anales de psicología, 2013, vol. 29, nº 3 (octubre), 1038-1059
http://dx.doi.org/10.6018/analesps.29.3.178511

ISSN edición impresa: 0212-9728. ISSN edición web (http://revistas.um.es/analesps): 1695-2294

A classification system for research designs in psychology

Manuel Ato*, Juan J. López y Ana Benavente

Departamento de Psicología Básica y Metodología, Universidad de Murcia

Título: Un sistema de clasificación de los diseños de investigación en psicología.
Resumen: En este trabajo se elabora un marco conceptual y se desarrollan unos principios básicos para fundamentar un sistema de clasificación de los diseños de investigación más usuales en psicología basado en tres estrategias (manipulativa, asociativa y descriptiva) de donde emanan varios tipos de estudios, tres para la estrategia manipulativa (experimentales, cuasiexperimentales y de caso único), tres para la asociativa (comparativos, predictivos y explicativos) y dos para la descriptiva (observacionales y selectivos).
Palabras clave: Metodología de la investigación; diseño de la investigación; diseño experimental; diseño no experimental.

Abstract: In this work we devise a conceptual framework and develop some basic principles to promove a classification system for the most usual research designs in psychology based on three strategies (manipulative, associative and descriptive) from which emerge different types of studies, three for manipulative strategy (experimental, quasi-experimental and single-case), three for associative strategy (comparative, predictive and explanatory) and two for descriptive strategy (observational and selective).
Key words: Research methodology; research design; experimental design; non experimental design.

Introduction

It is a essential for researchers in both basic and applied Psychology, to have a conceptual framework in order to properly place their projects, to know some of the basic principles underpinning a methodological review of their research articles and to manage an array of potential designs to plan their research appropriately. The main aim of this work is to present both a conceptual framework to evaluate the research process and a classification of the most common designs used in our research areas.

The three pillars of the research process

It is interesting to examine the general research process with the scheme proposed by Kline (2009, see also Pedhazur & Smelkin, 1991), who distinguish three methodological pillars supporting this process: design, measurement and analysis, closely related to four forms of validity of the research postulated by Campbell et al. (See Shadish, Cook & Campbell, 2002), namely: internal validity, statistical, construct and external validity (see Figure 1).

The first pillar of the research process is design, defined as a plan providing a framework for integrating all elements of an empirical study so that results are credible, unbiased and generalizable (Dannels, 2010). Research design is responsible for crucial aspects such as selection and assignment of participants, optimal deployment of experimental conditions (structure of treatments), control of extraneous variables that may be present in the research context (structure of control) and minimization of the error variance (structure of error). Two types of validity determine the quality of the design application: internal validity (the ability to control the effect of third variables that may be alternative causes to the investigated cause) and external validity (the ability to generalize the results to other participants, contexts and times). An appropriate balance between internal and external validity is one of the most desirable aims in an optimal research design.

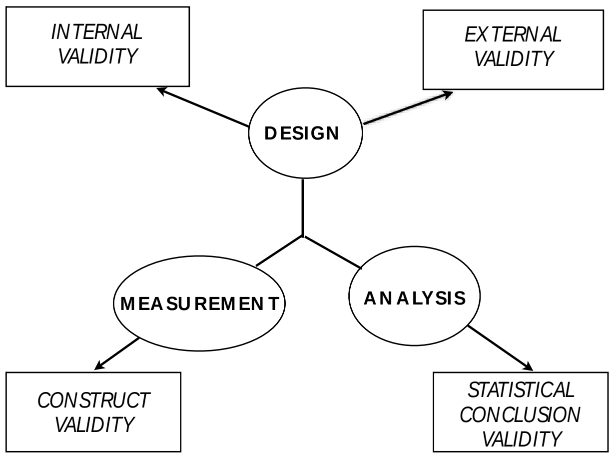

Figure 1. The three pillars in the research process

The second pillar of the research process is measurement, which begins with the identification, definition and measurement of variables, and ends with the generation of empirical data that are the input of statistical analysis procedures (see Martínez-Arias, Hernández & Hernández-Lloreda, 2006). It is critical that the empirical data have a high degree of reliability. The type of validity related to the measure is construct validity (the ability to properly define and operate research variables).

The third pillar of the process is analysis, which concerns the estimation of parameters and the hypothesis test regarding the proposed research objective (s) with the most appropriate statistical procedures. We must bear in mind that the statistical hypotheses test is not primary and it is currently common to attach greater importance to accuracy of parameter estimation and to the global fit of models than to hypothesis testing (Maxwell, Kelley And Raush, 2008; Maxwell

*** Dirección para correspondencia [Correspondence address]:**
Manuel Ato. Departamento de Psicología Básica y Metodología. Universidad de Murcia. Campus de Espinardo, Facultad de Psicología. 30100 Espinardo (Murcia, Spain). E-mail: matogar@um.es

& Delaney, 2004), and assess aspects such as effect size and the practical and clinical significance of results (Thompson, 2002a). The type of validity related to the pillar of analysis is validity of the statistical conclusion (whether the statistical analysis procedure used is correct and if the value of estimates approaches that of the population).

It is important to point out that, from a methodological viewpoint, even if the research is perfect in its substantive conception, it can be ruined if the methodological pillars of design, measurement and analysis are not properly used. Contrary to what many researchers believe, the use of sophisticated statistical techniques does not improve the research results if it was poorly designed or if appropriate measures were not used. It is then crucial for the researcher to focus on the selection and application of an appropriate design, valuing its potentialities and drawbacks to achieve the greatest degree of balance between internal and external validity.

Basic principles of the research design and evaluation process

There are some general principles that many reviewers of research articles often use as a guide to ensuring a coherent research process and can help the researcher to generate a report and objectively assess their work (see a more complete presentation of such principles in Light, Singer & Willet, 1990, Kline, 2009 and many chapters of the Hancock & Mueller Reviewer's Guide, 2010):

1) All research is designed to respond to one (or more) specific objectives. The reviewers of a research report hope to find direct correspondence between the research problem and the specific design used in its potential solution.
2) The most important methodological reason justifying the publication of a research article is to enable its replication. The reviewers expect that the research report will contain all relevant material to facilitate other researchers' replication of their work and to allow the application of methods of study integration (meta-analysis) in order to contribute to the accumulation of scientific knowledge.
3) With regard to the measurement of research variables, a research article should include precise information on the instruments of data collection and the metric nature of the empirical data, including the definition and operationalization of the variables, together with other technical aspects such as reliability, validity and cut-off scores of the data collection instruments used (Knapp & Mueller, 2010).
4) Regarding design, the research report should include a detailed description of the participants, the processes of selection and assignment or group membership, the context in which the work is performed and the procedures for controlling the potential foreign variables as well as an assessment of the generality of its findings, among other relevant issues. It should be noted that many aspects of design are not adequately addressed in research reports, especially in the case of quasi-experimental and non-experimental designs (see Vandenbroucke et al., 2007; Jarde, Losilla & Vives, 2012). The most interesting questions of research design that should grab the attention of researcher and evaluator are the following:

- The selection of the sample of participants, in particular their number and representativity regarding to the population, has important consequences for both the power (validity of the statistical conclusion) and the generalization of the results (external validity). The justification of optimal sample size has usually been treated within power analysis and the tradition of hypothesis testing (Bono & Arnau, 1995; Kraemer & Thieman, 1987; Lenth, 2001). It is most advisable to determine the optimal sample size by performing a prospective power analysis before using the software (eg G * Power3, see Faul, Erdfelder, Lang & Buchner, 2007), but the applied researcher must know that some statistical packages (e.g., SPSS) use retrospective, observed or post-hoc power analysis (after the study), a practice that some defend (e.g., Lenth, 2007; Onwuegbuzie & Leech, 2004), while others hold as unacceptable (e.g., Hoenig & Heisey, 2001; Levine & Ensom, 2001). It is highly advisable for the researcher to plan their study with a prospective power analysis and then check results with a retrospective power analysis, particularly where their hypothesis was not significant (Balkin & Sheperis, 2011).
- The definition of variables, independent (predictors or probable causes), dependent (responses or probable effects) and other additional variables (covariates) proposed to respond to the aims of the research is another key design issue, from where other important aspects of the process derive, such as the optimum number of independent variables to be included, their metrical (numerical or categorical) nature, whether they are manipulated or simply observed, the definition of groups or conditions of treatment, intervention or classification, if necessary, and in such case the assignment of participants to the groups, which may be random, not random but known and not random.
- The control of extraneous variables is another crucial question in a research design that receives insufficient attention in non-experimental studies, since in experimental studies manipulation of the independent variable/s and random assignment allow the researcher to balance the effect of third variables and analytically address cause and effect relationships. The random assignment of participants to treatment groups can control the influence of third variables, identified and unidentified when ideal conditions exist, but it is well known that when some foreign variable has a more powerful impact on the response variable rather than the causal, the randomization effect usually fades. It is therefore essential to strive to identify the third variables that can potentially cause confusion. The methodological status of a research design depends to a great extent on the degree of effec-

tive control of the foreign variables (Shadish, Cook & Campbell, 2002). Although there is an abundant array of techniques for controlling foreign variables (see Ato, 1991 and Ato & Vallejo, 2015), in general, experimental control techniques (elimination, constancy and various forms of equilibration, such as randomization, pairing and blocking), if they can be used, are preferable to statistical control techniques (in particular, standardization, adjustment and residualisation). There are other appropriate control techniques in research areas such as social psychology and clinical psychology (e.g., single / partial / double blind procedures or use of quasi-control groups, see Kirk, 2013, pp. 22-23), and other more promising specific techniques, such as propensity scores (see Austin, 2011; Shadish & Clark, 2004) and the technique of instrumental variables (Bollen, 2012).

- When experimental research and some non-experimental designs are applied it is common to distinguish three independent components of the variance. The primary variance is attributed to the independent variables that form the fundamental nucleus of the research, the secondary variance is attributed to other variables other than the fundamental that the researcher must strive to control, and the error variance represents the remainder of variation that is neither primary nor secondary variance and which may take simple forms (e.g., a level or a single error component) or complex (e.g., more than one level or several error components). The MAXMINCON principle (Kerlinger, 1985) is a general principle of research design whose goal is to achieve maximization of primary variance, minimization of error variance and control of secondary variance.

5) For statistical analysis, the statistical methods must be used in sufficient detail to be understood by other researchers, as well as the procedures used for the treatment of missing data (Graham, 2012) and compliance of assumptions on which they are based (Garson, 2015). It is important to note that, given a research design, multiple analytical procedures are possible in response to a research problem. In this case, the researcher must justify the selection of the procedure used. Hancock & Mueller (2010) proposed, for each of the most common statistical procedures in Psychology, a set of desiderata that should fulfill its application.

Referring to the rigorous use of statistical procedures in research reports to use in psychology, a study by Bakker and Wicherts (2011) used a random sample of 281 research articles published in impact journals and found that almost 20% of the statistical results are incorrectly reported and the error prevalence was significantly higher in low impact journals. In 15% of published articles, there was even significant evidence in favour of the author's (s) hypothesis which on being recalculated was found to be not significant.

It is also worth noting that many researchers, after applying statistical techniques, only give interest to p values of significance, within the tradition of the contrast of statistical hypotheses ("Null Hypothesis Significance Testing", NHST approach). This tradition has generated much controversy (see Balluerka, Gómez & Hidalgo, 2005; Harlow, Mulaik & Steiger, 1997; Kline, 2004; Pascual, Frías & García, 2004) and has led to the suspicion that it may even delay the accumulation of knowledge (See Schmidt, 1996; Gliner, Leech & Morgan, 2002). In 1996, the APA established a Committee to deal with questions on the application of statistical methods in psychological research. A paper by Wilkinson and the Task Force on Statistical Inference (1999) suggested including in any research report, in addition to the classic significance tests, measures of effect size, measurement indicators or uncertainties by means of confidence intervals and a selection of appropriate graphics. This has also been recommended in recent editions of the Publication Manual of the American Psychological Association (APA Manual, 2010), but most papers centralize all results around p values, and very few use point estimates with confidence intervals and appropriate graphics (Cumming et al., 2007; Cumming, 2012).

The inclusion of effect size indicators has been repeatedly recommended to assess the practical significance of a result against statistical significance and, in health sciences, to clinical significance (Kirk, 2005; Thompson, 2002a; Pardo & Ferrer, 2013). Although many have been developed (recent reviews can be found in Ellis, 2010 and Grissom & Kim, 2012), two basic types of indicators have been highlighted (Rosnow & Rosenthal, 2009): d-type (based on standardized mean difference) and r-type indicators (based on correlation and the ratio of explained variance). In both cases, the most usual interpretation utilizes the labels proposed by Cohen (1988), "low" ($r = 0.1$), "moderate" ($r = 0.3$) or "high" ($r = 0.5$), although this practice should be abandoned in favour of using confidence intervals and assessing the discrepancy between p-values along with magnitude effect indicators (see Sun, Pan & Wang, 2010), as well as trying to compare indicators that address a common problem within the same study and among different studies. In a recent review, Peng, Chen, Chiang & Chiang (2013) appreciate the significant increase in the use of indicators of magnitude of effect in recent years, but criticize the persistence of some inappropriate practices in research reports such as: (a) the use of unadjusted (or biased by sample size) indicators, coupled with a reduced use of other less biased and more desirable indicators such as omega square (ω^2) and intraclass correlation ρ_I (Ivarsson, Andersen, Johnson & Lindwall, 2013); (b) the use of Cohen's labels without any contextualisation or interpretation (Valentine & Coopers, 2003); (c) the persistent lack of clarity in the use of standard indicators η^2 and η_P^2 in the outputs of some statistical packages (eg, SPSS), which coincide in one-way ANOVA models but differ in factorial ANOVA models (Pierce, Block & Aguinis, 2004 and Richardson, 2011), (d) confidence intervals for both statistical test estimators and effect sizes (see Kelley & Preacher, 2012 and Thompson, 2007), and (d) the lack of integration between statistical significance tests and size effect indicators. In this respect, Levin & Robinson (2000) suggest using the so-called coherence of statistical conclusion when statistical significance and effect size coincide in the same result.

An analytical alternative to the NHST approach emerged as a silent revolution against the NHST approach at the end of the 20th century (Rodgers, 2010) and is gaining popularity among researchers (Judd, McClelland & Ryan, 2009; Maxwell & Delaney, 2004). Its aim lies in the comparison and adjustment of probability models, where the model as a whole (as opposed to the contrast of a particular hypothesis) plays a fundamental role (Kaplan, 2009). This new approach applies a process in four stages: 1) the specification of an equation (or a set of equations) with the basic ingredients to be included in the model, 2) the adjustment of the model according to pre-established criteria and its re-specification if is not properly adjusted, 3) the evaluation of its statistical assumptions, and 4) the interpretation of the finally accepted model. In contrast to the NHST approach, a model has only interpretative interest when it is acceptable to the empirical data (Ato & Vallejo, 2015, Losilla, Navarro, Palmer, Rodrigo & Ato, 2005). Table 1 presents a classification of the most common modeling structures in psychological research as a function of the number of equations required by their specification, the distribution that follows the response variable and the number of error terms needed.

