

Presentation Special Issue: Cognitive Science

A special issue in cognitive science and the implementation of a comprehensive reviewing system

FERNANDO MARMOLEJO-RAMOS*

University of Adelaide

Introduction

Cognitive science can broadly be defined as the interdisciplinary study of the mental operations displayed by a cognizer during its interactions with the internal and the surrounding environment (Marmolejo-Ramos, 2008; von Eckardt, 2006). The combination of explanations and methodological approaches offered by fields such as Psychology, Artificial Intelligence, Neuroscience, Linguistics, Anthropology, and Philosophy is one of the distinctive characteristics of cognitive science (Miller, 2003; Nickerson, 2006). Although cognitive science is part of the research agenda in various countries, there is evidence suggesting that this is not the case in Latin-America (Marmolejo-Ramos, 2008).

This special issue brings together the work that researchers around the world are currently carrying out on diverse topics in cognitive science and presents it to the research community in Latin America. The purpose behind this special issue is to motivate researchers on the continent to continue the studies presented herein and, ideally through networking with international researchers to initiate a rigorous research agenda in specific topics in cognitive sci-

ence. This special issue offers a qualitatively comprehensive reviewing system to qualitatively, and a quantitatively assess of the manuscripts submitted. The following sections consider these two aspects.

Firstly, the articles that compose the special issue will be briefly introduced so that the reader has a general idea of the topics and methodologies that characterize current cutting-edge research in cognitive science. Secondly, the specificities of the reviewing system used will be presented along with the results obtained for the current special issue. Finally, some suggestions for the potential implementation of a comprehensive reviewing system in the publication of high-quality research in cognitive science in Latin America will be discussed.

Current topics and methodologies in cognitive science research

The current areas being addressed by cognitive scientists include memory, attention, vision, emotions, language, neuroscience and mathematical psychology. A collection of papers published in this special issue presents empirical evidence in relation to specific topics in each of these areas.

It is well known that text comprehension requires cognitive processes such as inferences, simulation, and memory; there is research in relation to each of these processes. In relation to *memory* processes, it has been suggested that episodic (E) and semantic (S) memory need the operations of

* Fernando Marmolejo-Ramos, School of Psychology, Faculty of Health Sciences, University of Adelaide, Australia.

Correspondence concerning this article should be addressed to: School of Psychology, Faculty of Health Sciences, University of Adelaide, Adelaide, South Australia, Australia, 5005. Email: fernando.marmolejoramos@adelaide.edu.au, web page: <https://sites.google.com/site/fernandomarmolejoramos/>

the working memory (WM) in order to momentarily maintain active pieces of incoming information so that they can be linked to E and S memory systems. Elosúa et al. present a body of empirical evidence that suggests that a specific training in the executive functions of WM enhances reading comprehension in primary school children.

In broad terms, attention can be defined as the activity of focusing on one stimulus (or set of stimuli) while ignoring another (or another set of stimuli). For instance, participants can be asked to perform two sequential tasks and provide differential responses to each of them. In this special issue, Ruiz Fernández et al. report a study aimed at determining if the duration of response execution influences the processing of task order. The results indicate that participants optimize response scheduling in the case of dual-task situations.

Humans explore the environment via perceptual and motor senses such as touch, hearing, taste, smell, and sight. In particular, *vision* is one of the systems that has received considerable investigation due to its biological complexity and great effect on low- and high-order cognitive processes. Engmann and Cousineau show that stimuli with three redundant attributes in the visual modality (e.g., colour, form, and direction of movement) accelerate object recognition times. This is the first study of this kind to test how redundancy effect is increased (i.e. faster recognition times) due to more than two attributes that belong to the same modality. One of the core functions of the visual system is the use of gaze for the localization of objects. In two experiments, Yamada and Kawabe demonstrate that other person's gaze can affect the localization of objects and that attention deviates towards the gazed object or location. This is the first time that it has been suggested that mis-localisation of target objects is not only affected by distracter objects, but also by gaze cueing.

One of the greatest developments in cognitive science research is the recognition of *emotions* as a necessary component in human cognition. The recently formulated embodied cognition theory, argues that in metaphorical contexts emotionally valenced concepts are associated with sensorimo-

tor spatial representations. For example, the word "happy" is associated with the spatial location "up" and with body movements that match that physical direction. Ansorge and Bohner used an *Implicit Association Task* to demonstrate that positively laden concepts are semantically associated with "up" spatial locations, whereas negatively laden concepts are associated with "down" spatial locations. Most researchers in the embodied cognition framework argue that perceptual and motor systems are always activated during the processing of emotion concepts. However, other researchers argue that motor systems are not always activated during the processing of emotionally valenced stimuli and that perceptual systems can take sole control of the process. In a full-length article, Marmolejo-Ramos and Dunn report empirical evidence in this regard.

Research into *language* has been traditional in cognitive science for many years. However, most studies in psycholinguistics have focused on the English language in order to build theories and put forward claims concerning the interactions between language and cognition. In this issue, Mishra uses an eye-tracking method to demonstrate that children and adult speakers of Hindi (a free word order and verb final language) exhibit different shifts in visual attention during sentence generation. His results specifically indicate that while adults attend primarily to the verbs in a sentence, children attended mostly to the subjects and objects in a sentence. Various methods have been crafted to test the idea that language is embodied, given that its processing requires the activation of perceptual and motor brain systems. Milin and Zdravković examine a classical technique called semantic differential (SD) that measures commonly understood meanings that words carry. The authors propose a methodologically modified version of the SD technique to show how the abstract and concrete concepts have perceptual and motor properties that can be represented in the semantic space.

