Response to Reviewer Miguel Lázaro

I appreciate the opportunity to read this article. Once I read the title I was interested. I expected a theoretical manuscript in which different models of orthographic processing are depicted and some kind of reflections for educational purposes are proposed based on this theoretical background. It was fine as I think that educators and clinicians would benefit from this kind of job. However, this is only partly what the manuscript offers and, I have to admit, I am a little bit confused about the text. I do not exactly follow the rationale and therefore the goal of the paper.

This is an invited review paper for the Special Issue of "Psicológica" to celebrate the "25 anniversary of the Spanish Society of Experimental Psychology". It was our mistake not to stress that our focus was on computational models of visual-word recognition and we should have been more explicit on what this paper adds.

As such, we have prepared a review of the field of word recognition in which we are most familiar: (1) computational modelling and its constraining empirical evidence (including the current challenges); (2) educational implications; (3) methodological innovations. These are not intended to be related (i.e., computational models per se cannot be easily extended for educational purposes; instead, the subsection on the models served as an introduction to the current challenges faced by these models).

There are excellent edited books (and reviews) on all topics of cognitive psychology, and the field of visual-word recognition is no exception. Here our aim in the initial section was not to "try to reinvent the wheel" with just another review paper, Instead, our goal was to offer a brief (admittedly subjective) temporal trajectory of computational models of visual-word recognition, together with some of the current challenges (in particular as for the front-end of the models). We have now expanded some of the paragraphs so that there is a more comprehensive trajectory (in particular in the groundbreaking influence of the interactive activation model and its successors).

At the end, the choices of a review are multiple and depend on the preferences of the writers of the paper—this is a bit like the work of a coach in a football team (for sure, other players would do a good job in a match).

This is now the initial "outlining" paragraph, which we hope it is clearer. We have also expanded on the historical overview of the computational models, including the pioneering proposals (never implemented) made by David Rumelhart in the 70s of the past century, including from visual feature layers to syntactic and semantic layers. We have now provided more details on the interactive activation model and its successors.

A brief historical overview of computational models of visual-word recognition

The basis of the first mainstream of computational models of letter and visual-word recognition originated in the late 50s and 60s of the past century (e.g., letter recognition: pandemonium model, Selfridge, 1959; visual-word recognition: logogen model, Morton, 1969). In the pandemonium model, the recognition of letters was accomplished by a hierarchy of parallel, specialized units—the so-called "demons", each of which extracts a different feature of the letter stimulus. In the logogen model, the recognition of words is achieved through competition of lexical units—the "logogens", which are activated by the visual input and the one that reaches the threshold level of activation represents the identified word.

In the decade of the 70s, in an influential paper, Rumelhart (1977) described the layers of future computational models of visual-word recognition and reading: letter level, letter cluster level, lexical (word) level, syntactic level, and semantic level. The following groundbreaking step was the implementation of the first computational models of visual-word recognition (localist models): the interactive-activation model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982) and the activation-verification model (Paap et al., 1982), both having three layers of units: a visual letter feature level, a letter level, and an orthographic word level. While these two computational models were less ambitious than in the initial proposal by Rumelhart (1977), the implementation of layers for syntax and semantics would have been a Herculean task—and even today. The interactive-activation model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1981), in which excitatory and inhibitory connections operate across and within layers, highlights the importance of interactivity (see Carreiras et al., 2013, for review). When a printed word is presented, the model generates activation at the letter feature level, which in turn activates matching units at the letter and word levels—note that the units at the word level compete with each other (e.g., lexical unit for *mouse* would inhibit the lexical unit for *moose*). The interactive activation model is particularly effective in capturing benchmark effects of word context on letter perception, such as higher levels of letter activation for letters in orthographically legal words compared to orthographically illegal words.

The interactive-activation model was subsequently, in the spirit of nested modelling, at the core of more sophisticated models of visual-word recognition. First, the multiple-read out model (Grainger & Jacobs, 1996) extended the model to the lexical decision task. This model could make "yes" lexical decision responses when the activity of single word units reached a given threshold (the so-called "M" criterion) or when the overall degree of activation in the word layer reached a given threshold (the so-called "S" criterion)—this could explain why words with many orthographic "neighbors" (e.g., *blank*: *bland*, *blink*, *black*, *flank*, among others) produce faster lexical decision times than words with few orthographic "neighbors" (e.g., *harsh*). Furthermore, the multiple read-out model could also respond "no" to pseudowords, via a temporal deadline that was modulated on the degree of activation in the word level, thus capturing the phenomenon that lexical decision times are shorter and more error-prone for pseudowords with few orthographic "neighbors" (e.g., *cliud* [*cloud*]) than with (e.g., *blund* [*bound*, *bland*, *blend*, *blind*, *bold*, *blued*, *blunt*]). Jacobs et al. (1998) further expanded the multiple read-out model by adding a layer of sublexical phonological units (the so-called MROM-p) to the layer of sublexical orthographic units to capture phonological effects (e.g., the pseudohomophone *feal* [/fiːl/, as the word *feel*] producing longer lexical decision times than an orthographic control). In addition, Conrad et al. (2009) expanded the multiple-read out model in Spanish and German by adding an intermediate layer with the word's initial syllable between the letter level and the word level, thereby capturing the effects of syllable frequency in the lexical decision task (i.e., slower lexical decision times for those words with a frequent initial syllable). Second, the dual-route cascaded [DRC] model (Coltheart et al., 2001) extended the interactive-activation model not only to the lexical decision task—in a roughly similar manner as the multiple-read out model—but also to reading aloud tasks, thus providing a more explicit account of phonological processing rather than relying exclusively on orthographic units (see Frost, 1998, for a review of early research on phonological processing). The "lexical" route in the DRC model was composed essentially of the interactive activation model, and the "sublexical" route included a grapheme-to-phoneme rule system. Third, the Bilingual Interactive Activation model (Dijkstra et al., 1998) extended the interactive-activation model with two layers of words (i.e., one for each language) and an extra layer corresponding to the language nodes—this layer is connected to the two layers of word units (see van Heuven & Dijkstra, 2009, for an extension of this model [BIA+] including a more precise account of phonology and semantics).

