4. List below, from most fluent to least fluent, all of the languages (including your first language) to which you have been exposed. For example, you will write your first language in the first column, your second language in the second column as so on. Write in the box the age at which you first learned each language in terms of speaking, reading, and writing, the number of years you have spent learning each language, and the percentage of your exposure to each language every day.

	Language	Age (in years) first learned the language			Number of years spent learning (cumulative)	Everyday exposure (this should sum to 100% across the columns)
		Speaking	Reading	Writing		
1						
2						
3						
4						
5						

5. Indicate the age (in years) at which you started using each of the languages you have learned in the following contexts.

	Language	At home	At school	At work	At informal settings (e.g., friends or nannies)	After immigrating to a country where the language is spoken	Through software (e.g., Rosetta Stone)	Other (specify): ____________
1								
2								
3								
4								
5								

6. Please rate your current reading, writing, listening and speaking abilities for all languages you know (including your first language) according to the following scale:

----- 1 Very poor ----- 2 Poor ----- 3 Fair ----- 4 Functional ----- 5 Good ----- 6 Very good ----- 7 Native-like -----

	Language	Reading	Writing	Listening	Speaking
1		○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7
2		○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7
3		○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7
4		○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7
5		○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○7

7. Please rate your general language learning skills. In other words, how good do you feel you are at learning new languages, relative to your friends or other people you know?

Very poor	Poor	Fair	Neutral	Good	Very good	Excellent
○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7

8. Please estimate how many hours per day you spend engaged in the following activities in each of the languages you know.

   If you are not currently engaged in an activity using that language, write "0".

	Language	Watching television	Surfing the internet	Reading for fun	Reading for school/work	Writing email to friends	Writing for school/work	Other (specify): ____________
1								
2								
3								
4								
5								

APPENDIX D

STUDY 2 DEMOGRAPHIC BACKGROUND SURVEY

1. Age (in years): ______
2. Gender: ○ Male ○ Female
3. Handedness: ○ Left ○ Right
4. Country of origin (country in which you were born): ___________
5. Country of residence: ______________
6. If your country of origin and country of residence are different, how long have you been in the country of your current residence?
   ________(months)

7. What is the highest level of education you have completed? [choose a level of education]

(1) No degree (2) Secondary school (e.g., High School or GED) (3) Associate degree or progress toward Bachelor's

(4) Backelor's degree (e.g., BA, BS) (5) Master's degree (e.g., MA, MS, MBA) (6) Doctorate (e.g., PhD, MD, JD)

8. Do you have any known visual problems (either corrected or uncorrected)?

○ No ○ Yes (please specify) ___________________________

9. Do you have any known hearing problems (either corrected or uncorrected)?

○ No ○ Yes (please specify) ___________________________

10. Right now, are you doing anything else other than reading this page?

○ No, just this ○ Other activity (please specify) ___________________________

11. How loud are your surroundings right now?

○ Silent ○ Occasional noise ○ Frequent noise ○ Very loud

12. Do you have any comments about the study?

REFERENCES

Abdelhadi, S., Ibrahim, R., & Eviatar, Z. (2011). Perceptual load in the reading of Arabic: Effects of orthographic visual complexity on detection. *Writing Systems Research, 3*, 117-127.

Ager, S. 1998 Omniglot: a guide to writing systems. See http://www.omniglot.com.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science, 15*(2), 106–111.

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

Baker, C., Liu, J., Wald, L. L., Kwong, K. K., Benner, T. & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proc. Natl. Acad. Sci. U.S.A. 104*, 9087–9092

Barchard, K. A., & Williams, J. (2008). Practical advice for conducting ethical online experiments and questionnaires for United States psychologists. *Behavior Research Methods, 40*, 1111–1128.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.99375-39. Retrieved from http://CRAN.R-project.org/package-lme4.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing 27*(1), 101-118.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Brand, A., & Bradley, M. T. (2012). Assessing the effects of technical variance on the statistical outcomes of web experiments measuring response times. *Social Science Computer Review, 30*, 350–357.

Changizi, M. A. & Shimojo, S. (2005) Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society of London B: Biological Sciences, 272*, 267–275.

Changizi, M. A, Zhang, Q., Ye, H., & Shimojo, S. (2006). The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *The American naturalist*, *167*(5), E117–139.

Checkosky, S. F., & Whitlock, D. (1973). Effects of pattern goodness on recognition time in a memory search task. *Journal of experimental psychology*, *100*(2), 341–348.