Table 1. Modeling structures in psychological research

Modeling structure	Model name	Number of equations	Response variable distribution	Number of random terms
LM	Classical linear model	One	Normal	One
GLM	Generalized linear model	One	Exponential	One
LMM	Linear mixed model	One	Normal	More than one
GLMM	Generalized linear mixed model	One	Exponential	More than one
SEM	Structural Equation model	More than one	Exponential	More than one

The LM (classical linear model) structure models are the most well-known and are characterized by using one (or more) response variable (s) with normal distribution and specifying a single equation with one or more fixed components and a single error term (Agresti, 2015). This structure includes the analytical procedures of regression analysis, ANOVA and ANCOVA (for a response variable) and multivariate regression, MANOVA and MANCOVA (for more than one response variable). The GLM (Generalized Linear Model) structure models can use exponential family response variables (which include the normal distribution as a particular case), but also specify a single equation with an error term. This structure includes lesser known analytical procedures such as log-linear analysis, logistic regression, Poisson regression and logit analysis (Agresti, 2013, Ato & López, 1986; Hardin & Hilbe, 2007). LMM (mixed linear model) models also require response variables with normal distribution and an equation with all fixed and random components that are desired (West, Welch & Galecki, 2007). This structure also includes multi-level analysis procedures (Gelman, 2006; Hox, 2011; Snijders & Bosker, 2011). The generalization of the LMM structure for exponential family response variables defines the GLMM (generalized linear mixed model) structure models, which may include any combination of fixed and random effects (Hedeker, 2005). The SEM models also use any distribution of the exponential family but require the specification of several equations, each with its own error term (Kline, 2011; Kaplan, 2009).

Classification of Empirical Research Designs in Psychology

It is essential for a researcher to know the design very well in order to understand the principles of its application and the key aspects to highlight in their report (following the basic principles and conceptual framework developed above). Although several systems of classification of empirical research designs in Psychology have been proposed (e.g., Montero & León, 2007, Martínez-Arias, Castellanos & Chacón, 2014), and starting from the principle that any system of classification will be ambiguous, our proposal to classify research in psychology distinguishes between strategies, studies and research designs and is structured as follows :

- Theoretical Research
- Instrumental Research
- Methodological Research
- Empirical Research

Theoretical research

This category includes all papers compiling advances produced in substantive theory or methodology on a specific research topic, as well as research reviews or updates not requiring the use of original empirical data from primary studies. We have excluded theoretical subjective reflection works not based on a detailed review of other authors' findings.

Theoretical research can take one of three possible forms:

- The narrative review is a revision or theoretical update of primary studies on a research topic, rigorous but merely subjective, without any empirical contribution by the researcher (eg, Sánchez, Ortega & Menesini, 2012).
- The systematic review is a revision or theoretical update of primary studies, with a systematic development of the process of accumulation of data (selection of studies, codification of variables, etc.), but where statistical procedures are not used to integrate the studies (Eg, Orgilés, Méndez, Rosa & English, 2003; Rosa, Iniesta & Rosa, 2012).
- The quantitative systematic review or meta-analysis is an integration of primary studies with quantitative methodology (see Sánchez-Meca & Marín-Martínez, 2010), containing both a systematic development of the data accumulation process and the use of statistical methods to integrate the studies (e.g., Rosa, Olivares & Sánchez-Meca, 1999; Sánchez-Meca, Rosa & Olivares, 2004).

Instrumental research

This category includes all works analysing the psychometric properties of psychological measurement instruments, either new tests, for which it is recommended to follow the tests validation standards developed jointly by the American Educational Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurement in Education (NCME), published in their latest edition in 2014, or the translation and adaptation of existing tests for which it is recommended to follow the guidelines proposed by the International Test Commission (ITC, see Muñiz, Elosua & Hambleton, 2013).

It is strongly recommended that authors of instrumental research also read the "*Guide for the presentation of psychometric tests validation tests in Psychology, Education and Social Sciences*", which can be found on the website of the journal *Anales de Psicología / Annals of Psychology*.

Methodological research

This category includes all papers presenting new methodologies for the correct treatment of any topic related to the aforementioned three basic pillars of empirical research : design (e.g., Jarde, Losilla and Vives, 2012), measurement (Cuesta, Fonseca, Vallejo & Muñiz, 2013) and analysis (e.g., Cajón, Gervilla & Palmer, 2012, Vallejo, Ato, Fernández & Livacic, 1996) and review of methodological procedures in use (e.g., Barrada, 2012).

Empirical research

To respond to research problems in Psychology, three generally accepted strategies are commonly used (see Light, Singer & Willett, 1990; Arnau, 1995a): manipulative, associative and descriptive strategies. The first is the set of studies common in experimental research; the others comprise a second set of studies that together represent non-experimental research. These correspond in great measure to the three procedures developed in other sciences for testing theories through causal explanation, prediction, and description (Pearl, 2009).

The manipulative strategy aims at analysing a causal relationship (through the formulation of causal hypotheses) between two or more variables and can adopt one of three types of studies: experimental, quasi-experimental and single case. Experimental studies represent the ideal of research and must meet two requirements: 1) at least one variable must be manipulated and 2) the participants should be randomly assigned to the levels of the manipulated variable. The first requirement is essential in studies of manipulative strategy; the latter is often lacking in quasi-experimental studies and in most single case studies used in applied contexts.

The associative strategy seeks to explore the functional relationship between variables (formulation of covariation hypothesis) and can adopt three types of studies depending on whether the object of the exploration is the comparison of groups (a comparative study, also known as observational), the prediction of behaviours and / or group classification (predictive study) or the test of theoretical models for their integration into an underlying theory (explanatory study). The distinction between causal explanation and empirical prediction is treated in Shmueli (2010). Causal analysis is sometimes possible in some comparative and explanatory studies, but not in predictive studies (Schneider, Carnoy, Kilpatrick, Schmidt & Shavelson, 2007).

The descriptive strategy is intended to describe events as they occur, without any manipulation of variables, nor comparison of groups, nor prediction of behaviours, nor testing models, and can adopt two types of studies: observational and selective.

In the classification presented in this paper it is assumed that each research design corresponds to a study belonging to a specific strategy, depending on the aim. In their basic forms, taken in this work together with some of their most common generalizations, each design cited has its own characteristics distinguishing it from others. It should be noted, however, that in practice, designs are not usually presented in basic form and not often flexible to be adapted to particular research, but in this work we only deal with the basic forms, as presented in several entries in the Encyclopedia edited by Salkind (2010). Figure 2 summarizes our proposed classification of research designs, encompassed in a type of study that has been generated under a particular strategy.

A. *Manipulative strategy*

A.1. Experimental studies

Experimental research has evolved until now from the evolution of two great research traditions: the classic laboratory tradition, typical of the natural sciences and based on intraindividual variability, and the most modern statistical tradition of the field, typical of the Social sciences and based on interindividual variability (Ato, 1995a; Cook & Campbell, 1986). From the former there is a characteristic form of experimental research practised in some areas of applied psychology associated with single case designs. To the second belongs the experimental research methodology practised mainly in basic psychology and some applied areas linked to experimental designs. In those areas where it is not possible or ethical to apply experimental methodology, an alternative form of research has been developed with the name of quasiexperimental designs.

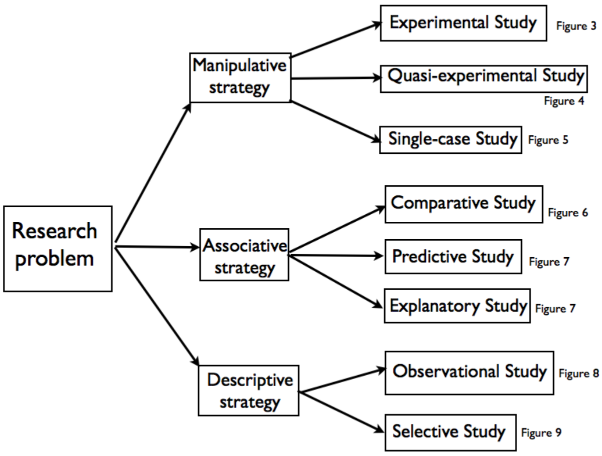

Figure 2. Strategies of research in Psychology

Experimental designs (or their equivalent in health sciences, randomized controlled trials) are characterized by the fulfillment of two essential requirements in a rigorous investigation: 1) manipulation of at least one treatment independent variable, and 2) control by randomization of the potential foreign variables in order to guarantee the initial equivalence of the groups. These two requirements are justified when the researcher pursues the analysis of cause and effect relationships between variables. Quasi-experimental designs only meet the first requirement (manipulation), but are usually applied in situations where it is not possible or is unethical to comply with the second (control by randomization), although they may also be proposed for the analysis of cause and effect relationships provided that alternative procedures for controlling foreign variables are used that fulfill their purpose. The credibility of the cause-effect relationship under analysis depends on the rigour and effectiveness of control procedures (Pearl, 2009). Three basic criteria are considered necessary to approach conclusions of the cause-effect type (Bollen, 1989; Cook & Campbell, 1979), namely the association criterion (existence of covariation between cause and effect), the criterion of direction of the cause) and the criterion of isolation (absence of confusion or spuriousness). But note that there are two different, though essentially compatible approaches to explain causal inference in applied research (Shadish, 2010): Campbell's classic tradition (see Shadish, Cook & Campbell, 2002), used in almost all Social Sciences and Psychology in particular, and Rubin's statistical tradition (see Rubin, 2004), more common in Health Sciences and other experimental sciences.

A.1.1 Experimental designs

The most popular classification of experimental designs used in Psychology is based on the comparison strategy that allows and distinguishes between subjects designs, which analyze the differences between averages of randomly administered treatments to groups with different participants, within subjects designs, that analyze the differences between averages of treatments administered to groups with the same participants, and mixed designs, which combine both forms of comparison.

An alternative classification, less popular but more ingrained from the model comparison approach (Ader & Mellenbergh, 1999; Bailey, 2008; McConway, Jones & Taylor, 1999; Milliken & Johnson, 2009) begins with the formulation of a structural model, each of whose elements is associated with one of three types of structure. The advantage of this classification lies in the optimum fusion of the pillars of the design and statistical analysis, allowing a better understanding of the nature of the design and clearly distinguishing two aspects often confused: the research design and the statistical model used to analyze it (Ato & Vallejo, 2015).

1) The first structure, called treatment structure, concerns the number and way of grouping treatments into a piece of research and allows distinguishing between a simple (a single factor of treatment) and factorial structure (more than one treatment factor), and within it, either a crossing between two or more factors (whether all levels of one factor can be combined with all levels of the other) or a nesting relationship (if all levels of a factor can only be combined with one and only one of the levels of the other).

2) The second, called control structure, refers to the number and way of grouping the experimental units in order to control potential foreign variables, and can use one of two possible options, experimental control (using control techniques such as randomization, pairing, and blocking) or statistical control (using statistical adjustment techniques such as residuals or the use of covariates). An essential requirement determining the combination of a treatment structure with a control structure is that any interaction between elements of the structure of the treatments and the structure of the control is assumed to be non-significant.

3) A third additional structure, called error structure, refers to the number of different sizes (or levels of aggregation) of the experimental units and associated error components that are postulated in the statistical model associated with the research design, and distinguishes between single error structure (the model has only one single size or aggregation level of experimental units and therefore one single error term) and multiple (the model has two or more sizes or aggregation levels and therefore several terms of error). For example, in research designed to evaluate the response of patients belonging to different diagnostic categories, it is necessary to distinguish the variation of the patients from the variation of the diagnostic categories. Both types of variation constitute different types of error and require different aggregation sizes / levels.

Given an error structure, the most common experimental designs in psychological research may be defined in this context as a peculiar combination of the treatment and control structures depicted in Figure 3. Assuming in principle a single error structure, the combination of a simple treatment structure (a treatment factor, defined with fixed or random effects) and an experimental control structure through full randomisation (where each subject in the sample randomly receives a treatment) represents the Completely Randomized Design (CRD, as in Briñol, Becerra, Gallardo, Horcajo & Valle, 2004), the most basic experimental design from which all others are constructed, and its statistical model has two sources of variation (a treatment effect and a single residual error component).

If the researcher also wishes to control a foreign variable through experimental control (e.g., pairing or blocking), the model must use restricted randomization (subjects in each block are randomly assigned to one treatment) and include an additional term for the factor to be controlled, and the result is the Randomized Block Design (RBD, as in Valiña, Seoane, Ferraces and Martín, 1995). Statistical analysis of the RBD requires pretesting the compliance of the assumption of additivity between the treatment and control structures. The test can be done directly, if it is a replicated design, or by the Tukey additive test (Alin & Kurt, 2006; Tukey, 1949) if it is a non-replicated design (Ato & Vallejo, 2015). It is important to bear in mind that in the RBD the blocking factor usually lacks substantive interest (i.e. it is proposed to minimise error variance making the treatment test effect more sensitive) but in any basic design, as well as the manipulated variable ,we can introduce other classification variables (the most common are sex, socioeconomic level, civil status, etc.) with a purely substantive interest, in these cases , they are not strictly blocking variables and therefore do not require verification of assumption of additivity.

Experimental control through blocking can be generalized by including two (or more) blocking factors, in which case the model must incorporate one term for each factor and another for interaction. If we wish to control two extraneous variables with the same number of levels, the RBD with two block variables can be simplified using experimental double-block control, which does not require any interaction terms, resulting in the Latin Square Design (LSD), rarely used in psychological research. However, if we wish to control a foreign variable by statistical control with a numerical covariant, the model must include an additional term for the covariant, and the result is the concomitant variables design (CVD as in Fernández-Berrocal & Extremera, 2006). Statistical control allows incorporating two or more numerical covariates and the corresponding statistical model must include an additional term for each covariant (Ato & Vallejo, 2015). The use of the RBD and LSD designs requires prior verification of the assumption of additivity between the treatment and control structures with the Tukey additive test (Alin & Kurt, 2006; Tukey, 1949); For its part, the use of the CVD design requires verification of the homogeneity assumption of the regression slopes (see Huitema, 2011).

A common feature of these four basic designs is that they are between subjects designs as each subject contributes one observation per response variable. Kirk (2013) considers CRD, RBD and LSD as the basic structures of between experimental design, for which, assuming a normally distributed response variable and a single error term, the procedures ANOVA and MANOVA are appropriate. In our opinion, CVD should also be considered as another basic design structure for which ANCOVA and MANCOVA procedures are appropriate. Both are still poorly understood and poorly used (Huitema, 2011; Miller & Chapman, 2001; Milliken & Johnson, 2009).

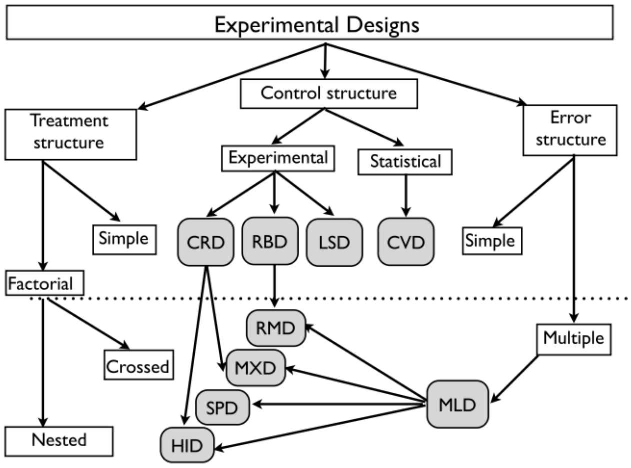

Figure 3. Experimental Designs

CRD: Completely randomized design.
RBD: Randomized blocks design.
LSD: Latin Square design.
CVD: Concomitant variables design.
RMD: Repeated measures design.
MXD: Between.within or mixed design
SPD: Split-plot design.
HID: Hierarchical design.
MLD: Multilevel design.

A simple extension of the RBD considers each experimental unit to be a level independent from a blocking factor, in which case a specific form of experimental control called within blocking is produced and the result is the Repeated Measures Design (RMD), where treatments are administered to all participants and each subject contributes more than one observation per response variable. The RMD with a treatment factor represents the basic within design as it combines a simple treatment structure with an experimental control structure by within subjects blocking and a structure with two error components referring to individual differences between the subjects and to the differences between intrasubject measures. In comparison to the basic CRD of similar conditions, the RMD tends to significantly reduce error variance and increase statistical power. However, since the different measures of the same experimental unit are not considered independent, statistical analysis should not be approached without checking whether sphericity assumption is verified (Ato & Vallejo, 2015).