Behavioural studies are the trademark of the experimental approach used by the researchers mentioned so far. However, other researchers resort to brain imaging techniques to address their questions of interest. Thus, Kondo and Watanabe

use a more recent technique in neuroscience called magnetoencephalography (MEG) to demonstrate that brain activity reflects a correlation between the subjective perception of brightness and luminance processing in the human visual system. Likewise, Tagai et al. use a method known as near infra-red spectroscopy (NIRS) to show how Japanese and non-Japanese participants exhibit different brain activity in relation to stimuli portraying politeness.

Finally, another approach requiring further examination for greater understanding of human behaviour is that of computational modelling. This method in turn relies on mathematical and statistical formulas in order to propose law-like rules that represent cognitive, perceptual, and motor processes and that can be verified quantitatively. Ultimately, the computational modelling of psychological processes via mathematical/statistical concepts and language is known as *mathematical psychology*. In this special issue, Gómez and Erber implement a computational model that tests a recent claim in social psychology regarding reciprocity in romantic desire. Their results not only provide a parsimonious explanation but a challenge to current research carried out in social psychology. Even the most well-designed behavioural or neuroscientific experiment or the most elegant computational model can go astray if the results obtained are not analysed properly. Scientists working in cognitive science, and many other fields, are based on statistical methods in order to make sense of the data obtained to render them into reportable results. Thus, Wilcox et al. propose two new statistical methods for pair-wise comparisons of data sets that are small and that resemble data obtained via Likert scales.

A comprehensive reviewing system for the publication of cognitive science research in Latin-America

As it is a current practice in all scientific journals, a peer-review system was used for the assessment of the manuscripts submitted to this special issue. In the traditional peer-review system, the guest editor assigns two to three reviewers whose expertise

matches the topic of the paper. In some cases, the authors can also submit a list of suggested reviewers and even reviewers who are opposed (for example, journals such as the Journal of Experimental Psychology: General and PLoS ONE welcome a list of (non) suggested reviewers). For this special issue, authors were asked to submit a list of four suggested reviewers; two of them were randomly chosen and asked to be official reviewers of the paper. Also, following the traditional reviewing system, both the reviewers' and the authors' identities were masked in order to minimise bias in the reviews. Additionally, authors were requested not to reveal in any form their identity in the paper.

It is also traditional that reviewers receive a template with which to assess the paper and perform a qualitative assessment of different sections of it, e.g., its abstract, introduction, etc. A standardised reviewing template (SRT) that combines a qualitative and a quantitative assessment of the papers was designed for this special issue. The quantitative assessment required that specific sections of the paper were evaluated on a Likert scale ranging from 1 (low quality) to 7 (high quality). Thus, the reviewer had to provide a quantitative assessment that reflected the qualitative appraisal given to specific sections of the paper. The sections of the paper evaluated were: *i*) title, *ii*) abstract, *iii*) keywords, *iv*) introduction, *v*) method, *vi*) results, *vii*) discussion, *viii*) conclusions, *ix*) and references. Reviewers were also asked to provide comments and ratings for the paper's *x*) originality and *xi*) relevance, and to give an *xii*) overall qualitative and quantitative appraisal of the paper.

Furthermore, reviewers could add additional general or specific comments. For each section reviewers were given specific aspects to consider as part of the evaluation based on the latest APA guidelines. A final average was computed using the mean score obtained across all twelve sections being assessed in order to decide whether the paper would be rejected, accepted with major modifications, or accepted with minor modifications (this point will be discussed later). Figure 1 illustrates sections of the SRT used for the evaluation of papers submitted to this special issue.

The results of the first reviewing round were sent to the authors so that they could address the comments raised by the reviewers. Once the authors returned their amended articles, these were sent back to their reviewers for a second reviewing round. Reviewers sent a modified SRT (SRTm) in order to provide a final assessment of the manuscript being evaluated. The SRTm required the reviewer to provide a qualitative and quantitative assessment of the revised manuscript taking into account the completeness of the amendments undertaken, the overall improvement, readability, scientific contribution, reception by readers and overall likelihood of publication. The results of the evaluation determined whether an article would be published in its revised form, published, with amendments or not published (see Figure 2). Decisions for each category were based on the SRT parameters (see details in later sections).

Validation of the reviewing template used for this special issue

Thirty one researchers, including a number who were neither reviewers nor authors in this special issue, assessed the standardised reviewing template (SRT) based on the instruction in Figure 1.

The aspects evaluated were the reviewing template's presentation, informativeness, clarity, comprehensiveness, and usability. In addition, researchers were asked to provide an overall appraisal of the SRT. Researchers were also asked to provide qualitative and quantitative assessments of each of these components, and were given the option to provide specific and general comments on the template. Table 1 presents the results of the assessment on the reviewing template.

The grand mean rating indicates that, among all things considered, the template is a rather useful tool for the reviewing of articles. Indeed, a robust estimator of central tendency suggests an even more promising picture, i.e., the grand median and its deviation (MAD) were 6 and 1.48 respectively. Thus, it seems that the SRT can be considered as a "highly complete" tool to be used for the qualitative and quantitative assessment of scientific papers.

