

Un sistema de clasificación de los diseños de investigación en psicología

Manuel Ato*, Juan J. López y Ana Benavente

Departamento de Psicología Básica y Metodología, Universidad de Murcia

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.

Title: A classification system for research designs in psychology.

Abstract: In this work we devise a conceptual framework and develop some basic principles to promote 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.

Introducción

Con el incremento en el número de manuscritos que reciben las revistas de psicología con factor de impacto, se ha convertido en una necesidad la normalización, estandarización y actualización de los criterios para la revisión metodológica de trabajos empíricos (Ramos y Catena, 2004). En respuesta a esta necesidad, el Comité Editorial de la revista *Anales de Psicología* se ha planteado actualizar sus criterios de revisión metodológica elaborando una clasificación de los diseños de investigación utilizados en sus publicaciones con el objeto de diagnosticar la presencia de errores comunes y proponer soluciones para mejorar la calidad metodológica de la revista.

El gran problema de las revisiones de artículos de investigación es que suele dejarse en manos del investigador experto en un tema sustantivo la decisión acerca de su publicación, quien en general confina sus comentarios críticos al área de contenido asumiendo que los aspectos metodológicos del trabajo en revisión están bien tratados y no contienen errores. Excepto en casos excepcionales, no suele haber revisores metodológicos de artículos de investigación sustantivos y sólo se proponen cuando un segundo proceso de revisión así lo requiere. Como consecuencia, los aspectos metodológicos se han convertido en la parte más vulnerable de los artículos de investigación, de ahí que con mucha frecuencia se requiera la ayuda de expertos en metodología, en la mayoría de los casos después de haber sido planificada y concluida la investigación.

Es por tanto un tema crucial que los investigadores dispongan de un marco conceptual donde situar sus proyectos de investigación, conozcan bien algunos de los principios básicos que sustentan una revisión metodológica y dispongan de un repertorio de diseños para planificar su investigación apropiadamente. El objeto fundamental de este trabajo es presentar tanto un marco conceptual para evaluar el proceso de investigación como una propuesta de clasificación de

los diseños de la investigación empírica más empleados en la psicología básica y aplicada.

Un marco conceptual para la investigación empírica en psicología

Tiene interés examinar el proceso general de investigación con el esquema propuesto por Kline (2009, p. 40, véase también Pedhazur & Smelkin, 1991), quienes distinguen tres pilares metodológicos que sustentan este proceso: el diseño, la medida y el análisis, íntimamente relacionados a su vez con las cuatro formas de validez de la investigación postuladas por Campbell y colaboradores (véase Shadish, Cook & Campbell, 2002), validez interna, de la conclusión estadística, de constructo y externa (véase Figura 1).

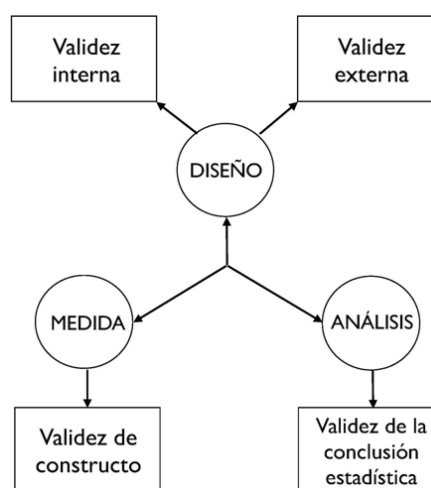


Figura 1. Los tres pilares del proceso de investigación.

El primer pilar del proceso de investigación es el diseño, que puede definirse como un plan que proporciona una estructura para integrar todos los elementos de un estudio empírico de modo que los resultados sean creíbles, libres de sesgo y generalizables (Dannels, 2010). Partiendo de un problema como objetivo, el diseño de investigación se encarga

*** 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, España). E-mail: matogar@um.es

de aspectos cruciales del proceso de investigación tales como la selección y asignación de los participantes y el control de las variables extrañas potenciales presentes en el contexto de investigación. Dos tipos de validez determinan la calidad de la aplicación del diseño: la validez interna (que se refiere a la capacidad de controlar el efecto de terceras variables que pueden ser causas alternativas a la causa investigada) y la validez externa (que se refiere a la capacidad de generalizar los resultados a otros participantes, a otros contextos y a otros momentos temporales). El equilibrio óptimo entre la validez interna y la validez externa es uno de los objetivos más deseables en un buen diseño de investigación, pero los estudios de la estrategia manipulativa priman la validez interna sobre la externa mientras que los estudios de la estrategia descriptiva priman la validez externa sobre la interna.

