APPENDIX 2

ARTICLES MINIMUM REQUIREMENTS

ABSTRACT	YES	NOT APPLICABLE
Is the abstract prepare according to APA Publication Manual (7th Ed.)? (Please check pp. 73-75)		
INTRODUCTION	**YES**	**NOT APPLICABLE**
Have the objectives been clearly defined and the working hypotheses been stated?		
Has it been clearly stated how the experimental design and data analyses will verify all hypotheses formulated to achieve the proposed objectives?		
MÉTHOD: Participants	**YES**	**N/A**
Has the sample been adequately described, especially with regard to characteristics relevant to generalization of results?		
If using groups, have they been proven equivalent in variables that may contaminate results?		
Has the sampling procedure been described?		
METHOD: Measurements	**YES**	**N/A**
Have the materials been described in detail, especially the original instrument's psychometric properties relevant to the research?		
Have the psychometric properties been adequately described with the sample's characteristics?		
If the instrument used has been adapted from a different language or culture, has the adaptation method been described?		
If the instrument has been created "ad hoc", has its need been rationalized?		
METHOD: Procedure (Detail of experimental design)	**YES**	**N/A**
In those experimental designs that entail manipulation or intervention, ...		
…..have the contents of such manipulation or intervention been clearly stated?		
…..has the procedure to perform such manipulation or intervention been explained?		
…..have the duration and intensity of exposure been specified?		
…..has the use of incentives been declared?		
…..has the assignment of participants to experimental groups been explained?		
MÉTHOD: Data Analyses	**YES**	**N/A**
Have the analyses that will be performed been explained?		
Has the program used to perform such analyses and its version been specified?		
Has compliance with precise assumptions for each specific analysis performed been informed?		
RESULTS: Basic information	**YES**	**N/A**
Has an adequate analysis of lost values been performed?		
Has all the descriptive statistics (simple size, mean, standard deviation, variance-covariance matrices, etc…) information necessary to understand the analyses been included?		
Has all the information necessary to include these results in meta-analyses of this field been included?		

RESULTS: Statistical significance tests	YES	N/A
Have statistical values, degrees of freedom and critical values (exact value of p) been included?		
Has the effect size and direction been stated, both in the original (with confidence intervals) and in standardized units (d, η^2, R^2,…)?		
Has the power been reported, especially if not rejecting the null hypothesis?		
Have the confidence intervals of the parameters been reported?		
If reporting point estimates of a parameter, have the standard errors been specified?		
RESULTS: Tables	**YES**	**N/A**
Are the tables necessary?		
Are all the tables consistent in format?		
Do all tables have a title and is this title short and informative?		
Do all columns have a heading?		
Have all abbreviations been explained?		
Can the probability levels be correctly identified?		
Are all tables referenced in the main body of the manuscript?		
RESULTS: Figures	**YES**	**N/A**
Are the figures necessary?		
Are the figures simple, clear and concise?		
Are all figures consistent in format?		
Are all concept figures in the same scale?		
Do all figures have a title and is this title short and informative?		
Are all figures referenced in the main body of the manuscript?		
DISCUSSIÓN	**YES**	**N/A**
Have you ensured that the discussion is NOT a simple repetition of the exposed results?		
When interpreting the results, have you taken into consideration the potential sources of bias, effect size, alternative explanations, other possible limitations, etc…?		
Have the results of this manuscript been discussed in relation to previous research?		
In case of discrepancy, have you offered some alternative explanation?		
APA ETHICAL ASPECTS: Data	**YES**	**N/A**
AVAILABILITY: The data must be available to the editor during both the review and the publication processes.		
Are your data available at the request of editor or the reviewers?		
DUPLICATION: It is prohibited to present data as original when they have been included in research that has already been published. This may distort the knowledge base by making it appear there is more information available tan really exists.		
Do you confirm that these data have not been used in any other publication?		
REANALYSIS: Data reanalysis is justified in light of new theories or methodologies that may offer a different view of the phenomena and in long–term, longitudinal or multidisciplinary studies. In such cases, authors are obliged to inform of previous publications.		
If you are presenting a data re-analysis, have you adequately justified its appropriateness and have you informed of previous publications?		
PIECEMEAL OR FRAGMENTED PUBLICATION: Information should be presented in the most parsimonious way possible, avoiding the distortion that may occur when presenting part of the data of a study in separate publications. However, in a few cases, this division may be desirable and it must be made known to the editor.		
If only part of the data of a study have been used, have you informed the editor and justified the appropriateness of such procedure?.		