Chen, Y. P., Allport, D. A., & Marshall. J. C. (1996). What are the functional orthographic units in Chinese word recognition: The stroke or the Stroke pattern? *The Quarterly Journal of Experimental Psychology, 49*, 1024-1043.

Chen, H. C., Chang, L. Y., Chiou, Y. S., Sung, Y. T., & Chang, K. E. (2011). Construction of Chinese orthographic database for Chinese character instruction. *Bulletin of Educational Psychology*, *43*, 269–290.

Chein, J. & Morrison, A. (2010). Expanding the mind’s workspace: Training and transfer effects with a complex working memory span task. *Psychomic Bulletin & Review, 17*, 193-199.

Cho, J. R. & Chen, H. C. (1999). Orthographic and phonological activation in the semantic processing of Korean Hanja and Hangul. *Language and Cognitive Processes, 14*, 481-502.

Cook, V. and Bassetti, B. (2005) An introduction to researching Second Language Writing Systems. In V. Cook and B. Bassetti (eds), *Second Language Writing Systems* (pp. 1-67). Clevedon, UK: Multilingual Matters.

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117*, 713-758.

Daniels, P. T., & Bright, W. (1996). *The World’s Writing Systems.* New York: Oxford University Press.

Dehaene, S., Pegado, F., Braga,L. W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science, 330*, 1359–1364.

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems.* Honolulu: University of Hawaii.

Demetriou, A., Kui, Z. X., Spandoudis, G., Christou, C., Kyriakides, L., & Platsidou, M. (2005). The architecture, dynamics, and development of mental processing: Greek, Chinese, or universal? *Intelligence,* 33, 109–141.

Duñabeitia, J.A., Orihuela, K., & Carreiras, M., (2014) Orthographic coding in illiterates and literates. *Psychological Science, 25*, 1275-1280.

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Daal, V. H. P., Polyzoe, N.,Tsipa, M., Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly, 39*(4), 438–468. doi:10.1598/RRQ.39.4.5

Fiset, D., Blais, C., Ethier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008). Features for identification of uppercase and lowercase letters. *Psychological science*, *19*(11), 1161–1168.

Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language, 62*, 1–18. http://dx.doi.org/10.1016/j.jml.2009.09.004

Frost, R. (2012). Towards a universal model of reading. *The Behavioral and brain sciences, 35*(5), 263–279. doi:10.1017/S0140525X11001841

Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Current Opinions in Neurobiology, 11*, 219-224.

Gelb, I. J. (1963). *A Study of Writing.* Chicago: University of Chicago Press.

Gibson, E.J. (1969). *Principles of perceptual learning and development.* New York: Meredith.

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115*, 577-601.

Goswami, U. (2004). Neuroscience and education. *British Journal of Educational Psychology, 74*, 1-14.

Gough, P. B. (1993). Context, form, and interaction. In K. Rayner (ed.), *Eye Movement in Reading: Perceptual and Language Processes.* New York: Academic Press.

Gough, P. B. (1996). How children learn to read and why they fail. *Annals of Dyslexia, 46*, 3-20.

Gough, P. B., & Juel, C. (1991). The first stage of word recognition. In L. Rieben and C. A. Perfetti (eds.), *Learning to read.* Hilsdale, NJ: Lawrence Erlbaum Associates.

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: from pixels to pandemonium. *Trends in cognitive sciences*, *12*(10), 381–387.

Gupta, A. (2004). Reading difficulties of Hindi-speaking children with developmental dyslexia. *Reading and Writing, 17*, 79-99.

Halliday, M. A. K. (1977). *Aims and perspectives in linguistics. Occasional papers number 1.* Applied Linguistics Association of Australia, Centre for Language Learning and Teaching, University of Southern Queensland, Australia.

Halliday, M. A. K., & Webster, J. (2003). Ideas about language. In J. Webster (Ed.), *On language and linguistics: Collected works of M. A. K. Halliday* (vol. 3, pp. 92–115). Continuum. (Essay originally published in 1977 in Aims and perspectives in linguistics. Occasional Papers, no. 1, pp. 32–55. Applied Linguistics Association of Australia)

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*, 491-528.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313*, 504-507.

Hirshorn, E., & Fiez, J. (2014). Using artificial orthographies for studying cross-linguistic difference in the cognitive and neural profiles of reading. *Journal of Neurolinguistics, 31*, 69-85.

Ho, C., Wong, Y.-K., Yeung, P.-S., Chan, D., Chung, K., Lo, S.-C., & Luan, H. (2012). The core components of reading instruction in Chinese. *Reading and Writing, 25*(4), 857–886.

