© International Journal of Clinical and Health Psychology

ISSN 1697-2600
2008, Vol. 8, Nº 3, pp. 751-764

Criteria of the peer review process for publication of experimental and quasi-experimental research in Psychology: A guide for creating research papers

Manuel M. Ramos-Álvarez[1], María M. Moreno-Fernández (*Universidad de Jaén, España*), Berenice Valdés-Conroy (*University of Northumbria, England*), and Andrés Catena (*Universidad de Granada, España*)

(Received April 4, 2008 / Recibido 4 de abril 2008)
(Accepted June 19, 2008 / Aceptado 19 de junio 2008)

ABSTRACT. Experimental research in Psychology is characterized by ensuring a method that guarantees objectivity, reliability, validity and replication of results. In this theoretical study we propose a set of criteria for the preparation and review of quasi-experimental and experimental research manuscripts, which follows such methodological objective dictates. These criteria are based on a review of structural aspects in experimental research, in the modern theory of psychological theorization, and in the validity theory of scientific research. All these aspects are complemented with those proposed in revisions about empirically-based, statistically-based peer-review systems, and recently refined according to the expert judgment approach. We distinguish between essential, obligatory, complementary, and methodological criteria. These norms are organized according to a measuring tool –the *ExperimenCheck2* system-, including report characteristics, antecedents, theoretical development, design, analysis and interpretation of results, format and bibliographical sources, also summarized in the form of a conceptual map. We also introduce the general guidelines of a reviewing process that fulfils scientific criteria, using the proposed evaluative guideline as the main organizing principle.

1 Correspondence: Metodología de las Ciencias del Comportamiento. Edificio D-2. Universidad de Jaén. Paraje las Lagunillas, s/n. 23071 Jaén (España). E-mail: mramos@ujaen.es

KEYWORDS. Criteria for manuscript creation. Norms for peer-review process of original papers. Experimental methodology. Quasi-experimental methodology. Psychology. Theoretical study.

RESUMEN. La investigación experimental en Psicología se rige por un método que garantiza la objetividad, la fiabilidad, la validez y la reproducción de los resultados. En este estudio teórico se propone un conjunto de criterios para la preparación y revisión de manuscritos de investigación experimental y cuasi-experimental que sigue precisamente tales dictados metodológicos de objetividad. Tales criterios se basan en una revisión de los aspectos estructurales propios de la metodología de investigación experimental y la teoría moderna de teorización psicológica, así como en la teoría de validez. Dichos aspectos se complementaron además con lo que se ha publicado sobre el propio proceso de revisión, tanto desde el punto de vista empírico como estadístico-metodológico y que recientemente se han depurado siguiendo la aproximación del juicio de expertos. Se distingue entre criterios básicos, obligatorios, criterios complementarios y del tipo metodológico. Los criterios normativos se organizan según el sistema *ExperimenCheck2*, un instrumento de medida que abarca las características del informe, los antecedentes, el desarrollo teórico, el diseño, análisis e interpretación de resultados, formato y fuentes documentales. Además, se incluye un mapa conceptual que servirá como guía del sistema propuesto. Se presentan también las directrices generales de un proceso de revisión basado en dicho sistema de evaluación.

PALABRAS CLAVE. Criterios de elaboración de artículos. Normas para el proceso de revisión de artículos originales. Metodología experimental. Metodología cuasi-experimental. Psicología. Estudio teórico.

Science is an organized body of knowledge acquired from a well defined and coherent scheme of activities, the scientific method. The scientific method is an auto-corrective, systematic and structured procedure (Keppel, 1991; Maxwell and Delaney, 2004), to evaluate empirical observations, minimizing bias from beliefs or faulted reasoning. As a consequence, scientific knowledge is noticeable, objective, accurate, reliable, public, and replicable. There is a wide agreement among psychologists in that scientific method criteria have to drive the research activity and the communication of the results. Making findings publicly available is a key aspect of research to allow its continuous review and replication.