REQUIREMENTS FOR PSYCHOMETRIC STUDIES

This document covers some of the standards for test scores to be met by manuscripts on psychometric studies being submitted for review in the journal. Even though the technical reasons that justify these criteria are not included in this document, they are based on consolidated advances, accepted by the scientific community, about how to proceed in studies of applied psychometrics.

Three types of psychometric studies can be differentiated

a) Those that report the construction and psychometric properties of A NEW TEST.
b) Studies that report the score properties of A TEST ON A DIFFERENT SAMPLE, but do not require the language translation of the items.
c) Work that focuses on the ADAPTATION OF A TEST to another language and sociocultural context.

There are psychometric studies whose objectives are different from the three mentioned above, and for which no requirements are set out in this document. Examples of such documents would be studies that establish a short version of a test, research orientated on the study of the psychometric properties of computerized versions, those that study the fit of several rival theoretical models or those that use Item Response Theory to estimate item properties and the accuracy of estimated trait. Other psychometric studies on the quality of scores not obtained by tests also exist.

The following criteria apply to the most common types of psychometric studies, where the study of validity evidence of scores is always a fundamental interest. This document specifies criteria for those studies that apply factor analysis models, firstly because virtually all manuscripts include a study of evidence of the internal structure of a test, and, secondly, because, in recent years, knowledge of the recommended conditions of application of such models when used to analyze the association between item matrices has greatly increased.

On the other hand, length restrictions for articles are taken into consideration. In most cases, a more comprehensive analysis does not necessarily imply a longer article. These criteria intend to increase the quality of this type of studies, provide researchers with guidelines that improve their work, and offer some of the quality indicators (mainly regarding methodological aspects) that will be used in the manuscript review process.

Although some technical criteria may be common to all manuscripts, others are particular to the type of study performed, especially for manuscripts that involve adaptations of tests to other cultures. However, all of them should include the characteristic expressions of current Psychometrics, as reflected in the standards of some professional associations (AERA, APA & NCME, 1999), in certain books (v.gr. Abad, Olea, Ponsoda & García, 2011; Martínez-Arias, Hernández-Lloreda & Hernánez-Lloreda, 2006) or in some recent paper (v.gr. Sireci & Padilla, 2014). The referred Standards are about to be changed, at which point this document will be updated if necessary.

A. STUDIES INVOLVING THE CONSTRUCTION AND PSYCHOMETRIC PROPERTIES OF SCORES OF A NEW TEST.

Introduction

A.1. The introduction must clearly explain the objectives of the new test (professional scope, targeted population, uses and interpretations that it intends to justify) and the reasons that have lead to the development of a new instrument.

A.2. The first evidence of validity has to do with the content of the test itself. The introduction must include the theoretical basis that justifies the type and content of the items created. Both, the construct (latent variable or domain) and its various components or sub-features must be precisely defined, clearly describing the items (or at least some representative items) written to measure each of these sub-features. When including expert judgment on the representation or relevance of the domain, the aggregate results of these judgments will be described in detail.

A.3. In typical performance tests, the complete test (or at least representative sample items) must be incorporated into the annex, along with its instructions and response format. In complex tests (due to number of items, its display format or the need to use alternative materials) the format of the items belonging to the various components must be described with examples.

Participants

A.4. The number of people, the type of sampling (specifying the incidental selection when applicable) and the descriptive statistics for the relevant sociodemographic variables (age, gender, educational level, profession,...) and other interesting variables related to the objectives of the study (e.g. clinical or psychosocial variables) must be reported.

Item analysis

A.5. The manuscript must include, at least, descriptive summaries of the means, standard deviations, skewness and kurtosis of the items, in addition to the discrimination index (usually item-total correlation). If the applied objectives suggest it, the Differential Items Functioning (DIF) indicators between the different groups must be reported.