However, the researchers provided comments on certain aspects of the SRT that need to be revised and that, if addressed, will enhance its effectiveness. The following list presents potential solutions to the current drawbacks of the SRT:

- That certain aspects of the article that are not totally decisive should be merged under other categories. For instance, "keywords" and "abstract" can be merged into one category.
- That a robust estimator is used to compute a final average along with its 95% confidence intervals. The use of a robust estimator (e.g. the median) ensures an unbiased average and the confidence intervals (e.g. non-parametric bootstrap confidence intervals) can be instrumental in making an educated-decision as to the fate of the article.
- That reviewers are requested to use decimal places in the Likert scale. Working with decimal places leads to more accurate results and enhances statistical computation of averages and confidence intervals. Indeed, if the SRT is presented as an on-line form, it would be ideal that the rating be performed via a slider rating scale that uses decimal places and that is anchored by bipolar opposite labels.
- That the SRT is used as a document that can be completed on-line, that has spelling capabilities, that can be opened using commercial web browsers (e.g, Firefox, Google chrome, etc.), and that works in different operating systems (Macintosh, Windows, etc.). By doing so, word-processing or spreadsheet applications that might cause technical problems could be bypassed (a similar approach is already being implemented in the online and interactive reviewing system used by the *Frontiers* journals).
- That other aspects that account for articles other than empirical (e.g. methodological articles, narrative review articles, and quantitative review articles) be included. For instance, the aspect of "accuracy/completeness of statistical/mathematical equations" could be added in the case of methodological articles; the "quality of tables and figures" could be used for empirical

TABLE 1

Results of the evaluation of the reviewing template used for the evaluation of the manuscripts submitted to the special issue in cognitive science for the journal *Universitas Psychologica*.

Template's aspect	Guiding question	Comments [qualitative assessment]	Mean rating (SD) [quantitative assessment]
Presentation	Is the current template complete regarding formatting, colours used, font face, font size, etc?	<p>"It is annoying to type comments into a box, within an excel sheet. It is hard to format these comments well."</p> <p>"Word wrapping is not consistently applied across all the cells; entering text into the single-line excel format is cumbersome especially line breaks etc. And where there is a lot of text the cells do not automatically resize to fit the content."</p> <p>"There really isn't sufficient space, and typing lots in a single cell is difficult."</p> <p>"There is no place to provide comments about format, readability, and completeness of tables and figures."</p> <p>"In general, I think that it is ok. However, I changed the font in my own copy since in the one I was sent, the letters appeared blurred."</p> <p>"I find the formatting uncondusive to easy work. The fonts change all the time (e.g. the letter size in sheet two and 3). Copying my substantive comments into the document ruins the formatting. Having to change something in an embedded piece of word document is especially laborious."</p> <p>"Font size for the comments is too small and I found no way of increasing the window's size to make it more comfortable to work on."</p> <p>"There is no problem on colours. However, at least in my environment, the texts are not easy to read. Text font is possibly not widely used. I hope that the font is Arial, Times New Roman, or Helvetica. In my sense, the font difficult to read occurs in pivot tables from a certain word file?"</p> <p>"I'm not quite sure what you mean by 'complete' (useful?), but some of the formatting makes it hard to read. I would also like to be able to make paragraphs in the comments boxes and got confused when it suddenly jumped from Excel to Word. I do use MS Office (for Mac), but would it be useable to people who do not have access to Office, or who choose not to use it, and can your guarantee that it works for all versions, operating systems and screen sizes?"</p>	5.27 (1.53)

Template's aspect	Guiding question	Comments [qualitative assessment]	Mean rating (SD) [quantitative assessment]
<i>Informativeness</i>	Is the template complete regarding the necessary information provided for an article's evaluation?	<p>"The 'methods' question could be broken up into design, procedure, clarity of communicating design & procedure."</p> <p>"I will rate and comment on whichever aspects you wish to hear about - perhaps add relevance to readership?"</p> <p>"No place to comment specifically about accuracy/ completeness of statistical equations."</p> <p>"I don't know what your exact goal with the template is. I would pay attention to the quality of writing. There was no information about the figures and tables. I don't think that the keywords can contribute equally to the overall grade as the other parts."</p> <p>"Too much attention is paid to details (e.g. are the keywords relevant?) and too little to the core quality of the paper: Aspects such as thoroughness, scholarship, quality and importance for the field."</p> <p>"It covers the necessary aspects of such an evaluation. However, I doubt whether all the different aspects (e.g., quality of abstract or key words and methods) should be influence the overall rating in a similar, non-weighted manner. I do not know how the single scores add into the overall score (whether they are weighted), but I would definitely suggest weighting the different single scores. For instance, the quality of the title is of much lesser importance than the quality of the method's section. I checked the formula for the cell with the average score and it seems as if really of the single scores are only averaged over. I think the single scores should be weighted (much smaller weight to title, abstract, key words, and references)."</p>	5.75 (1.67)
<i>Clarity</i>	Is the template completely clear regarding the presentation of instructions?	<p>"I think that the form is clear. I only have some problems with the three categories (reject, major revision and accept). You assign those categories boundary values, but when the mean score is for example 3.05, you don't have a category."</p> <p>"Not really clear. Personally, I would completely lose the excel multiple sheet format. It is not very pleasant to work in."</p> <p>"The example instructions referred to ratings with decimal points, but I seemed only to enter absolute numbers (which is probably sufficient, but I found it slightly confusing)"</p>	6.26 (0.9)

Template's aspect	Guiding question	Comments [qualitative assessment]	Mean rating (SD) [quantitative assessment]
<i>Comprehensiveness</i>	How complete is the template regarding the aspects evaluated in an article?	<p>“Perhaps overly so - some elements in the template seem to have little to do with the decision to accept, reject, etc. (for example the title - I may say ‘I strongly feel that the title needs to be changed’ but this does not affect my overall recommendation”</p> <p>“It seems strange to me that maximum points of all sections are same. For example, it might be better to place more importance on the originality than the keywords.”</p> <p>“I think that the form is complete (main aspects are covered). However, when you use a mean to evaluate an article, you weight for example the keywords equally with the results section. This would give to much weight to the keywords in my opinion. Furthermore, means might be more influenced by possible outliers than medians. Especially when combining different raters, it might be good to use a median instead of the mean.”</p> <p>“Given that all of the reviewers are most likely professional scientists, probably we don't need much explanation”</p> <p>“The template is comprehensive, but some reviewers might like a bit of flexibility to respond upon particular content in a manuscript.”</p> <p>“Not really comprehensive. I am missing a lot of more important features: Scholarship, soundness, quality, importance”</p> <p>“I think there should be a section about ‘novelty’ of the article”</p> <p>“If anything, I think it's too complete: I didn't know what was expected in the ‘relevance’ category at all; and while it was good to be able to rate things like key words and title, those ratings should not go into the main average with the same weighting as introduction and results, which are much more critical, and much harder to change. Maybe a further category to include would be Tables and Figures.”</p>	5.71 (1.5)