Disemination of scientific knowledge relies greatly on a publication process that requires an organized set of criteria not only for manuscript preparation but also for the peer-review process. The present proposal of norms for publications is designed to address three current difficulties in the peer-review process: lack of uniformity (Beyer, Chanove, and Fox, 1995), lack of delimitations (Gilliland and Cortina, 1997), and its lack of permeability against subjective evaluations (Gilliland and Beckstein, 1996; Hadjistavropoulos and Bieling, 2000). We propose a structured guideline for the reviewing and writing processes, both for scientists and young researchers.

The present set of norms concerns to experimental and quasi-experimental research publications -specific criteria can be found in Bobenrieth (2002), for Health Sciences; Buela-Casal and Sierra (2002) or Virués-Ortega and Moreno-Rodríguez (2008) for Clinical Psychology; Carretero-Dios and Pérez (2007), for instrumentation and measurement; and in Botella and Gambara (2006), for meta-analysis-. Criteria manuscripts for non-experimental, descriptive and qualitative studies can be seen in Campion (1993). In the following pages we will first justify the complexity of the proposed system of norms based on two major components: structure and validity. Then, we will explain the current additions to our previous set of norms (Ramos-Álvarez and Catena, 2004; Ramos-Álvarez, Valdés-Conroy, and Catena, 2006), and finally, we will present an evaluation instrument developed for experimental and quasi-experimental research publications.

The reviewing criteria

Structure

The research process follows three stages (Maxwell and Delaney, 2004): conceptual, methodological, and statistical. Science conveys a conceptual world (linked to theories and research hypothesis) and an empirical world (linked to observations and data). The connection between these two worlds is achieved by the method (linked to the hypothesis and able to obtain data that can contrast them). Combining this idea, along with the ideas developed in the American Psychological Association (APA) publication manual (2001), and the empirical review on the reviewing process of Beyer *et al*. (1995) and Gilliland and Cortina (1997), it can be concluded that theoretical components (introduction and discussion), experimental design (the method section) and the results section are essential for the review process.

Validity

The validity theory, recently reviewed by Shadish, Cook, and Campbell (2001), indicates that research conclusions can be seriously questioned by a series of validity threats. The value of the first part of the paper, included in the conceptual stage, relies on the validity of construct, that is, the reasons that may produce incorrect inferences on the construct explored in the study. Here we include problems with the definition of the construct, or with empirical definitions linked to the construct.

The value of the design, included in the methodological level, depends on two types of validity: internal (why the inferences on the effect of a given independent variable can be incorrect?), and external (how the inferences can be generalized across populations, contexts, etc.?). The control of relevant variables and the sampling of research units are the most important factors of this stage.

The value of the two last sections, results analysis and discussion, should be based on the statistical validity (why inferences from statistical analysis can be incorrect?). Problems with the analysis assumptions (*e.g.*, homogeneity of variances, sphericity, etc.) may increase the probability of Type I error particularly when increasing the number of contrasts.

These criteria from the validity theory have to be put together with those based on the modern theory of psychological theorization -see Ramos, Catena, and Trujillo (2004), for review- in which causality is the key concept. From this view, designing experiments is a two-stage activity: structural (statistical design) and strategic (manipulation and control of variables). These stages are connected by the plan of the research, where the problem under investigation is stated explicitly (Ramos *et al*., 2004). From a statistical point of view, the General Lineal Model and the Generalized Lineal Model (Estes, 1991; Harris, 1994; Judd and McClleland, 1989; Judd, McClelland, and Culhane, 1995; see, for example Minitab –Minitab Inc, 2008-, R –The R Project, 2008-/S-Plus –Insightful Inc, 2008-, or STATISTICA –StatSoft Inc, 2008-) are the frames of reference, being the multivariate linear regression the core technique. In this vein, it has been recommended (Wilkinson and the Task Force on Statistical Inference, 1999) to report additional statistics, mainly the size of the effect and the statistical power, in order to facilitate the evaluation of the results (Chow, 1998).

