Psicothema

Psicothema 2015, Vol. 27, No. 3, 283-289
doi: 10.7334/psicothema2014.276

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Guidelines for reporting evaluations based on observational methodology

Mariona Portell[1], M. Teresa Anguera[2], Salvador Chacón-Moscoso[3,4] and Susana Sanduvete-Chaves[3]
[1] Universitat Autònoma de Barcelona, [2] Universidad de Barcelona, [3] Universidad de Sevilla and [4] Universidad Autónoma de Chile (Chile)

Abstract

Background: Observational methodology is one of the most suitable research designs for evaluating fidelity of implementation, especially in complex interventions. However, the conduct and reporting of observational studies is hampered by the absence of specific guidelines, such as those that exist for other evaluation designs. This lack of specific guidance poses a threat to the quality and transparency of these studies and also constitutes a considerable publication hurdle. The aim of this study thus was to draw up a set of proposed guidelines for reporting evaluations based on observational methodology. **Method:** The guidelines were developed by triangulating three sources of information: observational studies performed in different fields by experts in observational methodology, reporting guidelines for general studies and studies with similar designs to observational studies, and proposals from experts in observational methodology at scientific meetings. **Results:** We produced a list of guidelines grouped into three domains: intervention and expected outcomes, methods, and results. **Conclusions:** The result is a useful, carefully crafted set of simple guidelines for conducting and reporting observational studies in the field of program evaluation.

Keywords: Program evaluation, observational methodology, designs, low intervention, reporting guidelines.

Resumen

Directrices para publicar evaluaciones basadas en metodología observacional. **Antecedentes:** la metodología observacional es una de las más apropiadas para la evaluación de la fidelidad de la implementación, especialmente en el caso de intervenciones complejas. Sin embargo, a diferencia de lo que ocurre con otros diseños evaluativos, en este caso no existe una guía que delimite los componentes necesarios a incluir en el reporte de estudios observacionales, con lo que su divulgación, transparencia y calidad podrían quedar mermadas. El objetivo de este trabajo es proponer un protocolo específico para el reporte de estudios evaluativos basados en la metodología observacional. **Método:** la idoneidad del protocolo propuesto se basa en información procedente de estudios observacionales en distintos ámbitos realizados por expertos consolidados en metodología observacional; guías generales para cualquier tipo de diseño y específicas para diseños similares a los observacionales; y propuestas de expertos recibidas en reuniones científicas. **Resultados:** se obtuvieron elementos a considerar para realizar un informe de metodología observacional, encuadrados en tres dominios: intervención y resultados esperados, método y resultados. **Conclusiones:** se presenta un protocolo útil y parsimonioso para el desarrollo y elaboración de reportes de evaluación de programas con metodología observacional.

Palabras clave: evaluación de programas, metodología observacional, diseños, baja intervención, directrices para la comunicación.

Evaluating a program requires taking evidence-based decisions, and there is an increasing call for evidence that not only focuses on the questions "What works?" and "What is the effect size?" but also seeks to explain how or why a particular program works in a particular context (Wong, Greenhalgh, Westhorp, Buckingham, & Pawson, 2013).

When observational designs based on observational methodology are used as a basis for decision-making in program evaluation, priority is given to collecting information on the behavior(s) of interest in a context of minimum intervention (Anguera, 2008; Chacón-Moscoso, Sanduvete-Chaves, Portell, & Anguera, 2013). Observational designs complement other program evaluation designs and offer several advantages:

(1) They build evidence through the close monitoring of behavior in context, providing descriptions of moment-by-moment changes and offering the possibility of drawing causal inferences based on regularities within causality models such as Mackie's INUS model (cf. Tacq, 2011).

(2) They prioritize the contextual representativeness of the data set (Anguera, 2003).

(3) They relinquish control over the experimental (high-intervention) or quasi-experimental (moderate-intervention) setting and minimize intervention by stakeholders other than program users or those close to them (Chacón-Moscoso et al., 2013).

(4) They are ideal for evaluating implementation fidelity (Dusenbury, Brannigan, Falco, & Hansen, 2003), which is the extent to which a program is delivered as intended (Durlak & Dupre, 2008). Implementation fidelity assessment is essential for

Received: December 11, 2014 • Accepted: April 28, 2015
Corresponding author: Mariona Portell
Facultad de Psicología - Edificio B
Universitat Autònoma de Barcelona
08193 Cerdanyóla del Vallès (Spain)
e-mail: mariona.portell@uab.cat

determining how and why a program works and for providing evidence on adherence to theoretically important components, such as completeness and dosage of implementation (e.g., number of sessions, duration, intensity); quality of program delivery (e.g., quality of interaction in the case of a teacher implementing a program); degree of participant engagement; aspects of program differentiation (Dusenbury et al., 2003); and fidelity of translation (adaptation of programs to the local context) (Lara et al., 2011).