El segundo pilar del proceso de investigación es la medida, que concierne a la identificación, definición y medición de las variables observables y no observables, y la generación de valores numéricos o datos empíricos que son la entrada de los procedimientos de análisis estadístico (véase Martínez-Arias, 1995d y Martínez-Arias, Hernández y Hernández-Lloreda, 2006). Es crítico que los datos tengan un alto grado de fiabilidad (o sea, que el error aleatorio se haya minimizado). El tipo de validez relacionada con la medida es la validez de constructo (que se refiere a la capacidad de definir y operar apropiadamente las variables de la investigación).

El tercer pilar es el análisis, que concierne a la estimación de parámetros y prueba de hipótesis acerca del objetivo de investigación planteado con los procedimientos estadísticos más apropiados. Pero es importante recordar que la prueba de hipótesis estadísticas no es lo primario y actualmente se demanda mayor énfasis en la exactitud de la estimación de parámetros que en la prueba de hipótesis (véase Maxwell, Kelley & Raush, 2008), y se valoran aspectos tales como el tamaño del efecto y la significación práctica y clínica de los resultados (Thompson, 2002a). El tipo de validez relacionada con el análisis es la validez de la conclusión estadística (que se refiere a si el método de análisis utilizado es correcto y si el valor de las estimaciones se aproxima al de la población).

Es importante resaltar que, desde un punto de vista metodológico, aunque una investigación sea perfecta en su concepción sustantiva, puede resultar arruinada si no se utilizan apropiadamente los pilares metodológicos del diseño, la medida y el análisis. Contrariamente a lo que muchos investigadores piensan, el empleo de técnicas estadísticas sofisticadas no mejoran los resultados de una investigación si se diseñó mal o no se utilizaron medidas apropiadas. Es crucial a este respecto que el investigador centre toda su atención en la selección de un diseño apropiado, valorando sus potencialidades y sus inconvenientes para conseguir el mayor grado de equilibrio entre validez interna y externa.

Principios básicos en el proceso de diseño y evaluación de una investigación

Hay algunos principios generales que muchos revisores de artículos de investigación suelen utilizar como guía para garantizar un proceso coherente de investigación y que pueden servir al investigador para valorar su trabajo (véase una exposición más completa de tales principios en Light, Singer & Willet, 1990, Kline, 2009 y Hancock y Mueller, 2010).

1) Toda investigación se diseña para responder a un objetivo específico. Los revisores de un informe de investigación esperan encontrar una correspondencia entre el problema de investigación y el diseño específico utilizado en su potencial solución. Un error bastante común es plantear objetivos diferentes para un mismo diseño, como sucede cuando se plantea un objetivo que se responde aplicando una prueba F global (ANOVA) y otro objetivo diferente al que se responde aplicando comparaciones múltiples después de la prueba F global.

2) La más importante razón que justifica la publicación de un artículo de investigación es posibilitar su replicación con la finalidad de desarrollar un conocimiento científico coherente. Los revisores esperan que el investigador se empeñe en conseguir que su informe contenga todo el material relevante para facilitar a otros investigadores la replicación de su trabajo y la aplicación de métodos de integración de estudios (meta-análisis) para contribuir a la acumulación del conocimiento científico.

3) En lo que respecta a la medida de las variables de la investigación, un artículo de investigación debe incluir información detallada de los instrumentos de recogida de información y de la naturaleza de los datos empíricos, incluyendo la definición y operacionalización de las variables. Un aspecto esencial para generar credibilidad en la medida es el informe e interpretación de la fiabilidad, la validez y las puntuaciones de corte de los instrumentos de recogida de la información utilizados (Knapp & Mueller, 2010). Y una vez obtenidos los datos empíricos, también deben tratarse otras cuestiones importantes relativas a la reducción o transformación de datos y a la potencial existencia de efectos de sesgo debidos al proceso de medición.

4) En lo referente al análisis estadístico, se deben describir los métodos estadísticos utilizados con el detalle suficiente para ser comprendidos por otros investigadores, así como los procedimientos para el tratamiento de datos perdidos (Graham, 2012), si se han utilizado. Además, si son posibles varios modelos alternativos para explicar un resultado, el investigador debe justificar la selección del más apropiado. Es muy recomendable a este respecto la monografía editada por Hancock & Mueller (2010), destinada principalmente a revisores de artículos de investigación en las Ciencias Sociales, donde se desarrolla, para cada uno de los procedimientos es-

tadísticos más comunes en psicología, un conjunto de requisitos (“*desiderata*”) que debería cumplir su aplicación. Información más detallada acerca de los métodos cuantitativos en psicología puede encontrarse en el Handbook editado por Little (2013a).