The criteria derived from the precedent ideas have been complemented with those proposed in empirically-based (Campion, 1993), and statistically-based review systems (Gore, Jones, and Thompson, 1992). Some items to evaluate the conceptual level, and the frequent biases in the analysis of results, specially related with the statistical validity of the research, have been added in the present proposal.

Some journals that evaluated their editorial policy (Beyer *et al*., 1995) have concluded that relevance-originality-novelty, technical and conceptual quality, and the suitability for the journal are the principal edges of a good reviewing process. In a similar approach, for Gilliland and Cortina (1997) the main edges were: design, method adequacy, theoretical and statistical quality, background literature, construct development, and writing style. In agreement with these ideas, we have reordered the items in our clusters, and added new items: writing style and suitability.

Also, we think that new experimental techniques (for example, Internet-based experiments) open a huge number of research opportunities, but also a great number of validity threats, so, we have added some items in agreement with the proposals of Birnbaum (2000) on Internet-based experimentation, and Shadish (2002) on field experiments.

Finally, in a recent revision, items were reorganized following three criteria: the suggestions of reviewers who had used these items, those of researchers, and those of typical didactic guide oriented to doctorate and post-graduate students.

The *ExperimenCheck2* system: criteria for the evaluation of experimental and quasi-experimental research

ExperimenCheck2 is a collection of recommendations and criteria for guiding manuscript writing and evaluation. The guide is composed of eight clusters: a) General, b) Literature antecedents and research rationale, c) Theoretical development, d) Experimental Design, e) Results, f) Interpretation of the results, g) Manuscript writing, h) Documentation sources.

Each cluster is displayed in the Appendix 1. The definition of the cluster and the section of the manuscript are in the heading of the Appendix. Most items are written

as an affirmative question, in which a parenthetical statement clarifies the evaluated content. Also, most items can be answered in a YES/NO fashion. Almost all items are useful for evaluating experimental and quasi-experimental research, so we will discriminate between them only if necessary.

Three types of criteria are included in each cluster. First, the items with one asterisk (basic items) have to be fulfilled completely in order to accept the paper. Second, the items with "+" symbol have to be used when evaluating the technical quality of the manuscript; especial attention must be paid to the design and statistical analysis. Third, the items without marks are complementary, and can be used for evaluating the general quality of the manuscript. Only papers with a high mark in the three types of criteria should be considered for publication.

The use of this tool could be improved using the conceptual map included in Appendix 2.

References

American Psychological Association (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Beyer, J.M., Chanove, R.G., and Fox, W.B. (1995). The review process and the fates of manuscripts submitted to AMJ. *Academy of Management Journal, 38,* 1219-1260.

Birnbaum, M.H. (Ed.) (2000). *Psychological experiments on the Internet*. San Diego: Academic Press.

Bobenrieth, M.A. (2002). Normas para revisión de artículos originales en Ciencias de la Salud. *International Journal of Clinical and Health Psychology, 2,* 509-523.

Botella, J. and Gambara, H. (2006). Doing and reporting a meta-analysis. *International Journal of Clinical and Health Psychology, 6,* 425-440.

Buela-Casal, G. and Sierra, J.C. (2002). Normas para la redacción de casos clínicos. *International Journal of Clinical and Health Psychology, 2,* 525-532.

Campion, M.A. (1993). Article review checklist: A criterion checklist for reviewing research articles. *Personnel Psychology, 46,* 705-718.

Carretero-Dios, H. y Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology, 7,* 863-882.

Chow, S.L. (1998). Précis of statistical significance: Rationale, validity, and utility. *Behavioral and Brain Sciences, 21,* 169-239.

Estes, W.K. (1991). *Statistical models in behavioral research*. Hillsdale, NJ: Erlbaum.

Faden, R.R., Beauchamp, T.L., and King, N.N. (1986). *A history and theory of informed consent*. Oxford, UK: Oxford University Press.

Gilliland, S.W. and Beckstein, B.A. (1996). Procedural and distributive justice in the editorial review process. *Personnel Psychology, 49,* 669-691.