(5) They are useful for evaluating complex interventions with various interacting components (Griffiths & Norman, 2013).

(6) They provide methodological solutions for obtaining quality data in the absence of standardized evaluation tools (Anguera, 2003).

Observational designs are well established in several fields, such as sport (Anguera & Hernández-Mendo, 2014), and their usefulness has been demonstrated in many others (e.g., Blanco, Sastre, & Escolano, 2010; Cerezo, Trenado, & Pons-Salvador, 2006; Gimeno, Anguera, Berzosa, & Ramírez, 2006; Herrero, 2000; Herrero & Pleguezuelos, 2008; Pérez-Tejera, Valera, & Anguera, 2011; Riberas & Losada, 2000; Roustan, Izquierdo, & Anguera, 2013). Nevertheless, there have been claims that little value is given to the extra effort expended in studies involving the direct observation of behavior (Patterson, 2008), and that it can be difficult to publish or procure funding for research on complex processes in natural settings, as this requires a shift from a standard of rigor based on experimental paradigms towards an approach favoring relevance (Rozin, 2009).

The application of reporting guidelines on observational studies would improve the completeness and transparency of reports and increase the chances of publication (Moher, Schulz, Simera, & Altman, 2010).

On reviewing the collection of reporting guidelines in the EQUATOR (Enhancing the Quality and Transparency of Health Research) Library for Health Research Reporting (Simera, Moher, Hoey, Schulz, & Altman, 2010; updated tables, 23 August 2013), we found numerous guidelines suited to moderate- and high-intervention evaluation designs, such as the CONSORT Statement for randomized clinical trials (Schulz, Altman, Moher, & CONSORT Group, 2010). We also found guidelines for methods with some similarities to observational designs because they deal with intensive repeated measurements in naturalistic settings (Stone & Shiffman, 2002), qualitative aspects (Blignault & Ritchie, 2009; Tong, Sainsbury, & Craig, 2007), or mixed methods research (Leech & Onwuegbuzie, 2010; Pluye, Gagnon, Griffiths, & Johnson-Lafleur, 2009). There were, however, no specific guidelines suited to the structural characteristics of observational methodology designs used for evaluation in low-intervention situations. The well-known STROBE (STrengthening the Reporting of Observational Studies in Epidemiology) guidelines (von Elm et al., 2007) are used for epidemiological studies such as cohort, case-control, and cross-sectional studies, and as such, are not suited to observational designs as we define them.

We believe that the lack of reporting guidelines specifically addressing evaluation studies based on observational methodology may constitute a publication hurdle, and that adherence to reporting guidelines created for other types of studies only serves to amplify the weaknesses and undermine the strengths of observational studies. The aim of this paper is to describe a set of guidelines we propose for specifically reporting evaluations based on observational methodology.

Method

To develop the proposed guidelines for reporting observational studies in the field of program evaluation, we drew on the experience of experts in the field of observational methodology and analyzed a wide range of studies that have used this methodology in different situations and contexts. We also reviewed the content and structure of (a) general reporting standards (i.e., standards that are not specific to any particular research design) (American Educational Research Association, 2006; American Psychological Association —APA—, 2010 —Journal Article Reporting Standards (JARS)—; Möhler, Bartoszek, Köpke, & Meyer, 2012; Zaza et al., 2000) and (b) reporting standards for research designs with some similarities to observational designs (Blignault & Ritchie, 2009; Leech & Onwuegbuzie, 2010; Pluye et al., 2009; Stone & Shiffman, 2002; Tong et al., 2007). Three drafts of the proposed guidelines were discussed (the first in 2011) by experts in methodological quality and observational designs at congresses of the Spanish Association of Methodology in Behavioral Sciences (AEMCCO) and the European Association of Methodology (EAM).

The main criteria underlying our guidelines are sufficiency of warrants and transparency of reporting (American Educational Research Association, 2006) to promote the presentation of sufficient, accurate, and transparent information that will enable other researchers to critically appraise and replicate the methodology used (Moher et al., 2010). The validity framework of our work was based on Chacón-Moscoso, Anguera, Sanduvete-Chaves, and Sánchez-Martín (2014).

Results

Table 1 summarizes the Guidelines for Reporting Evaluations based on Observational Methodology (GREOM), which contains 14 guidelines grouped into three domains. We deliberately omitted indications that are common to all types of evaluation designs (e.g., ethical considerations) and recommend referral to the JARS (APA, 2010) for guidance on these aspects. Below we provide a more detailed explanation of our proposed guidelines.