A variety of more complex between subjects experimental designs can be defined by assuming a single error structure and generalizing the four between subject basic designs (CRD, RBD, LSD and CVD) to factorial treatment structures in crossover relationship, also allowing the use of fixed treatments (treatments are chosen arbitrarily by the researcher ignoring other possible treatments of the population), random (treatments are randomly selected from the set of treatments of the population) or mixed (some fixed and other random treatments). In all cases, generalization presents greater difficulties in the understanding of interactions, which are generally prone to error of interpretation by researchers (Meyer, 1991; Pardo, Garrido, Ruiz & San Martín, 2007).

The basic repeated measures design can also be generalized to factorial treatment structures, where each experimental unit represents a block receiving all treatment combinations and the set of units constitutes a unique group (e.g., Blanca, Luna, López, Rando & Zalabardo, 2001). When the set of experimental units is also divided into groups based on some variable of interest for research, the result is the between-within or mixed design (MXD), so called as it consists of between subject and within subject parts, each representing a different aggregation size or level of experimental unit and therefore a proper error term (e.g., Ruiz-Vargas & Cuevas, 1999). The MXD design is very similar to that in agricultural and biological research called Split Plot Design (SPD). In the experimental forms of RMD and MXD designs it is assumed that the order of treatment of within subject levels is established randomly or at least using some assignment sequence that guarantees group equivalence. Where the latter cannot be randomly assigned or guaranteed (for example, when a temporal sequence is involved) RMD should be treated as a quasi-experimental design or, if none of the variables are considered manipulated, as a comparative design.

A different factorial structure is required when factors are in a nesting relationship, which can occur with both treatment and control structures. In both cases, experimental units of some level of aggregation are nested within other higher level units, with their corresponding error terms and the result is the Hierarchical Design (HID, e.g., García-Sánchez & Rodríguez, 2007). The generalization of this situation to any number of variables employing two or more experimental unit sizes or some nesting relationship leads to the vast family of multilevel designs (MLD), which also include the two forms of repeated measures design (RMD and MXD), as they employ two levels of aggregation of experimental units and are therefore multiple error structures. It is possible to distinguish two basic multilevel designs where smaller units (e.g., participants, patients, customers) are nestled within larger units (e.g., classes, clinics, companies): in the first, subjects within each group are randomly assigned to treatments; in the second, groups as a whole are randomly assigned to treatments. For the treatment of multilevel data with experimental designs the most recommended references are Dziak, Nahum-Shani & Collins (2012), Hoffman & Rovine (2007) and Milliken & Johnson (2009).

The type of statistical analysis applied to experimental designs is another issue of interest to the applied researcher. Within the context of statistical modeling, several analytical approaches are distinguished. For intersubjective designs the most popular is the univariate approach, which uses the procedures of the classical linear model (regression, ANOVA and ANCOVA), where all factors are assumed fixed. In very simple situations and high regularity conditions, some statistical packages also include random factors (although the resulting model would rather have the structure of a mixed linear model), but for other more complex situations, such as those derived from hierarchical and in general multilevel designs, the univariate approach is not recommended due to the problems it brings (see Ato, Vallejo & Palmer, 2013; Quené & van den Bergh, 2004). For within subject designs the univariate approach is valid only if the sphericity assumption is met (Ato & Vallejo, 2015). The alternative to the univariate approach to analyze within subjects designs was in its time the multivariate approach, but although it is still quite popular today , in academic circles the mixed approach is considered more appropriate (Maxwell & Delaney,

2004; Milliken & Johnson, 2009; Vallejo & Fernández, 1995), a more complex approach that is already available to users of all professional statistical packages (particularly, GENSTAT, SAS, SPSS, STATA and R), but is more suitable for analyzing between and within subjects experimental designs (Ato & Vallejo, 2015).

A.1.2. Quasi-experimental designs

Quasi-experimental designs have the same aim as those which are experimental, i.e. establishing causal relationships, and meeting the requirement of manipulation of at least one VI, but it is not possible (nor ethical) to meet the requirement of random assignment to ensure there are no differences between groups before assigning a treatment or program. As a consequence, we cannot guarantee the equivalence of groups before treatment (selection bias) or the exclusion of third variables that may explain the treatment effect (spurious or confounding). To compensate for this bias, it is common to turn to the use of treatment groups, control groups, pretest and posttest measures and other experimental control techniques (e.g., matching methods, see Stuart & Rubin, 2008) and statistical control techniques (e.g., use of covariates, See Steiner, Cook, Shadish & Clark, 2010) in order to balance preexisting differences between groups.

However, in certain areas of applied psychology quasi-experimental designs are more commonly used than other design alternatives. Obviously, as a consequence of the initial non-equivalence of the groups, causal analysis presents many more difficulties in quasi-experimental designs than in experimental. Even so, there are some statistical procedures, such as propensity scores (Rosenbaum & Rubin, 1983; see also Luellen, Shadish & Clark, 2005), whose purpose is to construct a function of all explanatory variables estimating probability for each participant to belong to a group, in order to achieve (through pairing or statistical adjustment) what would have been obtained had randomization been used.

There are many varieties of quasi-experimental designs. Shadish, Cook and Campbell (2002) present a detailed classification of these. A shorter presentation can be found in Ato (1995a), Vallejo (1995a) and Ato & Vallejo (2015). Following Judd and Kenny (1981), the classification used here requires the researcher to first decide between two forms of comparison of treatments, cross-sectional (where the essential comparison is between subjects or between non-equivalent groups) or longitudinal (essential comparison is within subjects or between multiple measures) and secondly on how to use non-random assignment (known or unknown).

It is possible to distinguish two basic modules, generally called preexperimental designs, from which all quasi-experimental ones are constructed. The first module is the posttest only design (POD), which has an experimental group that the program, intervention or treatment is administered to, and another control group, but lacks pretest measures, and therefore only uses between subject comparisons. The second module is the pretest-posttest design (PPD), which has a single pretest and posttest group of treatment, and so only uses within subject comparisons.

Among the quasi-experimental cross-sectional designs, the two prototypical cases are the non-equivalent groups design (NEGD), representing a special combination of both preexperimental modules cited, where assignment is neither random nor known (eg Pons, González and Serrano, 2008), and the regression discontinuity design (RDD), similar to the NEGD design but where allocation to treatment is at least known, based on the adoption of a cut-off point (eg Carreño, 2015). Both designs can be significantly improved by including control techniques such as the use of covariates or group matching by propensity scores. An extension of the NEGD design with two additional groups that do not include pretest is the Solomon four group design (S4GD), which is often presented with random assignment as an experimental design (e.g. Whitman, Van Rooy, Viswesvaran & Alonso, 2008) or when not possible with non-random assignment such as a NEGD design (eg, Flórez-Alarcón & Vélez-Botero, 2010). Its main advantage is the ability to test sensitization to the pretest and to control the reactivity of the measuring instrument (García, Frías and Pascual, 1999).

Among the quasi-experimental longitudinal designs we highlight the repeated measures quasi-experimental design (QRMD), where the intrasubject variable is manipulated and whose administration has not been randomized (e.g., Labrador & Alonso, 2007), the interrupted time series design (ITSD), an extrapolation of the pre-experimental PPD module with a group and multiple measures before and after introducing an intervention or program (eg, Alvira, Blanco & Torres, 1996) and can be generalized to several groups and several response variables (e.g., Escudero & Vallejo, 2000) and the longitudinal panel design (LPD), which in its simplest form requires a group with two variables measured at various time points, but can also be generalized to multiple variables. Figure 4 summarizes the preexperimental and quasi-experimental designs used in Psychology.

Considering that all quasi-experimental designs can be analyzed with different statistical procedures, the most common analytical options are, for the cross-sectional designs (NEGD, RDD and S4GD), some of the alternatives of the generalized linear model in any of its modalities: regression, ANOVA, ANCOVA or analysis of change scores (Ato, 1995b,c; Ato & Vallejo, 2015; Ato, Losilla, Navarro, Palmer and Rodrigo, 2005; Spector, 1981). As for longitudinal designs, for the QRMD Arnau, (1995g), and Vallejo and Fernández, (1995) can be consulted . For the ITSD design, time series analysis is most appropriate in the case of a high number of pretest and posttest measurements (Arnau, 1995e,f ; Vallejo, 1995a,b; 1996; Vallejo, Arnau, Bono, Cuesta, Fernández & Herrero, 2002) or if there is a reduced number of measures, some of the alternative procedures are discussed in detail in Arnau (1995c), Arnau & Bono (2004) and Bono & Arnau (2014). For the LPD design, the analytical procedures discussed in Arnau and Gómez (1995) are appropriate. In addition, intensive longitudinal methods are today a recommended class of models proposed to analyse mass data collected over time (Bolger & Laurenceau, 2013).

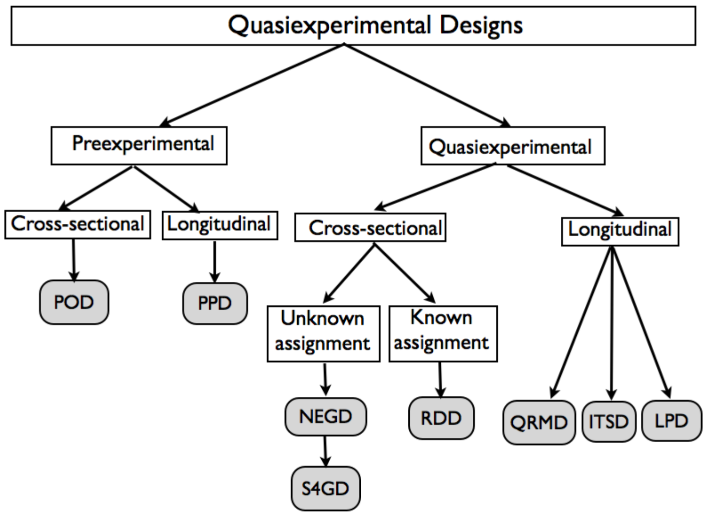

Figure 4. Quasiexperimental Designs

POD: Postest only design.
PPD: Pretest-Postest design.
NEGD: Non equivalent groups design.
S4GD: Solomon four groups design.
RDD: Regression discontinuity design.
QRMD: Quasiexperimental repeated measures design.
ITSD: Interrupted time series design.
LPD: Longitudinal panel design.

A.1.3 Single Case Designs

Single case designs were developed in the tradition of experimental control and initially systematized by Sidman (1960). As with manipulative strategy designs, causal analysis is only possible when the two basic requirements (variable manipulation and control by randomization) are met, but since there is only one unit in most cases, or at most a small number, the randomization principle is uncommon in single case designs. However, some schemes have recently been proposed to introduce randomization into many single-case designs (see Reichardt, 2006; Kratochwill & Levin, 2010a), and a set of standard criteria has been developed by a panel of experts to determine whether the application of a single case design may be interpretable in causal terms (see Byiers, Reichie & Symons, 2012; Kratochwill & Levin, 2010b; Kratochwill, Hitchcock, Horner, Levin, Odom, Rindskopf & Shadish, 2010). A review of 409 papers published in the 2000-2010 period further suggests that much of the current research with these procedures satisfies many of the quality criteria required for the experimental methodology (Smith, 2012).

In the typical application of a single case design, a behaviour is first identified and measured over a period of time (baseline phase) and then an intervention or treatment is applied and changes in behaviour are observed (intervention phase). The intervention can be eliminated or altered in successive phases and the changes in the behaviour are again observed.

There are two dimensions in the classification of single case designs which, following Arnau (1995b), Bono and Arnau (2014) and Ato & Vallejo (2015), allow us to distinguish them in a first dimension, depending on the reversibility of response to baseline levels after removal or alteration of the treatment (reversible and non-reversible designs) and, in a second dimension, depending on the comparison strategy used (between series, within series and mixed designs). They are represented in Figure 5.

The most basic single-case design is the basic two-phase A-B design, which may be intraseries (basic design with N = 1 or SABD) or mixed (basic design with multiple N or MABD).

Reversible designs are classified into two large groups, based on the reversal process, in simple or complex. Depending on the number of phases, simple reversion designs can be of two types: 1) intraseries designs, which include the three-phase design (ABAD), the four-phase design (ABABD) and the withdrawal design (BABD), and 2) mixed designs, which also have four phases, and can be inversion or more commonly generalization (four phase design or 4PD). Among the complex reversion designs, currently popular in applied areas are the multilevel design (MUD), the multiple treatment design (MTD) and the interaction design (IND). There are many works that use reversion designs in the literature (eg, Tincani, Crozier & Alazetta, 2006; Munro & Stephenson, 2009).

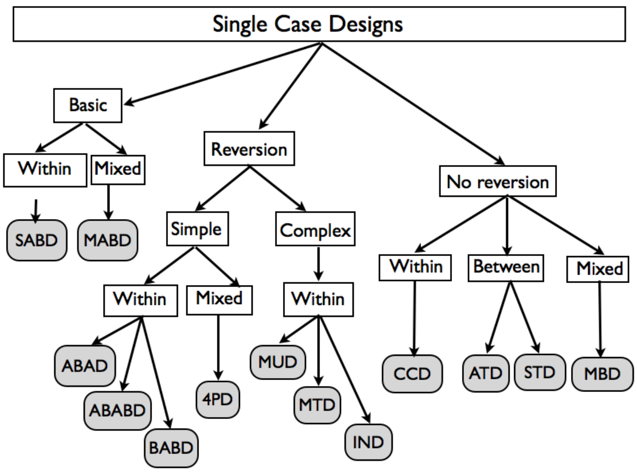

Figure 5. Single Case Designs

SABD: Simple AB design.
MABD: Multiple AB design.
ABAD: Within series design.
ABABD: Within series ABAB design.
BABD: Within series BAB design.
4PD: 4 phases design.
MUD: Multilevel design.
MTD: Multiple treatment design.
IND: Interaction design.
CCD: Changing criteria design.
ATD: Alternating treatment design.
STD: Simultaneous treatment design.
MBD: Multiple baseline design.

The most interesting intraseries non-reversion design is the changing criterion design (CCD, e.g. Foxx & Rubinoff, 1979), of easy application and high flexibility. Among the non-reversion between-series designs, the alternating treatments design (ATD, e.g. Barlow & Hayes, 1979) and the simultaneous treatments design (STD, eg McCullough, Cornell, McDaniel & Mueller, 1974) are also known in the literature as multi-element designs. Among those with mixed comparisons the most popular is the multiple baseline design (MBD, e.g., Bornstein, Bellack & Hersen, 1977; McClannaham, McGee, McDuff & Krantz, 1990), which has several meanings depending on which baseline data are recorded between behaviours, between subjects or between contexts. More detailed information on single case designs can be found in Ato & Vallejo (2015) Bono & Arnau (2014), Gast & Ledford (2014), Kazdin (2011), Kennedy (2005) and Kratochwill & Levin (2010a).

Many different procedures have been proposed for the statistical analysis of single case designs, ranging from simple descriptive and non-parametric procedures to those more complex and parametric that address the problem of serial dependency. The most popular choice (but also the most likely to commit type I errors when practised by non-experts) is visual scanning, which evaluates the change based on three essential factors: *level* of the response variable from condition to condition, the *trend* that follows in the set of observations and the *latency* that requires the response variable to change after a change in conditions. The most common nonparametric procedures are randomization tests (Edgington, 1992, 1995; Ferron & Fosters-Johnson, 1998) and several statistical effect size indices based on non-overlapping data (Parker, Vannest & Davis, 2011). Parametric procedures allow the statistical problem of serial dependence, in particular time series analysis, to be treated statistically with applications of ARIMA and AR models (see Velicer & Molenaar, 2013), linear using generalized least squares estimation (see Swaminathan, Rogers, Horner, Sugai & Smolkowski, 2014), and the most flexible multilevel analysis capable of modeling diverse problems such as serial dependency, regression analysis for autocorrelated data nonlinear trends, and between- and within-subject heterogeneity (see Baek & Ferron, 2013, Shadish & Rindskopf, 2007 and Shadish, Kyse & Rindskopf, 2013).

