

## The consequences of literacy and schooling for parsing strings

*Saima Malik-Moraleda*<sup>1</sup>

*Karla Orihuela*<sup>2</sup>

*Manuel Carreiras*<sup>1,3,4</sup>

and

*Jon Andoni Duñabeitia*<sup>1,\*</sup>

<sup>1</sup>. *BCBL, Basque Center on Cognition, Brain and Language; Donostia, Spain*

<sup>2</sup>. *Université de Toulouse II; Toulouse, France*

<sup>3</sup>. *Ikerbasque, Basque Foundation for Science; Bilbao, Spain*

<sup>3</sup>. *University of the Basque Country; Bilbao, Spain*

\* Contact information:

**Jon Andoni Duñabeitia, PhD.**

*j.dunabeitia@bcbl.eu*

*Basque Center on Cognition, Brain and Language (BCBL)*

*Paseo Mikeletegi 69, 0*

*20009 Donostia – Spain*

*+34 943309300 (ext. 208)*

### ACKNOWLEDGEMENTS AND FUNDING

This research has been partially funded by grants PSI2015-65689-P and SEV-2015-0490 from the Spanish Government, AThEME-613465 from the European Union, and ERC-AdG-295362 grant from the European Research Council. This project has been also supported by a 2016 BBVA Foundation Grant for Researchers and Cultural Creators awarded to the last author. The authors wish to thank the Instituto Nacional Para la Educación de los Adultos (INEA) for granting access to a subgroup of Mexican participants.

## **ABSTRACT**

Word processing initially occurs through letter-by-letter parsing at early stages of reading development. Here we investigated the role of literacy in parsing both linguistic and non-linguistic strings, as well as the consequences of the possible changes brought by the acquisition of reading. Illiterates matched with schooled literates on socio-demographic and cognitive measures were presented with a character search task. Overall, literates performed better than illiterates in identifying constituents in linguistic and non-linguistic strings. Illiterates showed a similar performance for all types of strings. In contrast, literates showed a graded pattern with non-linguistic strings being processed much worse than linguistic strings. These results support domain-specific models of orthographic processing, and they suggest that visual word recognition is not fully parallel to visual object recognition. Most importantly, they demonstrate the impact of literacy on the ability of breaking down a word into its constituents.

**KEYWORDS:** literacy; reading; orthographic coding; letter processing; illiterates

**RUNNING HEAD:** Literacy and character processing

**WORD COUNT:** 3191

Studies comparing literate, ex-illiterate and illiterate participants provide growing evidence on substantial cognitive and brain changes as a result of reading acquisition (Carreiras et al., 2009; Dehaene et al., 2010; Dehaene, Cohen, Morais, & Kolinsky, 2015; Pegado et al., 2014). In particular, it has been argued that literacy improves analytical strategies when attending to different types of stimuli (Dehaene, Cohen, Sigman, & Vinckier, 2005; Szwed, Ventura, Querido, Cohen, & Dehaene, 2012). For instance, when being presented with the top part of an image, literates seem to demonstrate an advantage in deciding whether it is the same as an image previously shown (Ventura et al., 2013). Yet, some studies suggest that illiterates are not devoid of all types of analytic processing (Lachmann, Khera, Srinivasan, & van Leeuwen, 2012), and in certain tasks requiring part verification, illiterate adults perform on par with ex-illiterates that learned to read during adulthood, and only schooled literates showed enhanced perceptual processing skills (Kolinsky, Morais, & Content, 1987). Thus, the extent to which analytic processing is effectually influenced by literacy acquisition remains uncertain.

The term analytic processing might refer to different types of skills, making it of paramount importance to firstly establish a divide on what constitutes analytic processing and what does not and secondly narrow down the precise skillset acquired through literacy. Perhaps the most elemental divide one can make in visual recognition is a division of labor between holistic and part-based recognition systems (Farah, 1990, 1992). A part-based system parses the object or image into its constituents and integrates them in order to match them with the parts parsed out of other objects or images. On the other hand, in the holistic recognition system parsing mechanisms are not employed and recognition is done as a whole following Gestalt principles.