Domain A: Intervention and expected outcomes

(1) Fitting intervention-observational method. The aims related to this guideline are to describe the intensity of the intervention (low, moderate, or high) and justify the use of observational methodology (i.e., rationale and benefits) in the given study. Observational designs applied to program evaluation are usually associated with low-intensity programs (Anguera, 2008). They are used in situations in which a program, or part of it, is implemented (without the manipulation of specific orders or instructions) in an everyday context in which users continue with their daily activities or in which new (but non-disruptive) activities are generated. Manipulation of variables does not form part of observational methodology. Observational design elements, however, can be incorporated into moderate or high-intensity programs, but in all cases, access must be provided to a description of the intervention and its theoretical background and supporting evidence (Craig et al., 2013; Möhler et al., 2012).

(2) Outcomes. The structure of the expected outcomes has to be described and justified (Schünemann, Oxman, & Fretheim, 2006). This requires justification of the response levels chosen

Table 1
Guidelines for Reporting Evaluations based on Observational Methodology (GREOM)

Guideline number	Description
Domain A: Intervention and expected outcomes	
1. Fitting intervention-observational method	Justify choice of observational method in the context of the intervention (low-, moderate-, or high-intensity)
2. Outcomes	Describe structure of the expected outcomes in relation to program components and clarify link between outcomes and specific study measures. Justify choice of response level(s)
Domain B: Method	
3. Design	Describe study design using the extensive/intensive sub-criterion (Figure 1). Check consistency between study design and information related to guidelines 2 and 7
B1: Samples	
4. Study units	Indicate study units (participants, groups, response levels, etc.), eligibility and exclusion criteria, participant selection criteria, intended and actual sample size, and participant characteristics
5. Times	Indicate number of sessions. Specify criteria for starting/ending a session. Describe method used for within-session sampling and, in the case of follow-up designs, between-session sampling
6. Contexts	Describe attributes of research context in relation to its applicability to other contexts. Define context and setting selection criteria
B2: Instruments	
7. Observation instrument	Describe observation instrument and rationale behind its structure. Provide access to observation instrument
8. Primary recording parameters	Indicate recording units used for each code (occurrence, position within a sequence, duration). Specify type of behavioral indicator: static (e.g., frequency and duration) and, where applicable, dynamic (e.g., frequency of transition or indictors related to the sequential structure of behavior and/or the detection of T-patterns)
9. Recording instrument	Describe recording instruments and procedures
B3: Data quality control	
10. Session acceptance criteria	Specify factors taken into account to justify within- and between-session consistency and maximum allowable time-related disruptions for each session
11. Observer characteristics	Describe any observer characteristics that might have influenced observations (e.g., training/competence) and indicate relationship between the observer and the person being observed
12. Reliability	Demonstrate reliability of data set and give details of coefficient of agreement and, where applicable, generalizability theory used
Domain C: Results	
13. Flow of study units	Show flow of participants throughout the study and include information on response levels and, where appropriate, within and/or between-session monitoring
14. Analyses	Explain rationale for analyses of associations between overall measures and/or analyses aimed at identifying response patterns (e.g., lag sequential analysis, T-pattern detection and polar coordinate analysis). Specify the software used

Note. These guidelines are complementary to the Journal Articles Reporting Standards (JARS, American Psychological Association, 2010)

(perceivable components of the target behavior), which should be guided by the literature or created ex novo based on experience accumulated in the relevant context (Anguera, 2008). This information will help the reader to identify the most meaningful outcome(s) for decision-makers (Green & Glasgow, 2006) and establish a link between these outcomes and the specific measures that will be obtained with the observational study.

Domain B: Method

(3) Design. The study design must be clearly described. In observational methodology, the term *design* refers not only to what units are going to be observed and when, but also to how the data are going to be collected, organized, and analyzed (Anguera, 2008). Figure 1 shows the three dichotomous criteria that give rise to eight possible observational designs (Anguera, Blanco, & Losada, 2001). The first criterion relates to level of response or dimensionality and differentiates between unidimensional and multidimensional designs (single vs multiple levels of response). The second criterion relates to number of units and differentiates between idiographic studies, which focus on a single user or a natural group of users (e.g., a family) and nomothetic studies, which focus on groups of users. The third criterion relates to time and differentiates between single-session (point) studies and multiple-session (follow-up) studies. We recommend complementing the description of the design with a fourth criterion related to sequential data (Figure 1). This criterion distinguishes between extensive studies (focusing on static behavioral indicators, such as frequency or duration) and intensive studies (focusing on dynamic behavioral indicators or sequential data, such as frequency of transition, relative frequency of transition, or detection of T-patterns).

B1: Samples. In observational studies, it is essential to distinguish between three samples: units, times, and context.