Un estudio de Bakker y Wicherts (2011) utilizó una muestra aleatoria de 281 artículos publicados en revistas de impacto y encontró que casi un 20% de los resultados estadísticos son incorrectamente reportados y que la prevalencia de error es sensiblemente mayor en revistas de bajo impacto. En un 15% de los artículos publicados se encontraron incluso pruebas significativas en favor de la hipótesis de su/s autor/es que después de recalcularlas no resultaron ser significativas.

Conviene además notar que muchos investigadores, después de aplicar las técnicas estadísticas, solamente conceden interés a los valores p de significación, dentro de la tradición del contraste de hipótesis estadísticas (“*Null Hypothesis Significance Testing*”, NHST). Pero desde hace algún tiempo esta tradición ha generado mucha controversia (véase Kline, 2004, Pascual, Frías y García, 2004, Balluerka, Gómez e Hidalgo, 2005), llevando a algunos incluso a sospechar que puede retrasar la acumulación del conocimiento (véase Schmidt, 1996, Gliner, Leech & Morgan, 2002). En 1996, la APA creó un Comité para tratar las cuestiones sobre aplicaciones de los métodos estadísticos en la investigación psicológica. Un trabajo de Wilkinson y la Task Force on Statistical Inference (1999) sugirió incluir en todo informe de investigación, además de las pruebas de significación clásicas, medidas de tamaño del efecto, indicadores de medida o incertidumbre mediante intervalos de confianza y una selección de gráficos apropiados. Así se viene recomendando también en recientes ediciones del *Publication Manual of the American Psychological Association* (APA Manual, 2010), pero todavía son mayoría los trabajos que centralizan todos los resultados en torno a valores p , y muy escasos los que utilizan estimaciones puntuales con intervalos de confianza y gráficos apropiados (Cumming et al., 2007), lo que hoy se postula como reforma necesaria para una nueva disciplina estadística en la que se complemente y amplíe la información de las pruebas de significación (Thompson, 2002b, Harlow, Mulaik & Steiger, 1997, Kline, 2004, Cumming, 2012).

5) En lo que concierne al diseño, el informe de investigación debe incluir una detallada descripción de los participantes, de las condiciones experimentales (si se introduce un tratamiento o programa) o del contexto donde se realiza el estudio (si no hay tratamiento) y de los procedimientos de control de las variables extrañas (validez interna) así como una valoración de la generalidad de sus hallazgos (validez externa), entre otras cuestiones relevantes que se tratan con detalle en varias de las entradas de la *Encyclopedia of Research Design* editada por Salkind (2010). Es necesario destacar una vez más que muchas cuestiones relativas al diseño no se tratan adecuadamente en los informes de investigación, que en general se considera de baja calidad metodológica, en particular

cuando el diseño es no experimental y el área de investigación son las ciencias de la salud (véase Vandembroucke et al., 2007, Jarde, Losilla y Vives, 2012).

En un trabajo de Harris, Reeder & Hyun (2011) se solicitó a editores y revisores de revistas de psicología de alto impacto que describieran los errores de diseño y análisis estadístico más frecuentes, y se encontraron problemáticos, entre otros, los siguientes aspectos: (a) insuficiente descripción del diseño, (b) falta de congruencia entre objetivo de investigación y diseño, (c) problemas de potencia y tamaño muestral, (d) insuficiente descripción del análisis estadístico, incluyendo violaciones de los supuestos de las pruebas y ausencia de indicadores de tamaño del efecto (e) confusión entre diseño y modelo estadístico empleado para analizarlo, (f) utilización de pruebas estadísticas múltiples sin ajuste de probabilidad, (g) inadecuado tratamiento de los valores perdidos y (h) interpretación inadecuada de los resultados de la investigación. Respuestas muy similares se encontraron en trabajos paralelos del mismo grupo para medicina (Fernandes-Taylor, Hyun, Reeder & Harris, 2011), psiquiatría y salud pública (Harris, Reeder y Hyun, 2009a,b). En un trabajo anterior de Byrne (2000) utilizando el mismo procedimiento con revistas de ciencias de la salud se demostró que la causa más relevante para rechazar un artículo fue el diseño de investigación (más del 70%) y que la sección que registró la mayor tasa de rechazos fue la de método (más de un 52%).