The next part of the manuscript is devoted to the experimental evidence that should support educational praxis. The crowding effect, the potential role of colors… are relevant aspects as they are many other issues. To my understanding, this is again a very shallow approach to the issue. I really do not thing educators will learn much about how to proceed with struggled children based on the paper. I believe the authors are really experts not only in the theoretical part of the manuscript but also in this applicated one, so I can not understand why the paper does not go much deeper to offer useful information to professional.

We agree with the Reviewer that these two paragraphs could be expanded. First, on the use of colors when learning an unspaced writing system, the paragraph was too concise. We have expanded that paragraph, also with examples, and we believe that it will be much easier to read:

Interestingly, research on visual word recognition also provided some ideas to enhance learning to read. For instance, Perea and Wang (2017) proposed an innovative method to learn Chinese that can be extended to other writing systems that do not employ interword spaces: colors. The logic was that, at the early stages of learning to read in unspaced writing systems, color information provides a useful visual cue to help to segment the words (e.g., 大象打算在森林开一家商店 [The elephant plans to open a store in the forest]), facilitating the reading process. Perea and Wang (2017) found that alternating colors across words in Chinese facilitated the process of word identification for young readers—they also found a parallel advantage for adult readers when the text contained unfamiliar words. Subsequent research has generalized this finding to adult learners of Chinese as L2 (see Zhou et al., 2020). In a similar vein, Pan et al. (2021) showed that, in Chinese children, the benefit of the sentences with alternating colors decreased as a function of Grade (i.e., a strong benefit in Grades 2 and 3, but not on Grades 4 and 5). Furthermore, alternating the colors across words in Chinese may help eye guidance during reading (i.e., location closer to the optimal viewing position; see Zhou et al., 2018). Thus, using colors to separate words could be helpful for children or adult individuals who are learning to read and write in unspaced writing system (e.g., Chinese, Japanese, Thai, Javanese, among others).

As for the paragraph on interletter spacing (following the suggestions from the other Reviewer), we have also expanded it (as shown below, the evidence is not fully conclusive, so we prefer to be cautious here):

Another avenue in which research of visual-word recognition has an educational side is designing fonts to help special populations when reading. For instance, a number of studies highlighted the need for dyslexic-friendly fonts to facilitate the word processing in dyslexic populations (see Bachmann & Mengheri, 2018; Marinus et al., 2016; Perea et al, 2012; Zorzi et al., 2012; Benmarrakchi & El Kafi, 2021). Generally, these studies showed that reading performances for individuals with reading impairments decline when letters (and words) are presented closely together or when the font has a difficult design. Thus, setting inter-letter spacing and using a simple design would improve reading performance in individuals with dyslexia. Note, however, that the empirical evidence is not particularly conclusive (see Slattery et al., 2016, for a cautionary note). (…) In a recent study on eye movements during reading, Łuniewska and colleagues (2022) found no significant impact of inter-letter spacing on reading speed or comprehension in readers with dyslexia, a result consistent with Hakvoort et al.'s (2017) earlier findings. However, it is possible that increased inter-letter spacing only benefits a subset of individuals with dyslexia who are particularly susceptible to visual crowding, as suggested by Joo et al. (2017). More multi-laboratory research is needed to settle the role of crowding and inter-letter (or inter-word) spacing during reading.

Lastly, there are some reflections about the linear mixed models, and I can subscribe all of them,but I do not understand why it is there.

Our understanding is that the Reviewer was expecting something else, such as discussing the educational implications on the basis of the computational models; we are not sure one can easily do that, perhaps with “verbal” models, but our focus was on computational models and an overview of the field as a whole.

Let us reiterate that the take-home message in the current paper is that the field of visual-word recognition has: (1) a long tradition of computational models (while facing some challenges, those stated in the manuscript and probably many more), (2) a long tradition of transfer knowledge to nearby areas (i.e., applied implications), and (3) a long tradition of innovative methodological approaches that have later adopted in other fields. Thus, we believe that three subsections of the paper are important—of course, each reader may have her/his preferred subsection. In the end, the idea is that more and more readers are attracted to the field—to that end, they have tens of useful references in the present paper that can help them find lags in knowledge at theoretical, educational, and methodological levels.

In sum, we hope that the Reviewer will find this revised version publishable in Psicológica. We have answered the two specific comments made by the Reviewer on the applied side (interletter spacing, color alternation), and we have provided some background and rationale for this paper.