Gilliland, S.W. and Cortina, J.M. (1997). Reviewer and editor decision making in the journal review process. *Personnel Psychology, 50,* 427-452.

Gore, S.M, Jones, G., and Thompson, S.G. (1992). The Lancet's statistical review process: Areas for improvement by authors. *The Lancet, 340,* 100-102.

Hadjistavropoulos, T. and Bieling, P.J. (2000). When reviews attack: Ethics, free speech, and the peer-review process. *Canadian Psychology, 41,* 152-159.

Harris, R.J. (1994). *ANOVA: An analysis of variance primer*. Itasca, IL: F.E. Peacock.
Insightful, Inc. (2008). *S-Plus, Release 8*. www.insightful.com
Judd, C.M. and McClelland, G.H. (1989). *Data analysis: A model comparison approach*. San Diego, CA: Harcout, Brace, Jovanovich.
Judd, C.M., McClelland, G.H., and Culhane, S.E. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review or Psychology, 46*, 433-465.
Keppel, G. (1991). *Design and analysis. A researcher's handbook* (3rd ed.). New Jersey: Prentice Hall.
Maxwell, S. E. and Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ.: Lawrence Erlbaum Associates.
Minitab, Inc. (2008). *MINITAB (Statistical software). Release 15*. www.minitab.com
Ramos-Álvarez, M.M. and Catena, A. (2004). Normas para la elaboración y revisión de artículos originales experimentales en Ciencias del Comportamiento. *International Journal of Clinical and Health Psychology, 4*, 173-189.
Ramos, M.M., Catena, A., and Trujillo, H.M. (2004). *Manual de métodos y técnicas de investigación en ciencias del comportamiento*. Madrid: Biblioteca Nueva.
Ramos-Álvarez, M.M., Valdés-Conroy, B., and Catena, A. (2006). Criteria of the peer-review process for publication of experimental and quasiexperimental research in Psychology. *International Journal of Clinical and Health Psychology, 6*, 773-787.
Shadish, W.R. (2002). Revisiting field experimentation: Field notes for the future. *Psychological Methods, 7*, 3-18.
Shadish, W.R., Cook, T.D., and Campbell, D.T. (2001). *Experimental and Quasi-experimental designs for generalized causal inferences*. Berkeley: Houghton Mifflin.
StatSoft, Inc. (2008). *Statistica (data analysis software system), Release 8*. www.statsoft.com
The R Project (2008). *R Project*. www.r-project.org
Virués-Ortega, J. and Moreno- Rodríguez, R. (2008). Guidelines for clinical case reports in behavioural clinical Psychology. *International Journal of Clinical and Health Psychology, 8*, 765-777.
Wilkinson, L., and the Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and Explanations. *American Psychologist, 54*, 594-604.

APPENDIX 1. *ExperimenCheck2*: Evaluation guideline.

A. General (beliefs about the general quality of the manuscript after a reading)

- A.1. What is the strongest point of the manuscript?
- A.2. What is the weight (0-100) of the A1 point? (0: you think that the point is positive, but can be suppressed of the paper without loss; 100: you think that this is a core for the paper).
- A.3. What is the weakest point of the manuscript?
- A.4. What is the weight (0-100) of the A3 point? (0: you think the point is completely negligible; 100: you think the point makes the paper non publishable).
- A.5. Suitability of the manuscript? (The topic is interesting for the general reader of the journal, and it will be frequently cited, or, by the contrary, can have interest for a scarce number of potential readers).

B. Antecedents and motivation of the research (Introduction section)

- B.1. * Is the description of the research problem clearly stated? (The main goal of the research is clearly stated at the beginning of the paper. The reader should not have to work out a lot in order to determine what the research question is).
- B.2. * Is there enough background literature to justify the importance of the research problem? (According to the literature, is the research question a major or a minor one?).
- B.3. * If the research problem has been solved, can the paper guide future research? (Is there a literature agreement on the importance of the research question?).
- B.4. Is the theoretical frame clearly stated? (The research problem should be embedded within the frame of a psychological theory and not stated merely on empirical terms).
- B.5. Are the empirical referents of the problem clearly stated? (The rationale should be based on a clear sequence of effects-data).
- B.6. Is the writing of the manuscript coherent? (The reading of the manuscript should follow a clear and coherent course and not moving from point to point without apparent connection between ideas. Introduction should be from general to specific).