B. **Associative strategy**

B.1. Comparative Studies

Comparative studies are those that analyse the relationship between variables by examining the differences between two or more groups of individuals, taking advantage of the differential situations created by nature or society. Although they share the same aim as experimental studies, the establishment of causal relationships, they are non-experimental studies as they do not use manipulated variables (or if so these have not been manipulated for ethical or administrative reasons), nor random assignment of participants (Anderson et al., 1980). In their most characteristic form, the independent variables (IV) of these studies are attributive (classification variables), as opposed to active (manipulated variables), which are the VIs characteristic of the experimental and quasi-experimental studies, and to measurement variables, which are most common in predictive and explanatory studies.

As with quasiexperimental studies, the absence of randomization and the use of non manipulated variables requires an additional effort on the part of the researcher to try to reproduce the optimal conditions of an experimental study, avoiding the presence of selection biases (differences introduced in the selection of groups), information bias (differences introduced in data collection and treatment) and confounding (existence of third variables that may be potential explanations of dependent variable, DV).

For these two basic forms of bias and confusion, that reduce the internal validity of the research, there are many experimental and statistical control procedures to avoid its influence. There are many experimental and statistical control procedures to circumvent these two basic forms of bias (see Grimes, 2002, Gelman & Hill, 2007, Gennetian, Magnuson & Morris).

A comparative study usually takes one of several possible temporal approaches. It is called retrospective (or ex-post-facto) when the independent variable has occurred before the beginning of the study, cross-sectional when the definition of IV and DV is performed concomitantly in time, and prospective (or longitudinal) when observed at the beginning of the study and prolongs the DV record over time. Within this general structure several research designs in Psychology can be defined (see Figure 6).

B.1. 1. Retrospective or ex post facto studies

Retrospective studies use historical information to return in time by examining previous events (hence its alternative designation of ex-post-facto studies, see Suchman, 1967). Although the time period covered may span many years, an advantage of retrospective studies is the limited time it takes to complete, since it only requires recording and analysing data. For this reason they are particularly useful for studying the association between uncommon or unpredictable variables or when there is time lag between an alleged cause and its possible effect. These are typical of research in health sciences (in particular, Mental Health and Epidemiology, see Mann, 2003), where they have been imported to related disciplines such as Psychopathology and Clinical Psychology. It is precisely in the clinical context where retrospective studies are commonly used instead of experiments to propose cause-effect hypotheses in situations where manipulated variables are not used.

Two types of retrospective studies can be differentiated, depending on whether the definition of the groups of participants is based on the independent or dependent variable (Cohen, Manion & Morrison, 2012; Kirk, 2013). In the retrospective cohort design (RCD, also called a historical cohort), records are used to identify two groups of subjects based on whether they have been exposed to the independent variable, and then compared in the dependent variable/s, its differential characteristic being the existence of a single independent variable and one or more dependent ones (e.g., Rios, Godoy and Sánchez, 2011). A cohort in this context represents a group of who have experienced a significant event (e.g., birth, divorce or illness) during a given time interval. In contrast, in the case-control design (CCD), records are used to identify two groups based on whether they show evidence of the dependent variable (cases) or not (controls) and then compare them in terms of their previous exposure to variables (e.g., Gruber, Pope, Hudson & Yurgelun-Todd, 2003).

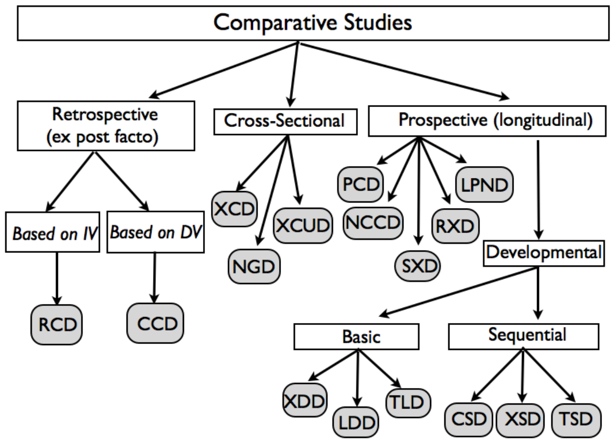

Figure 6. Comparative Studies

RCD: Retrospective cohort design.
CCD: Case-control design.

XCD: Cross-sectional cohort design.
NGD: Natural groups design.
XCUD: Cross cultural design.

PCD: Prospective cohort design.
NCCD: Nested case-control design.
SXD: Simultaneous cross-sectional design.
LPD: Longitudinal panel design.
RXD: Repeated cross-sectional design.

XDD: Cross-sectional developmental design.
LDND: Longitudinal developmental design.
TLD: Time-lag design.
CSD: Cohort sequential design.
XSD: Cross sequential design.
TSD: Time sequential design.

Let us imagine that a researcher is interested in studying the link between attendance at early childhood centres (IV) and academic performance in primary education (DV). If a retrospective cohort design was used, the researcher could use two groups of children who differ by whether they attended nursery schools and then comparing the average academic achievement attained in primary school. That is to say, the IV is first defined retrospectively and the DV is then measured, but both variables occur before the study begins. If the case-control design was used, the researcher could use two groups of children that show differences in academic achievement in primary (cases and controls) and would then retrospectively examine each case and control the attendance at children's education courses. That is the DV is first defined and the IV is then retrospectively evaluated, which occur prior to the study.

A feature of retrospective studies is their explicit interest in formulating cause-effect hypotheses in order to reproduce the characteristics of the experimental methodology in a non-experimental context and with variables not susceptible to manipulation. In fact, in educational research, retrospective or ex post facto designs are also called causal-comparative designs (see Brewer & Kuhn, 2010; Gay, Mills & Airasian, 2012), while in statistical literature they correspond to some extent to inverse experiments (see Loy, Goh & Xie, 2002). However, there are numerous potential threats to internal validity in retrospective studies. Shadish, Cook, and Campbell (2002, p.131-133) discuss up to 53 general threats to validity in case-control design, making causal relationships practically unfeasible in these studies (see Farrant, 1977). However, some retrospective or ex post facto studies often find interesting relationships that may subsequently be the subject of experimental research and we can often use sophisticated control procedures to achieve conditions somewhat similar to those obtained through randomization that allow analysis of cause-effect relationships (see Vélez, Egurrola and Barragán, 2013).

B.1.2. Cross-sectional Studies

Unlike retrospective studies, cross-sectional ones are defined at a specific time point and follow an eminently associative tradition where interest in establishing cause-and-effect relationships is secondary. While retrospective and prospective cohort studies are the best means of assessing incidence issues (i.e., the number of new cases being developed during a specific time interval), cross-sectional studies are primarily used to assess prevalence issues (they aim to determine the number of cases that exist in a given population at a specific time point). Cross-sectional studies are also suitable for the study of DVs that remain stable over time, i.e. they are not susceptible to change.

In Epidemiology and Psychopathology, cross-sectional studies use one or more groups of participant subjects evaluated at a given time in one or more DVs. The simplest is the cross-sectional cohort design (XCD, see Hudson, Pope & Glynn, 2005), which uses cross-sectional cohort samples and then retrospectively evaluates the history of exposures (IVs) and outcomes (DVs) in cohort members in a specific period of time (e.g., Stefani and Feldman, 2006). Obviously, due to the sampling process, some participants will not have had exposure (IV) or have experienced any result (DV). This is an essential difference with cohort studies (retrospective and prospective), where the history of exposures and / or outcomes affects all subjects in the cohort (Mann, 2003).

In some areas of applied psychology the use of the natural groups design (NGD, see Shaughnessy, Zechmeister & Zechmeister, 2012) whose aim is the comparison in one or more DVs of preexisting groups, where groups are selected using participants that belong at levels of variables that are sources of individual differences (e.g., sex, intelligence, ethnic group, psychopathological disorder, etc.), and also to the same culture. When natural groups belong to different cultures the result is cross-cultural design (XCUD, see Matsumoto & van Vijver, 2011). The development of influential transcultural questionnaires, such as the European Social Survey or the Program for International Assessment of Students' Achievements (PISA), has popularized this type of research and allowed the exploration of new analytical options that had not previously been considered (Davidov, Schmidt & Billiet, 2011).

Cross-sectional studies require minimal effort in cost and time and are quite efficient when proposing associative hypotheses, but they pose serious validity problems proposing causal-effect hypotheses.

B.1.3. Prospective (or longitudinal) studies

In prospective or longitudinal studies, IVs and DVs are seen after the initiation of research (although IV may sometimes refer to an earlier situation). Hence, it is much more costly and time-consuming to complete a prospective study than a retrospective one, but it does however appreciably improve reliability. And due to their longitudinal status, some potential threats to the internal validity of retrospective studies are not considered in prospective studies, although the drawbacks to establishing cause-and-effect relationships remain patent, mainly problems of bias selection and confusion, which also aggravate as a result of the abandonment of subjects (attrition) that often occurs the longer the research proceeds.

One of the most common prospective designs is the prospective cohort design (PCD), whose methodology is similar to the retrospective cohort design, but the cohort is evaluated forward, at various points over time, using one or more relevant DVs. This design usually includes a single cohort and extends over time to determine the outcome of interest (for example, the incidence of depression), in which case cohort members who do not develop depression serve as internal controls. However it may also include more than one cohort, one exposed and one unexposed, in which case the latter group serves as an external control (Mann, 2003). A design with a similar structure to the retrospective case-control one is the nested case-control design (NCCD), in which cases and controls are selected from within a prospective study, allowing a significant reduction in time and cost compared to the case-control design (see Ernster, 1994).

There are several prospective or longitudinal designs of some interest in several applied areas of Psychology (Arnau, 1995c; Bijleveld et al., 1998; Taris, 2000). Although it does not have a strictly longitudinal orientation, the most basic is the simultaneous cross-sectional design (SXD), where several groups that differ in IV (e.g., years from graduation) are measured at the same time in a response variable (e.g., employment of graduates in Psychology). Its generalization consists of repeated cross-sectional design (RXD), comprising at least one SXD administered at two or more time points (e.g., in 2010 and 2015). In both cases, the researchers' interest is to study change at the individual level in some variable of interest, but not change due to age, which is the subject of developmental studies. The longitudinal panel design (LPND), where one or more groups of subjects are evaluated in a limited number of temporal moments, is a generalization of the SXD design (if multiple cohorts are used), of the RXD (if each of the different cohorts constitute different age groups) and the LPD (if the sample of participants making up the cohort is evaluated at different times). Arnau (1995d, e, f) and Gómez (1995) deal in detail with some analytical questions of many longitudinal designs. Bijleveld et al. (1998), Fitzmaurice, Laird & Ware (2009), Menard (2008) and Singer & Willett (2003) cover all relevant issues of longitudinal data design and in-depth analysis.

B.1.4. Developmental studies

In the context of developmental psychology, comparative studies have traditionally been concerned with analyzing differences or changes in behaviours or abilities of evolutionary interest in terms of age or development. It is very important to determine if the interest of the researcher is age/maturation differences or non-age related changes to decide whether a study is developmental (Schaie & Caskie, 2006).

Three basic developmental designs have been defined within this tradition (Baltes, Reese & Nesselroade, 1977; Schaie, 1965; Rosel, 1995). First, cross-sectional developmental design (XDD), which uses samples from different age groups evaluated with between subjects level at the same time point, aims to find differences according to age in the measurement of one or more DVs. Despite its name, the longitudinal dimension in this design is due to the comparison between different age groups. The differential characteristic that justifies the inclusion of this design in developmental studies is that the key variable is age, while in cross-sectional studies and simultaneous cross-sectional design, age is a control variable (Taris, 2000). The design is quite efficient, as it allows a longitudinal study to be approached from a cross-sectional perspective, but it is not possible to evaluate the intraindividual change and the age effects are confounded with the developmental effects. Second, longitudinal developmental design (LDD), is concerned with the analysis of change that occurs in a group of individuals observed repeatedly over time. Third is the time-delayed developmental design (TLD), where observations are recorded at various time points, such as in the longitudinal design, but a sample of individuals of the same age group is used for each temporal moment.

These three designs are special cases of the general developmental model of Schaie (1965), who postulated that any developmental change may be associated with one (or more) of three independent dimensions chronological age (the number of years from birth to the chronological moment in which the behavior is evaluated), cohort or generation (referring to the group of individuals that participate from a common context in the same chronological moment) and period (the time the measurement is made). The three dimensions are confounding so once two of them are specified, the third is determined. For example, an individual from a cohort born in 1950 and evaluated in the 2010 period is known to have an age of 60 at the time of measurement. Thus, in XDD design, age and cohort are confounding and therefore no effects due to the period can be observed, since the studied age groups should belong to different cohorts. The LDD design is less affordable in terms of time and cost than the XDD, but age and period are confounding and therefore effects due to the cohort cannot be observed. On the other hand, in the TLD design, period and cohort effects are confounding (Schaie, 1994; Schaie & Caskie, 2006).

However, by imposing appropriate constraints on one of the three dimensions of the model, it is possible to analyze the other two components and their interaction (Arnau, 2005; Bijleveld et al., 1999; Menard, 2008). Three strategies have been proposed to analyse all combinations of two of the three components, producing three alternative design forms to classic developmental designs called generically sequential designs (see Schaie, 1994): cohort-sequential design (CSD), cross-sequential design (XSD), and time-sequential design (TSD). In general, these are much more efficient than the basic developmental designs. For example, a longitudinal design with children aged 5 to 13 observed every 2 years requires completion of 8 years, while an equivalent sequential design reduces by half the time required for the study, using a group of 5-year-olds (Observed at ages 5, 7 and 9) and another group of 9-year-olds (observed at age 9, 11 and 13). More information on research methods in developmental psychology can be found in Laursen, Little & Card (2013) and Teti (2006).

B.2. Predictive and explanatory studies.

Another type of non-experimental research deriving from associative strategy is studies whose main purpose is to explore relationships between variables to predict or explain their behaviour. They have also largely been known under the generic name of correlational studies, but this term is now considered incorrect as the type of statistical analysis is not exactly the crucial subject in this type of study (Cook & Campbell, 1979). However, for those predictive studies that only explore relationships between variables using correlation coefficients we will continue using the term correlation designs in this paper.

Three distinctive features allow us to define these types of studies operationally in relation to other non-experimental studies:

- The existence of a single sample of participants not usually randomly selected;
- The measure of each participant in the sample in two or more variables usually of a quantitative nature (i.e. measured variables, not manipulated) but occasionally also categorical in nature, and
- Availability, as a starting point to address interpretation of an array of correlations (or covariance) between variables.

Following Pedhazur & Smelkin (1990), we have classified the works that follow this line of research, depending on the complexity of the aim, in predictive and explanatory studies, and represented in Figure 7.

B.2.1. Predictive studies

The most basic form usually adopted by these types of studies occurs when the aim of the research is simply to explore a simple functional relationship between two or more variables, with no distinction between them. Since no form of control of extraneous variables is used over the functional relationship, the resulting design is called simple correlational design (SCD). The statistical analysis of the association between variables often utilizes a correlation coefficient or a correlation coefficient matrix appropriate to the metric nature of the variables (e.g., Sicilia, Aguila, Muyor, Orta & Moreno, 2009). But other alternative analytical techniques are also possible.