Template's aspect	Guiding question	Comments [qualitative assessment]	Mean rating (SD) [quantitative assessment]
<i>Usability</i>	How complete is the level of usability of the template?	<p>“Typing into excel makes this very user-unfriendly.”</p> <p>“I found it very difficult, I need to adjust my reviewing practices substantially to complete the reviewing format. In a usual review I state my most important concerns (or positive statements) first, in what I see as descending order of importance. This section includes everything that, in my opinion is my basis for recommending ‘accept’, ‘reject’ or choices in between. Then afterwards, a list of less significant points, issues or questions that I feel should be dealt with but are not as important. In fact I wrote out my main points in this format before transferring them into the reviewing section by section. As a result this removes my personal weighting of the issues that are more or less relevant (for example I would never reject a paper for lacking keywords)”</p> <p>“Difficult to enter information into the form. Couldn't enter title, etc. in the space provided. Hard to edit comments”</p> <p>“For me, it was terrible, but it is because I don't have ms excel installed in my computers”</p> <p>“I would say the template is not very usable. Having to write in excel is uncomfortable. Many normal formatting commands don't work properly, and it is hard to keep an overview. Moreover, some of the phrases are ambiguous.”</p> <p>“In this review, I am so content with this area of the space. However, when we review research papers, some reviewers may need plenty of room for their comments. So I think it is better that a more wide space is prepared or the space is changeable.”</p>	5.43 (1.92)

Template's aspect	Guiding question	Comments [qualitative assessment]	Mean rating (SD) [quantitative assessment]
<i>Overall appraisal</i>	All in all, how would you rate the overall completeness of the template?	<p>“Complete but hard to use, increased my reviewing time by approximately one hour”</p> <p>“I prefer other formats, in which I can submit a pdf with responses or online submission, to this excel spreadsheet.”</p> <p>“The format lacks spell-checking capabilities.”</p> <p>“Doing it in a excel spreadsheet was really cumbersome when more than one sentence was needed in the comments section”</p> <p>“Although I support the underlying idea of structuring the reviewing process, I do not think the template is a success. I assume ‘completeness’ here means ‘quality’? In any case, I would rather not work with the template in its current form. I don’t think the categories reflect what is important in a paper, and I don’t really like working in an excel sheet. An important flaw is the underlying assumption that every aspect on the evaluation sheet is weighed equally in the calculation of the rating, and that this average reflects the quality of the paper. For example, I fail to see how the relevance of the keywords could ever be equally important as anything else.”</p> <p>“Overall, although some points have to be improved, almost all of this system satisfied me.”</p>	5.48 (1.42)
<i>Grand mean rating (SD) ^a =</i>			5.65 (1.53)
Specific and general comments added based on the qualitative and quantitative assessments of the template			
<p>“I have to say that I found typing review comments into the template quite annoying. Formatting my comments into separate paragraphs was impossible, and this format also encourages exceptionally short reviews which may not be so useful to authors.”</p> <p>“Using the numeric scales to rate each individual section gives a possibly incorrect assessment of my overall judgment of the work. For example in the ms I just reviewed there are many positive aspects to the method and results sections including sufficient descriptions of items selected, procedures etc. So I could not justify giving the minimum scores on these sections, even though I feel that there are deeply fundamental flaws related to the assumptions and approach taken, which would probably lead me to make a recommendation of ‘reject’ under more typically used reviewing process. Also I feel this work is highly original and possibly relevant, although this depends again on whether the authors can overcome the significant flaws.”</p> <p>“the reviewing time should be extended to one month (not 15 days as it is now)”</p> <p>“Apart from minor points that should be considered as augmentation, the template is an excellent tool for making reviewing efficient.”</p> <p>“I think the rating on the evaluation sheet is not optimal. The ‘grades’ that are given for each section are given the same weight. This does not seem appropriate to me. For example, I think that the keywords are less important than the methods section. Personally, I prefer to write my own review without a rating form such as this one.”</p> <p>“I did find it very helpful to structure my review in this way and differentiate between different aspects. Good tool!”</p> <p>“The weighting could be more elaborated because the score for originality, relevance, overall appraisal should be weighted more than those of title, keyword, and references. How to assign weights would need some discussions, though.”</p> <p>“Maybe you could also include one question concerning the quality of Figures. But this is a minor suggestion.”</p>			

Note. The ratings were performed using a Likert scale ranging from 1 (low completeness) to 7 (high completeness).

^a The 95% bias-corrected and accelerated confidence intervals (with 2000 bootstrap replicates) were 5.41 (lower confidence interval) and 5.87 (upper confidence interval).

Source: Own work

The following SRT is designed to provide a qualitative and a quantitative assessment of the paper.

- The qualitative assessment consists of a narrative assessment of the paper regarding its structure and content. This type of assessment should be typed under the section "comments". The length of the comments depends only on the reviewer's judgment.
- The quantitative assessment consists of the rating of different aspects of the article. This type of assessment is done by providing a number based on a 7-point Likert scale (see scale below). The purpose of this rating is to provide a numeric evaluation of the section being examined and should be used to complement the qualitative assessment.