C. Theoretical considerations (Introduction section)

- C.1. * Are the hypotheses clearly stated and is well justified? (Hypotheses should be based on a clearly stated theory and not empirically based only).
- C.2. * Does the hypothesis follow from the theory? (Hypotheses should be derived from the rationale).
- C.3. * Is the hypothesis a causal one?
- C.4. *Are there clear empirical predictions derived from the hypothesis?
- C.5. * Is the conceptual system relevant? (Constructs and concepts should be clearly related and relevant to the research problem).
- C.6. * Is there a presentation of alternative points of view on the research problem? (The manuscript should not be centred on author's beliefs).
- C.7. Is the cited literature coherent with the manuscript's theoretical foundation?
- C.8. * Is there some theoretical originality in the manuscript? (Is the manuscript based on other authors' predictions?).

D. Experimental & Quasi-Experimental design (Method section)

D.1. * Ethics. Is there a clear statement on subjects' rights (for example, according to Helsinki declaration)?
* Animals: Statements on their treatment and suffering.
* Humans: All research with humans must obtain informed consent (Faden, Beauchamp, and King, 1986).
* Therapy: Have the number of sessions been optimized to obtain the therapeutic benefit? Are all the participants (including the control conditions) going to receive the benefits of the therapy?

D.2. * Is the method section complete and properly organized? (Participants, instruments, design, procedure, etc.). Experimental units, instruments, apparatus, psychometric features, and procedure should be described up to a degree to allow replication.

D.3. * Is the hypothesis testable? The hypothesis must contain clear predictions on the direction of the results.

D.4. * From a structural point of view: Are the independent variables adequate to contrast the hypothesis?
· For each independent variable: The selected levels must be well described (number, spacing between levels and range of the variable).
· For factorial designs: When using a complete factorial, all conditions should be included and, if some experimental conditions are omitted, there should be a clear justification of why.

D.5. * For dependent variables: Keeping in mind the measurement instruments, are the psychometric properties of the variable (sensibility, reliability, and validity) adequate?

D.6. * Internal validity (for experiments only): How good is the control of threats against the cause-effect inferences? (As a minimum caution, were the subjects randomly assigned to the experimental conditions?).

D.7. Are the characteristics of the participants fully described? Recruitment procedure, number, population characteristics, other features relevant for the research.

D.8. Is the subjects sample defined up the point to allow replications of the experiment?

D.9. * Do the sampling of the experimental units and the characteristics of the obtained samples guarantee the validity of the research?

D.10. * Is the procedure (*e.g.*, apparatus and instruments, stimuli, subjects, task) appropriate for reaching the aims of the research?

D.11. Are the instruments/apparatus described to a good level to facilitate replication? When available, cite commercial product and model. When non available, describe the main features regarding the study. Do not mention details of common apparatus (for example, compatible PC).

D.12. * When novel methods are used: Are they sufficiently described and justified?

D.13. Does the procedure section contain enough details to allow replication of the study? (Instructions, practice trials, experimental trials, etc.).

D.14. * Are the terms "experiment" and "quasi-experiment" correctly used?
· Experiments fulfil two criteria: first, is there a direct manipulation of the independent variable or it is rather indirect (subjects are members of a group because they have the attribute of interest up to the desired degree)? Second, is there a good control of relevant variables?
· Quasi-experiments: It is clearly stated the design (*e.g.*, pre-post measurements with non-equivalent control group).

D.15. * Design quality:
· Experiments: Does the design optimize the possibilities of rejecting the null-hypothesis? (Is it better a between or within subjects design? Is it better a factorial than a unifactorial design (one-way design)? Is it better a multivariate than a univariate design? Is it preferable a balanced or an unbalanced design?, and so on.
· Quasi-experiments: Are the two main characteristics of quasi-experimentation: pre-post measures, and non-equivalent control group present? Are there others features that enhanced the possibility of rejecting the null hypothesis (for example, cohorts, several pre-post measures, covariates, discontinuity of the regression, etc.)?