(4) Information must be provided on the *study units* (participants, groups, response levels, or other units), the inclusion and exclusion criteria used, and the participants' characteristics (Lapresa, Álvarez, Anguera, Arana, & Garzón, 2015). Details should be given on the intended and actual sample size, including information on individuals who refuse to participate or withdraw from the study (APA, 2010; Pluye et al., 2009).

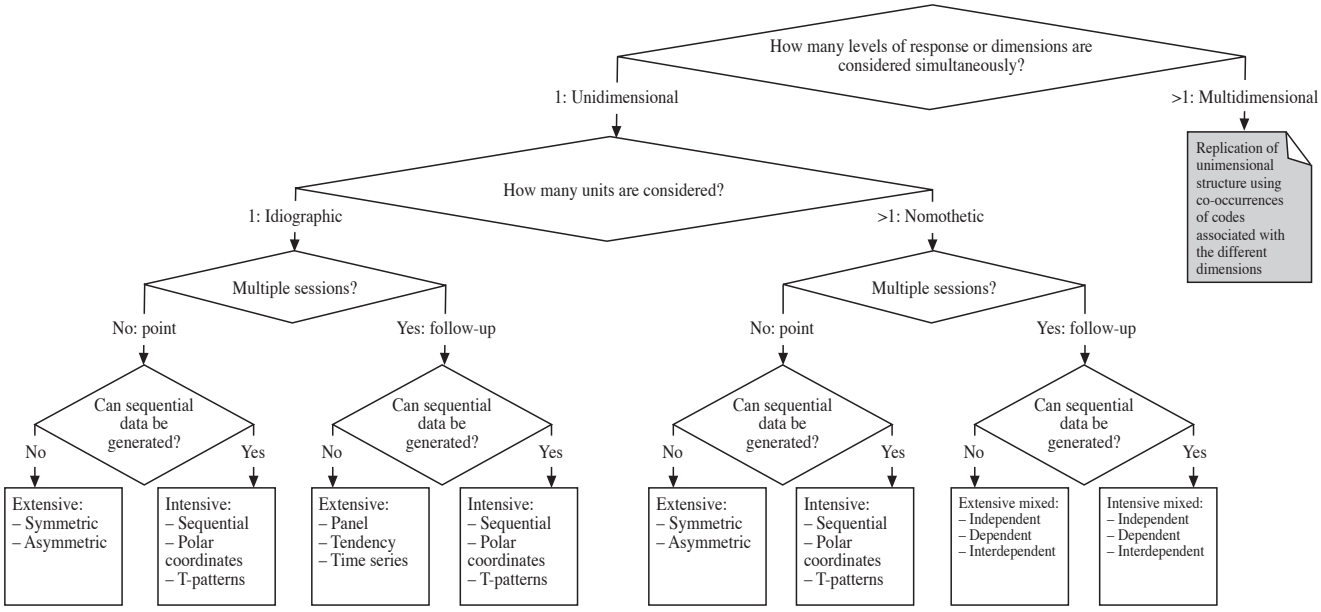

Figure 1. *Identification of observational designs. The extensive / intensive sub-criterion complements the basic observational design taxonomy that distinguish between eight designs: (1) Point/Idiographic/Unidimensional; (2) Point/Nomothetic/Unidimensional; (3) Follow-up/Idiographic/Unidimensional; (4) Follow-up/Nomothetic/Unidimensional; (5) Point/Idiographic/Multidimensional; (6) Point/Nomothetic/Multidimensional; (7) Follow-up/Idiographic/Multidimensional; (8)Follow-up/Nomothetic/Multidimensional*

(5) Times. The study report must include clear information on what moments of a session are observed and for how long (within-session sampling). Observational studies with a follow-up design can provide what Moskowitz, Russell, Sadikaj, and Sutton (2009) call *intensive repeated measures in naturalistic settings*. In this case, it is important to describe the between-session sampling method (number of sessions and criteria for starting/ending each session). The best way of obtaining representative between-session samples is through random sampling (Stone & Shiffman, 2002), and new technologies offer interesting resources for performing non-participative observational studies with randomized inter-session sampling (e.g., Mehl & Robbins, 2012).

(6) Contexts. It is essential to describe the context in which the data are recorded, with coverage of demographic, socioeconomic and cultural aspects, to justify the criteria used to choose this context and explain similarities between the study context and the context of interest. According to Pawson, Greenhalgh, Harvey, and Walshe (2005), evaluative research should answer the following questions "WHAT is it about this kind of intervention that works, for WHOM, in what CIRCUMSTANCES, in what RESPECTS, and WHY?" (p. S1:31). Such an approach places context in a central role and highlights the importance of the concept "mechanism", which is defined as the underlying processes operating in a particular context to generate outcomes of interest (Wong et al., 2013). Observational methodology includes techniques and procedures designed to capture these mechanisms in daily interactions between stakeholders (Anguera, 1999).

