Legibility Variations of Chinese Characters and Implications for Visual Acuity Measurement in Chinese Reading Population

Jun-Yun Zhang,[1,2] *Ting Zhang*,[1,2] *Feng Xue*,[3] *Lei Liu*,[*,4] *and Cong Yu*[*,1,2]

Purpose. Written Chinese contains an enormous number of characters with a wide range of spatial complexities. Consequently, the legibility of Chinese characters is expected to vary significantly, and this variability offers the challenge of deriving a simple visual function measurement for the Chinese reading population. The purpose of this study was to suggest a solution to the challenge through psychophysical studies of Chinese character legibility.

Methods. To illustrate legibility variations in Chinese characters, visual acuities for six groups of Chinese characters from low to high spatial complexities and one group of Sloan letters were determined in six normal-sighted Chinese observers. The relationship between legibility and optical defocus were then determined for the Landolt C, the Snellen E, and three groups of Chinese characters representing low, medium, and high spatial complexities in 26 normal-sighted Chinese readers.

Results. The acuity size of Chinese characters increased steadily with stimulus complexity, though at a slower rate than would be expected if visual acuity were based on the finest details of the stimuli. The acuity size versus optical defocus functions of three Chinese character groups and the Snellen E had similar slopes and differed only by a vertical shift, depending on the optotype spatial complexity. The function of the Landolt C was significantly steeper.

Conclusions. The findings indicate that visual acuity assessment in Chinese readers is complicated by the spatial complexity of Chinese characters, but the fact that the Snellen E, which is the current national standard of acuity measurement in China, and Chinese characters showed similar dependence on optical defocus may indicate a potentially valid way to infer functional vision in Chinese readers with Snellen E acuity. (*Invest Ophthalmol Vis Sci.* 2007;48:2383–2390) DOI: 10.1167/iovs.06-1195

Visual acuity is the most frequently used vision test in both clinical practice and basic research. Though it is not always explicitly stated, the fact is that visual acuity is measured for two related but different purposes. The first is to determine the refractive error of the eye as well as the lens correction that makes the stimulus optically conjugate to the retina.[1] For this purpose, visual acuity is usually defined as a measure of spatial resolution of the visual system and is described by the finest details (minimum angle of resolution, MAR) that can be resolved. International[2] and U.S.[1] standards recommend the eight- and four-orientation Landolt C as the primary visual acuity test optotype. Both standards stipulate that the gap of the C target is the detail to be resolved and that the width of the gap, which is one fifth of the target height, is the measure of visual acuity. In China, the standard optotype is the Snellen E,[3] whose stroke width is one fifth of the optotype height. For the purpose of refraction, visual acuity measurement should use a universal standard and should be free from the observers' visual environment and cultural background.

However, there are situations in which the optical conjugation between image plane and the retina is not the main concern.[4] Even when refractive errors are optimally corrected, visual performance can be impaired by diseases and trauma of the eye or of the visual neural pathway or by demanding operating conditions, such as low luminance, low contrast, and disabling glare. Therefore, for purposes such as diagnosis of ocular diseases, visual rehabilitation, job qualification, and disability benefit, visual acuity is used as a measurement of functional vision—that is, how well a person can perform vision-related activities under certain conditions. For a visual acuity measurement to be functionally relevant, the stimuli should be closely related to the observer's visual tasks. For literate observers, the most important visual task is undoubtedly reading text of the observers' native language.

Because of the importance of reading text in functional vision, letter charts have become the dominant way of clinical evaluation of functional vision around the world. However, letters are complex spatial patterns, and what constitutes the "finest detail" in letters cannot be precisely defined. Therefore, uniting visual acuity for refraction and for functional vision has always been a challenge.[1] Both international and U.S. standards stipulate that alternative optotypes, such as letters, should be equivalent to the Landolt C in test results. ISO 8597 (International Organization for Standardization) [2] stipulates that a set of optotypes is equivalent to the Landolt C if they differ less than 0.05 log unit. In the U.S. standard, the allowed difference is within 5%. If an alternative set of optotypes is not equivalent to the Landolt C, a size conversion factor should be determined to scale the optotypes. In phonics-based languages, at least in those using Roman alphabets, an agreement between refraction oriented Landolt C acuity and functional vision oriented letter acuity is relatively easy to achieve. Sloan et al.[5] tested acuity in 214 eyes with various refractive errors by using the Landolt C and the uppercase English letters CDHKNORSVZ and found that the two measurements correlated highly (Pearson $r = 0.90$). When such agreement between primary and alternative optotypes can be established, the same unit for visual resolution, MAR (minimum angle of resolution) or logMAR, can be used for functional vision measurement, even though for