The holistic and part-based division takes a new dimension when it comes to linguistic stimuli as printed words are a specific type of visual stimuli where the individual constituents (i.e., the letters) require their identity to be computed alongside their relative position in it – a process called orthographic processing (Humphreys, Evett, & Quinlan, 1990). Word

recognition has been extensively shown to rely on a part-based visual recognition system (Osswald, Humphreys, & Olson, 2002; Pelli, Farell, & Moore, 2003; Pelli & Tillman, 2007; Suchow & Pelli, 2013), yet a more holistic processing of linguistic stimuli is also endorsed by classic effects observed in literates (e.g., Reicher, 1969; Wheeler, 1970).

One of the most commonly reported examples of a flexible orthographic coding system that lies in between analytic and holistic types of processing is the transposed-letter effect (henceforth, TL; e.g., Perea, Duñabeitia, & Carreiras, 2008). In many cases readers process words with two transposed letters (e.g., *apartment*) as if they were the normal word (*apartment*), due to the flexibility of the orthographic parser and its tolerance to small positional variations (Perea & Carreiras, 2006; Perea & Lupker, 2004; Schoonbaert & Grainger, 2004). As a result, in a same-different task where readers are presented with the word *apartment*, they will be more likely to incorrectly produce a same response to *apartment* than to *apandment*, where the characters have been replaced rather than transposed (see Duñabeitia, Dimitropoulou, Grainger, Hernández, & Carreiras, 2012; Massol, Duñabeitia, Carreiras, & Grainger, 2013, for evidence gathered with different types of character-transpositions in this paradigm). Nonetheless, these TL effects which partially rely on some form of holistic processing necessarily require a previous step from the orthographic parser by which all the parts constituting the whole are recognized. In order to produce a “same” response to *apartment* and *apartment*, one needs to a) recognize both *t* and *r* as individual characters and b) flexibly code the position within the string occupied by both *t* and *r*.

Interestingly, illiterate adults and pre-literate children show an erratic performance and lack of TL effects when perceptually matching linguistic and non-linguistic strings (Duñabeitia, Lallier, Paz-Alonso, & Carreiras, 2015; Duñabeitia, Orihuela, & Carreiras, 2014). This critical finding seems to suggest that the type of analytic processing that mediates visual word recognition and grants access to sub-lexical units is acquired through literacy during schooling. The aim of the current study thus was to investigate how literacy improves analytic skills in general, and part-based recognition in particular, in both linguistic and non-linguistic domains. To

do so, the ability of literates and illiterates in recalling an individual constituent of a linguistic and non-linguistic string was assessed in a character search task in which participants had to decide whether a given element had been previously presented within a string. Participants were presented with strings made of letters constituting either words or pseudowords (orthographically and phonologically legal strings), and with strings of pictograms. As literacy (together with schooling) is taunted to improve analytic processing, we expected literates to be better than illiterates mainly in accessing the constituents of the linguistic strings. Moreover, domain-general models of orthographic processing would predict that schooled literates display clear enhancement of analytic processing of both linguistic and non-linguistic strings to a similar extent (e.g., Gomez, Ratcliff, & Perea, 2008), while domain-specific accounts would predict that the advantage of schooled literates over illiterates is significantly higher when presented with linguistic than with non-linguistic strings (see Duñabeitia et al., 2015, for a summary) .