D.16. Design quality (simple versus complex designs):
· Is it adequate to use a unifactorial and/or univariate design? (Are all relevant factors treated in the right way?).
· Is it adequate to use a factorial and/or multivariate design? (Should a simpler design be more efficient than the one selected?).

D.17. + Is the between/within subjects manipulation the correct one for the research purposes?

D.18. In general, is the experimental design the most suitable for contrasting the hypothesis?

D.19. Construct validity:
· Have the reactivity changes been avoided?
  b. Auto-reports.
  c. Context reactivity.
  d. Experimenter expectations.
  e. Participants' diffusions of treatment.
· Have the main biases been avoided?
  b. Novelty.
  c. Compensatory equalization (for example, subjects in the control group can be treated especially because they are in disadvantage).
  d. Compensatory competence (for example, subjects in the control group can do their best, because they think they are as good as those of the experimental group).
  e. Frustration (for example, untreated subjects can be frustrated and change their behaviour accordingly).

D.20. + Internal validity: aside from randomization, is there any further explicit control of threats against interval validity?
· Between subjects' designs: Which kind of homogenization techniques have been used?
· Repeated measures designs: Which kind of techniques to control sequential effects have been applied (counterbalance, Latin square, etc.)?
· What controls of context variables (i.e., instructions, experimenter, ambient, etc.) have been applied?
· Quasi-Experiments: Are there considerations about validity threats as ambiguous causal order, selection, history, etc.?

D.21. Internal validity: Is the attrition problem presented and discussed? Are there some indications of a possible relation of attrition and manipulation of the variables?

D.22. External validity:
· Subjects/Independent variables:
  a. Is the sample representative of the population?
  b. Is there any bias in the sampling process?

· Context:
  a. Is the context (lab/field/Internet, etc.) the adequate? Should be better to use another research context?
  b. If the context is uncommon (*i.e.*, Internet), is there any mention of its adequacy to the research purposes?

E. Results

E.1. * Do the novelty and meaning of the results make a significant contribution to the literature? Do they justify the research?

E.2. * Are the statistic-research hypothesis coherent?
- Have the authors included the relevant analysis for contrasting their hypothesis? (a general analysis for a general hypothesis, contrast analysis for specific hypothesis, trends analysis for functional relationship hypothesis, and so on). Did the authors check whether the data meet the analysis' assumptions?
- In factorial design, is the analysis of the interaction based on simple effects or interaction of contrast? Which one should be the best for contrasting the hypothesis on the interaction?
- When multiple dependent variables have been measured, is the right alternative a MANOVA or multiple ANOVA?
- When relevant variables have been considered, have they being considered for an ANCOVA?

E.3. * Data description. Are the main descriptive statistics (averages, standard deviations or standard errors of mean displayed in a Table or a Figure?

E.4. * Is there an adequate description of the analysis techniques?

E.5. * Is there a complete description of the main parameters of the analyses? (For example, if ANOVA was used, are the *F* value, the degrees of freedom, the mean square error, and the significance level reported? Do all the parameters agree? (For example, are the degrees of freedom in accordance with the number of subjects/ treatments?).

E.6. Following APA statistical recommendations: Is the results section adequate?
- Starting with a qualitative description, and detailing after the inferential description of the statistical results.
- Descriptive stats data including (mean, standard deviation or standard error of the mean). Details of contrast analyses should be included (*e.g.*, $F_{(3, 126)} = 6{,}35$; $p < 0{,}05$; $MSE = 425{,}657$).
- When a large number of data is reported, it is convenient the use of tables.

E.7. + The units of the analysis have been clearly stated and justified (subjects, groups, etc.)?

E.8. + Are the power and size of the effect computed and discussed?

E.9. * Can the error variances be the cause of null results?