From the [1]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China; the [2]Institute of Neuroscience, Chinese Academy of Sciences, Shanghai, China; the [3]EENT Hospital, Fudan University, Shanghai, China; and the [4]School of Optometry, University of Alabama at Birmingham, Birmingham, Alabama.

Supported by Beijing Normal University and the Chinese Academy of Sciences.

Submitted for publication October 5, 2006; revised December 31, 2006 and January 17, 2007; accepted March 16, 2007.

Disclosure: **J.-Y. Zhang**, None; **T. Zhang**, None; **F. Xue**, None; **L. Liu**, None; **C. Yu**, None

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Each of the following is a corresponding author: Cong Yu, State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China; yucong@bnu.edu.cn.
Lei Liu, School of Optometry, University of Alabama at Birmingham, Birmingham, AL; liul7788@uab.edu.

Copyright © Association for Research in Vision and Ophthalmology

In the rest of the article, unless specifically indicated, “letter” refers to a Sloan letter or a Chinese character or both.

Apparatus

The stimuli were generated by a MatLab-based (The MathWorks, Natick, MA) WinVis program (Neurometrics Institute, Oakland, CA) and were presented on a 21-in. color monitor (2048 × 1536 pixels, 0.189 × 0.189 mm/pixel, 75-Hz frame rate). The maximum and minimum luminance of the monitor was 89 and 0.02 cd/m^2. All optotypes were minimal-luminance, black figures on a maximum-luminance white background. Observers viewed the displays binocularly in a dimly lit room. An instrument table with a chin–headrest combo was used to maintain correct viewing distance.

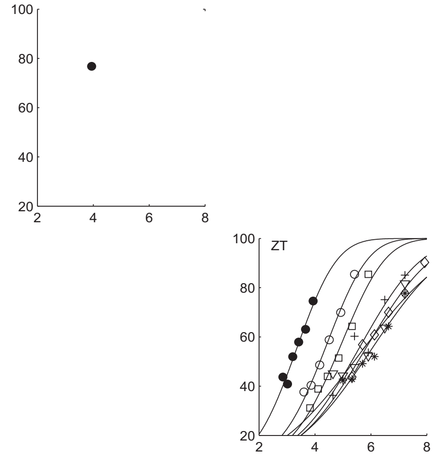
Observers and Procedures

In the two parts of the study, determination of the legibility of Chinese characters and the measurement of the effects of optical defocus on Chinese character legibility, different sets of observers and different procedures were used.

Chinese Character Legibility

Six young (mean age, 22.8 years) native Chinese speakers with normal or corrected-to-normal vision served as observers. All observers had a college education and at least 7 years of training in reading and writing English. Except ZJ, a coauthor, all were new to psychophysical observations and were unaware of the purpose of the study. Each observer first underwent refractive testing by a trained technician using a Snellen E light box at the designated viewing distance of 5 m. Most observers were slightly myopic and wore corrections. The average best corrected acuity was −0.114 ± 0.049 logMAR.

Sloan letters and the Chinese optotypes CC1 through CC6 were used. A method of constant stimuli was used to measure the acuity size of a stimulus group. A single letter stimulus was presented at the center of the screen with unrestricted duration. The observer’s task was to report the stimulus letter from a 10-letter list with a key-press (0–9).



was halved. The experiment was conducted in a dimly lit room. A front surface mirror was used to increase optical distance.

The observer's task was to report the five optotypes from left to right. The stimulus was shown continuously until all five optotypes were reported. The observers knew the tested letters well, and they had a list of the tested optotypes in large print at hand. The test began with a large optotype size (usually 0.2 or 0.3 log unit larger than the threshold size estimated in pilot studies). If the observer reported four or five optotypes of a possible five correctly, the optotype size was reduced by approximately 0.05 log unit. If the observer reported three or fewer optotypes correctly, the same optotype size was repeated with a new line of optotypes. If the observer was then able to report four or five optotypes correctly, the optotypes size was reduced, and the test continued. Otherwise, the mean number of correctly read optotypes of this size was recorded, and their contribution to acuity was considered according to the letter-by-letter scoring principle. Acuity was recorded in terms of one fifth of the threshold letter size.