B2: Instruments. A transparent report will include information on the observation and recording instruments used and on the primary recording parameters.

(7) Observation instrument. The aim here is to justify the use of the observation tool (explain why it is suited to the goals of the study) and provide the reader with access to the full coding manual (e.g., in an appendix or as supplemental material). Observational methodology prioritizes the use of observation instruments that are fully adapted to the context of interest, and this generally requires the design of ad hoc tools (Anguera, 2003). The category system is the basic tool used in observational methodology, but the field format system is being increasingly used (Anguera, Magnusson, & Jonsson, 2007) to meet the needs of multidimensional designs.

(8) Primary recording parameters. When a report deals with observations made by humans (as opposed to automatic devices) using categorical codes, it is essential to clearly specify the recording units used (Anguera, 2008). For each code, the observer can record their occurrence, their position within a sequence, and/or their duration. The type of recording unit used will determine the type of behavioral indicator that the data set will produce. It is important to specify whether the study deals only with static behavioral indicators (frequency and duration) or also with dynamic indicators (e.g., frequency of transition, relative frequency of transition, or other indicators related to the sequential structure of behavior and/or the detection of T-patterns) (Bakeman & Quera, 2011; Casarrubea et al., 2015).

(9) Recording instrument. The study report must contain a description of the tools (software, etc.) and the procedures used to record the data. Several open-source software applications are now available that greatly simplify the recording of observational data. Two examples are LINCE (Gabin, Camerino, Anguera, & Castañer, 2012) and HOISAN (Hernández-Mendo, López-López, Castellano, Morales-Sánchez, & Pastrana, 2012).

B3: Data quality control. The data set can only be analyzed for the intended purpose once its quality has been established (Anguera, 2003). Researchers therefore need to report on how they controlled for factors that might affect the quality of the data set

by describing session acceptance criteria, observer characteristics, and reliability analyses.

(10) Session acceptance criteria. It is important to indicate the factors taken into account to justify within and between-session consistency and the maximum allowable time-related disruptions established for each session (Anguera, 1990).

(11) Observer characteristics. The relationship between the observer and the person being observed is one of the most important aspects that need to be described when characterizing the observer. It is essential to identify the observer within the hierarchy of stakeholders and describe the type of observation (participative, non-participative, participation-observation, or self-observation) (Anguera, 1979). Finally, information should be given on observer training and competence (Losada & Manolov, 2015).

(12) Reliability. An instrument is reliable if it produces few measurement errors and demonstrates stability, consistency, and dependency across individual scores (Blanco, 1989). Data set reliability can be analyzed qualitatively (e.g., the consensus agreement method; Anguera, 1990) or quantitatively. Interobserver agreement indices are the most widely used quantitative measures of reliability in observation studies. Additionally, the generalizability theory can be used to analyze multiple sources of variance (observers, occasions, tools, etc.) simultaneously (Blanco, 2001). These analyses can be performed using SAGT (Software for the Application of the Generalizability Theory) (Hernández-Mendo et al., 2014).

Domain C: Results

(13) Flow of study units. The study flow chart should depict the flow of participants throughout the study (including information on discontinuations and withdrawals), in addition to response levels and, where appropriate, the times at which these were studied within and between sessions.

(14) Analyses. The logic of observational methodology combines qualitative perspectives (more common in the early stages of a study) and quantitative perspectives (more common at later stages) (Portell, Anguera, Hernández-Mendo, & Jonsson, 2015; Sánchez-Algarra & Anguera, 2013). Accordingly, many JARS (APA, 2010) recommendations for the reporting of quantitative studies also apply here. The purpose in this guideline is to highlight aspects that are specific to evaluation studies based on observational designs. Thus, it is essential to differentiate between and justify analyses of relationships between overall measures and analyses designed to identify response patterns. The options available for analyzing observational data vary according to the study design and the nature of the data (Blanco, Losada, & Anguera, 2003). Sequential analysis is particularly relevant in observational methodology because it can uncover "hidden" response patterns and help to better understand the mechanisms involved in an intervention. Sequential techniques include lag sequential analysis (Bakeman & Quera, 2011), T-pattern detection (Magnusson, 2000), and polar coordinate analysis (Sackett, 1980). Both the rationale behind the analyses chosen and the software used should be specified. Open-source programs include GSEQ (Bakeman & Quera, 2011) for lag sequential analysis, Theme (Magnusson, 2000) for T-pattern detection, and HOISAN (Hernández-Mendo et al., 2012) for polar coordinate analysis.