## Methods

**Participants.** A group of 26 Mexican unschooled illiterate adults was tested (3 males, mean age= 40.19, SD=10.57), chosen from a limited pool of illiterate participants available. This group was matched with a group of 26 Mexican adult schooled literates (2 males, mean age =39.04, SD=9.94) on their gender, age, socio-economic status (SES), working memory skills (WM) and intelligence (see Table 1). All the literate adults had attended school regularly during childhood, and none of the illiterate adults had been formally educated. The illiterate and literate participants were recruited at the exact same neighborhoods from the Mexican state of Morelos to minimize the effects resulting from the different socio-demographic origins. SES was measured with a normative questionnaire from the Mexican National Statistical Institute that ranks the citizens in SES groups according to basic socioeconomic indicators. WM was measured using the digits subtest of the Wechsler Adult Intelligence Scale (Wechsler, 2008). Intelligence was measured with the non-verbal subtest of the

K-BIT test (Kaufman & Kaufman, 1990), and the resulting data were transformed to their corresponding centiles according to the norms. All participants were right handed, as measured by an abridged version of the Edinburgh Handedness Inventory (Oldfield, 1971), and all of them had Spanish as their native language. Familiarity with written language was also assessed with the pseudoword reading subtest of the PROLEC-R test (Cuetos, Rodríguez, Ruano, & Arribas, 2007), wherein participants are presented with a list of pseudowords and the total amount of strings successfully read represents the score. Results showed that none of the illiterate participants could read any of the strings in the list, while literate participants were able to correctly read on average 38.12 strings ( $SD=2.36$ ) out of the 40 items from the test (95.29% of accuracy; see Table 1). The BCBL Ethics Committee approved all the experimental protocols and participants gave informed consent prior to any data collection.

*- Please insert Table 1 around here -*

**Materials.** Items could either be linguistic (words and pseudowords) or non-linguistic elements. Linguistic strings were created by mixing uppercase Latin letters, and non-linguistic strings were created by mixing icons corresponding to different pictograms (see Figure 1 for examples). A total of 150 different words, 150 pseudowords and 150 pictograms strings were created. Each element appeared twice as a target, once requiring a “Yes” response when the target had been presented in the reference, and once requiring a “No” response when it had not. All individual items appeared an equal amount of times in each of the five positions of the strings in the trials requiring a “Yes” response, thus controlling for position throughout the experiment. Trials of pseudowords, words and pictures were presented separately in blocks of 300 trials each, with each participant completing one block of each type of stimuli. “Yes” and “No” trials were intermixed so that each block had 150 trials of each type. In total, each participant completed 900 experimental trials.

- Please insert Figure 1 around here -

**Procedure.** Each participant was tested individually on a computer placed at an approximate distance of 70cm, with the linguistic stimuli presented in black Courier New font on a white background in a monitor set to 1024x768 resolution at 90Hz. The experiment was run using Experiment Builder<sup>®</sup>. Participants were presented with a short practice prior to the experimental trials. In each experimental trial, participants paid attention to a reference, which consisted of a string composed by a series of 5 individual elements (characters/pictograms) displayed in the center of the screen for 1000ms. They were subsequently presented with a target item that consisted of one single element presented in red, horizontally centered in the screen and vertically aligned in the lower part of the screen. The target remained on the screen until the participant gave a response, or for a maximum of 3000ms. Participants were required to respond whether the target item had been presented in the reference by pressing two buttons on a numeric keyboard (numbers 7 and 9, with the left and right thumbs, respectively). Auditory feedback was given on the accuracy of the response: a high pitch tone was played after incorrect responses, and a low-pitch tone after correct ones. The inter-trial interval was set to 1000ms. Item presentation within a block was randomized across participants, and block order was also randomized. A short break was placed in the middle (after 150 trials) and at the end of each block. The whole experimental session lasted around 45 minutes.

## **Results and Discussion**

The design of the experiment included Group (literate|illiterate) as a between-participants factor and Type of String (words|pseudowords|pictures) as a within-participant

factor. Mean error rates were calculated for each group and experimental condition<sup>1</sup> (see Table 2 for descriptive data). One participant from the illiterate group was excluded from the analysis given that she only pressed one button throughout the experiment. The whole set of data were analyzed following the principles of Signal Detection Theory (Macmillan & Creelman, 2004) by calculating the sensitivity index ( $d'$  statistic) across conditions and groups. The  $d'$  provides a statistical calculation of the signal-to-noise ratio in a distribution of responses that allows for determining participants' sensitivity to the presence of an individual element within a string in the different stimuli conditions. For each Type of String,  $d'$  indices were calculated by considering each individual's normalized hit rate (proportion of correct responses in 'Yes' trials) and the normalized false alarm rate (proportion of incorrect responses in 'No' trials). In cases in which the false alarm rate was 0, it was set to  $1/(2N)$ . An ANOVA was conducted following the  $2 \times 3$  design described above, and given that Mauchly's test showed significant departure from sphericity [ $\chi^2(2)=6.93, p<.04$ ], the degrees of freedom were adjusted using the Huynh-Feldt estimation of sphericity ( $\epsilon=.93$ ).