Written informed consent was obtained from all observers. The research adhered to the tenets of the Declaration of Helsinki.

Results

Legibility of Chinese Characters

Figure 2a shows each observer's psychometric function and Weibull fitting for all stimulus groups (each data point represents the average correct responses for 10 letters). Acuity sizes estimated from Weibull fittings for all individual optotypes were plotted against stroke frequencies in Figure 2b, shown as small symbols, and the average acuity sizes for each stimulus group were shown as large bold symbols. ANOVA suggested a significant stimulus group effect ($F_{6,30} = 89.88$, $P < 0.0005$), in that more complex optotypes had larger acuity sizes. The acuity size of the Sloan letters, 3.68 arcmin, was significantly lower than that of all CC groups—a topic that will be discussed later. After the Sloan letter group was removed, the stimulus group effect among the six CC groups was still significant ($F_{5,25} = 29.62$, $P < 0.0005$). For Chinese stimuli, acuity character size increased linearly as a function of the stroke frequency (Fig. 2b). However, the best-fitting line (the straight line in Fig. 2b) had a shallow slope (0.435). From the simplest to the most complicated CC groups (CC1 vs. CC6), a 2.5-fold increase in mean stroke frequency (2.22–5.52 strokes/letter) was accompanied with a 1.28-fold, or 0.1-log-unit, increase in acuity size (4.68–5.99 arcmin), which was equivalent to one full line on the Bailey-Lovie acuity chart.[14] These data indicate that character recognition may not be totally based on discrimination of the finest details. Otherwise, the acuity size would have increased with a much steeper slope.

Figure 2b shows that Sloan letters had significantly smaller acuity sizes than CC1 stimuli (average 3.41 vs. 4.68 arcmin; $F_{1,5} = 96.29$, $P < 0.0005$), even though both groups had similar stroke frequencies (2.02 vs. 2.22 strokes/letter). This difference may be accounted for by thicker strokes of Sloan letters (Fig. 1). To test this hypothesis, eight Chinese characters were selected that had structures similar to that of eight of the Sloan letters, with the exception of maybe a rotation and some minor differences, such as the straightness of strokes or stroke endings. These Chinese characters were rendered in two ways, with normal bold Heiti strokes and with graphically thickened strokes similar to those of Sloan letters (CC_{thin} and CC_{thick} in Fig. 3a). CC_{thin} and CC_{thick} stimuli had stroke frequencies similar to those of the corresponding Sloan letters.

Acuity sizes of these groups were obtained from three young observers in the same way as in the main experiment, and the average acuity sizes of individual optotypes are shown in Figure 3b. The acuity size difference between Sloan letters and regular Chinese bold Heiti characters was reproduced (mean = 3.88 and 4.66 arcmin for Sloan and CC_{thin}, respectively; $F_{1,2}=30.09$, $P = 0.032$). Thickening the strokes reduced the threshold size (mean acuity size from 4.66 arcmin for CC_{thin} to 4.31 arcmin for CC_{thick}; $F_{1,2} = 17.87$, $P = 0.052$). Moreover, the mean acuity size of CC_{thick} was not significantly different from that of Sloan (3.88 arcmin; $F_{1,2} = 9.35$, $P = 0.092$). These results indicate that the smaller acuity sizes of Sloan letters shown in Figure 2b were at least partially due to their thicker stroke width. Other factors that may have contributed to the acuity size difference will be analyzed in the Discussion section.