Hocking, R. R. (1976). The Analysis and Selection of Variables in Linear Regression. *Biometrics, 32*, 1-49.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14*, 3990425.

Huang, H. S., & Hanley, J. R. (1995). Phonological awareness and visual skills in learning to read Chinese and English. *Cognition*, *54*, 73-98.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformations or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007.

Jiang, Y. V., Shim, W. M., & Makovski, T. (2008). Visual working memory for line orientations and face identities. *Perception, & Psychophysics*, *70*(8), 1581–1591.

Katz, L., & Frost, R. (1992). Reading in different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), Orthography, phonology, morphology, and meaning (pp. 67-84). Amsterdam: Elsevier North Holland Press.

Landerl, K. (2000). Influences of orthographic consistency and reading instruction on the development of nonword reading skills. *European Journal of Psychology of Education, 3*, 239–257.

Lanthier, S. N., Risko, E. F., Stolz, J. a, & Besner, D. (2009). Not all visual features are created equal: Early processing in letter and word recognition. *Psychonomic bulletin & review*, *16*(1), 67–73.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd Edition. SpringerVerlag, New York. MR0852406

Liu, T., Chen, W., Liu, C. H., & Fu, X. (2012). Benefits and costs of uniqueness in multiple object tracking: the role of object complexity. *Vision research*, *66*, 31–38.

Mann, V. (1985). A cross-linguistic perspective on the relation between temporary memory skills and early reading ability. *Remedial and Special Education, 6,* 37–42.

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

Morrison, A., & Chein, J. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin* & Review, *18*, 46-60.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30,* 7–22.

Nag, S (2011). The akshara languages: What do they tell us about children's literacy learning? In R. Mishra and N. Srinivasan, (Eds.), *Language-Cognition: State of the Art* (pp. 291 – 310). Lincom Publishers, Germany

Nag, S. (2014). Alphabetism and the science of reading: from the perspective of the akshara languages. Frontiers in Psychology, 5:866, 1-3. doi: 10.3389/fpsyg.2014.00866

Nag, S., Caravolas, M., & Snowling, M. J. (2010). Beyond alphabetic processes: literacy and its acquisition in the alphasyllabic languages. *Reading and Writing*, *24*(6), 615–622.

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

Nag, S., & Snowling, M. J. (2011). Cognitive profiles of poor readers of Kannada. *Reading and Writing, 24*(6), 657–676.

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an alphasyllabary: The case of Kannada. *Writing Systems Research, 2*(1), 41–52.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Norris, D., Kinoshita, S., & van Casteren, M. (2010). A stimulus sampling theory of letter identity and order. *Journal of Memory and Language, 62*, 254-271.

Perfetti, C, A., & Harris, L. N. (2013). Reading universals are modulated by language and writing system. *Language Learning and Development, 9*, 296-316. doi:10.1080/15475441.2013.813828

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

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review, 12*, 43-59.

Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., & Tan, L.-H. (2007). Reading in two writing systems: Accommodation and assimilation of the brain's reading network. *Bilingualism: Language and Cognition, 10*(2), 131-146.

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a "universal" phonological principle. In R., Frost. & L., Katz. (Eds.), *Orthography, phonology, morphology and meaning* (pp. 227-248). Amsterdam: North-Holland.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46*(28), 4646–4674. doi:10.1016/j.visres.2006.04.023

Plaut, D. C. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science, 23*, 543-568.

Plaut, D.C. (2005). Connectionist approaches to reading. In M.J. Snowling and C. Hulme (Eds.), *The science of reading: A handbook* (pp. 24-38). Blackwell: Oxford.

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.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533-536.

Rao, C., Vaid, J., Srinivasan, N., & Chen, H. C. (2011). Orthographic characteristics speed Hindi word naming but slow Urdu naming: Evidence from Hindi/Urdu biliterates. *Reading and Writing: An Interdisciplinary Journal, 24*, 679-695.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*, 31-74.

Scheerer, E. (1986). Orthographic and lexical access. In G. August (Ed.), *New trends in graphemics and orthography* (pp. 262-286). Berlin:De Gruyter.

Seidenberg, M. S. (2006). Connectionist Models of Reading In Gareth Gaskell. (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 235-25). Oxford University Press.

Seidenberg, M. S. (2011). Reading in different writing systems: One architecture, multiple solutions. In P. McCardle., J. Ren., & O. Tzeng. (Eds.), *Across languages: Orthography and the Gene-Brain-Behavior Link.* Paul Brooke Publishing.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(2), 143–174.