1	2	3	4	5	6	7
Low quality						High quality

***** Each section of the article should be assessed according to the APA (6th Edition) standards (e.g., Chapter 2). A guiding question is provided only as an example *****

Article section	Comments	Rating
Title [E.g., Does the title encapsulate the core idea of the article?]		
Abstract [E.g. Does the abstract summarise the paper's core idea?]		
Keywords [E.g., Do the keywords reflect the topics addressed in the paper?]		

Overall appraisal [What's the overall quality of the article?]		
Average rating of the paper		##
Based on the previous qualitative and quantitative assessments of the paper, are there other specific and general comments you'd like to add?		

Figure 1. Snapshot of some of the sections of the reviewing template (SRT) used for the evaluation of the manuscripts submitted to the special issue in cognitive science for the journal *Universitas Psychologica*.

Source: Own work

The current form is intended to give a general qualitative and a quantitative assessment of the revised version of the present article in order to determine its probability of being published.

Average rating of the paper and decision ranges ^(A).

If the overall average rating falls within the following ranges then:

1 – 3: not publish

3.1 – 5: publish if and only if further recommendations are addressed

5.1 – 7: publish in the current form

^A An average rating based on the mean is automatically computed in order to facilitate the categorisation of the paper within the ranges suggested above. **NOTE = ratings aren't limited to integers; decimal places can be used (e.g., a rating can have the value of 5.7)**

revised manuscript's aspects	Comments	Rating
Addressing comments Have the authors addressed all the comments satisfactorily? 1 2 3 4 5 6 7 not very satisfactorily very satisfactorily		
Overall improvement How much improvement does the revised version of the manuscript have? 1 2 3 4 5 6 7 not very improved very improved		
Readability How readable is the revised version of the manuscript? 1 2 3 4 5 6 7 not very readable very readable		
Scientific contribution How likely is the revised version of the manuscript to contribute to research on the topic? 1 2 3 4 5 6 7 not very likely very likely		

Readership's reception How well you consider the revised version of the manuscript will be received by its readership? 1 2 3 4 5 6 7 not very well received very well received		
Overall publication likelihood All in all, how recommendable is the publication of the revised version of the manuscript? 1 2 3 4 5 6 7 not very recommended very recommended		
Average rating of the revised manuscript		#DIV/0!

Based on the previous qualitative and quantitative assessments of the revised version of the article, are there other recommendations you'd like to add? (particularly if the average rating is ≤ 5). If you don't recommend this paper for its publication (i.e., average rating ≤ 3), please explain why.

Figure 2. Snapshot of the modified reviewing template (SRTm) used for the second reviewing round of the manuscripts submitted to the special issue in cognitive science for the journal *Universitas Psychologica*. Source: Own work

and review articles, and the “relevance to readership”, “level of scholarship”, “overall quality”, “novelty”, and “soundness” could be added for all types of articles.

- That some minor aspects of an article, e.g., “title” and “references”, be down-weighted. However, it could be argued that two approaches might help in this case: *i*) the use of robust estimators to determine the final average rating and *ii*) providing more guiding and precise questions for seemingly minor aspects of an article. For instance, “references” might seem a minor aspect if the reviewer has to assess their APA appropriateness. However, if the reviewer is questioned as to the relevance of the literature used, the number of references consulted/cited, how recent they were, etc., the relevance of the “reference” could increase.

How decisions were made as to the rejection/acceptance of manuscripts

The SRT devised for this special issue relied on the quantitative averaged results of the ratings performed on various aspects of the manuscripts to determine their fate. For this purpose, the categories “reject”, “accept with major modifications”, and “accept with minor modifications” were conceived. In order to make the reviewing process clear and transparent, both authors and reviewers had access to the SRT form.

The average ratings and the three categories were paired with three quality levels as follows: 1 – 3: low quality [reject], 3.1 – 5: average quality [accept with major modifications], and 5.1 – 7: high quality [accept with minor modifications]¹. The ranges used for the SRT aimed at giving similar probabilities to an article of being rejected

(*R*) (0.33), accepted with major (*AM*) and minor modifications (*Am*) (0.316 each). Overall, the probability of an article being accepted outweighed its chances of being rejected, 0.63 vs 0.33. The reason for this tendency towards acceptance is mainly based on the fact that experienced researchers were selectively invited to contribute to papers. Thus, it was assumed, much as priors are set in Bayesian statistics, that contributions would have a high chance of being of good quality.

The acceptance ranges of the present SRT can be altered to suit different acceptance/rejection probabilities. For instance, a range from 1-4 could represent *R*, 4.1-5.5 could represent *AM*, and 5.6-7 could represent *Am*. In this case, the probabilities of *R* would be 0.5, *AM* would be 0.233, and *Am* would be 0.233. Indeed, if some extra degree of strictness is sought in the case of acceptance, the span of the acceptance ranges could be further modified so that the range 4.1-5.9 represents *AM* and the range 6-7 represents *Am*. Under these new acceptance ranges, articles would have a 0.5 probability of being rejected, a 0.3 probability of being *AM*, and a 0.166 probability of being *Am*. Whatever the acceptance/rejection rates are, it is essential that authors are informed of the reasons leading to the rejection of a paper (see Gernet, 2008).