Effects of Optical Defocus on Chinese Character Legibility

The variation of legibility of Chinese characters shown in Figure 2b was wide. However, we argue (see the Discussion section) that the task of deriving a simple visual function

in strokes. In contrast, Chinese characters are either pictographs (single-body) or compounds of pictographs. Because pictographs are more realistic depictions of natural objects or events, they do not always have the regularities of Roman letters. As a consequence, the stroke types and their placement in Chinese characters are much less predictable. From a Gestalt psychology point of view, Chinese characters are not as good patterns as Roman alphabets, because Chinese characters offer many more alternatives, and good patterns have few alternatives.[21] Not-so-good patterns with more unique features should be harder to recognize than good patterns with fewer unique features. Indeed, Pelli et al.[22] found that human efficiency for recognizing Chinese characters was only one third of the efficiency for recognizing regular English letters. In our study, however, CC1 was approximately one-third less legible than Sloan letters (Fig. 2b), rather than two-thirds less legible if predicted by the efficiency difference. The acuity difference even became statistically nonsignificant at equal stroke widths (Fig. 3b). The vanishing advantage of English letters over Chinese characters can be explained by the nature of the acuity task. Acuity is a unique task in which subjects are forced to perform recognition or identification based solely on the global features of the stimulus, because fine features have been significantly attenuated or completely removed by ocular optics. This low-pass filtering becomes an ultimate equalizer that wipes out most of the graphical differences between stimulus groups. In a separate theoretical study, we demonstrated that the distances among optotypes in a space defined by a few low-order geometric moments, which captured the global characteristics of two-dimensional images, could account for most of the errors made by human observers in recognizing near-acuity Sloan letters and Chinese characters.

A Chinese character expresses a meaning and is thus functionally equivalent to an English word. Chinese can be read twice as fast as English.[23] Recognition of Chinese characters is four times as efficient as recognition of five-letter English words (five letters is the average length of English words).[22] However, this advantage of Chinese character over English words does not hold at acuity. Our result showed that CC1 was 37.2% less legible than Sloan letters. Meanwhile, Sheedy et al.[24] compared acuities of single Sloan letters and lowercase English words (five to six letters) in four font faces and found that lowercase words were 4.5% to 7% less legible than Sloan letters. These data together indicate that CC1 is approximately 33% less legible than lowercase English words. We speculate that in an acuity test, the global properties of the physical structure of the stimulus, whether it is a letter, a character or a word, determine the acuity size. Familiarity with the stimulus or the meaning of the stimulus may have little effect.

Deriving a Visual Function Measurement from Snellen E Acuity

The difference between the easiest among our 60 Chinese characters (人, 3.77 arcmin) and the most difficult (验, 6.90 arcmin) was 0.263 log unit. The legibility variability of characters in real Chinese text is likely to be even larger. Can we designate one CC group of intermediate legibility and use its acuity size as the functional measure for Chinese readers? Before we can speculate on a solution, we must know how the legibility of characters of different complexities change when viewing condition changes. If they all change in proportion, then the acuity of one group may be designated as the functional measure for all conditions, which was exactly what we demonstrated when we used plus lenses to simulate refractive errors. The parallel straight lines of the Snellen E, CC1, CC3, and CC6 shown in Figure 4b enable a practitioner to use a patient's Snellen E acuity to infer the patient's performance with Chinese characters of different spatial complexities. The next question is which CC group best represents the visual demand for reading Chinese? One solution would be to use the number of strokes that occur most frequently in daily Chinese text. Shu et al.[25] studied properties of the 2570 Chinese characters listed in the official elementary school textbooks in Beijing. The distribution of stroke numbers ranged from 1 to 24 and could be fitted with a Gaussian with the mean at 9.10 ± 0.09 strokes. Our CC3 group (eight to nine strokes) is the closest to this mean. The acuity size for CC3 is 0.210 log unit larger than that of the Snellen E (Fig. 4b), which converts to a scaling factor of 1.622. Therefore, if a road sign is designed to be read 100 m away by a driver with 20/20 vision, the Chinese characters on the sign have to be at least $1.622 \times 100 \times \tan(5$ arcmin$) = 0.236$ m, or 23.6 cm tall. Because the driver's vision is determined by a Snellen E chart, the 1.622 factor compensates for the acuity size difference between the Snellen E and average Chinese characters.

It is worth noting that the Chinese characters used in this study were "simplified" characters, which are standard in mainland China and Singapore. Many simplified characters have fewer strokes than the corresponding traditional characters. For instance, the simplified character 无 (none, void) in CC2 would be 無 as a traditional character. The distribution of the number of strokes and thus the mean acuity size are likely to be larger for traditional Chinese characters, and for Kanji in Japanese, because most of them are traditional Chinese characters.