Discussion

From a perspective that advocates methodological complementarity (Chacón-Moscoso et al., 2013, 2014), we have proposed a set of simple guidelines for conducting and reporting evaluations based on observational methodology that we hope will become a standard tool for researchers and practitioners. We believe that this step was necessary to increase awareness of the contribution of observational methodology to program evaluation, to improve the completeness and transparency of reports, and to increase the chances of publication.

Our proposal requires the inclusion of information that is not traditionally reported (e.g., the full coding manual). Reporting transparency has been limited by space constraints in journals for many years, but this is no longer a problem in many journals thanks to online supplements.

Finally, we would like to stress that our proposed GREOM guidelines constitute an initial step in a process that we hope will be enriched by contributions from other experts and studies that provide empirical evidence on the usefulness of these guidelines.

Acknowledgments

We gratefully acknowledge the support of 1) the Spanish government project *Observación de la interacción en deporte y actividad física: Avances técnicos y metodológicos en registros automatizados cualitativos-cuantitativos* (Secretaría de Estado de Investigación, Desarrollo e Innovación del Ministerio de Economía y Competitividad) for the period 2012-2015 [Grant DEP2012-32124]; 2) the Generalitat de Catalunya Research Group [*GRUP DE RECERCA E INNOVACIÓ EN DISSENYS (GRID). Tecnología i aplicació multimedia i digital als dissenys observacionals*], Grant number 2014 SGR 971; and 3) the Spanish government project PSI2011-29587, funded by Spain's Ministry of Science and Innovation, and the Chilean National Fund of Scientific and Technological Development (FONDECYT) within the project *Methodological quality and effectiveness from evidence* [reference number 1150096].

We are also extremely grateful for the helpful comments and input from 1) members of the research group *Observación de la interacción en deporte y actividad física* and the GRID group; and 2) participants at the AEMCCO (Spanish Association of Methodology in Behavioral Sciences) and EAM (European Association of Methodology) conferences who reviewed and provided feedback on the preliminary drafts of the GREOM guidelines. We also thank the two anonymous peer reviewers for their constructive comments and suggestions.

References

American Educational Research Association (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher, 35*(6), 33-40.

American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Anguera, M.T. (1979). Observational typology. *Quality & Quantity, 13*, 449-484.

Anguera, M.T. (1990). Metodología observacional [Observational methodology]. In J. Arnau, M.T. Anguera & J. Gómez (Eds.), *Metodología de la investigación en Ciencias del Comportamiento* [Methodology of the research in Behavioral Sciences] (pp. 125-236). Murcia: Secretariado de Publicaciones de la Universidad de Murcia.

Anguera, M.T. (1999). *Hacia una evaluación de la actividad cotidiana y su contexto: ¿presente o futuro para la metodología?* [Towards an evaluation of the daily activity and its context: Present or future for the methodology?]. Lecture of admission to the Reial Acadèmia de Doctors, Barcelona (1999). Reprinted in A. Bazán Ramírez & A. Arce Ferrer (Eds.), *Estrategias de evaluación y medición del comportamiento en Psicología* [Strategies of evaluation and measurement of the behavior in Psychology] (pp. 11-86). México: Instituto Tecnológico de Sonora y Universidad Autónoma de Yucatán.

Anguera, M.T. (2003). Observational methods (General). In R. Fernández-Ballesteros (Ed.), *Encyclopedia of behavioral assessment, vol. 2* (pp. 632-637). London, UK: Sage.

Anguera, M.T. (2008). Diseños evaluativos de baja intervención [Low-intervention evaluation designs]. In M.T. Anguera, S. Chacón-Moscoso & A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* [Social and health program evaluation: A methodological approach] (pp. 153-184). Madrid: Síntesis.

Anguera, M.T., Blanco, A., & Losada, J.L. (2001). Diseños observacionales, cuestiones clave en el proceso de la metodología observacional [Observational designs, key issues in the process of observational methodology]. *Metodología de las Ciencias del Comportamiento, 3*(2), 135-160.

Anguera, M.T., & Hernández-Mendo, A. (2014). Metodología observacional y psicología del deporte: estado de la cuestión [Observational methodology and sport psychology: State of the art]. *Revista de Psicología del Deporte, 23*(1), 103-109.

Anguera, M.T., Magnusson, M.S., & Jonsson, G.K. (2007). Instrumentos no estándar [Nonstandard instruments]. *Avances en Medición, 5*(1), 63-82.

Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge: Cambridge University Press.

Blanco, A. (1989). Fiabilidad y generalización de la observación conductual [Reliability and generalization of behavioral observation]. *Anuario de Psicología, 43*(4), 5-32.

Blanco, A. (2001). Generalizabilidad de observaciones uni y multifaceta: estimadores LS y ML [Generalizability of mono and multifaceted observations: LS and ML estimators]. *Metodología de las Ciencias del Comportamiento, 3*(2), 161-193.