Although most journals report on impact factor (*IF*), not all of them report the acceptance probability (usually called the “rejection rate”, *RR*). Thus, the present SRT might assist in minimising the *RR/IF* ratio of a journal by setting explicit rejection probabilities (as discussed above) in order to prevent cases in which journals with low *IF*s have very high *RR*s (see Aarssen et al., 2008). For example, a journal whose *RR* is 0.6 and *IF* is 1.283 would be preferable to one that has the same *IF* but an *RR* of 0.8 since chances of publication would be higher in the former while the *IF* of both journals is the same. Furthermore, Aarssen et al. (2008) suggest that journals should publish their *RR*s along with their *IF*s since they can also be an indicator of the quality of the papers published and, by default, the journal’s quality. Note, however, that although relying exclusively on *IF*s has been a trend in the sciences for a long time, it has recently been demonstrated

1 In the SRTm these ranges were: 1 – 3: not publish, 3.1 – 5: publish if and only if further recommendations are addressed, and 5.1 – 7: publish in the current form. Whenever a reviewer gave an average rating between 3.1 and 5 to an article, it was sent back again to the author to address the comments. Once the author returned the reviewed article, it was sent back to the reviewer. The reviewer, in turn, was required to use the SRTm for a third assessment of the manuscript.

how misleading IFs can be (see Abin & Mabe, 2000; Mayor, 2010).

Overall quality of the articles published in the current special issue

One key advantage of the SRT proposed herein is that it requires that the qualitative and the quantitative assessments of the article are coherent. That is, a low qualitative assessment implies a low rating, whereas a high qualitative assessment requires a high rating. However, whereas the qualitative reports are prone to different interpretations, quantitative results lead to more precise conclusions. As each manuscript was evaluated by two reviewers, a two-sample permutation resample test (without replacement and using 10,000 simulations) was used to determine whether there was a significant difference between the mean ratings given by the two groups of reviewers in the first reviewing round (see chapter 3 in Chihara & Hesterberg, 2011 for a description of this type of tests).

The results showed that the observed difference between means ($Or1_{g1-g2}$) was not significant, $Or1_{g1-g2} = 0.36$, $p = 0.21$ (see Figure 3A). These results thus indicate that there was an overall agreement between reviewers on the quality of the articles. Thus, averaging the ratings given by the two groups of reviewers to each manuscript suggested that all articles were of a rather high quality, $M_{\text{first round}} = 5.25$, $95\% CI_{\text{bca}} = [4.78, 5.63]$, $SD_{\text{first round}} = 1.14$, $\text{range}_{\text{first round}} = 3-7$, $n = 28^2$. The same procedure was performed on the results of the second reviewing round. The results showed that the observed difference between means ($Or2_{g1-g2}$) was not significant, $Or2_{g1-g2} = -0.34$, $p = 0.72$ (see Figure 3B). As in the first reviewing round, these results suggest an overall agreement between reviewers

on the quality of the amended articles. An average of the ratings given by all the reviewers indicated that articles were of a high quality, $M_{\text{second round}} = 5.68$, $95\% CI_{\text{bca}} = [4.93, 6.06]$, $SD_{\text{second round}} = 1.18$, $\text{range}_{\text{second round}} = 2-7$, $n = 24$. There was an improvement in the average rating of the articles' quality from round one to round two, i.e. from 5.25 to 5.68; however, the observed difference, $O_{rr2-rr1} = 0.43$, was not statistically significant as suggested by a permutation test, $p = 0.1$ (see Figure 3C).

It should be noted that the current analysis was intended to provide a comprehensive quantification of the quality of the manuscripts during the two reviewing rounds. One advantage of the reviewing template, however, is that it requires that the reviewers provide qualitative and quantitative assessments of various sections of a manuscript. Thus, a robust correlation between both reviewers' ratings of each of the articles' sections can be expected, in order to obtain levels of agreement between reviewers. The correlation value obtained and the associated p -value can be used as indicators of the level of agreement between reviewers on an article, and the significance of such agreement, respectively. Alternatively, a permutation test can be performed between the ratings given by two reviewers to a specific article in order to determine whether reviewers have significant differences between their average scores.

Discussion and conclusions

This special issue is aimed to motivate cutting-edge research in selected topics in cognitive science in Latin America and to promote the use of a comprehensive reviewing template (labelled herein as SRT and SRTm) to construct it. Marmolejo-Ramos (2008) has already provided recommendations as to how a rigorous research agenda in cognitive science could be carried out in Latin America. However, it is traditional that the research results be published in order to make them available for replication and extension; thus, ultimately, leading to an expansion of knowledge. A very important and decisive stepping stone in this broadening of knowledge is having access to unrestricted and sound research. The

2 There were 14 manuscripts submitted upon invitation and two reviewers examined each of them. For the second reviewing round, two submissions were withdrawn after the results of the first reviewing round. Additionally, in some cases, some reviewers did not provide a quantitative assessment of the articles for the second reviewing round. Therefore, the ratings' sample size used in this analysis was larger than that one used in the subsequent analyses.

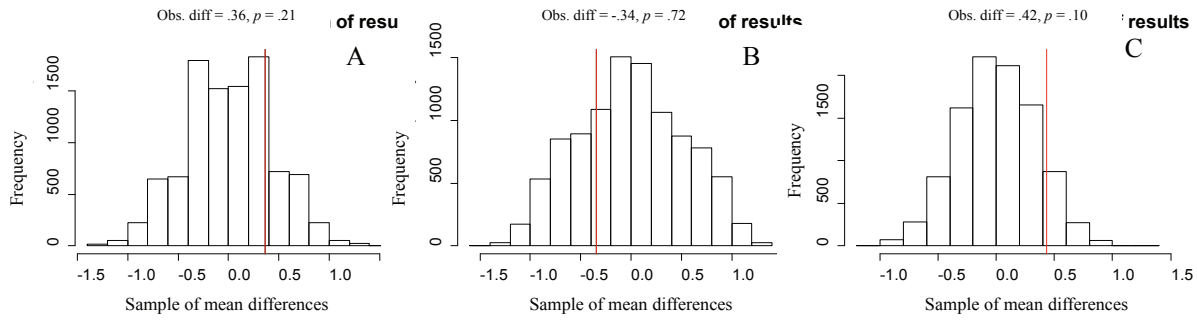


Figure 3. Histograms representing the results of the permutation test performed on the observed mean difference between groups of reviewers in reviewing round one (A) and the observed mean difference between groups of reviewers in reviewing round two (B). Figure C represents the results of the permutation test performed on the observed mean difference between all the ratings obtained in reviewing round one and all the ratings obtained in reviewing round two. A red vertical line represents the actual observed difference in each case.