The ANOVA on the  $d'$  scores showed a significant Group effect [ $F(1,49)=90.16, p<.001, \eta^2_{\text{partial}}=.65$ ], a significant effect of Type of String [ $F(1.86,91.19)=67.82, p<.001, \eta^2_{\text{partial}}=.58$ ], and a significant two-way interaction between the two factors [ $F(1.86,91.19)=70.17, p<.001, \eta^2_{\text{partial}}=.59$ ]. The significant interaction suggested that the discriminability of characters in strings was different for literates and illiterates as a function of the type of visual material that was being presented. Post-hoc analyses demonstrated that illiterates had equal sensitivity to the

---

<sup>1</sup> An analysis of the reaction times was also performed, although these were not expected to be especially illuminating given the low accuracy rates. After removing RTs associated with erroneous responses and trials with no responses (2.5% of timeouts), trials with latencies above and below 2.5 standard deviations from the standardized mean of each condition per participant were also excluded (2.91% of the data). An ANOVA was performed with Group (illiterate/literate) as a between-participants factor and Type of String (word|pseudoword|picture) and Type of Answer (yes|no) as within-participant factors. Mauchly's test for sphericity violation was significant for Type of String [ $\chi^2(2)=12.75, p<.01$ ], and the corresponding degrees of freedom were adjusted using the Huynh-Feldt estimation ( $\epsilon=.85$ ). The main effect of Group was not significant ( $F<1$  and  $p>.94$ ), but the main effect of Type of String was significant [ $F(1.70,83.47)=6.99, p<.01, \eta^2_{\text{partial}}=.12$ ], and so it was the main effect of Type of Answer [ $F(1,49)=35.74, p<.001, \eta^2_{\text{partial}}=.42$ ]. A significant three-way interaction was found [ $F(2,98)=3.11, p=.05, \eta^2_{\text{partial}}=.06$ ]. Post-hoc analysis looking at the pairwise comparisons between each level of the factor Type of String for each level of Group and Type of Answer showed that the RTs of illiterate participants were comparable across conditions (all  $t_s(24)<1.7$  and all  $p_s>.11$ ). In contrast, literate participants were significantly faster in responding to pseudowords and words than to picture strings (all  $t_s(25)>2.35$  and  $p_s<.03$ ), while showing comparable RTs to pseudowords and words (all  $t_s(25)<1$  and  $p_s>.45$ ).

three types of strings, and that their  $d'$  indices were not statistically different from each other (all  $t_{s(24)} < 1$  and  $p_{s} > .50$ ). Still, their  $d'$  scores were statistically different from zero in all cases, demonstrating that their performance was not erratic despite being relatively poor (all  $t_{s(24)} > 4.50$  and  $p_{s} < .001$ ). In contrast, the performance of the literate group was significantly better for words than for pseudowords (namely, a word superiority effect) and for picture strings, and it was also better for pseudowords than for pictures (all  $t_{s(25)} > 2.50$  and  $p_{s} < .02$ ). All the  $d'$  indices for literates were also significantly different from zero (all  $t_{s(25)} > 8.85$  and  $p_{s} < .001$ ). Finally, literate participants performed significantly better than illiterate participants with all types of strings (all  $t_{s(49)} > 2.90$  and  $p_{s} < .01$ ).