It is obvious that in SCD the degree of control exercised over the third variables that can potentially affect the analysed relation is null and that all variables have the same methodological status, this being the reason why this design is highly prone to the most threats to internal and external validity. Fortunately, it can be significantly improved when the functional relationship between the variables is scanned by controlling one (or more) third variable (s). The most appropriate control procedure in this case is statistical control through residualisation or partialisation, in which case the resulting design is called a controlled correlational design (CCD). Statistical analysis usually uses partial or semi-partial correlation coefficients of k order, where k refers to the number of controlled variables, or a matrix of partial or semi-partial correlation coefficients (eg, Del Rey, Elipe & Ortega, 2012; Piamontesi, Esteban, Furlan, Sánchez-Rosas & Martínez, 2012).

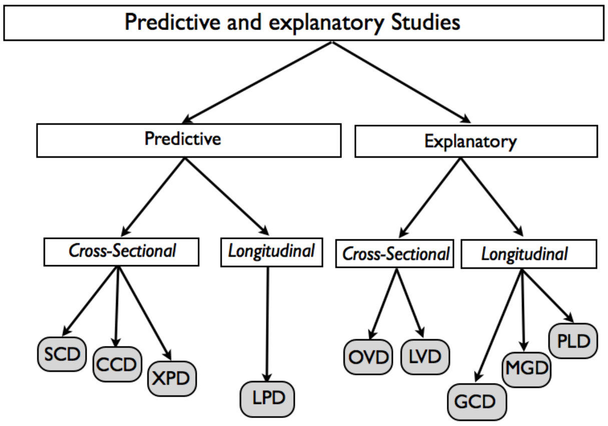

Figure 7. Predicitive and explanatory Studies

SCD: Simple correlational design.
CCD: Controlled correlational design.
XPD: Cross-sectional predictive design.
LPD: Longitudinal predictive design.
OVD: Observed variables design.
LVD: Latent variables design.
GCD: Growth curve design.
MGD: Multiple groups design.
PLD: Panel longitudinal design.

When the aim of the research is to explore a functional relationship through prediction of some criterion variable from one or more predictors, a cross-sectional predictive design (XPD) is applied, where it is common to use the terms predictor, independent variable, and criterion, replacing those of dependent variable. XPD does not usually employ control procedures and the most common statistical procedure is to specify an appropriate linear or non-linear regression model, depending on the nature and distribution of the criterion variable, and may also be treated from a multilevel perspective. The generalized linear models (Ato, Losilla, Navarro, Palmer & Rodrigo, 2005; Dobson, 2002; Hardin & Hilbe, 2007) and multilevel models (Hox, 2011) are especially recommended in this context. When the criterion variable is numerically and normally distributed, a normal regression model is usually specified (e.g., Jiménez, Martínez, Miró and Sánchez, 2012), when the variable is binary or dichotomous a logistic regression model (eg, Planes, Prat, Gómez, Gras and Font-Mayolas, 2012) and when a hierarchical structure presents a multilevel model (e.g. Núñez, Vallejo, Rosario, Tuero & Valle, 2014). In many cases it is possible, and often convenient, to evaluate the efficacy of the prediction through a classification of participants. In addition to logistic regression, it is useful in this context to apply other multivariate techniques such as discriminant analysis, hierarchical classification and segmentation, and neural networks (see Levy and Varela, 2003).

If the researcher has many predictors, there are two alternative means of introducing them into a regression equation: either all at once (simultaneous) or through a series of steps (sequential). In the latter case, there are two general options for determining a logical order of introduction of predictors: either through an automatic stepwise selection procedure, or a hierarchical procedure, directed by the researcher according to the aim of the research. There are several reasons for preferring the latter to the former (Snyder, 1991; Thompson, 1995, 2001), mainly that with hierarchical selection one can incorporate variables in the first stages of the process control that the investigator considers pertinent, to analyze the predicted object of the research from the influence of the control variables in the final stages. There are many examples in the psychological literature on applications of stepwise and hierarchical selection (eg, García-Izquierdo, García-Izquierdo & Ramos-Villagrasa, 2007 and Cejudo, López-Delgado & Rubio, 2016).

Both correlational designs and cross-sectional predictive design are essentially single-measurement. A generalization of the XPD to the longitudinal case leads to the longitudinal predictive design (LPD), which is characterized by recording an intensive measurement of the criterion variables and by attempting to eliminate many methodological artifacts (Mindick & Oskamp, 1979), although it also suffers from problems associated with longitudinal research in terms of cost and time. LPD applications are less abundant and use more sophisticated statistical procedures than XPD (e.g., Rosenberg, Frees, Sun, Johnson & Robinson, 2006).

B.2.2. Explanatory studies

Another type of research deriving from the associative strategy is explanatory studies, whose essential aim is the testing of models regarding the existing relationships between a set of variables, as they derive from an underlying theory. In a regression model, the basic statistical model of predictive designs, we clearly define predictor and criterion roles, but postulate very simple modeling structures (usually one criterion and several predictors) and consider all variables as manifest or observable and measurements without error, which considerably limits its applicability. Its generalization to more advanced models, including simultaneous modeling structures and unobserved or latent variables, allows us to define a system of regression equations where variables can exchange their predictor and criterion roles. To further clarify their role in the context of each regression equation of the system, the terms of exogenous variables (outside the system) and endogenous variables (inside the system) are used alternately, instead of the terms of predictor variables and criterion variables.

Two research situations are contemplated here (see Figure 7). Firstly, the observed variables design (OVD), characterized by defining a structural network of relations between variables that can be represented by a system of regression equations, assuming that all variables are manifest or observable. The resulting model belongs to the SEM structure (see Table 1). The most appropriate statistical procedure to analyze this type of design data starts from an array of correlations (or covariates) and then applies the path analysis. The adjustment of the empirical data to the proposed model is evaluated using appropriate adjustment measures for structural models (see Kline, 2011). However, as noted above, researchers should always be aware that this analysis assumes that the overt variables are reliable manifestations of their corresponding constructs, which cannot always be justified (see Cole & Preacher, 2013).

Given a functional relationship between a predictor and a criterion, one of the most interesting applications of the OVD design is the evaluation of the effects of one or more third variables (Ato & Vallejo, 2011), justifying the significant increase of research that submit to mediation and moderation test models (Jose, 2013, Hayes, 2013). In its simplest form, a mediation model is concerned with intervening processes that produce a functional relationship or a treatment effect (eg, Gartzia et al., 2012), while a moderation model is concerned with processes that affect the magnitude of the relationship or effect (e.g., Rodrigo, Molina, García-Ros and Pérez-González, 2012).

Second, the latent variables design (LVD) distinguishes a structural part in a model (representing a structural model of relations between variables, as in the OVD design) and a part of measurement (including the different indicators that define a latent construct or variable) and is also represented by a system of structural equations, where some variables are observable and others are latent. They are Structural Equation Models (SEM) with latent variables. There are at least two appropriate statistical approaches to estimating the SEM model parameters, a covariance-based SEM procedure which is most widespread (Kline, 2011; Levy and Varela, 2006). That specific analysis programs (embodied in the AMOS, CALIS, EQS, LISREL, MPLUS and R packages) have been designed, and a variance-based SEM method, called partial least squares path modeling (PLS-PM), usually applied in cases where the severe conditions required by the former regarding sample size, multivariate normality and independence of observations are not met (Esposito Vinci et al., 2010; Hair, Hult, Ringle & Sarstedt , 2013), and for which there are also appropriate analysis programs (e.g. the PLS-PM program included in package R).

The LVD design allows controlling any foreign variable simply by including it in the structural model as observable variables and specifying relationships with the rest of variables, observable and latent (e.g., Rosario, Lourenço, Paiva, Núñez, González-Pienda & Valle, 2012). In addition, mediation and moderation effects can also be tested often by comparing the fit of two alternative models that differ by either including effects of the third variables or not (e.g., Buelga, Cava and Musitu, 2012).

An interesting generalization of LVD designs that can be used to model longitudinal data in which repeated measurements of at least one variable are observed very often leads to the growth in curve design (GCD). An affordable introduction to this type of design can be found in Arnau and Balluerka (2004) and Preacher, Wichman, McCallum & Briggs (2008). Other explanatory longitudinal designs employing SEM models, such as the multiple groups design (MGD) and the panel longitudinal design (PLD), can be found in Little (2013b).

C. **Descriptive strategy**

The descriptive strategy represents, together with the associative strategy, one of two characteristic forms of non-experimental studies as it does not fulfill any of the two basic criteria of experimental research (manipulation of variables and control by random assignment). But while in the associative strategy the research objectives are translated into hypotheses seeking comparisons between groups or the prediction or explanation of behaviours or processes, in the descriptive strategy the aim of the research is the definition, classification and / or categorization of events to describe mental processes and manifest behaviours, which in essence do not usually require the use of hypotheses.

There are two major types of studies included in the descriptive strategy, which also represent alternative forms of data collection used in most non-experimental studies: observational studies, where behaviours are observed and classified in accordance with arbitrary codes, and selective, where opinions or attitudes are recorded on a response scale usually by questionnaire or interview. Following Kish (1987), the essential characteristic of an observational study is the realism with which the behaviours , object of the research are investigated, whereas the basic characteristic of a selective study is the representativeness of the selected sample regarding the target population. In contrast, the essential feature of an experimental study is random assignment.

It is important to note that the data analysis that is expected in a descriptive study is not necessarily descriptive; in many cases; a more complex analysis is performed using a wide repertoire of univariate and multivariate analytical techniques (e.g. classification, dimensionality or scaling techniques, but also procedures derived from the classic or generalized linear model). The type of research does not therefore depend on the statistical analysis carried out, but on the aim pursued. When the aim of the research is descriptive, the researcher does not usually pose specific hypotheses for empirical testing and will present results that appropriately describe behaviour or mental processes, either with descriptive statistics or with more sophisticated analytical techniques.

C.1. Observational studies

An observational study is a research plan of great procedural versatility to record the spontaneous behaviour of a unit (participant, dyad, team, etc.) using specific observation techniques and following a sampling plan of behaviours in natural contexts (Rabadán and Ato , 2004).

According to Anguera (1995) and Shaughnessy, Zechmeister and Zechmeister (2012), observational studies are generally distinguished according to the method of observation used, or indirect observation (non-reactive) of behaviours that occurred previously, as when a crime scene is analyzed for evidence or archival records are consulted on behaviours of interest, or by direct observation. In the latter case, the researcher must create an observation context that can vary along a continuum from the use of methods without intervention (naturalistic observation) in the context of observation to the use of intervention methods, among which (the observer plays a dual role, observing behaviours and participating in the observed situation) and structured observation (where the observer intervenes in the situation in order to facilitate the recording of behaviours Through naturalistic observation as in the studies of Piaget). In all these situations, the researcher must take into account the specific limitations of this methodology, which include: 1) the scarce control usually obtained from foreign variables (in fact, the variable tag is not normally used in this context); 2) possible dependence on subjective judgments and 3) reactivity problems arising from observation.

Observational studies are well suited to many situations of interest to qualitative research (Marshall & Rossman, 2006), such as narrative studies, exploring a single participant, phenomenological studies, exploring a small group of participants and ethnographic studies, interested in a culture. Especially appropriate is observational methodology in the case study, where the researcher intensively studies specific aspects of the behaviour of an entity (which may include one or more subject (s), dyad, school / s, hospital / (See Woodside, 2010) and also in theory-generating studies (based on the "grounded theory" (Strauss & Corbin, 1998) that proposes generating or discovering a theory or pattern analysis of a process, action or interaction based on data collected from participants (eg, Fernández-Alcántara et al., 2013).

Despite the diversity of possible studies, the methodology applied in observational studies is unique and "... it is characterized basically by the perceptivity of behaviour, habituality in the context, the spontaneity of observed behaviour and the elaboration measure of observation instruments "(Anguera, Blanco, Hernández and Losada, 2011, 64). In this sense, observational design can be defined as a flexible guide to making decisions about data collection (including the construction of observation instruments) as well as data management and analysis, highlighting three key criteria to distinguish them (Anguera, Blanco and Losada, 2001; Anguera, Blanco, Hernández and Losada, 2011):

- The observed units of study are distinguished according to their level of integration into ideographic (a single unit of observation, individual or aggregate) and nomothetic (a plurality of observation units);
- temporality of the study allows distinguishing between observations made in a cross-section (at a specific point in time) or longitudinal (follow-up), and
- dimensionality of the study, referring to the levels of response recorded, facilitates the distinction between investigations requiring a single level of response (one-dimensional) and multidimensional (multidimensional) research.

Crossing the two possible options for each of the three criteria cited produces eight types of observational designs depicted in Figure 8. Each of the four quadrants in the figure also corresponds to the categories of a taxonomy proposed by Sakett (1978) on the nature of data obtained from observational designs, in which we distinguish between type I data (sequential and base-event), type II data (concurrent and event-based), type III data (Sequential and time-base) and type IV data (concurrent and time-base).

Anguera, Blanco and Losada (2001) and Blanco, Losada and Anguera (2003) illustrate the nature of each observational design and present some examples of use in numerous applications of clinical , school and sports psychologies, ethology (animal and human) and observation in natural contexts, which have been collected in a series of publications edited by Anguera (1999a, b, c, d, e).

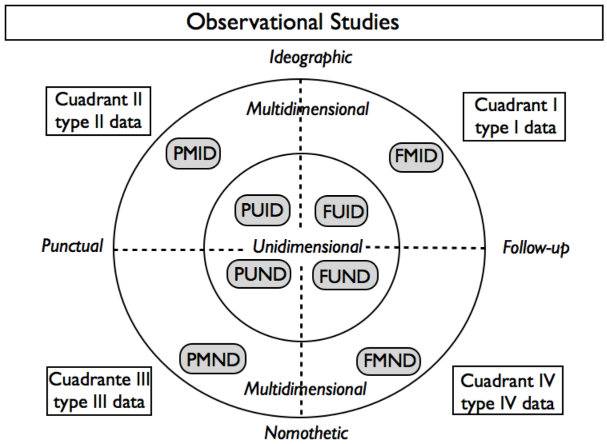

Figure 8. Observational Studies

PMID: Punctual multidimensional ideographic design.
FMID: Follow-up multidimensional ideographic design.
PMND: Punctual multidimensional nomothetic design.
FMND: Follow-up multidimensional nomothetic design.

PUID: Punctual unidimensional ideographic design.
FUID: Follow-up unidimensional ideographic design.
PUND: Punctual unidimensional nomothetic design.
FUND: Follow-up unidimensional nomothetic design.

The observational designs described here will be classified as descriptive when their objective fits within a descriptive research strategy. However, many representative studies of what is now known as observational methodology transcend the descriptive framework and have no easy fit in this category. Using the observational designs discussed here, it is sometimes suggested analysing functional relationships between behaviours, using very similar techniques to single case designs (but generally emphasizing external validity over internal validity) and using data analysis procedures that depend on the nature of the data recorded, through a subtle transformation of the qualitative record into quantitative data, and therefore must be located within the associative strategy, although using the same observational designs described here. The common feature of all observational designs, fit or not within the descriptive strategy, is the use of observation as a data collection technique.

C.2. Selective studies

Selective studies are one of the most common forms of research in psychology based on the use of the sampling survey method, and its fundamental distinctive feature is the use of the self-report technique to collect empirical information (in particular, through interviews and questionnaires) In a sample of participants, an assumed representative of a population, whose elements are determined by some sampling plan, to investigate a characteristic of the population (Martínez-Arias, 1995a). Within this definition, many national surveys (for example, the five-yearly National Health Survey and international surveys (the biennial European Social Survey about the opinions and attitudes of European citizens). Selective surveys using non-behavioural research are not included here.

An essential question in the application of a selective study is the representativeness of the sample in relation to the population, a question that concerns the external validity of the research. Represtantiveness should be established using a set of sociodemographic variables unrrelated to the subject of the survey and verifying that the sample estimates do not differ from the population parameters within a pre-established margin of error. The highest degree of representativeness is reached when probabilistic sampling is used, but it is very common in applied psychology to use non-probabilistic sampling, and in particular subjective sampling methods, with the lowest degree of representativeness.