A.6. When necessary, the missing data rates and imputation methods applied must be indicated.

A.7. In multidimensional tests, the discrimination indices must be obtained in relation to the expected theoretical component or after running the factor analysis, i.e., relative to the corresponding total of the subtest or dimension.

A.8. When discarding items, the criteria used must be described and validated.

Validity Evidence: types and evidence of relationships with other variables

A.9. Studies must include at least three types of validity evidence: about the test content, about its internal structure (using factor analysis techniques) and about relationships with other variables (convergent or predictive validity). The A.2 point refers to the criteria for content evidence.

A.10. In the studies on evidence of relationships with other variables, the hypotheses must be validated (theoretically or through previous studies).

A.11. The different scores obtained in other instruments used must be briefly described, including the item response format and some indicator of score accuracy.

A.12. Even though multiple designs and statistical analyzes can be performed to obtain evidence of relationships with other variables, only those most appropriate according to the validation objectives, the number of variables involved and their metric status must be applied.

A.13. The application of multivariate techniques in order to verify the predicted relationships when considering all variables together will only be applied when the study's objectives, as well as the number of dimensions and external variables, indicate it.

Evidence on Internal structure: Exploratory Factor Analyses

When applying an exploratory factor model (EFA), the recommended choices must be made according to type of data available. These choices are briefly collected in the following criteria.

UNIVERSIDAD COMPLUTENSE
DE MADRID
FACULTAD DE PSICOLOGÍA

Colegio Oficial
de la Psicología
de Madrid

A.14. The type of software used and the type of correlations analyzed must be reported. Polychoric correlations must be used when the response format is below 5 ordered categories or when there is a significant number of items whose responses are not normally distributed (as shown by indicators of skewness and kurtosis).

A.15. Regarding the extraction method, no studies using the Principal Component method will be accepted, unless its application is justified conveniently. When the application of an estimation method requires the fulfillment of certain assumptions, information on such compliance must be provided.

A.16. At least two methods will be applied to decide the number of retained factors, none of which should be the K1 Kaiser rule. One of them must be Parallel Analysis method.

A.17. In most cases, an oblique rotation should be used, unless the theoretical predictions recommend independent factors, which is not usually common in psychological research.

A.18. The Results section must include a complete factorial weights matrix, indicating in the case of oblique rotations whether it is the configuration matrix or the structure matrix. The percentage of total variance explained jointly by the common factors, the percentage of variance explained by each factor (in the case of orthogonal rotations) and the sum of squared loadings (for oblique rotations) must be reported. In the latter case, the estimated correlations between factors must be included. The method used to estimate the factor scores must also be specified.

A.19. The reliability indicators of the various scores derived from the factorial study must be provided, reporting at least the α coefficient and the standard error of measurement. If rating these indicators with adverbs is desired, the ordinal criteria established by the European Federation of Psychologists Associations (http://www.efpa.eu/professional-development/assessment) must be applied. The estimation of other reliability coefficients and some derivations of the factorial model (e.g. omega coefficient) are recommended.

Evidence on Internal structure: Confirmatory Factor Analysis

A.20. Since it is only recently that the internal structure of tests is studied in this type of study, previous studies on internal structure are not available. Therefore, the application of CFA is recommended only when there are very specific predictions about the number of factors, the items that should load or not for each one, and the relationships between factors.

A.21. In any case, confirmatory factor analysis will not be applied to the same data that have been subjected to EFA. Both models can be applied only if applied to different sub-samples (cross-validation studies).

A.22. The type of association matrix (variance-covariance or correlation) subjected to analysis must be specified.

A.23. Regarding the extraction method, the reason for choosing one in particular according to the type of data available must be justified. In any case, the maximum likelihood method will not be applied if the items have less than 5 response categories or the assumption of normal distribution of scores is not met.

A.24. The method used to set the metric should be specified. The factor solution must be reported with the standardized weights and the different variances and covariances estimated for the items and factors.

A.25. Although some adjustment indicators are part of the software and the method of estimation used, various adjustment measures must be incorporated. A minimum of three of the following descriptive indices must be included: RMSEA, SRMR, NNFI (TLI) and CFI (RNI). References of the cutoff points used to assess the level of adjustment must be reported.