*- Please insert Table 2 around here -*

Reading is an empowering ability that starts with the small act of decoding the letters that constitute a word, requiring a set of analytic skills which invoke part-based visual recognition processes (Lachmann & van Leeuwen, 2014; Pelli et al., 2003). Our results showed a clear-cut dissociation in the specific analytic capacities enhanced or developed as a consequence of reading acquisition during schooling. Overall, illiterates showed a much poorer performance than literates in accessing the constituent characters of linguistic and non-linguistic strings, committing a significantly higher percentage of errors. Besides, illiterates showed an equally poor performance with all types of strings, while literates performed significantly better with letter strings than with picture strings. Furthermore, literate adults showed a better performance with words than with pseudowords (i.e., a word superiority effect).

The main findings of the current study suggest that reading acquisition promotes a type of analytic, part-based processing that does not seem to be inherent to the visual system. Schooled literate adults showed better perceptual analytic skills than illiterates with all types of visual materials, being this difference higher for linguistic than for non-linguistic stimuli. These

results are in line with previous studies suggesting that schooling instills part-based processing in non-linguistic stimuli too (e.g., Kolinsky, Morais & Content, 1987; Ventura et al, 2008), given that schooled literates also showed better performance than unschooled illiterates with non-linguistic picture strings (see Table 2). But considering that the differences between groups were maximal with linguistic stimuli, these data suggest a certain extent of domain specificity prompted by schooling (see Carreiras et al., 2014; Duñabeitia et al., 2014, 2015). In light of these results, we tentatively conclude that the processing of linguistic strings is primarily part-based, and that this skill is developed solely after the acquisition of reading during schooling.

Different types of expertise have been associated with holistic processing (Richler, Tanaka, Brown, & Gauthier, 2008; Wong et al., 2011). However, and despite the vast reading expertise accumulated by literates, it is remarkable that reading heavily relies on a part-based analytic system (Pelli, Farell, & Moore, 2003). As analytic processing is mainly improved for linguistic material, it could be tentatively suggested that literacy acquisition yields qualitatively different processing routines for letters than from other types of visual stimuli. A plausible reason for this change might lie on the specific grapheme-to-phoneme mappings and phonological recoding that takes place during letter-learning, which can trigger an analytic style of reading that cannot occur otherwise (Fernandes, Vale, Martins, Morais, & Kolinsky, 2014), and that cannot extend to non-linguistic materials lacking phonological representations. An alternative approach to explain the advantage shown for literate adults with linguistic strings would be to interpret these differences as a consequence of the ease for integrating linguistic items in the phonological working memory buffer. For obvious reasons, phonological recoding strategies cannot be used to retain in memory picture strings, and one could tentatively argue that this could be the underlying reason for the differences observed in the literate group. However, if the advantage were exclusively due to phonological recoding and integration in phonological working memory, both pseudowords and words would have benefitted equally from this strategy, and thus one would not have expected the word superiority effect observed in the current study. Hence, even if this is

a mechanism that could have been mediating in the current task, we believe that it is not the main strategy at play.

We interpret these findings as supportive of models of orthographic coding that are domain-specific (see also Duñabeitia et al., 2014, 2015). Considering these results from character-identity processing together with preceding data from studies tackling on character-position processing, we propose that the most plausible interpretation of the origins of orthographic processing – a basic initial step in the stream of processes involved in reading – lies mainly on domain-specific coding skills promoted by schooling-mediated literacy. The general picture of the whole set of studies on this matter suggests that this may be the case, given that illiterates consistently perform worse than schooled literates in coding the identity and position of all types of characters, and that literates consistently perform better with linguistic than with non-linguistic stimuli. Nonetheless, this does not necessarily invalidate domain-general approaches to orthographic coding, which in conjunction with expertise-based interpretations of parsing skills, suggest that the underlying mechanisms guiding these processes are not letter-specific, and that the advantage typically seen for letter strings could be due to the increased experience with the print (see Perea et al., 2015; Perea, Jiménez, & Gómez, 2016). It is important to stress that a letter-based word identification mechanism is not inconsistent with the view that a supra-segmental holistic level also governs later stages of word recognition. In fact, both analytic and holistic processing are essential for efficient reading, insofar letter-by-letter coding and holistic recognition all contribute to reading rate (Pelli & Tillman, 2007). Thus, it is possible that the domain-specific analytic processing skills attained through literacy interact with domain-general visuo-orthographic mechanisms. Further research is clearly needed in order to shed light on this issue, using both linguistic and non-linguistic stimuli to concurrently study character-position and character-identity coding in pre-literate and neo-literate children, on the one hand, and in unschooled illiterate, ex-illiterate and schooled literate adults, on the other.