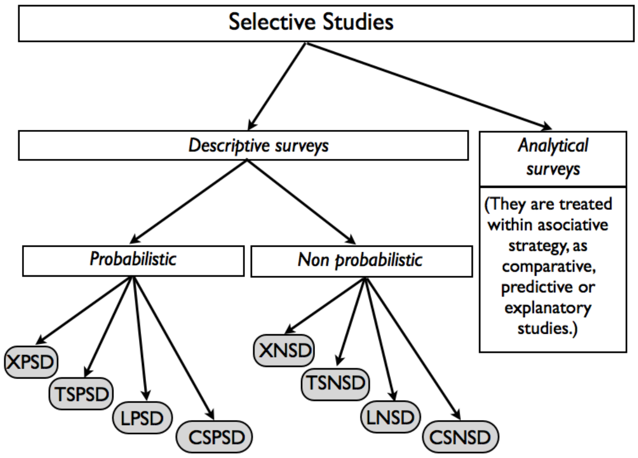

Figure 9. Selective Studies

XPSD: Cross-sectional probabilistic survey design.
TSPSD: Time series probabilistic survey design.
LPSD: Longitudinal probabilistic survey design.
CSPSD: Cohort sequential probabilistic survey design.

XNSD: Cross-sectional non probabilistic survey design.
TSNSD: Time series non probabilistic survey design.
LNSD: Longitudinal non probabilistic survey design.
CSNSD: Cohort sequential non probabilistic survey design.

In addition to the aim of the research, the design of a selective study should take into account the information collection procedure (probabilistic versus non probabilistic) and adopt a number of essential decisions prior to sampling, during sampling and post-sampling (Martínez-Arias, 1995c). Selective designs can be classified according to various perspectives. From a methodological perspective they can be descriptive (when the aim of the study is to obtain particular information from populations by placing emphasis on precision in parameter estimates) and analytical (when the objective is group comparison, prediction or explanation of behaviour). From a temporal perspective, selective studies may be cross-sectional probabilistic survey design (XPSD), where one or more samples from the same population are used, keeping time constant, time series probabilistic survey design (TSPSD), which are cross-sectional surveys with a new sample using the same instrument for each wave, usually keeping the age constant, longitudinal sequential probabilistic survey design (LPSD), where a sample is used at various time points to evaluate the change, keeping the cohort constant, and longitudinal cohort sequential probabilistic survey design (CSPSD), which combines elements of the three previous designs by studying cohorts longitudinally as new cohorts are added sequentially.

Within the descriptive strategy, the selective studies of interest are obviously descriptive and not analytical. The latter must be considered within the associative strategy. However, the common feature is the use of the survey. The selective designs, whether descriptive or analytical, constitute an arsenal of procedures of the so-called survey methodology that differ according to representativity (probabilistic vs. non-probabilistic) and the time of data collection (see Figure 9).

The entire process of application and estimation of a survey can be consulted in Gómez (1990) and Martínez Arias (1995a, b, c). More detailed information regarding the survey research can be found in a two-volume Encyclopedia edited by Lavrakas (2008).

Conclusions

Based on a contextual framework and postulating some basic principles to support the methodological aspects of development, articulation and evaluation of a research report, this paper addresses a classification of the most common research designs in psychology where methodological and analytical criteria are unified to facilitate their identification.

It is obvious that any classification will be ambiguous by its very nature, and that in some cases we will find designs that do not have an easy location. This paper presents a total of 70 basic designs, which can be easily generalized to cover a wide spectrum of all the design options available to a researcher. The future intention of its authors is to update this preliminary presentation in successive re-editions, including new basic designs that may not have been taken into account, eliminating designs that may be obsolete and developing a more detailed presentation where each design is treated with possible extensions, analytical options and some examples of application in practice.

Given the problem in our country of needing to publish to achieve new academic and professional goals, our strong recommendation is to encourage closer collaboration between substantive experts and methodological experts to strengthen the development of more competitive research in psychology.

Acknowledgments.- Our special thanks to the editorial team of the journal *Anales de Psicología / Annals of Psychology*, for giving us the opportunity to develop this work, and also to fifteen experts on psychological issues and methodology, who have responded to our call to review a first draft and have provided us with very valuable to improve it. The final version contains many of the suggestions received. The errors that still persist in the work should only be imputable to their authors. This English translation was published in December 2017.

References

Ader, H.J. & Mellenbergh, G. (1999). *Research methodology in the life, behavioral and social sciences.* London, UK: Sage Publications.

Agresti, A. (2013). *Categorical data analysis, 3rd Edition.* New York, NY Wiley.

Agresti, A. (2015). *Foundations of linear and generalized linear models.* New York, NY Wiley.

Aiken, L.S. & West, S.G. (1996). *Multiple regression: testing and interpreting interactions.* Thousand Oaks, CA: Sage Publications.

Alin, A. & Kurt, S. (2006). Testing non-additivity (interaction) in two-way ANOVA tables with no replication. *Statistical Methods in Medical Research, 15*, 63-85.

Alvira, F.; Blanco F. & Torres, M. (1996). El empleo de diseños de series temporales en la evaluación de intervenciones publicadas: un ejemplo aplicado. *REIS, 76,* 173-192.

Anderson, S., Auquier, A., Hauck, W.W., Oakes, D., Vandaele, W. & Weisberg, H.I. (1980). *Statistical methods for comparative studies: techniques for bias reduction.* New York, NY: Wiley.

Anguera, M.T. (1995). La metodología cualitativa. En M.T. Anguera, J. Arnau, M. Ato, M.R. Martínez, J. Pascual y G. Vallejo, *Métodos de la investigación en Psicología* (pp. 513-22). Madrid: Síntesis.

Anguera, M.T. (1999a). *Observación en la escuela: Aplicaciones.* Barcelona: Edicions de la Universitat de Barcelona.

Anguera, M.T. (1999b). *Observación en deporte y conducta cinético-motriz: Aplicaciones.* Barcelona: Edicions de la Universitat de Barcelona.

Anguera, M.T. (1999c). *Observación de conducta interactiva en contextos naturales: Aplicaciones.* Barcelona: Edicions de la Universitat de Barcelona.

Anguera, M.T. (1999d). *Observación en etología (animal y humana: Aplicaciones.* Barcelona: Edicions de la Universitat de Barcelona.

Anguera, M.T. (1999e). *Observación en Psicología Clínica: Aplicaciones.* Barcelona: Edicions de la Universitat de Barcelona.

Anguera, M.T., Blanco, A. y Losada, J.L. (2001). Diseños observacionales, cuestión clave en el proceso de la metodología observacional. *Metodología de las Ciencias del Comportamiento, 3,* 135-160.

Anguera, M.T., Blanco, A., Hernández, A. y Losada, J.L. (2011). Diseños observacionales: ajuste y aplicación en psicología del deporte. *Cuadernos de Psicología del Deporte, 11,* 63-76.

APA (2013). *Publication Manual of the American Psychological Association, 6th edition.* Washington, DC: American Psychological Association.

Arnau, J. (1995a). Diseños experimentales de sujeto único. En Anguera, M.T., Arnau, J., Ato, M., Martínez-Arias, R., Pascual, J. y Vallejo. G., *Métodos de investigación en psicología* (pp. 179-194). Madrid: Síntesis.

Arnau, J. (1995b). Análisis estadístico de datos para los diseños de sujeto único. En Anguera, M.T., Arnau, J., Ato, M., Martínez-Arias, R., Pascual, J. y Vallejo. G., *Métodos de investigación en psicología* (pp. 195-243). Madrid: Síntesis.

Arnau, J. (1995c). Diseños de investigación longitudinal. En J. Arnau (ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento* (pp. 35-53). Méjico: Limusa.

Arnau, J. (1995d). Diseños longitudinales de un solo sujeto y una sola variable. En J. Arnau (ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento* (pp. 79-141). Méjico: Limusa.

Arnau, J. (1995e). Diseño de medidas repetidas de un solo grupo de sujetos. En J. Arnau (ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento* (pp. 191-221). Méjico: Limusa.

Arnau, J. (1995f). Diseño de medidas repetidas II: diseños de dos grupos y diseños de estructuras más complejas. En J. Arnau (ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento* (pp. 223-253). Méjico: Limusa.

Arnau, J. y Balluerka, N. (2004). Análisis de datos longitudinales y de curvas de crecimiento. Enfoque clásico y propuestas actuales. *Psicothema, 16*, 156-162.

Arnau J. y Bono, R. (2004). Evaluating effects in short-time series: alternative models of analysis. *Perceptual and Motor Skills, 98*, 419-432.

Arnau, J. y Gómez, J. (1995). Diseños longitudinales en panel. En J. Arnau (ed.). *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento* (pp. 341-400). Méjico: Limusa.

Ato, M. (1991). *Investigación en Ciencias del Comportamiento. I: Fundamentos.* Barcelona: PPU/DM.

Ato, M. (1995a). Tipología de los diseños cuasiexperimentales. En Anguera, M.T., Arnau, J., Ato, M., Martínez-Arias, R., Pascual, J. y Vallejo. G., *Métodos de investigación en psicología* (pp. 245-270). Madrid: Síntesis.

Ato, M. (1995b). Análisis estadístico I: diseños con variable de asignación no conocida. En Anguera, M.T., Arnau, J., Ato, M., Martínez-Arias, R., Pascual, J. y Vallejo. G., *Métodos de investigación en psicología* (pp. 271-304). Madrid: Síntesis.

Ato, M. (1995c). Análisis estadístico II: diseños con variable de asignación conocida. En Anguera, M.T., Arnau, J., Ato, M., Martínez-Arias, R., Pascual, J. y Vallejo. G., *Métodos de investigación en psicología* (pp. 305-320). Madrid: Síntesis.

Ato, M. y López, J.J. (1985). *Análisis estadístico para datos categóricos.* Madrid: Síntesis.

Ato, M. y Vallejo, G. (2015). *Diseños de investigación en Psicología*. Madrid: Pirámide.

Ato, M. y Vallejo, G. (2011). Los efectos de terceras variables en la investigación psicológica. *Anales de Psicología, 27*, 550-561.

Ato, M., Vallejo, G. y Palmer, A. (2013). The two-way mixed model: a long and winding controversy. *Psicothema, 25*, 130-136.

Ato, M., Losilla, J.M., Navarro, B., Palmer, A. y Rodrigo, M.F. (2005). *El modelo lineal generalizado*. Girona: Documenta Universitaria - EAP, S.L.

Austin, P.C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46,* 399-424.

Baek E. K., Ferron J. M. (2013). Multilevel models for multiple-baseline data: Modeling across participant variation in autocorrelation and residual variance. *Behavior Research Methods*, 45, 65–74.

Bailey, R.A. (2008). *Design of comparative experiments.* Cambridge, UK: Cambridge University Press.

Bakker, M, & Wicherts, J.M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavioral Research Methods, 43*, 666-678.

Balkin, R.S. & Sheperis, C.J. (2011). Evaluating and reporting statistical power in counseling research. *Journal of Counseling & Development, 89,* 268-272.

Balluerka, N.; Gómez, J.; Hidalgo, D. (2005). The controversy over null hypotesis significance
revisited. *Methodology, 1,* 55-70.

Baltes, P.B., Reese, H.W. & Nesselroade, J.R. (1977). *Life-span developmental psychology: introduction to research methods* (traducción español: Ed. Morata, 1981). Monterey, CA: Brooks/Cole.

Barrada, J.R. (2012). Tests adaptativos informatizados: una perspectiva general. *Anales de Psicología, 28,* 289-302.

Bijleveld, C.J.H.; van der Kamp, L.J.; Moijaart, A.; van der Kloot, A.; van der Leeden, R. & van der Burg, E. (1998). *Longitudinal data analysis: designs, models and methods.* Newbury Park, CA: Sage.

Blanca, M.J., Luna, R.; López, D., Rando, B. y Zalabardo, C. (2001). Procesamiento global y local con tareas de categorización de la orientación. *Anales de Psicología, 17*, 247-254.

Blanco, A., Losada, L. y Anguera, M.T. (2003). Data analysis techniques in observational designs applied to the environment-behavior relation. *Medio Ambiente y Comportamiento Humano, 4*, 111-126.

Bolger, N. & Laurenceau, J.P. (2013). *Intensive longitudinal methods: an introduction to diary and experience sampling research*. New York, NY: The Guilford Press.

Bollen, K.A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K.A. (2012). Instrumental variables in Sociology and the Social Sciences. *Annual Review of Sociology, 38,* 37-72.

Bono, R. y Arnau, J. (1995). Consideraciones generales en torno a los estudios de potencia. *Anales de Psicología, 11*, 193-202.

Bono, R. y Arnau, J. (1996). Análisis de la potencia del estadístico C mediante simulación. *Psicothema, 8,* 699-708.

Bono, R. y Arnau, J. (2014). *Diseños de caso único en Ciencias Sociales y de la Salud.* Madrid: Editorial Síntesis.

Bornstein, M.R.; Bellack, A.S. & Hersen, M. (1977). Social-Skills Training for Unassertive Children: A Multiple-Baseline Analysis. *Journal of Applied Behavior Analysis, 10,* 183-195.

Brewer, E.W. & Kuhn, J. (2010). Casual-comparative design. In N.J. Salkind (Ed), *Encyclopedia of Research Design,*Vol. I (pp. 124-131). Thousand Oaks, CA: Sage Publications.

Buelga, S., Cava, M.J. y Musitu, G. (2012). Reputación social, ajuste psicosocial y victimización entre adolescentes en el contexto escolar. *Anales de Psicología, 28,* 180-187.

Cajal, B.; Gervilla, E. y Palmer, A. (2012). When the mean fails, use an M-estimator. *Anales de Psicología, 28*, 281-288.

Carreño, R. (2015). Efecto del programa BETA-PUCV sobre la conducta prosocial y la responsabilidad social de sus alumnos: un análisis con regresión por discontinuidad. *Estudios Pedagógicos XLI, 41-53*.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, 2nd ed.* Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, L.; Manion, L. & Morrison, K. (2012). *Research Methods in Education, 6th Edition.* New York, NY: Routledge.

Cole, D.A. & Preacher, K.J. (2014). Manifest variable path analysis: potencially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19,* 300-315.

Cook, T.D. y Campbell, (1979). *Quasiexperimentation: Design and analysis issues for field settings*. Chicago, IL: Rand-McNally.

Cook, T.D. & Campbell, D.T. (1986). The causal assumptions of quasi-experimental practice. *Synthese, 68,* 141-180.

Cejudo, J., López-Delgado, M.L. & Rubio, M.J., P. (2016). Inteligencia emocional y resiliencia: su influencia en la satisfacción con la vida en estudiantes universitarios. *Anuario de Psicología, 46*, 51-57.

Cuesta, M., Fonseca, E., Vallejo, G. y Muñiz, J. (2013). Datos perdidos y propiedades psicométricas en los test de personalidad. *Anales de Psicología, 29,* 285-292.

Cumming, G. (2012). *Understanding the new statistics: effect sizes, confidence intervals and meta-analysis.* New York, NY: Routledge.

Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, P., Kleinig, A., Lo, J., McMenamin, N. & Wilson, S. (2007). Statistical reform in psychology: is anything changing? *Psychological Science, 18,* 230-232.

Dannels, S.A. (2010). Research design. In Handcock, G.R. & Mueller, R.O., eds., *The reviewer's guide to quantitative methods in the social sciences* (pp. 343-355). New York, NY: Routledge.

Davidov, E., Schmidt, P. & Billiet, J., eds. (2011). *Cross-cultural analysis: Methods and applications*. New York, NY: Routledge.

Del Rey, R.; Elipe, P. & Ortega, R. (2012). Bullying and cyberbullying: overlapping and predictive value of the co-ocurrence. *Psicothema, 24*, 608-613.