En nuestra opinión, las cuestiones de mayor interés de un informe de investigación que deben acaparar la atención de investigador y evaluador son las siguientes:

- La selección de una muestra de participantes, a ser posible representativa, de su población es una cuestión central que tiene importantes consecuencias tanto para la potencia (validez de la conclusión estadística) como para la generalización de los resultados (validez externa), a lo que debe seguir el procedimiento de asignación de participantes a tratamientos, si procede. La justificación del tamaño muestral óptimo se ha tratado usualmente dentro del análisis de potencia y de la tradición del contraste de hipótesis (Bono y Arnau, 1995, Kraemer & Thieman, 1987, Lenth, 2001). Lo más recomendable es determinar el tamaño muestral realizando un análisis de potencia prospectivo o *a priori* (o sea, antes de realizar el estudio), utilizando alguno de los programas de software actualmente disponibles (el más popular es *G*Power3* -Faul, Erdfelder, Lang & Buchner, 2007-, véase algún ejemplo de uso en Balkin & Sheperis, 2011), pero el investigador aplicado debe conocer que algunos paquetes estadísticos como SPSS y SAS realizan análisis de potencia retrospectivo, observado o *post-hoc* (o sea, después de haber realizado el estudio), una práctica que algunos defienden (Lenth, 2007, Onwuegbuzie & Leech, 2004), mientras otros la denuncian como inaceptable (Hoenig & Heisey, 2001, Levine & Ensom, 2001). En nuestra opinión, sería altamente recomendable que el investigador planificara su estudio con un análisis de potencia *a priori* y com-

probara después sus resultados con un análisis de potencia observado, en particular cuando no resultó significativo.

- La inclusión de indicadores de magnitud del efecto se ha recomendado reiteradamente para valorar la significación práctica de un resultado frente a la significación estadística y, en ciencias de la salud, de la significación clínica (Thompson, 2002a, Pardo y Ferrer, 2013). Aunque se han desarrollado muchos indicadores de magnitud del efecto (véanse revisiones recientes en Ellis, 2010 y Grissom & Kim, 2012), se han destacado dos tipos básicos (Rosnow & Rosenthal, 2009): indicadores tipo d (que se basan en la diferencia de medias estandarizada) e indicadores tipo r (que se basan en la correlación y la proporción de varianza explicada). En ambos casos, la interpretación más usual emplea las etiquetas de Cohen (1988), “pequeño”, “moderado” o “grande”, pero esta práctica debería ser abandonada en favor de la utilización de intervalos de confianza y la evaluación de la discrepancia entre valores de p e indicadores de magnitud del efecto (véase Sun, Pan & Wang, 2010), además de intentar la comparación de indicadores que tratan un problema común dentro de un mismo estudio y entre diferentes estudios. En una reciente revisión, Peng, Chen, Chiang & Chiang (2013) valoran el notable incremento que se ha producido en la utilización de los indicadores de magnitud del efecto en años recientes, pero denuncian la subsistencia de algunas prácticas inadecuadas en informes de investigación tales como: (a) el uso de indicadores tipo r (R^2 y η^2) no ajustados, además de un casi nulo uso de otros indicadores menos sesgados como omega cuadrado (ω^2) y la correlación intraclase (véase Ivarsson, Andersen, Johnson & Lindwall, 2013), (b) el empleo de las “etiquetas” de Cohen (alto, moderado y bajo) sin ninguna contextualización ni interpretación, (c) la persistente falta de claridad en el uso de los indicadores eta cuadrado y eta cuadrado parcial, que coinciden en modelos ANOVA de un factor pero difieren en modelos ANOVA factoriales (véase Pierce, Block & Aguinis, 2004 y Richardson, 2011), (d) el escaso uso de los intervalos de confianza tanto para los estimadores de las pruebas estadísticas como para los tamaños del efecto (véase Kelley & Preacher, 2012, Thompson, 2007), y (d) la falta de integración entre las pruebas de significación estadística y los indicadores de magnitud del efecto. A este respecto, Levin y Robinson (2000) sugieren emplear la denominada coherencia de la conclusión estadística cuando apuntan en la misma línea la significación estadística y la magnitud del efecto.
- El control de las variables extrañas es otra cuestión crucial de la validez interna de un diseño de investigación que recibe una atención insuficiente en los estudios no experimentales, ya que en los estudios experimentales la manipulación de alguna variable independiente y la aleatorización permiten al investigador controlar aceptablemente las variables extrañas y abordar analíticamente relaciones causa-