Blanco, A., Losada, J.L., & Anguera, M.T. (2003). Analytic techniques in observational designs in environment behavior relation. *Medio Ambiente y Comportamiento Humano, 4*(2), 111-126.

Blanco, A., Sastre, S., & Escolano, E. (2010). Desarrollo ejecutivo temprano y teoría de la generalizabilidad: bebés típicos y prematuros [Executive function in early childhood and generalizability theory: Typical babies and preterm babies]. *Psicothema, 22*(2), 221-226.

Blignault, I., & Ritchie, J. (2009). Revealing the wood and the trees: Reporting qualitative research. *Health Promotion Journal of Australia, 20*(2), 140-145.

Casarrubea, M., Jonsson, G.K., Faulisi, F., Sorbera, F., Di Giovanni, G., Benigno, A., Grescimanno, G., & Magnusson, M.S. (2015). T-pattern analysis for the study of temporal structure of animal and human behavior: A comprehensive review. *Journal of Neuroscience Methods, 239*, 34-46.

Cerezo, M.A., Trenado, R.M., & Pons-Salvador, G. (2006). Interacción temprana madre-hijo y factores que afectan negativamente a la parentalidad [Early mother-infant interaction and factors negatively affecting parenting]. *Psicothema, 18*(3), 544-550.

Chacón-Moscoso, S., Anguera, M.T., Sanduvete-Chaves, S., & Sánchez-Martín, M. (2014). Methodological convergence of program evaluation designs. *Psicothema, 26*, 91-96.

Chacón-Moscoso, S., Sanduvete-Chaves, S., Portell, M., & Anguera, M.T. (2013). Reporting a program evaluation: Needs, program plan, intervention, and decisions. *International Journal of Clinical and Health Psychology, 13*, 58-66.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2013). Developing and evaluating complex interventions: The new Medical Research Council guidance. *International Journal of Nursing Studies, 50*(5), 587-592.

Durlak, J.A., & Dupre, E.P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting the implementation. *American Journal of Community Psychology, 41*, 327-350.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W.B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, *18*, 237-256.

Gabin, B., Camerino, O., Anguera, M.T., & Castañer, M. (2012). Lince: Multiplatform sport analysis software. *Procedia - Social and Behavioral Sciences, 46*, 4692-4694.

Gimeno, A., Anguera, M.T., Berzosa, A., & Ramírez, L. (2006). Detección de patrones interactivos en la comunicación de familias con hijos adolescentes [Interactive patterns detection in family communication with adolescents]. *Psicothema, 18*(4), 785-790.

Green, L.W., & Glasgow, R.E. (2006). Evaluating the relevance, generalization, and applicability of research: Issues in external validation and translation methodology. *Evaluation & the Health Professions, 29*(1), 126-153.

Griffiths, P., & Norman, I. (2013). Qualitative or quantitative? Developing and evaluating complex interventions: Time to end the paradigm war. *International Journal of Nursing Studies, 50*, 583-584.

Hernández-Mendo, A., Castellano, J., Camerino, O., Jonsson, G.K., Blanco, A., Lopes, A., & Anguera, M.T. (2014). Programas informáticos de registro, control de calidad del dato y análisis de datos [Observational software, data quality control and data analysis]. *Revista de Psicología del Deporte, 23*(1), 111-121.

Hernández-Mendo, A., López-López, J.A., Castellano, J., Morales-Sánchez, V., & Pastrana, J.L. (2012). HOISAN 1.2: programa informático para uso en metodología observacional [Software to use in observational methodology]. *Cuadernos de Psicología del Deporte, 12*(1), 55-78.

Herrero, M.L. (2000). Utilización de la técnica de coordenadas polares en el estudio de la interacción infantil en el marco escolar [Use of the technique of polar coordinates in the study of children interaction in school]. *Psicothema, 12*(2), 292-297.

Herrero, M.L., & Pleguezuelos, C. (2008). Patrones de conducta interactiva en contexto escolar multicultural [Patterns of interactive behavior in a multicultural school context]. *Psicothema, 20*(4), 945-950.

Lapresa, D., Álvarez, I., Anguera, M.T., Arana, X., & Garzón, B. (2015, in press). Comparative analysis of the use of space in 7-a-side and 8-a-side soccer: A specific illustration of how to determine the minimum sample size when using observational methodology. *Motricidade*.

Lara, M., Bryant-Stephens, T., Damitz, M., Findley, S., Gavillán, J.G., Mitchell, H., ..., Woodell, C. (2011). Balancing "fidelity" and community context in the adaptation of asthma evidence-based interventions in the "real world". *Health Promotion Practice, 12*(6 Suppl. 1), 63S-72S.