Dobson, A.J. (2002). *Introduction to generalized linear models, 2nd Edition*. Boca Ratón, FL: Chapman & Hall.

Dziak, J.J.; Nahum-Shani, I. & Collins, L.M. (2012). Multilevel factorial experiments for developing behavioral interventions: power, sample size and resource considerations. *Psychological Methods, 17*, 153-175.

Edgington (1992). Nonparametric tests for single-case experiments. In T. R. Kratochwill & J. R. Levin (Eds.). *Single-case research design and analysis: New directions for psychology and education* (pp. 133–157). Hillsdale, NJ: Erlbaum

Edgington, E. S. (1995).*Randomization tests* (3rd ed.). New York: Marcel Dekker.

Ellis, P.D. (2010). *The essential guide to effect sizes: statistical power, meta-analysis and the interpretation of research results.* Cambridge, UK: Cambridge University Press.

Ernster, V.L. (1994). Nested case-control studies. *Preventive Medicine, 23*, 587-590.

Escudero, J.R. & Vallejo, G. (2000). Comparación de tres métodos alternativos para el análisis de series temporales interrumpidas. *Psicothema, 12*, 480-486.

Esposito Vinzi, V., Chin, W.W., Henseler, J. & Sang, H., eds. (2010). *Handbook of partial least squares: concepts, methods and applications.* New York, NY: Springer.

Farrant, R.H. (1977). Can after-the-fact designs test functional hypotheses, and are they needed in psychology? *Canadian Psychological Review, 18*, 359-364.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Fernández-Alcántara, M., García-Caro, M.P., Pérez-Marfil, N. y Cruz-Quintana, F. (2013). Experiencias y obstáculos de los psicólogos en el acompañamiento de los procesos de fin de vida. *Anales de Psicología, 29,* 1-8.

Fernández-Berrocal, P. y Extremera, N. (2006). Emotional intelligence and emotional reactivity and recovery in laboratory context. *Psicothema, 18*, 72-78.

Ferron, J. & Foster-Johnson, L. (1998). Analyzing single-case data with visually guided randomization tests. *Behavior Research Methods, Instruments & Computers, 30*, 698-706.

Fitzmaurice, G.; Laird, N. & Ware, J. (2009). *Applied longitudinal analysis.* New York, NY: Wiley.

Florez-Alarcón, L. y Vélez-Botero, H. (2010). Efecto de la sensibilización pretest de prevención selectiva del abuso alcohólico. *Revista Latinoamericana de Medicina Conductual, 1*, 2-35.

Foxx, R.M. & Rubinoff, A. (1979) Behavioral treatment of caffeinism: reducing excessive coffee drinking. *Journal of Applied Behavior Analysis, 12*, 335-344.

Frías, M.D, Pascual, J. y García, J.F. (2000). Tamaño del efecto de tratamiento y significación estadística. *Psicothema, 12, sup.2*, 236-240.

García, J.F., Frías, M.D. y Pascual, J. (1999). Potencia estadística del diseño de Solomon. *Psicothema, 11*, 431-436.

García-Izquierdo, A.L.; García-Izquierdo, M. y Ramos-Villagrasa, P.J. (2007). Aportaciones de la inteligencia emocional y la autoeficacia: aplicaciones para la selección de personal. *Anales de Psicología, 23,* 231-239.

García-Sánchez, J.N. y Rodríguez, C. (2007). Influencia del intervalo de registro y del organizador gráfico en el proceso-producto de la escritura y en otras variables psicológicas. *Psicothema, 19*, 198-205.

Garson (2015). *Missing values analysis & data imputation, 2015 Edition.* Asheboro, NC: Statistical Publishing Associates.

Gartzia, L., Aritzeta, A., Balluerka, N. y Barberá, E. (2012). Inteligencia emocional y género: más allá de las diferencias sexuales. *Anales de Psicología, 28,* 567-575.

Gast, D.L. & Ledford, J.R. (2014). *Single case research methodology: applications in special education and behavioral sciences, 2nd Ed.* New York, NY: Routledge.

Gay, L.R., Mills, G.E. & Airasian, P.W. (2012). *Educational research: competencies for analysis and applications*, *10th Edition*. NJ: Pearson.

Gelman, A. (2006). Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics, 48*, 432-435.

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge University Press.

Gennetian L.A.; Magnuson, K. & Morris, P.A. (2008). From statistical association to causation: What developmentalists can learn from instrumental variables techniques coupled with experimental data. *Developmental Psychology*, 44, 381-384.

Gómez, J. (1995). Análisis de datos de medidas repetidas mediante modelos de ecuaciones estructurales. En J. Arnau (ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento* (pp. 255-304). Méjico: Limusa.

Gliner, J.A., Leech, N.L. & Morgan, G. A. (2002). Problems with null hypothesis significance testing (NHST): what do the textbooks say? *The Journal of Experimental Education, 7*, 83-92.

Graham, J.W. (2012). *Missing data: analysis and design*. New York, NY: Springer.

Grimes, D.A. (2002). Bias and causal associations in observational research. *Lancet, 359 (9302),* 248-252.

Grissom, R.J. & Kim, J.J. (2012). *Effect sizes for research: a broad practical approach, 2nd Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Gruber, A.J., Pope, H.G., Hudson, J.I. & Yurgelun-Todd, D. (2003). Attributes of long-term heavy cannabis users: a case-control study. *Psychological Medicine, 33,* 1415-22.

Hair, J.F., Hult, G.T.M., Ringle, C.M. & Sarstedt, M. (2013). *A primer on partial least squares structural equation modeling (PLS-SEM).* Thousand Oaks, CA: Sage.

Hancock, G.R. & Mueller, R.O., eds. (2010). *The Reviewer's Guide to Quantitative Methods in the Social Sciences.* New York, NY: Routledge.

Hardin, J. & Hilbe, J. (2007). *Generalized linear models and extensions*, 2nd Edition. College Station, TX: Stata Press.

Harlow, L.L., Mulaik S.A. & Steiger, J.H. (1997). *What to do if there were no significance tests?* Mahwah, NJ: Erlbaum.

Hayes, A.F. (2013). *Introduction to mediation, moderation and conditional process analysis: a regression-based approach.* New York, NY: Guilford Press.

Hedeker, D. (2005). Generalized linear mixed models. In B. Everitt & D. Howell (eds): *Encyclopedia of Statistics in Behavioral Science.* New York, NY: Wiley.

Hoenig, J.M. & Heisey, D.M. (2001): The abuse of power: the pervasive fallacy of power calculations in data analysis. *The American Statistician, 55,* 19-24.

Hoffman, L. & Rovine, M.J. (2007). Multilevel models for the experimental psychologist: foundations and illustrative examples. *Faculty Pblications, Department of Psychology.* Paper 421.

Hox, J. (2011). *Multilevel analysis: Techniques and applications*, *2nd Ed.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hudson J.L, Pope H.G, Jr, & Glynn R.J. (2005). The cross-sectional cohort study: an underutilized design. *Epidemiology, 16*, 355–359.

Huitema, (2011). *The analysis of covariance and alternatives: Statistical Methods for experiments, quasi-experiments and single-case studies, 2nd Ed*. New York, NY: Wiley.

Ivarsson, A., Andersen, M.B., Johnson, U. & Lindwall, M. (2013). To adjust or not adjust: nonparametric effect sizes, confidence intervals and real-world meaning. *Psychology of Sport and Exercise, 14*, 97-102.

Jarde, A.; Losilla, J.M. y Vives, J. (2012). Instrumentos de evaluación de la calidad metodológica de estudios no experimentales: una revisión sistemática. *Anales de Psicología 28*, 617-628.

Jiménez, M.G., Martínez, M.P.; Miró, E. y Sánchez, A.I. (2012). Relación entre estrés percibido y estado de ánimo negtivo: diferencias según el estilo de afrontamiento. *Anales de Psicología, 28*, 28-36.

Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher, 30*, 3-13.

Jose, P.E. (2013). *Doing statistical mediation and moderation*. New York, NY: The Guilford Press.

Judd, C. & Kenny, D. (1981). *Estimating the effects of social interventions*. Cambridge, MA: Cambridge University Press.

Judd, C.M, McClelland, G.H. & Ryan, C.S. (2009). *Data Analysis: a model comparison approach*. New York, NY: Routledge.

Kaplan, D. (2009). *Structural Equation Modeling: foundations and extensions, 2nd ed.* Los Angeles, CA: Sage Publications, Inc.

Kazdin AE. (2011). *Single case research designs: Methods for clinical and applied settings, 2nd. ed*. NewYork, NY: Oxford University Press.

Kelley, K. & Preacher, .J. (2012). On effect size. *Psychological methods, 17*, 137-152.

Kennedy, C.H. (2005). *Single-case designs for educational research*. Boston, MA: Allyn and Bacon.

Kerlinger, F.N. (1985). *Foundations of behavioral research, 3rd. Ed*. New York, NY: Holt, Rinehart & Winston.

Kirk, R.E. (2005). The importance of effect magnitude. En S.F. Davis (ed.): *Handbook of research methods in Experimental Psychology,* 83-105. Oxford, UK: Blackwell.

Kirk, R.E. (2013). *Experimental design: procedures for the behavioral sciences, 3rd Edition*. New York, NY: Wadsworth Publishing.

Kish, L. (1987). *Statistical design for research*. New York, NY: Wiley.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

Kline, R.B. (2009). *Become a behavioral science researcher. A guide to producing research that matters*. New York, NY: The Guilfond Press.

Kline, R.B. (2011). *Principles and practice of structural equation modeling, 3rd Edition*. New York, NY: The Guilford Press.

Knapp, T.R. & Mueller, R.O. (2010). Realibility and validity of instruments. In G.R. Hancock & R.O. Mueller, eds.: *The reviewer guide to quantitative methods in the social sciences (pp. 337-341)*. New York, NY: Routledge.

Kraemer, H. & Thiemann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage Publications.

Kratochwill & Levin (2010a). *Single-case intervention research: methodological and statistical advances*. Washington, DC: American Psychological Association.

Kratochwill & Levin (2010b). Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. *Psychological Methods, 15,* 124-144.

Kratochwill, T.R.; Hitchcock, J.H.; Horner, R.H.; Levin, J.R.; Odom, S.L.; Rindskopf, D.M. & Shadish, W.R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*, 26-38.

Labrador, F.J. & Alonso, E. (2007). Evaluación de la eficacia a corto plazo de un programa de intervención para el trastorno de estrés postraumático en mujeres mexicana víctimas de violencia doméstica. *Revista de Psicopatología y Psicología Clínica, 12*, 117-130.

Lavrakas, P.J., ed. (2008). *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage Publications.

Laursen, B.; Little, T.D. & Card, N.A., eds. (2013). *Handbook of developmental research methods*. New York, NY: Guilford Press.

Lenth, R.V. (2001). Some practical guidelines for effective sample size determination. *American Psychologist, 55*, 187-193.

Lenth, R.V. (2007). *Post hoc power: tables and commentary*. Technical Report nº 378. The University of Iowa: Department of Statistics and Actuarial Science.

Levin, J. R., & Robinson, D. H. (2000). Statistical hypothesis testing, effect-size estimation, and the conclusion coherence of primary research studies. *Educational Researcher, 29*, 34–36.

Levine, M. & Ensom, M.H.H. (2001). Post hoc power analysis: an idea whose time has passed?. *Pharmacotherapy, 21,* 405-409.

Levy, J.P. y Varela, J., eds. (2003). *Análisis multivariable para las Ciencias Sociales*. Madrid: Pearson Educación, S.A.

Levy, J.P. y Varela, J., eds. (2006). *Modelización con estructuras de covarianzas en Ciencias Sociales*. La Coruña: Netbiblo, S.L.

Light, R.J., Singer, J.D. & Willett, J.B. (1990). *By design: planning research on higher education*. Cambridge, MA: Harvard University Press.

Little, T.D., ed. (2013a). *The Oxford Handbook of quantitative methods*. 2 vls. New York, NY: Oxford University Press USA.

Little, T.D. (2013b). *Longitudinal structural equation modeling*. New York, NY: The Guilford Press.

Losilla, J.M., Navarro, B., Palmer, A., Rodrigo, M.F. y Ato, M. (2005). *Del contraste de hipótesis al modelado estadístico*. Girona: Documenta Universitaria - EAP, S.L.

Loy, C.; Goh, T. & Xie, M. (2002). Retrospective factorial fitting and reverse design of experiments. *Total Quality Management 13*, 589-602.

Luellen, J.K., Shadish, W.R. & Clark, M.H. (2005). Propensity scores: an introduction and experimental test. *Evaluation Review, 29*, 530-538.

Mann, C.J. (2003). Observational research methods. Research design II: cohort, cross-sectional and case-control studies. *Emergency Medical Journal, 20*, 54-60.

Marshall, C. & Rossman, G.B. (2006). *Designing qualitative research, 4th edition*. Thousand Oaks, CA: Sage.

Martínez-Arias, M.R. (1995a). El método de las encuestas por muestreo: conceptos básicos. En M.T. Anguera, J. Arnau, M. Ato, M.R. Martínez, J. Pascual y G. Vallejo, *Métodos de la investigación en Psicología* (pp. 385-432). Madrid: Síntesis.

Martínez-Arias, M.R. (1995b). Diseños muestrales probabilísticos. En M.T. Anguera, J. Arnau, M. Ato, M.R. Martínez, J. Pascual y G. Vallejo, *Métodos de la investigación en Psicología* (pp. 433-484). Madrid: Síntesis.

Martínez-Arias, M.R. (1995c). Las decisiones posteriores al muestreo. En M.T. Anguera, J. Arnau, M. Ato, M.R. Martínez, J. Pascual y G. Vallejo, *Métodos de la investigación en Psicología* (pp. 485-510). Madrid: Síntesis.

Martínez-Arias, M.R.; Castellanos, M.A. y Chacón, J.C. (2014). *Métodos de investigación en Psicología*. Madrid: EOS Universitaria.

Martínez-Arias, M.R.; Hernández, M.V. y Hernández-Lloreda, M.J. (2006). *Psicometría*. Madrid: Alianza Universidad.

Matsumoto, D. & van de Vijver, F.J.R., eds. (2011). *Cross-cultural research methods in psychology*. New York, NY: Cambridge University Press.

Maxwell, S.E. & Delaney, H.D. (2004). *Designing experiments and analyzing data: a model comparison perspective. 2nd Edition*. New York, NY: Psychology Press.

Maxwell, S.E.; Kelley, K. & Rausch, J.R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *American Review of Psychology, 59,* 537-563.

McClannahan, L.E.; McGee, G.G.; McDuff, G.S. & Krantz, P.J. (1990). Assessing and improving child care: A personal appearance index for children with autism, *Journal of Applied Behavior Analysis, 23*, 469.482.

McConway, K.J.; Jones, M.C. & Taylor, P.C (1999). *Statistical modeling using Genstat*. Oxford, UK: Oxford University Press.

Menard, S. (2008). Separating age, period, and cohort effect in developmental and historical research. In S.Menard (ed.), *Handbook of longitudinal research: design, measurement, analysis* (pp. 219-232). San Diego, CA: Academic Press.

Meyer (1991). Misinterpretation of interacton effects: A reply to Rosnow and Rosenthal. *Psychological Bulletin, 110*, 571-573.

Miller, G.A. & Chapman, J.P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology,* 110, 40-48.

Milliken G.A. & Johnson, D.E. (2002). *Analysis of messy data. II: analysis of covariance*. New York, NY: Van Nostrand.

Milliken G.A. & Johnson, D.E. (2009). *Analysis of messy data. I: designed experiments*. 2nd edition. New York, NY: Van Nostrand.

Mindick, B. & Oskamp, S. (1979). Longitudinal predictive research: an approach to methodological problems in studying contraception. *Journal of Population, 2,* 259-264.