efecto. La técnica de la aleatorización consigue controlar la influencia de terceras variables, identificadas y no identificadas, cuando se dan las condiciones idóneas, pero es bien conocido que cuando alguna variable extraña tiene un impacto más potente sobre la variable de respuesta que la variable causal, el efecto de la aleatorización se desvanece. Es por tanto esencial identificar éste y otros tipos de terceras variables que generan numerosas fuentes de confundido. De hecho el estatus metodológico de un diseño de investigación depende del grado de control efectivo de las variables extrañas (Shadish, Cook & Campbell, 2002). Aunque hay un abundante repertorio de técnicas de control de variables extrañas (algunas se tratan en Ato, 1991, pp. 155-175), en general, las técnicas de control experimental (eliminación, constancia y varias formas de equilibración, como la aleatorización, el emparejamiento y el bloqueo), si se pueden utilizar, son preferibles a las técnicas de control estadístico (en particular, estandarización, ajuste y residualización). En algunas ocasiones, incluso la interpretación causal de algunos diseños no experimentales como los diseños comparativos *ex-post-facto* y algunos diseños explicativos puede facilitarse cuando se emplean sofisticadas técnicas de control tales como las puntuaciones de propensión (“*propensity scores*”, véase Shadish & Clark, 2004) o la técnica de las variables instrumentales (Gennetian, Magnuson & Morris, 2008).

Un sistema de clasificación de los diseños de la investigación empírica en psicología

Es conveniente resaltar la importancia de conocer bien el diseño que un investigador utiliza en su investigación para comprender los principios de su aplicación (dentro del marco conceptual descrito anteriormente) y los aspectos clave a destacar en su informe (siguiendo los principios básicos desarrollados más arriba). Aunque se han propuesto varios sistemas de clasificación de los diseños de investigación empírica en psicología (Johnson, 2001 y Montero y León, 2007), y partiendo del incontestable principio de que todo sistema de clasificación será necesariamente ambiguo, nuestra propuesta para clasificar la investigación en psicología se vertebrará en las cuatro clases siguientes, y se desarrolla a continuación:

- Investigación teórica
- Investigación instrumental
- Investigación metodológica
- Investigación empírica

Investigación teórica

En esta categoría se incluyen todos aquellos trabajos que recopilan los avances producidos en la teoría sustantiva o en la metodología sobre una temática de investigación específica, y también las revisiones o actualizaciones de investigación que no requieran emplear datos empíricos originales, procedentes de estudios primarios. Se excluyen aquí los trabajos

de reflexión teórica subjetiva que no se basan en una revisión detallada de los hallazgos de otros autores.

Una investigación teórica puede adoptar una de tres formas posibles:

- 1) La revisión narrativa es una revisión o actualización teórica de estudios primarios sobre una temática de investigación, rigurosa pero meramente subjetiva, sin ningún aporte empírico por parte del investigador (p.e., Sánchez, Ortega y Menesini, 2012).
- 2) La revisión sistemática es una revisión o actualización teórica de estudios primarios, con un desarrollo sistemático del proceso de acumulación de datos (selección de estudios, codificación de las variables, etc.), pero donde no se utilizan procedimientos estadísticos al uso para integrar los estudios (p.e., Orgilés, Méndez, Rosa e Inglés, 2003, Rosa, Iniesta y Rosa, 2012).
- 3) La revisión sistemática cuantitativa o meta-análisis es una integración de los estudios primarios con metodología cuantitativa (véase Sánchez-Meca y Marín-Martínez, 2010), que contiene tanto un desarrollo sistemático del proceso de acumulación de datos como la utilización de métodos estadísticos para integrar los estudios (p.e., Sánchez-Meca, Rosa y Olivares, 2004).

Investigación instrumental

En esta categoría se incluyen todos aquellos trabajos que analizan las propiedades psicométricas de instrumentos de medida psicológicos, ya sea de nuevos tests, para los que se recomienda seguir los estándares de validación de tests desarrollados conjuntamente por la *American Educational Research Association* (AERA), la *American Psychological Association* (APA) y el *National Council on Measurement in Education* (NCME), que se publicaron en su última edición en 1999 (se espera una nueva edición para 2013 ó 2014), o de la traducción y adaptación de tests ya existentes para los que se recomienda seguir las directrices propuestas por la *International Test Commission* (ITC) - véase Muñiz, Elosua y Hambleton, 2013-.

Se recomienda encarecidamente que los autores de trabajos de investigación instrumental lean también la “*Guía para la presentación de trabajos psicométricos de validación de tests en Psicología, Educación y Ciencias Sociales*”, que puede encontrarse en la web de la revista *Anales de Psicología*.