Leech, N.L., & Onwuegbuzie, A.J. (2010). Guidelines for conducting and reporting mixed research in the field of counseling and beyond. *Journal of Counseling and Development, 88*(1), 61-69.

Losada, J.L., & Manolov, R. (2015). The process of basic training, applied training, maintaining the performance of an observer. *Quality & Quantity, 49*(1), 339-347.

Magnusson, M.S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments, & Computers, 32*(1), 93-110.

Mehl, M.R., & Robbins, M.L. (2012). Naturalistic observation sampling: The Electronically Activated Recorder (EAR). In M.R. Mehl & T.S.

Conner (Eds.), *Handbook of research methods for studying daily life*. New York, NY: Guilford Press.

Moher, D., Schulz, K.F., Simera, I., & Altman, D.G. (2010). Guidance for developers of health research reporting guidelines. *PLOS Medicine, 7*(2).

Möhler, R., Bartoszek, G., Köpke, S., & Meyer, G. (2012). Proposed criteria for reporting the development and evaluation of complex interventions in healthcare (CReDECI): Guideline development. *International Journal of Nursing Studies, 49*(1), 40-46.

Moskowitz, D.S., Russell, J.A., Sadikaj, G., & Sutton, R. (2009). Measuring people intensively. *Canadian Psychology*, *50*, 131-140.

Patterson, M.L. (2008). Back to social behavior: Mining the mundane. *Basic and Applied Social Psychology, 30*, 93-101.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review - a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy, 10* (Suppl. 1), 21-34.

Pérez-Tejera, F., Valera, S., & Anguera, M.T. (2011). Un nuevo instrumento para la identificación de patrones de ocupación espacial [A new instrument to identify spatial occupancy patterns]. *Psicothema, 23*(4), 858-863.

Pluye, P., Gagnon, M.P., Griffiths, F., & Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *International Journal of Nursing Studies, 46*(4), 529-546.

Portell, M., Anguera, M.T., Hernández-Mendo, A., & Jonsson, G.K. (2015, in press). Quantifying biopsychosocial aspects in everyday contexts: An integrative methodological approach from the behavioral sciences. *Psychology Research and Behavior Management*.

Riberas, G., & Losada, J.L. (2000). Aplicación de un diseño mixto en la evaluación de la interacción comunicativa en un centro de acogida [Application of a mixed design in the evaluation of talkative interaction in a center of welcome]. *Psicothema, 12*(2), 470-473.

Roustan, M., Izquierdo, C., & Anguera, M.T. (2013). Sequential analysis of an interactive peer support group. *Psicothema, 25*(3), 396-401.

Rozin, P. (2009). What kind of empirical research should we publish, fund and reward? A different perspective. *Perspectives on Psychological Science, 4*, 435-439.

Sackett, G.P. (1980). Lag sequential analysis as a data reduction technique in social interaction research. In D.B. Sawin, R.C. Hawkins, L.O. Walker & J.H. Penticuff (Eds.), *Exceptional infant. Psychosocial risks in infant- environment transactions* (pp. 300-340). New York: Brunner/ Mazel.

Sánchez-Algarra, P., & Anguera, M.T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: Impact of recording and coding among predominating perspectives. *Quality & Quantity, 47*(2), 1237-1257.

Schulz, K.F., Altman, D.G., Moher, D., & CONSORT Group (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ, 340*, 698-702.

Schünemann, H.J., Oxman, A.D., & Fretheim, A. (2006). Improving the use of research evidence in guideline development: 6. Determining which outcomes are important. *Health Research Policy and Systems, 4*(18).

Simera, I., Moher, D., Hoey, J., Schulz, K.F., & Altman, D.G. (2010). A catalogue of reporting guidelines for health research. *European Journal of Clinical Investigation, 40*(1), 35-53. Updated tables, 23 August 2013, retrieved from: http://www.equator-network.org/wp-content/uploads/2011/10/Catalog-of-RG-update-23-August-2013.pdf.

Stone, A.A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*(3), 236-243.

Tacq, J. (2011). Causality in qualitative and quantitative research. *Quality & Quantity, 45*, 263-291.

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality and Health Care, 19*(6), 349-357.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gotzsche, P.C., & Vandenbroucke, J.P. (2007). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *PLOS Medicine, 4*(10), e296.

Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J., & Pawson, R. (2013). RAMESES publication standards: Meta-narrative reviews. *Journal of Advanced Nursing, 69*(5), 987-1004.

Zaza, S., Wright-De Agüero, L.K., Briss, P.A., Truman, B.I., Hopkins, D.P., Hennessy, M.H., ..., Pappaioanou, M. (2000). Data collection instrument and procedure for systematic reviews in the guide to community preventive services. Task Force on Community Preventive Services. *American Journal of Preventive Medicine, 18* (Suppl. 1), 44-74.