Shannon, C.E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Shen, H. H., & Ke, C. (2007). Radical awareness and word acquisition among nonnative learners of Chinese. *Modern Language Journal, 91,* 97–111.

Shu, H., Chen, X., Anderson, R., Wu, N., & Xuan, Y. (2003). Properties of school Chinese: Implications for learning to read. *Child Development, 74*, 27-47.

Simcox, T., & Fiez, J. (2013). Collecting response times using Amazon Mechanical Turk and Adobe Flash. *Behavior research methods*. doi:10.3758/s13428-013-0345-y

Simpson, I. C., Mousikou, P., Montoya, J. M., Defior, S. (2013). A letter visual-similarity matrix for Latin-based alphabets. *Behavior Research Methods, 45*, 431-439.

Sprouse, J. (2011). A validation of Amazon mechanical turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods, 43*, 155–167.

Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association, 69*(347), 730–737.

Taft, M. (2006). Processing of characters by native Chinese readers. In P. Li., E. Bates., L. H. Tan., & O. J. L. Tzeng (Eds.), *Handbook of East Asian Psycholinguisitcs* (Vol. 1, pp. 237-249). Cambridge: Cambridge University Press.

Tamaoka, K., & Kiyama, S. (2013). The effects of visual complexity for Japanese kanji processing with high and low frequencies. *Reading & Writing, 26,* 205-223. DOI 10.1007/s11145-012-9363-x

Tamaoka, K., Kirsner, K., Yanase, Y., Miyaoka, Y., & Kawakami, M. (2002). A Web-accessible database of characteristics of the 1,945 basic Japanese kanji. *Behavior Research Methods, Instruments & Computers, 34*, 260-275.

Tan, L. H., Spinks, J. A., Eden, G. F., Perfetti, C. A. & Siok, W. T. (2005). Reading depends on writing, in Chinese. *Proceedings of the National Academy of Sciences, USA, 102*, 8781–8785.

Tokowicz, N., Michael, E., & Kroll, J. F. (2004). The roles of study abroad experience and working memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition, 7*, 255-272.

Treiman, R., & Kessler, B. (2005). Writing systems and spelling development. In M. Snowling, & C. Hulme (Eds.). *Science of reading: A handbook.* Oxford, England: Blackwell.

Treiman, R., & Kessler, B. (2011). Similarities among the shapes of writing and their effects on learning. *Written Language and Literacy, 14*(1), 39–57.

Vellutino, F. R., Steger, J. A., Moyer, B. M., Harding, S. C., & Niles, C. J. (1977). Has the perceptual deficit hypothesis led us astray? *Journal of Learning Disabilities, 10*, 54-64.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance, 27*, 92-114.

Watson, A. (2012). Perimetric complexity of binary digital images. *The Mathematica Journal, 14*, 1–41. doi:10.3888/tmj.14-5

White, T. G., Graves, M. F., & Slater, W. H. (1990). Growth of reading vocabulary in diverse elementary schools: Decoding and word meaning. *Journal of Educational Psychology, 82* (2), 281-290.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin and Review, 8*, 221-243.

Winskel, H. (2010). Spelling development in Thai children. *Journal of Cognitive Science, 11*, 7-35.

Woollams, A., Lambon Ralph, M. A., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review, 114*, 316-339. doi:10.1037/0033-295X.114.2.316

Wu, N., Zhou, X., & Shu, H. (1999). Sublexical processing in reading Chinese: A development study. *Language & Cognitive Processes, 14*, 503-524.

Wydell, T. N (2012). Cross-cultural/linguistic differences in the prevalence of developmental dyslexia and the hypothesis of granularity and transparency. In T. N. Wydel, & L. Fern-Pollak (Eds.), *Dyslexia e A comprehensive and international approach* (pp. 1e14). Rijeka, Croatia: InTech.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440*, 91-95.

Yang, L. Y., Guo, J. P., Richman, L. C., Schmidt, F. L., Gerken, K. C., & Ding, Y. (2013). Visual skills and Chinese reading acquisition: A meta-analysis of correlation evidence. *Educational Psychology Review*, *25*, 115–143

Yang, J. F., McCandliss, B. D., Shu, H., & Zevin, J. D. (2009). Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. *Journal of Memory and Language, 61*, 238-257.

Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language, 54*, 145-160.

Ziegler, J. C. & Goswami, U. (2005) Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3–29.