Montero, I. y León, O.G. (2007). A guide for naming research studies in psychology. *International Journal of Clinical and Health Psychology, 7,* 847-862.

Muñiz, J., Elosúa, P. y Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema, 25*, 151-157.

Núñez, J.C., Vallejo. G., Rosário, P., Tuero, E. y Valle, A. (2014). Variables del estudiante, del profesor y del contexto en la predicción del rendimiento académico en Biología: análisis desde una perspectiva multinivel. *Revista de Psicodidáctica, 19(1)*, 145-171.

Onwuegbuzie, A.J. & Leech, N.L. (2004). Post hoc power: a concept whose time has come. *Understanding Statistics, 3,* 201-230.

Orgilés, M., Méndez, F.X., Rosa, A.I. e Inglés, C. (2003). La terapia cognitivo-conductual en problemas de ansiedad generalizada y ansiedad por separación: un análisis de su eficacia. *Anales de Psicología, 19,* 193-204.

Pardo, A. y Ferrer, R. (2013). Significación clínica: falsos positivos en la estimación del cambio individual. *Anales de Psicología, 29*, 301-310.

Pardo, A., Garrido, J., Ruiz, M.A. y San Martín, R. (2008). La interacción entre factores en el análisis de varianza: errores de interpretación. *Psicothema, 19,* 343-349.

Parker, R.I., Vannest, K.J. & Davis, J.L. (2011). Effect size in single-case research: a review of nine nonoverlap techniques. *Behavior Modification, 35*, 303-322.

Pascual, J., Frías, M.D. y García, J.F. (2004). Usos y abusos de la significación estadística (NHST): propuestas de futuro (¿Necesidad de nuevas normativas editoriales?). *Metodología de las Ciencias del Comportamiento, número especial,* 465-469.

Pearl, J. (2009). *Causality: models, reasoning and inference, 2nd Edition*. Cambridge, MA: Cambridge University Press.

Peng, C-Y., Chen, L-T., Chiang, H-M., Chiang, Y-C. (2013). The impact of APA and AERA guidelines on effect size reporting. *Educational Psychological Review, 25*, 157-209.

Pedhazur, E.J. & Smelkin, L.P. (1991). *Measurement, Design and Analysis: an integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Piamontesi, S.E.; Esteban, D., Furlan, L.A., Sánchez-Rosas, J. y Martínez, M. (2012). Ansiedad ante los exámenes y estilos de afrontamiento ante el estrés académico en estudiantes universitarios. *Anales de Psicología, 28,* 89-96.

Pierce, C.A., Block, R.A. & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA. *Educational and Psychological Measurement, 64*, 915-923.

Planes, M., Prat, F.X., Gómez, A.B., Gras, M.E. y Font-Mayolas, S. (2012). Ventajas e inconvenientes del uso del preservativo con una pareja afectiva heterosexual. *Anales de Psicología, 28,* 161-170.

Pons, R.M., González, M.E. y Serrano, J.M. (2008). Aprendizaje cooperativo en matemáticas: un estudio intracontenido. *Anales de Psicología, 24,* 253-261.

Preacher, K.J., Wichman, A.L., McCallum, R.C. & Briggs, N.E. (2008). *Latent growth curve modeling.* Quantitative Applications in the social sciences, 157. Thousand Oaks, CA: Sage.

Quené, H. & Van Den Bergh (2004). On multi-level modeling of data from repeated measured designs: a tutorial. *Speech Communication,* 43, 103-121.

Rabadán, R. y Ato, M. (2003). *Técnicas cualitativas para investigación de mercados*. Madrid: Pirámide.

Reichardt, C.S. (2006). The principle of paralellism in the design of studies to estimate treatment effects. *Psychological Methods, 11*, 1-18.

Richardson, J.T.E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*, 135-147.

Ríos, M.I., Godoy, C. y Sánchez, J. (2011). Síndrome de quemarse por el trabajo, personalidad resistente y malestar psicológico en personal de enfermería. *Anales de Psicología, 27*, 71-79.

Rodgers, J.L. (2010).The epistemology of mathematical and statistical modeling: a quiet methodological revolution. *American Psychologist, 65*, 1-12.

Rodrigo, M.F., Molina, J.G., García-Ros R. y Pérez-González, F. (2012). Efectos de interacción en la predicción del abandono en los estudios de Psicología. *Anales de Psicología, 28,* 113-119.

Rosa, A.I., Iniesta, M. y Rosa, A. (2012). Eficacia de los tratamientos cognitivo-conductuales en el trastorno obsesivo-compulsivo en niños y adolescentes: una revisión cualitativa. *Anales de Psicología, 28,* 313-326.

Rosa, A.I., Olivares, J. y Sánchez, J. (1999). Meta-análisis de las intervenciones conductuales de la enuresis en España. *Anales de Psicología, 15,* 157-167.

Rosário, P., Lourenço, A., Paiva, M.O.; Núñez, J.C., González-Pienda, J.A. y Valle, A. (2012). Autoeficacia y utilidad percibida como condiciones necesarias para un aprendizaje académico autorregulado. *Anales de Psicología, 28*, 37-44.

Rosel, J. (1995). Problemas metodológicos de muestreo en los diseños longitudinales. En J. Arnau (ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento* (pp. 55-75). Méjico: Limusa.

Rosenbaum, P.R. & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.

Rosenberg, M.A.; Frees, E.W.; Sun, J., Johnson, P.H. & Robinson, J.M. (2006). Predictive modeling with longitudinal data: a case study of Wisconsis Nursing Homes. *North American Actuarial Journal, 11,* 54-69.

Rosnow, R.L & Rosenthal, R. (2009). Effect sizes: why, when and how to use them. *Zeirschrift für Psychologie, 217*, 6-14.

Rubin, D.B. (2004). Teaching statistical inference for causal effects in experiments and observational studies. *Journal of Educational and Behavioral Statistics, 29*, 343-367.

Ruiz-Vargas, J.M. y Cuevas, I. (1999). Priming perceptivo versus priming conceptual y efectos de los niveles de procesamiento sobre la memoria implícita. *Psicothema, 11*, 853-871.

Sakett, G.P., ed. (1978). *Observing behavior: Vol II. Data collection and analysis methods.* Baltimore: University Park Press.

Salkind, N.J., Ed. (2010). *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage.

Sánchez, V. Ortega, R. y Menesini, E. (2012). La competencia emocional de agresores y víctimas de bullying. *Anales de Psicología, 28,* 71-82.

Sánchez-Meca, J. y Marín-Martínez, F. (2010). Meta-analysis in psychological research. *International Journal of Psychological Research, 3,* 151-163.

Sánchez-Meca, J., Rosa, A.I. y Olivares, J. (2004). El tratamiento de la fobia social específica y generalizada en Europa: un estudio meta-analítico. *Anales de Psicología, 20*, 55-68.

Schaie, K.W. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64,* 9-107.

Schaie, K.W. (1994). Developmental designs revisited. In S.H.Cohen & H.W.Reese (eds), *Life-span developmental psychology* (pp.45-64). Hillsdale,NJ: Erlbaum.

Schaie, K.W. & Caskie, G.I.L. (2006). Methodological issues in aging research. In D.M. Teti (ed.), *Handbook of Research Methods in Developmental Science*, 21-39. Oxford, UK: Blackwell Publishing.

Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: implications for the training of researchers. *Psychological Methods, 1*, 115-129.

Schmidt, K. & Teti, DM. (2006). Issues in the use of longitudinal and cross-sectional designs. In D.M. Teti (ed.), *Handbook of Research Methods in Developmental Science*, 3-20. Oxford, UK: Blackwell Publishing.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W.H. & Shavelson, R.J. (2007). Estimating causal *effects using experimental and observational designs: A think tank white paper*. Washington, D.C. American Educational Research Association.

Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: implications for the training of researchers. *Psychological Methods, 1,* 115-129.

Shadish, W.R. (2010). Campbell and Rubin: a primer and comparison of their approaches on causal inference in field settings. *Psychological Methods, 15*, 3-17.

Shadish, W.R & Clark, M.H. (2004). An introduction to propensity scores. *Metodología de las Ciencias del Comportamiento, 4,* 291-300.

Shadish, W.R. & Rindskopf, D. (2007). Analyzing data from single-case designs using multilevel models: new applications and some agenda items for future research. *Psychological Methods, 18,* 385-405.

Shadish, W.R., Cook, T.D. & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Shadish, W., Cook, T.D. & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin Company.

Shadish, W., Kyse, E.N. & Rindskopf, D.M. (2013). Analyzing data from single.case designs using multilevel models: new applications and some agenda items for future research. *Psychological Methods, 18,* 385-405.

Shaughnessy, J.; Zechmeister, E. & Zechmeister, J. (2012). *Research methods in psychology, 9th edition*. New York, NY: McGraw-Hill.

Shmueli, G. (2010). To explain or to predict. *Statistical Science, 25*, 289-310.

Sicilia, A.; Aguila, C.; Muyor, J.M., Orta, A. y Moreno, J.A. (2009). Perfiles motivacionales de los usuarios en centros deportivos municipales. *Anales de Psicología, 25*, 160-168.

Sidman, M. (1960) *Tactics of scientific research: evaluating experimental data in psychology.* Boston, MA: Authors Cooperative Inc.

Simonton, D.K. (1997). Creative productivity: a predictive and explanatory model of career trajectories and landmarks. *Pychological Review, 104,* 66-89.

Singer, J.D. & Willett, J.B. (2003). *Applied longitudinal data analysis: modeling change and event occurrence.* New York, NY: Oxford University Press.

Smith, J.D. (2012). Single-case experimental designs: a systematic review of published research and current standards. *Psychological Methods, 17*, 510-550.

Snijders, T.A.B. & Bosker, R.J. (2011). *Multilevel analysis. An introduction to basic and advanced multilevel modeling, 2nd Edition.* Newbury Park, CA: Sage.

Snyder, P. (1991). Three reasons why stepwise regression methods should no be used by researchers. En B. Thompson, ed., *Advances in Social Science Methodology* (vol.1, pp. 99-105). Greenwich, CT: JAI Press.

Spector, P.E. (1981). *Research Designs.* Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage.

Stefani, D. y Feldberg, C. (2006). Estrés y estilos de afrontamiento en la vejez: un estudio comparativo en senescentes argentinos institucionalizados y no institucionalizados. *Anales de Psicología, 22,* 267-272.

Steiner, P.M.; Cook, T.D.; Shadish, W.R. & Clark. M.H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods, 15*, 250-267.

Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: grounded theory procedures and techniques.* Newbury Park, CA: Sage.

Stuart, E.A. & Rubin, D.B. (2008). Best practices in quasi-experimental designs: matching methods for causal inference. In Osborne, J., ed., *Best practices in quantitative methods*, (p. 155-176). Thousand Oaks, CA: Sage.

Suchman, E.A. (1967). *Evaluative research: principles and practices in public service and social action programs.* New York, NY: Russell Sage.

Sun, S., Pan, W. & Wang, L. (2010). A comprehensive review of effect size reporting and interpreting practices in academic journals in Education and Psychology. *Journal of Educational Psychology, 102*, 989-1004.

Swaminathan, H.; Rogers, H.J.; Horner, R.H.; Sugai G. & Smolkowski, K. (2014). Regression models and effect size measures for single case designs. *Neuropsychological Rehabilitation, 24,* 554-571.

Taris, T.W. (2000). *A Primer in Longitudinal Data Analysis.* Newbury Park, CA: Sage.

Teti, D.M., Ed. (2006). *Handbook of Research Methods in Developmental Science.* New York, NY: Wiley.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: a guidelines editorial. *Educational and Psychological Measurement, 55*, 525-534.

Thompson, B. (2001). Significance, effect sizes, stepwise methods and other issues: strong arguments move the field. *Journal of Experimental Education, 70,* 80-93.

Thompson, B. (2002a). “Statistical,” “practical,” and “clinical”: How many kinds of significance do counselors need to consider? *Journal of Counseling and Development, 80*, 64-71.

Thompson, B. (2002b). What future quantitative social science research could look like: confidence intervals for effect sizes. *Educational Researcher, 31,* 24.31.

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools, 44,* 423-432.

Tincani M, Crozier S, Alazetta L. (2006). The Picture Exchange Communication System: Effects on manding and speech development for school-aged children with autism. *Education and Training in Developmental Disabilities,41,*177–184.

Tukey, J. (1949). One degree of freedom for non-additivity. *Biometrics*, *5*, 232–242.

Valentine, J.C. & Coopers, H. (2003). *Effect size substantive interpretation guidelines: issues in the interpretation of effect sizes*. Washington, DC: What Works Clearinghouse.

Valiña, M.D.; Seoane, G.; Ferraces, M.J. y Martín, M. (1995). Tarea de selección de Wason: un estudio de las diferencias individuales. *Psicothema, 7,* 641-653.

Valentine, J. C. & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. Washington, DC: What Works Clearinghouse.

Vallejo G. (1995a). Diseños de series temporales interrumpidas. En Anguera, M.T., Arnau, J., Ato, M., Martínez-Arias, R., Pascual, J. y Vallejo. G., *Métodos de investigación en psicología* (pp. 321-334). Madrid: Síntesis.

Vallejo G. (1995b). Análisis de los diseños de series temporales interrumpidas. En Anguera, M.T., Arnau, J., Ato, M., Martínez-Arias, R., Pascual, J. y Vallejo. G., *Métodos de investigación en psicología* (pp. 335-352). Madrid: Síntesis.

Vallejo, G. y Fernández, P. (1995). Diseño de medidas parcialmente repetidas con ausencia de independencia en los errores. En J. Arnau (ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento* (pp. 223-253). Méjico: Limusa.

Vallejo, G., Ato, M., Fernández, P., & Livacic-Rojas, P. E. (2013). Multilevel bootstrap analysis with assumptions violated. *Psicothema*, *25*. 520-528.

Vallejo, G.; Arnau, J.; Bono, R.; Cuesta, M.; Fernández, P. y Herrero, J. (2002). Análisis de diseños transversales de series temporales cortas mediante procedimientos paramétricos y no paramétricos. *Metodología de las Ciencias del Comportamiento, 4,* 301-323.

Vandenbroucke, J.P.; von Elm, E., Altman, D.G., Gotzshe, P.C., Mulrow, C.D., Pocock, S.J., Poole, C.; Schlesselman, J.J. & Egger, M. (2007), Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Epidemiology, 18*, 805-835.

Vélez, M.; Egurrola, J. y Barragán, F.J. (2013). Ronda clínica y epidemiológica: uso de la puntuación de propensión (*propensity score*) en estudios no experimentales. *Iatreia, 26*, 95-101.

Velicer WF, Molenaar P. ( ). Time Series Analysis. In Schinka J. & Velicer W.F. (eds.) *Research Methods in Psychology*, *Handbook of Psychology vol. 2* (I. B. Weiner, Editor-in-Chief), 2nd Ed. (pp. 628-660). New York: John Wiley & Sons.

West, B.T.; Welch, K.B. & Galecki, A.T. (2014). *Linear mixed models: a practical guide using statistical software, 2nd. Ed*. Boca Raton, FL: Chapman and Hall/CRC.

Whitman, C., Van Rooy, D.L., Viswesvaran, C. & Alonso, A. (2008). The susceptibility of a mixed model measure of emotional intelligence to faking: a Solomon four-group design. *Psychology Science Quarterly, 40*, 44-63.

Wilkinson, L. and the *Task Force on Statistical Inference* (1999). Statistical methods in psychology journals: guidelines and explanations. *American Psychologist, 54*, 594-604.

Woodside, A.G. (2010). *Case study research: theory, methods and practice*. Bingley, UK: Emerald.

(Article received: 3-6-2013; accepted in Spanish: 4-7-2013; translated: 7-12-2017)