Finally, it is also worth pointing out that the parallel relationship between the Snellen E and the Chinese characters shown in Figure 4b was obtained by introducing optical defocus. While this method (dioptric blur) has been widely used to simulate refractive errors in the eye, subtle differences may exist. We are currently working on a clinical population to see whether the same relationship holds in naturally occurring refractive errors.

Conclusions

Measurements of visual acuities of Sloan letters and Chinese characters of different numbers of strokes revealed that more complex optotypes had larger acuity sizes. The increase of acuity size with optotype spatial complexity, however, was not fast enough to support the notion that visual acuity was determined by discrimination of the smallest details. Sloan letters had significantly smaller acuity size than the simplest group of Chinese characters, even though the two groups had comparable spatial complexity. This difference, however, could partially be explained by the stroke width difference between these optotypes. When optical defocus was introduced, the acuity sizes of Chinese characters increased in the same way as Snellen E optotypes. Such simple relationships may help to derive a functional measure from Snellen E acuity that is relevant to Chinese reading.

Appendix

Selection of Chinese Optotypes

First, the 500 most frequently used Chinese characters (CCs) were selected from an official character-frequency table,[26] which was compiled based on a linguistic corpus of 138 million CCs. Six groups of CCs were selected based on the number of strokes (i.e., 2–4, 5–6, 8–9, 11–12, 13–15, and 16–18 strokes). To reduce testing characters to a manageable number, a computer analysis of physical similarity among characters in each group was conducted. The 50×50 bitmap of each character was considered as a point in a 50×50-dimension space, with the coordinates $x_1, x_2, x_3, \ldots, x_{2500}$). These coordinates were either 0 (a black pixel) or 1 (a white pixel). The

Euclidean distance between the *i*th and the *j*th characters in this space was $d_{ij}=[\sum_{k=1}^{2500}(x_{ik}-x_{jk})^2]^{1/2}$. All pair-wise Euclidean distances among representing points of all characters within each group were calculated. Several studies[15–17] have shown that Euclidean distances correlated with perceived similarity between English letters. From each CC group, 12 to 14 characters with intermediate Euclidean distances from each other were selected. This procedure excluded characters that were either physically too similar or too different. Additional considerations of pronunciation and spatial configurations further reduced the number of characters in each group to 10. The resultant six sets of Chinese optotypes are shown in Figure 1a (CC1–CC6).

Calculation of Optotype Stroke Frequency

Although the number of strokes has been used in many studies of letter recognition to index the complexity of stimuli, it is not a good measure of spatial complexity because the total number of strokes ignores the spatial arrangement of strokes, and thus may not provide a good measurement of stroke density. For example, although the character 口 contains four straight line segments, there are only two line segments in the horizontal or vertical direction. Characters 三 (three) and 川 (river) have only three strokes each, but the stroke density in the vertical or horizontal direction is 1.5 times of that of 口. A more objective measurement of spatial complexity of optotypes is stroke frequency.[27] It was originally defined as the average number of strokes crossed by a slice through the letter width. Because many Chinese characters have unbalanced top–bottom or left–right configurations (for example, the second and third characters in CC4 in Fig. 1a), and because some Chinese characters have predominantly oblique strokes (the last character of CC3), a more sophisticated method was used to calculate stroke frequencies of our optotypes. As shown in Figure 1b, each letter was sliced in one-pixel steps into six direction–position combinations: horizontally on the upper and lower halves, vertically on the left and right halves, and obliquely at 45° and 135° on the central portion of the optotype. A mean of crossed strokes was obtained from each slicing, and the maximum of the six slices was taken as the stroke frequency of the optotype. The average stroke frequency for the Sloan letters was 2.0 strokes/letter, slightly larger than that obtained in one horizontal slice (1.6 strokes/letter).[27] The average stroke frequencies in the six groups of Chinese characters increased monotonically from 2.2 to 5.5 strokes/letter.

Another measurement of spatial complexity of patterns is perimetric complexity.[22,28] Perimetric complexity is a size-invariant measure of dispersion and is defined as the square of inside-and-outside perimeter of a pattern, divided by the ink area. We calculated perimetric complexities of all the optotypes used in our study and found they correlated highly with stroke frequencies ($r = 0.956$). Because stroke frequency is the more intuitive of the two measurements, we chose to use stroke frequency as our measurement of optotype complexity.