E.10. + Can null results be the consequence of the poor qualities of the dependent variable (for example, ceiling effects)?

E.11. + Are the inferential techniques used for contrasting the hypothesis appropriate? For example:
- Is the between/within manipulation considered in the analysis?
- When non-quantitative dependent variables have been measured, do the categorical data analysis perspective was used?

- · In repeated measures designs, is there evidence that assuming non-sphericity null hypothesis can still be rejected? (*i.e*., is there an alternative MANOVA, or a Geisser-Greenhouse/Huynh-Feldt approach?
- · Given the data features (for example, a huge number of zeros, a scarce number of subjects), a non-parametric ANOVA was considered?
- · When data do not meet some assumptions of the general linear model, and transformations of data can be an alternative, is the transformation well described and supported in the literature?
- · When non-balanced designs, do the authors used the correct Sum of Squares (Type I, II, or III)?
- · When incomplete factorial designs, was the analysis appropriate for nested, fractional, etc.?

E.12. + Statistical validity. Is there some evidence that data cannot meet the assumptions of the statistical test? (For example, means and variances are related) If so, is there some description of the results of assumptions test?

E.13. + Statistical validity. Which kind of correction of Type I error has been done? (Bonferroni, Sidak, Tukey test, and son on).

E.14. + Are there any testable predictions that have been neglected?

E.15. + When attrition is bigger than expected, is there any explanation? Is attrition a possible source of observed results?

E.16. + Are the statistical-methodological techniques used in the right way? Are there more analyses than needed?

E.17. + Have statistical test been used in a comprehensive way? Are they applied in a stereotyped way?

E.18. + When novel statistical techniques were used, are there literature references to justify their use?

E.19. + Is the statistical-methodological literature adequate for justifying the reported data analyses?

E.20. How original are the empirical findings reported in the paper?

E.21. + When statistical innovations are the key point of the paper, how original they are?

E .22. + Do the authors propose a significant statistical-methodological improvement? (new control techniques, new dependent variables, new analyses).

F. Discussion (interpretation of results)

F.1. * Is the interpretation of the results and the statistical analysis congruent with the research problem as explained in the introduction section?

F.2. Are all the relevant results discussed?

F.3. * Is the discussion of statistical results in accordance with the limitations of the design (especially when the sample characteristics, the control of relevant variables, and so on, are a matter)? Are there any signs of overgeneralization of the results?

F.4. Are the relevant results summarized at the beginning of the discussion?

F.5. Are the alternative interpretations of the results presented and discarded in accordance with the statistical tests?

F.6. Does the paper place in context the results with those relevant in the literature (especially when contradictory data have been noticed)? Authors should not select only confirmatory evidence from the literature.

F.7. + Do the authors appear to avoid noisy results? (For example, they take as reliable results that in other context should be rejected: positive results because “outliers”, subjects that do not meet the learning criterion, null results, and so on).

F.8. Are there any considerations about the limitations of the study (for example, because the selected dependent variable), and/or their implication for future research?
F.9. Is the discussion fluid, it connects the introduction and observed results? Are those arguments *ad hoc*?
F.10. How original and relevant for the literature is the interpretation of the results? Is there any substantive theoretical or empirical contribution?

G. APA format (Considering all the paper)

G.1. * Do, at least, a 70% of APA norms are covered in the paper?
G.2. * Is the format the correct one? (Consider the general size of the sections, the spacing between text lines, the numbering of pages, the size of page, the margins, the letter type -CG Times).
G.3. * Are the sections of the paper in the correct order?
G.4. * Are all the sections complete? Do all information needed for understanding the paper is present and in the right order?
G.5. * Is the paper clear, structured in a comprehensive fashion? (Organized around a main point, with a clear rationale, and a good level of coherence between the sections).
G.6. The writing style is direct, simple, with technical terms described the first time they appear in the text. Not difficult to follow for a standard reader (non-specialist).
G.7. Is the paper complemented with figures and/or tables to explain complex parts (for example, the procedure)?
G.8. Regarding APA norms, are title, filiations and abstract amenable for publication?
- Title is short and clear, and it contains the main point of the paper.
- Authors name, without reference to their academic position.
- Complete address for correspondence.
- Abstract, in a single paragraph, 100-200 words length, presenting the important parts of the paper: motivation, rationale, main results and main conclusion.