In a nutshell, the findings in this study suggest that reading acquisition instills part-based processing of strings. Furthermore, these results demonstrate that schooling-mediated literacy improves analytic processing of linguistic strings more than that of non-linguistic strings, providing literates with a set of domain-specific analytic skills that are essential for efficient visual word identification.

## REFERENCES

- Carreiras, M., Quiñones, I., Hernández-Cabrera, J. A., & Duñabeitia, J. A. (2014). Orthographic coding: brain activation for letters, symbols and digits. *Cerebral Cortex*, *bhu163*, 4–59.
- Carreiras, M., Seghier, M. L., Baquero, S., Estévez, A., Lozano, A., Devlin, J. T., & Price, C. J. (2009). An anatomical signature for literacy. *Nature*, *461*(7266), 983–986.
- Cuetos, F., Rodríguez, B., Ruano, E., & Arribas, D. (2007). PROLEC-R. Batería de evaluación de los procesos lectores, revisada. Madrid: TEA.
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, *16*(4), 234–244.
- Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in Cognitive Sciences*, *9*(7), 335–341.
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., ... Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, *330*(6009), 1359–64.
- Duñabeitia, J. A., Dimitropoulou, M., Grainger, J., Hernández, J. A., & Carreiras, M. (2012). Differential sensitivity of letters, numbers, and symbols to character transpositions. *Journal of Cognitive Neuroscience*, *24*(7), 1610–1624.
- Duñabeitia, J. A., Lallier, M., Paz-Alonso, P. M., & Carreiras, M. (2015). The impact of literacy on position uncertainty. *Psychological Science*, *26*(4), 548–50.
- Duñabeitia, J. A., Orihuela, K., & Carreiras, M. (2014). Orthographic coding in illiterates and literates. *Psychological Science*, *25*(6), 1–23.
- Farah, M. J. (1990). *Visual Agnosia: Disorders Of Object Recognition And What They Tell Us About Normal Vision*. MIT Press.
- Farah, M. J. (1992). Is an object an object an object? Cognitive and neuropsychological investigations of domain specificity in visual object recognition. *Current Directions in Psychological Science*, *1*(5), 164–169.
- Fernandes, T., Vale, A. P., Martins, B., Morais, J., & Kolinsky, R. (2014). The deficit of letter processing in developmental dyslexia: combining evidence from dyslexics, typical readers and illiterate adults. *Developmental Science*, *17*(1), 125–141.
- García-Orza, J., Perea, M., & Muñoz, S. (2010). Are transposition effects specific to letters? *The Quarterly Journal of Experimental Psychology*, *63*(8), 1603–18.
- Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: a model of letter position coding. *Psychological Review*, *115*(3), 577–600.