Acknowledgments

The authors thank Patti Fuhr (VA Medical Center, Birmingham, AL) for careful reading of the manuscript and for lending her clinical insight and Shu-Guang Kuai for helping to write software programs for the study.

References

1. National Academy of Science National Research Council (NAS-NRC). Recommended standard procedures for the clinical measurement and specification of visual acuity. Report of Working Group 39. *Adv Ophthalmol.* 1980;41:103–148.
2. International Organization for Standardization (ISO). Visual acuity testing: standard optotype and its presentation. *Optics and Optical Instruments.* Geneva, Switzerland: International Organization for Standardization; 1986.
3. National Standard of People's Republic of China. *Standard Logarithmic Visual Acuity EChart* (GB 11533-1989). Beijing: China Standard Press; 1990.
4. Sloan L. Measurement of visual acuity: a critical review. *Arch Ophthalmol*. 1951;45:704–725.
5. Sloan L, Rowland WM, Altman A. Comparison of three types of test target for the measurement of visual acuity. *Q Rev Ophthalmol*. 1952;8:4–15.
6. Grimm W, Rassow B, Wesemann W, Saur K, Hilz R. Correlation of optotypes with the Landolt ring: a fresh look at the comparability of optotypes. *Optom Vis Sci*. 1994;71:6–13.
7. Khamar BM, Vyas UH, Desai TM. New standardized visual acuity charts in Hindi and Gujarati. *Indian J Ophthalmol*. 1996;44:161–164.
8. Ruamviboonsuk P, Tiensuwan M. The Thai logarithmic visual acuity chart. *J Med Assoc Thai*. 2002;85:673–681.
9. Al-Mufarrej MM, Abo-Hiemed FA, Oduntan AO. A new Arabic distance visual acuity chart. *Optom Vis Sci*. 1996;73:59–61.
10. Woo G, Lo P. A Chinese word acuity chart with new design principles. *Singapore Med J*. 1980;21:689–692.
11. Cheng AS. Relative legibility study using Chinese optotypes. *Singapore Med J*. 1991;32:38–40.
12. Hao YM, Johnston AW. An evaluation of logMAR vision test charts for near vision using Chinese characters. *Clin Exp Optom*. 1997;80:178–186.
13. Ferris FL 3rd, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. *Am J Ophthalmol*. 1982;94:91–96.
14. Bailey IL, Lovie JE. New design principles for visual acuity letter charts. *Am J Optom Physiol Opt*. 1976;53:740–745.
15. Coffin S. Spatial frequency analysis of block letters does not predict experimental confusions. *Percept Psychophys*. 1978;23:69–74.
16. Gervais MJ, Harvey LO Jr, Roberts JO. Identification confusions among letters of the alphabet. *J Exp Psychol Hum Percept Perform*. 1984;10:655–666.
17. Loomis JM. A model of character recognition and legibility. *J Exp Psychol Hum Percept Perform*. 1990;16:106–120.
18. Roethlein BE. The relative legibility of different faces of printing types. *Am J Psychol*. 1912;23:1–36.
19. Ferris FL 3rd, Freidlin V, Kassoff A, Green SB, Milton RC. Relative letter and position difficulty on visual acuity charts from the Early Treatment Diabetic Retinopathy Study. *Am J Ophthalmol*. 1993;116:735–740.
20. von Benda H. *Dimensionsanalyse Der Statischen Sehscharfe* [*Dimension Analysis of Static Visual Acuity*]. Gottingen, Germany: Verlag fur Psychologie; 1981.
21. Garner WR. Good patterns have few alternatives. *Am Sci*. 1970;58:34–42.
22. Pelli DG, Burns CW, Farell B, Moore-Page DC. Feature detection and letter identification. *Vision Res.* 2006;46:4646–4674.
23. Lu X, Zhang J. Reading efficiency: a comparative study of English and Chinese orthographies. *Read Res Instruct*. 1999;38:301–317.
24. Sheedy JE, Subbaram MV, Zimmerman AB, Hayes JR. Text legibility and the letter superiority effect. *Hum Factors*. 2005;47:797–815.
25. Shu H, Chen X, Anderson RC, Wu N, Xuan Y. Properties of school Chinese: implications for learning to read. *Child Dev*. 2003;74:27–47.