Investigación metodológica

En esta categoría se incluyen todos aquellos trabajos que presentan nuevas metodologías para el correcto tratamiento de cualquier temática relativa a los tres pilares básicos de la investigación empírica que desarrollamos más arriba: el diseño (p.e., Jarde, Losilla y Vives, 2012), la medida (p.e., Cuesta, Fonseca, Vallejo y Muñiz, 2013) y el análisis (p.e., Cajal,

Gervilla y Palmer, 2012, Vallejo, Ato, Fernández & Livacic-Rojas, en prensa), así como estudios de simulación (p.e., Bono y Arnau, 1996) y de revisión de los procedimientos metodológicos en uso (p.e., Barrada, 2012).

Investigación empírica

Para dar respuesta a los problemas de investigación que se plantean en psicología suelen utilizarse tres estrategias de investigación generalmente aceptadas (véase Light, Singer & Willett, 1990, Arnau, 1995a): estrategia manipulativa, asociativa y descriptiva. La primera configura el conjunto de estudios que son habituales en la investigación experimental; las otras dos componen un segundo conjunto de estudios que conjuntamente representan la investigación no experimental.

Con la estrategia manipulativa se persigue analizar la relación causal existente (hipótesis causales) entre dos o más variables y puede adoptar tres tipos de estudios: experimental, cuasiexperimental y caso único. Los estudios experimentales representan el ideal de investigación pero deben cumplir dos requisitos: 1) al menos una variable debe ser manipulada y 2) los participantes se asignan al azar a los tratamientos. El primer requisito es imprescindible; el segundo suele faltar en los estudios cuasiexperimentales y la mayoría de los estudios de caso único usados en contextos aplicados.

Con la estrategia asociativa se persigue explorar la relación funcional existente entre variables (hipótesis de covariación) y puede adoptar tres tipos de estudios en función de que el objeto de la exploración sea la comparación de grupos (estudio comparativo), la predicción de comportamientos y/o la clasificación en grupos (estudio predictivo) o la prueba de modelos teóricos (estudio explicativo) para su integración en una teoría subyacente. El análisis causal es también posible en algunos estudios comparativos y explicativos, pero no suele ser posible en estudios predictivos.

Con la estrategia descriptiva se pretende describir las cosas tal y como ocurren, sin ningún tipo de manipulación de variables, ni comparación de grupos, ni predicción de comportamientos ni prueba de modelos, y puede adoptar dos tipos de estudios: observacional y selectivo.

En la clasificación que se presenta en este trabajo se asume que los diseños de investigación se corresponden con una estrategia concreta, en función del objetivo que persigue. En las formas básicas del diseño de investigación que se tratan aquí, cada diseño tiene unas características propias que lo distinguen de los demás. Pero conviene no obstante advertir que en la práctica los diseños no siempre se observan en sus formas básicas, y que en muchas ocasiones es posible flexibilizar alguna/s de sus características para adaptarlas a una situación de investigación particular, aunque en este trabajo se tratan sólo las formas básicas junto a algunas de sus más comunes generalizaciones. La Figura 2 resume nuestra propuesta de clasificación de los diseños de investigación.

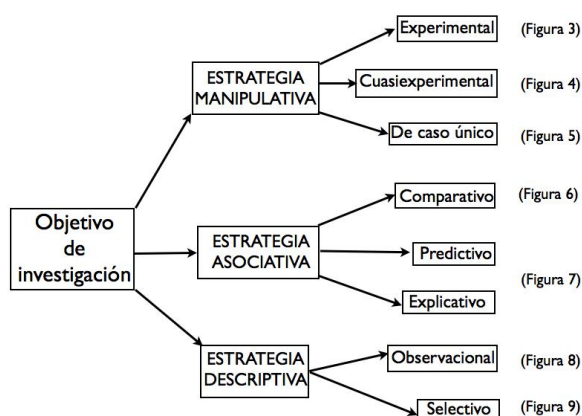


Figura 2. Clasificación de los diseños de investigación en psicología.