Source: Own work

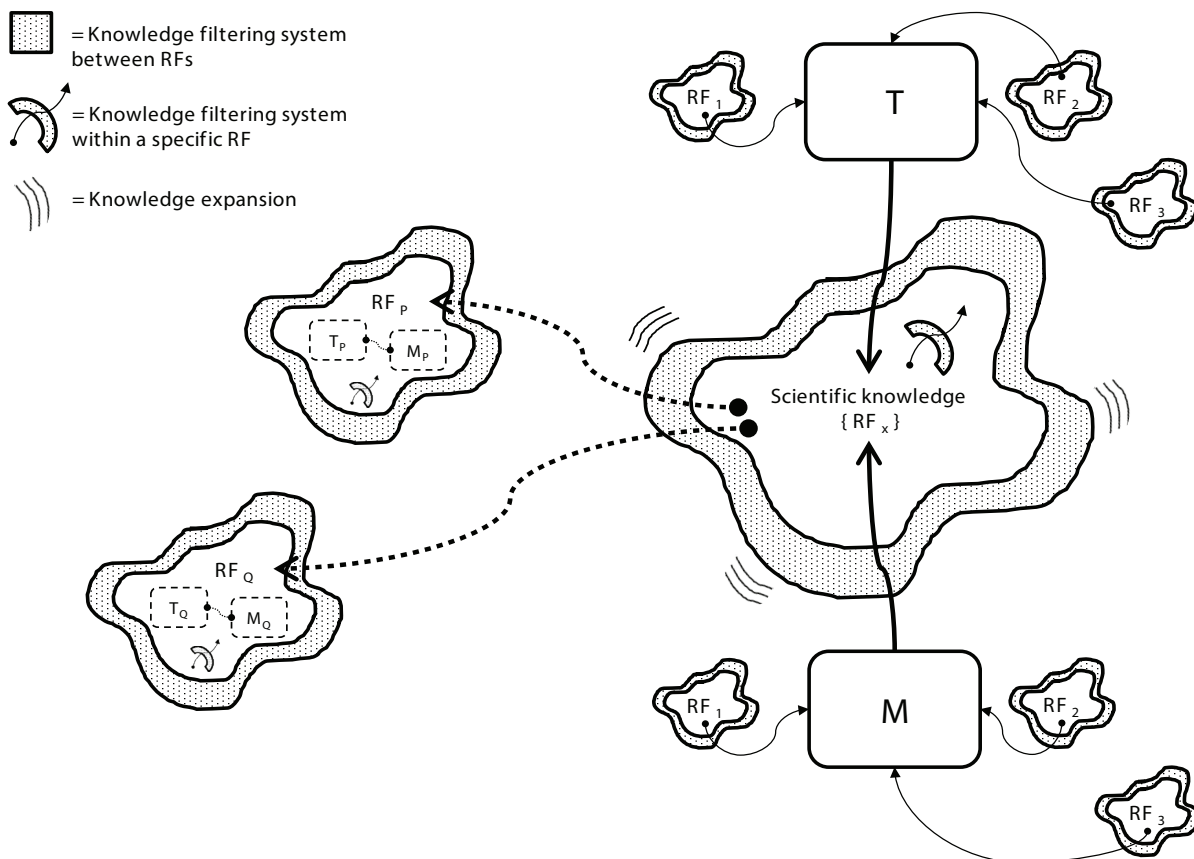


Figure 4. Illustration of the relationship between knowledge expansion in any research field and knowledge filtering systems (KFS). A transparent reviewing process, peer-reviewed peer review approaches (Wicherts et al., 2012), and comprehensive reviewing templates characterise KFS. RF = research field, T = theories and concepts, and M = statistical and research methods.

Source: Own work

term ‘unrestricted’ refers explicitly to research made public via open source journals and the term ‘sound’ refers to research that is thoroughly and transparently peer-reviewed. Next section will consider ideas regarding the role of a comprehensive reviewing system for the publication of sound research.

The standardised reviewing system (SRT) used for the construction of this special issue can be instrumental in the publication of research that must undergo a comprehensive review process. Specifically, given that the SRT proposed herein requires both a qualitative and a quantitative assessment of specific aspects of a manuscript, the implementation of a rigorous, balanced, and educated review of manuscripts is required from reviewers.

The construction and expansion of scientific knowledge is determined by the quality of research that is produced within specific research fields and the quality of research that is cross communicated between different research fields. A way to secure high-quality research is by filtering the work that is being published; the current peer-review systems aim at doing so. However, traditional reviewing systems have serious drawbacks that prevent the publication of excellent work or allow the publication of faulty research. Perhaps one of the major reasons for this is the lack of transparency used in the peer-review system (see Wicherts, Kievit, Bakker, & Borsboom, 2012). Another cause could be the templates used for the manuscripts review.