- Humphreys, G. W., Evett, L. J., & Quinlan, P. T. (1990). Orthographic processing in visual word identification. *Cognitive Psychology*, 22(4), 517–560.
- Kaufman, A. S., & Kaufman, N. L. (1990). K-BIT: Kaufman brief intelligence test. American Guidance Service.
- Kolinsky, R., Morais, J., Content, A., & Cary, L. (1987). Finding parts within figures: A developmental study. *Perception*, 16(3), 399-407.
- Lachmann, T., Khera, G., Srinivasan, N., & van Leeuwen, C. (2012). Learning to read aligns visual analytical skills with grapheme-phoneme mapping: evidence from illiterates. *Frontiers in Evolutionary Neuroscience*, 4(June), 8.
- Lachmann, T., & van Leeuwen, C. (2007). Paradoxical enhancement of letter recognition in developmental dyslexia. *Developmental Neuropsychology*, 31(1), 61–77.
- Lachmann, T., & van Leeuwen, C. (2014). Reading as functional coordination: not recycling but a novel synthesis. *Frontiers in psychology*, 5, 1046.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.
- Massol, S., Duñabeitia, J. A., Carreiras, M., & Grainger, J. (2013). Evidence for letter-specific position coding mechanisms. *PloS One*, 8(7), e68460.
- Osswald, K., Humphreys, G. W., & Olson, A. (2002). Words are more than the sum of their parts: evidence for detrimental effects of word-level information in alexia. *Cognitive Neuropsychology*, 19(8), 675–695.
- Pegado, F., Comerlato, E., Ventura, F., Jobert, A., Nakamura, K., Buiatti, M., ... Dehaene, S. (2014). Timing the impact of literacy on visual processing. *Proceedings of the National Academy of Sciences*, 111(49), E5233–42.
- Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature*, 423(6941), 752–756.
- Pelli, D. G., & Tillman, K. a. (2007). Parts, wholes, and context in reading: a triple dissociation. *PloS One*, 2(8), e680.
- Perea, M., & Carreiras, M. (2006). Do transposed-letter similarity effects occur at a prelexical phonological level? *Quarterly Journal of Experimental Psychology*, 59(9), 1600–1613.
- Perea, M., Duñabeitia, J. A., & Carreiras, M. (2008). Transposed-letter priming effects for close versus distant transpositions. *Experimental Psychology*, 55(6), 384–93.
- Perea, M., Jiménez, M., & Gomez, P. (2016). Does location uncertainty in letter position coding emerge because of literacy training? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 996-1001.
- Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, 51(2), 231–246.

- Perea, M., Winkler, H., Mallouh, R. A., Barnes, L., & Gomez, P. (2015). In defense of position uncertainty: a reply to Duñabeitia, Orihuela, and Carreiras (2014). *Psychological Science*, 26(4), 545–547.
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81(2), 275–280.
- Richler, J. J., Tanaka, J. W., Brown, D. D., & Gauthier, I. (2008). Why does selective attention to parts fail in face processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1356.
- Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: effects of repeated and transposed letters. *Language and Cognitive Processes*, 19(3), 333–367.
- Suchow, J. W., & Pelli, D. G. (2013). Learning to detect and combine the features of an object. *Proceedings of the National Academy of Sciences of the United States of America*, 110(2), 785–90.
- Szwed, M., Ventura, P., Querido, L., Cohen, L., & Dehaene, S. (2012). Reading acquisition enhances an early visual process of contour integration. *Developmental Science*, 15(1), 139–49.
- Ventura, P., Fernandes, T., Cohen, L., Morais, J., Kolinsky, R., & Dehaene, S. (2013). Literacy acquisition reduces the influence of automatic holistic processing of faces and houses. *Neuroscience Letters*, 554, 105–9.
- Ventura, P., Pattamadilok, C., Fernandes, T., Klein, O., Morais, J., & Kolinsky, R. (2008). Schooling in western culture promotes context-free processing. *Journal of Experimental Child Psychology*, 100(2), 79–88.
- Wechsler, D. (2008). Wechsler adult intelligence scale—Fourth Edition (WAIS—IV). San Antonio, TX: NCS Pearson.
- Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, 1(1), 59–85.
- Wong, A. C. N., Bukach, C. M., Yuen, C., Yang, L., Leung, S., & Greenspon, E. (2011). Holistic processing of words modulated by reading experience. *Holistic Processing of Words Modulated by Reading Experience*, 6(6), e20753.