Estrategia manipulativa

Estudios experimentales

La investigación experimental se ha desarrollado hasta nuestros días desde la evolución de dos grandes tradiciones de investigación: la tradición clásica de laboratorio, propia de las ciencias naturales y basada en la variabilidad intraindividual, y la más moderna tradición estadística de campo, propia de las ciencias sociales y basada en la variabilidad interindividual (Ato, 1995, Cook & Campbell, 1986). De la primera subsiste una forma característica de investigación experimental que se practica en algunas áreas de la psicología aplicada con la denominación de diseños de caso único. A la segunda pertenece la metodología de investigación experimental que se practica fundamentalmente en la psicología básica y algunas áreas aplicadas con la denominación de diseños experimentales. En aquellas áreas donde no es posible o no es ético aplicar la metodología experimental se ha desarrollado una forma alternativa de investigación con el nombre de diseños cuasiexperimentales.

Los diseños experimentales y algunos diseños de caso único (o su equivalente en ciencias de la salud, los ensayos controlados aleatorizados) se caracterizan por el cumplimiento de los dos requisitos que se consideran esenciales en una investigación rigurosa: 1) manipulación de al menos una VI de tratamiento y 2) control mediante aleatorización de las potenciales variables extrañas, lo que garantiza la equivalencia inicial de los grupos. Estos dos requisitos se justifican cuando el investigador persigue el análisis de las relaciones causa-efecto entre variables. Los diseños cuasiexperimentales sólo cumplen el primero de los requisitos (manipulación), pero suelen aplicarse en situaciones donde no es posible o no es ético cumplir con el segundo (control mediante aleatorización), aunque también pueden proponerse para el análisis de las relaciones causa-efecto siempre que se utilicen procedimientos eficientes de control de variables extrañas alternativos. Del rigor y la efectividad de los procedimientos de control depende la credibilidad de la relación causa-efecto

objeto de análisis. Tres criterios básicos se consideran necesarios para abordar conclusiones del tipo causa-efecto (Bollen, 1989, Cook & Campbell, 1979), a saber: el criterio de *asociación* (existencia de covariación entre causa y efecto), el criterio de *dirección* (precedencia temporal de la causa) y el criterio de *aislamiento* (ausencia de espuriedad). Los fundamentos estadísticos de la inferencia causal en estudios experimentales y no experimentales pueden consultarse en Shadish, Cook & Campbell (2002) y en Pearl (2009).

Diseños experimentales

La clasificación más popular de los diseños experimentales utilizados en psicología se fundamenta sobre la estrategia de comparación que permiten y distinguen entre diseños intersujetos, que analizan las diferencias entre medias de tratamientos administrados al azar a grupos con diferentes participantes, diseños intrasujetos, que analizan las diferencias entre medias de tratamientos administrados a grupos con los mismos participantes, y diseños mixtos, que combinan ambas formas de comparación.

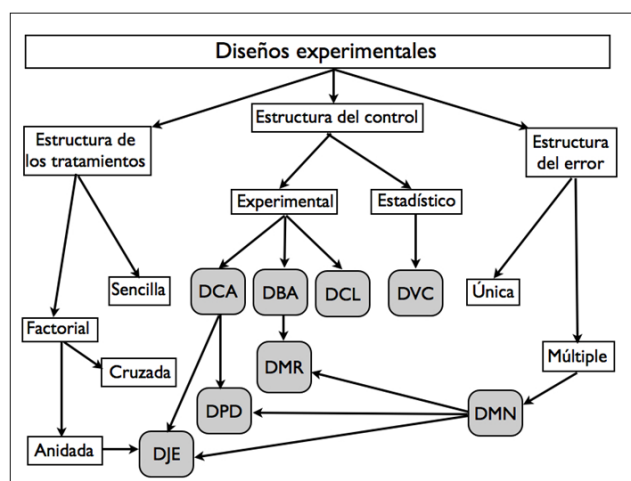
Una clasificación alternativa, menos popular pero más arraigada en la tradición estadística (Ader & Mellenbergh, 1999, Bailey, 2008, McConway, Jones & Taylor, 1999, Milliken y Johnson, 2009), parte de la formulación de un modelo estructural cada uno de cuyos elementos se asocia con uno de tres tipos de estructura. La ventaja de esta clasificación reside en la fusión óptima de los pilares del diseño y análisis estadístico, lo que permite comprender mejor la naturaleza del diseño y distinguir con claridad dos aspectos que frecuentemente se confunden: el diseño de investigación y el modelo estadístico utilizado para analizarlo (véase Ato y Vallejo, 2007).