The goal of a comprehensive reviewing system is, therefore, to regulate the construction and expansion of knowledge by filtering the research that is submitted for potential publication. As Figure 5 illustrates, a knowledge filtering system (\square) should be in place in order to filter research across different research fields (RF) and within themselves (\mathfrak{S}). Such a knowledge filtering system (KFS) should be characterized by having a transparent reviewing process, by using peer-reviewed approaches (Wicherts et al., 2012), and implementing comprehensive reviewing templates. Different RFs (e.g. RF_1 , RF_2 , and RF_3) can contribute independently to the construction of theories and concepts (T) and the development of statistical and methodological approaches (M). A specific RF (RF_X) can, in turn,

adopt Ts and Ms developed by other RFs, in order to build their own Ts and Ms.

The construction of Ts and Ms within a specific RF leads to the expansion (\mathfrak{S}) of scientific knowledge within that field. More importantly, knowledge expansion in a RF has repercussions in the scientific knowledge being developed in other (emerging and traditional) RFs. For instance, the expansion in RF_X can have effects on the Ts and Ms being constructed in RF_P and RF_Q .

The KFS, illustrated in Figure 5, constitutes a comprehensive reviewing system that enables the dissemination of high-quality research, and it is the task of the journals in any field to implement it. The *Frontiers* journals are, to the best of my knowledge, the only ones implementing one of the characteristics of this system, i.e. they use a transparent reviewing system. Most journals do not provide the specificities of the reviewing process to the author and he/she rarely knows who reviewed his/her article; i.e. the reviewing process traditionally occurs “behind closed curtains” (Wicherts et al., 2012). However, in a transparent reviewing system this is not the case. Thus, it is suggested that an interactive and transparent reviewing system be put into practice.

For instance, in the first reviewing round, two to three reviewers, out of a list of nominated reviewers, could be randomly selected to review the paper. In the second reviewing round, both authors and reviewers could be put in contact to work together on the amendment of the article. Indeed, it would be ideal if the identities of both authors and reviewers were revealed from the moment a manuscript is submitted to a journal, so that transparency in the reviewing process is enhanced.

Two other desirable characteristics of a KFS are the use of a peer-reviewed peer review approach and the implementation of comprehensive reviewing templates. As originally proposed by Wicherts et al (2012), there should be put into practice a ‘peer-reviewed peer review’ approach in which the examinations of an article are further examined by another set of reviewers. Comprehensive reviewing templates should be characterised by using both qualitative and quantitative assessments of different

sections of an article in order to produce a complete evaluation. The SRT and the SRTm presented herein aim at contributing in this direction. Indeed, it is conceivable that the SRT and the SRTm be combined in order to produce a single reviewing template that can be used throughout different evaluation rounds.

It is thus the intention of this editorial to promote the adoption of a KFS, similar to the one proposed herein, by Psychology (and other scientific) journals in Latin America. As there is no journal in Latin America dedicated to publishing research in cognitive science, it would be ideal to see that this special issue motivates its gestation and that the KFS is at its core.

Acknowledgments

Two researchers reviewed each manuscript. Since 12 papers were part of this issue, 24 researchers should be acknowledged for their role as reviewers. However, one researcher requested not to have his name disclosed. The following researchers participated as reviewers in this special issue (no order implied): Jazmín Cevalco, Helge Strømsø, Iring Koch, Hartmut Leuthold, Gernot Horstmann, Sascha Topolinski, Mario Fific, Michael Zehetleitner, Yuki Yamada, Helene Kreysa, Katsumi Watanabe, Vassilis Pelekanos, Art Glenberg, Petar Milin, David Vinson, Pablo Gómez, Hiroyuki Tsubomi, Atsunori Ariga, Fuminoro Ono, Hanako Ikeda, Siva Tian, Lisa Lix, and Rogier Kievit. The following researchers participated as reviewers of two papers that were withdrawn from this special issue (no order implied): Helen Mitchell, Sergio Chaigneau, Julio Santiago, and Caitlin Fausey.

Finally, I would like to thank Robyn Groves for proofreading this editorial.

References

- Aarseen, L. W., Tregenza, T., Budden, A. E., Lortie, C. J., Koricheva, J., & Leimu, R. (2008). Bang for your buck: rejection rates and impact factors in ecological journals. *The Open Ecology Journal*, 1, 14-19.
- Abin, M., & Mabe, M. (2000). Impact factors: use and abuse. *Perspectives in Publishing*, 1, 1-6.
- Chihara, L., & Hesterberg, T. (2011). *Mathematical statistics with resampling and R*. N. J.: Wiley.
- Gernert, D. (2008). How to reject any scientific manuscript. *Journal of Scientific Exploration*, 22 (2), 233-243.
- Marmolejo-Ramos, F. (2008). A call to arms: Time to do cognitive science in Latin-America. *International Journal of Psychological Research*, 1 (2), 41-52.
- Mayor, J. (2010). Are scientists nearsighted gamblers? The misleading nature of impact factors. *Frontiers in Psychology*, 1 (215), doi: 10.3389/fpsyg.2010.00215
- Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, 7, 141-144.
- Nickerson, R. S. (2006). Cognitive science: experimental methods. In L. Nadel (Ed.), *Encyclopaedia of cognitive science*. Doi: 10.1002/0470018860
- Von Eckardt, B. (2006). Cognitive science: philosophical issues. In L. Nadel (Ed.), *Encyclopaedia of cognitive science*. Doi: 10.1002/0470018860
- Wicherts, J. M., Kievit, R. A., Bakker, M., & Borsboom, D. (2012). Letting the daylight in: Reviewing the reviewers and other ways to maximise transparency in science. *Frontiers in Computational Neuroscience*, 6 (20), Doi: 10.3389/fncom.2012.00020