### **AUTHOR CONTRIBUTIONS STATEMENT**

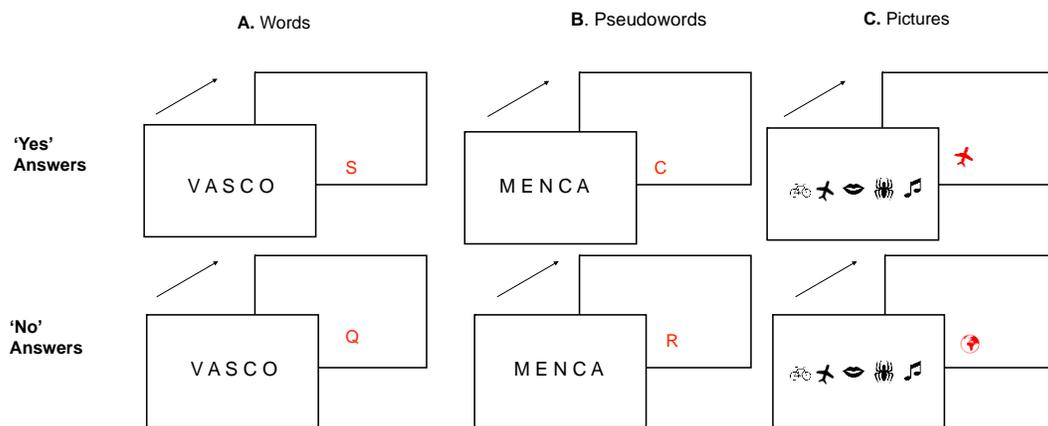
J.A.D. and M.C. designed the study. K.O. collected the data, and S.M-M. and J.A.D. completed the data analysis and drafted the manuscript. All authors agreed on the content prior to publication.

### **COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

## FIGURES

**Figure 1.** Experimental design. In separate blocks, the participant could encounter three different types of references – words (A), pseudowords (B) and picture (C) strings – that would remain on the screen for 1000ms, followed by the target word or pictogram, which remained on the screen until the participant responded whether the target was present (‘Yes’ answers) or not (‘No’ answers) in the reference string, or until 3000ms if no response was produced.



## TABLES

**Table 1.** Profile of the schooled literate and illiterate participants. Standard deviations are provided within parentheses. Age corresponds to the chronological age in years. SES represents the mean score in the National tests for socio-economic status. Working Memory was obtained by using the Digits subtest from the WAIS. The score of the IQ test corresponds to the centile in the non-verbal subtest of the K-BIT. Reading accuracy represents the percentage of correctly read pseudowords from the PROLEC-R test.

	Age	SES	Working Memory	IQ	Reading Accuracy
<i>Literates</i>	39.04 (9.94)	29.77 (13.88)	7.23 (3.21)	73.62 (11.48)	95.29 (5.89)
<i>Illiterates</i>	40.19 (10.57)	29.77 (13.88)	5.88 (3.03)	69.81 (8.86)	0 (0)

**Table 2.** Reaction times (in ms), accuracy rates (in percentage of errors) and sensitivity rates ( $d'$  scores) for the schooled literate and illiterate groups as a function of the type of response and the type of string, with standard deviations presented within parentheses.

	WORDS		PSEUDOWORDS		PICTURES	
	YES	NO	YES	NO	YES	NO
<b><i>Reaction Times</i></b>						
<i>Literates</i>	953 (162)	1094 (189)	954 (158)	1077 (143)	1028 (163)	1223 (197)
<i>Illiterates</i>	963 (230)	1070 (381)	957 (241)	1137 (368)	1020 (303)	1159 (409)
<b><i>Error Rates</i></b>						
<i>Literates</i>	9.83 (6.73)	9.61 (14.64)	11.73 (10.57)	13.12 (15.15)	29.23 (13.55)	40.31 (19.13)
<i>Illiterates</i>	32.39 (16.46)	51.63 (21.02)	31.60 (17.02)	53.86 (20.45)	31.15 (16.71)	52.68 (22.14)
<b><i>d'</i></b>						
<i>Literates</i>	3.09 (1.18)		2.70 (1.17)		.85 (.49)	
<i>Illiterates</i>	.45 (.50)		.41 (.44)		.46 (.48)	