A.26. In general, model re-specifications inconsistent with theoretical predictions will not be accepted. If they have been carried out, each re-specification will be indicated in the starting model and their relevance will be theoretically validated.

A.27. Regarding the accuracy of the extracted scores, provisions of criterion A.18. must be applied.

Interpretation of scores

A.28. When the objectives of the study include the use of the test as a diagnostic or criterion-referenced interpretations (assigning the person to a certain category) purposes, the recommended cutoff points must be empirically validated and the procedures used must be specified.

A.29. When there are predictions of different performance in different samples, the variability and central tendency statistics obtained for each sample will be attached. When necessary, the possible "floor and ceiling effects" will be reported, i.e., high rates of participants who obtain minimum or maximum scores.

B) ADDITIONAL CRITERIA FOR STUDIES SHOWING TEST SCORE PROPERTIES ON A DIFFERENT SAMPLE, WHICH DO NOT REQUIRE THE IDIOMATIC TRANSLATION OF ITS ITEMS.

B.1. Most of the criteria established for category A studies are valid when attempting to study the psychometric properties of score on a different sample. It is particularly relevant to fulfill the A.9 criteria.

B.2. An important and specific key for this second type of study is solidly validate in the Introduction section the theoretical or applied reasons that justify a new psychometric study on a new sample.

B.3. The introduction must summarize the main results of psychometric studies of the test carried out on other samples, including specific provisions on the expected convergence or lack thereof for the new study.

B.4. Studies must compare, preferably using statistical techniques, the differences found in the new sample in relation to the internal structure of the test, incorporating analysis of congruence between factors (when EFA is applied) or metric invariance (when CFA is applied). In any case, the choice of factor analysis used must be explained.

B.5. Indicators of score accuracy obtained in the current sample and in other samples will be statistically compared by estimating confidence intervals or by performing the appropriate statistical contrasts.

B.6. When the metric invariance is ensured, the performance of the relevant samples (original and current) in the estimated scores will be statistically compared with the test

B.7. The manuscript will include the main findings on the use and interpretation of the scores that are advisable in light of the evidence obtained in the new sample.

C) ADDITIONAL CRITERIA FOR STUDIES THAT INTEND TO ADAPT A TEST TO A DIFFERENT LANGUAGE AND SOCIOCULTURAL BACKGROUND.

C.1. In general, the guidelines proposed by the International Test Commission for the translation and adaptation of psychological and educational tests (ITC, http://www.intestcom.org/upload/sitefiles/40.pdf) must be met.

C.2. When the objective of the study includes the development of a version of the test in another language, the translation method and the procedures of semantic analysis and item relevance used by experts must be described. The procedure used to reach the final translated version and, if applicable, the changes incorporated in the new version must be clearly described.

C.3. While criteria belonging to paragraph B of this document are still in force, in this third type of studies, confirmatory factor models are applicable in most cases, and therefore, the specific requirements described above must be met, in order to study the relevant metric equivalence

C.4. The study must also include validity evidence for the relationships with other variables, as performed in the original version of the test.

C.5. If different adaptations of the test (to different languages and sociocultural contexts) are used in cross-cultural studies comparing performance in different contexts or countries, the different levels of measurement equivalence (configuration, metric and scalar) must be explored using multigroup confirmatory models.

REFERENCES

Abad, F.J., Olea, J., Ponsoda, V. & García, C. (2011). *Medición en Ciencias Sociales y de la Salud* [Measurement in Social Sciences and Health Sciences]. Madrid, Spain: Síntesis.

American Educational Research Association, American Psychological Association y National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Borsa, J. C., Damásio, B. F., & Bandeira, D. R. (2012). Adaptação e validação de instrumentos psicológicos entre culturas: Algumas considerações. *Paidéia, 22*(53), 423-432.

Martínez-Arias, R., Hernández-Lloreda, V. & Hernández-Lloreda, M.J. (2006) *Psicometría* [Psychometrics]. Madrid, Spain: Alianza.

Sireci, S. & Padilla, J.L. (2014). Validating assessment: Introduction to the Special Section. *Psicothema, 26*, 97-99.