G.9. Following APA criteria: Is the reference section appropriate?
- Containing only references to work previously cited in the text.
- When writing papers form journals: Author/s. (year). Title. *Journal, volume*, pages. (Journal title and volume should be in italics).
- For books: Author/s. (year). *Title*. Place of edition: Editorial. (The title of the book should be in italics).
- For book's chapters: Author/s. (year). Title of chapter. Book Editors. *Title of the book.* (chapter, pages). Place of edition: Editorial. (The title of the book should be in italics).
- WWW based material: Author/s. (year month). Title. *Web organization*. Retrieved from "Web address" (Web organization in italics and web address as hypervincule). *E.g*., Ramos-Álvarez, M.M. (2000, October). Integration of Multiple Cues in Learning. *The Brunswik Society Newsletters, 15 (Millenium Issue)*, Article#17. Retrieved from: http://brunswik.org/newsletters/newsletter2000/2000news.html.

G.10. Following APA norms: Are the citations appropriate in the text?
- Second name of authors and year between parentheses if the citation is part of the text; *e.g*., "Ramos and Catena previous study (2004)".
- When cited work is a reference for a statement in the text, second name of the authors and year must be between parentheses separated by commas; *e.g*., "It has been demonstrated that coffee ingestion alters mood in children, (Ramos, Catena, and Perez, 2004)".

- · If more than one study is referenced, citations must be ordered alphabetically.
- · Citations with more than two authors that appear several times in the text must be abbreviated after the first time (*e.g.*, Ramos *et al*., 2004).
- · Citations from different publications with same the same first author and same year must contain lowercase letters to distinguish between them (*e.g.*, Ramos *et al*., 2004a; Ramos *et al*., 2004b).
- · These codes must be kept in the reference section.

G.11. Following APA criteria: Are the tables, graphs and illustrations adequate? (Authors should respect journal format and unit measures). Are the figures and tables informative and necessary?

G.12. Are there any original, creative or substantial contributions to the vocabulary? If new terms have been introduced, empirical, theoretical or methodological, are they truly needed?

G.13. Are there any creative and original contributions to the format? Are these contributions substantial, useful and justified?

H. Documentary sources (citations and references)

H.1. * Is there evidence of a good knowledge and management of the background literature, both in conceptual and technical aspects? (It is clear that authors know previous work related with the theoretical framework, experimental methodology and effects, or the citations seem to be irrelevant and stereotyped?).

H.2. * Are the citations of previous work accurate? (The information and the citations correspond or at the contrary, cited work seem incompatible with the ideas in the text).

H.3. * Are the documental sources adequate? There is a good amount of work related with the problem under investigation and relevant areas or citations are dispersed.

H.4. * Are the referenced works, scientifically credited? (Most of the citations are from scientific and academic publications and not from other non-scientific resources).

H.5. * Are documental resources updated? (A huge number of the cited papers have been published during the last 5 years).

H.6. Is the cited literature enough and appropriate for the aims of the manuscript? (There should be references to previous work to justify the experiments. The background literature referenced should have a close link with the aims of the research).

H.7. Are the referenced works specific? (Most of the citations come from credited journals specialized on the topic of the research problem).

H.8. Is there evidence of a critical review of previous literature? (There are citations of previous work with hypotheses that are different to those proposed by the authors. There should be no sign of deliberate bias to cite previous research that shares results or hypothesis with the author's point of view).

H.9. Is there a good deliberate selection of documental resources? (References are original and not just copied from other sources).

H.10. Is there a good balance of cited work? (National, international, empirical, theoretical, etc).

H.11. Are there any original contributions related to previous work? (Cited work is highly useful with references that might direct to other kind of resources; *e.g.*, www, technical, material, etc.).

APPENDIX 2. Conceptual map.