- 1) La primera estructura, denominada *estructura de los tratamientos*, concierne al número y forma de agrupar los tratamientos en una investigación y permite distinguir entre estructura sencilla (un solo factor de tratamiento) y factorial (más de un factor de tratamiento), y dentro de ésta puede utilizarse o bien una relación de cruce entre dos o más factores (si los niveles de un factor pueden combinarse con todos los niveles del otro factor) o una relación de anidamiento (si los niveles de un factor sólo pueden combinarse con uno y sólo uno de los niveles del otro).
- 2) La segunda, denominada *estructura del control*, se refiere al número y forma de agrupar las unidades experimentales con el objeto de controlar las potenciales variables extrañas, y puede utilizar una de dos opciones posibles, el control experimental (que emplea técnicas de control tales como la aleatorización, el emparejamiento y el bloqueo) o bien el control estadístico (que emplea técnicas de ajuste estadístico tales como el uso de covariantes). Dos requisitos esenciales determinan la combinación de una estructura de tratamientos con una estructura de control: a) se asume como no significativa cualquier interacción entre la estructura de los tratamientos y la estructura del control, y b) los

factores de la estructura de control asumen efectos aleatorios.

- 3) Una tercera estructura adicional, la *estructura del error*, se refiere al número de diferentes tamaños (o entidades) de unidad experimental y de sus componentes de error existentes en el modelo estadístico asociado al diseño de investigación, y distingue entre estructura de error única (el modelo sólo tiene un único tamaño de unidad experimental y por tanto un término de error) y múltiple (el modelo tiene dos o más tamaños de unidad experimental y por tanto más de un término de error). Por ejemplo, en una investigación diseñada para evaluar la respuesta de participantes pertenecientes a diferentes grupos es preciso distinguir la variación de los participantes de la variación de los grupos. Ambos tipos de variación constituyen diferentes tipos de error y tamaños diferentes de unidad experimental (participantes y grupos).

Dada una estructura de error determinada, los diseños experimentales más comunes en la investigación psicológica pueden definirse en este contexto como una combinación peculiar de las estructuras de tratamiento y de control y se representan en la Figura 3. Asumiendo una estructura de



DCA: Diseño completamente aleatorio. DBA: Diseño de bloques aleatorios. DCL: Diseño de cuadrado latino. DVC: Diseño con variables concomitantes. DMR: Diseño de medidas repetidas. DPD: Diseño de parcela dividida. DJE: Diseño jerárquico. DMN: Diseño multinivel.

Figura 3. Diseños experimentales.

error única, la combinación de una estructura de tratamientos sencilla (un factor de tratamiento, definido con efectos fijos o aleatorios) y una estructura de control experimental mediante aleatorización representa el diseño completamente aleatorio (DCA), el diseño experimental más básico a partir del cual se construyen todos los demás, y su modelo estadístico tiene dos fuentes de variación (un efecto de tratamiento y un componente de error residual). Si el investigador desea además controlar una variable extraña mediante control experimental (p.e., emparejamiento o bloqueo), el modelo debe

incluir un término adicional para el factor a controlar, que suele ser una variable categórica con efectos aleatorios, y el resultado es el diseño de bloques aleatorios (DBA). Es importante tener en cuenta que en el DBA el factor de bloqueo carece de interés sustantivo (o sea, se propone sólo para reducir varianza), pero en cualquier diseño básico se pueden introducir, además de la variable manipulada, otras variables categóricas (p.e., sexo, nivel socioeconómico, categorías diagnósticas, etc.) con un interés puramente sustantivo, que reducen también varianza y suelen ser de efectos fijos, pero no son en sentido estricto variables de bloqueo.

El control experimental mediante bloqueo puede generalizarse incluyendo dos (o más) factores de bloqueo, en cuyo caso el modelo debe incorporar un término para cada factor y otro término para su interacción. Si se desean controlar dos variables extrañas con el mismo número de niveles, el DBA con dos variables de bloqueo puede simplificarse utilizando control experimental mediante doble bloqueo, que no requiere ningún término de interacción, y el resultado es el diseño de cuadrado latino (DCL). Si por el contrario se desea controlar una variable extraña mediante ajuste estadístico con una covariante, el modelo debe incluir un término adicional para la covariante numérica, y el resultado es el diseño con variables concomitantes (DVC). El control estadístico permite incorporar dos (o más) covariantes numéricas, y el correspondiente modelo estadístico debe incluir un término adicional para cada covariante.

Una característica común a estos cuatro diseños básicos es que son diseños intersujetos. Kirk (2013) considera el DCA, el DBA y el DCL como las estructuras básicas de diseño experimental para cuyo análisis, asumiendo variables distribuidas normalmente, son apropiados los modelos ANOVA/MANOVA y el DVC como la estructura básica de diseño para el que son apropiados los modelos ANCOVA/MANCOVA, una forma de análisis que requiere supuestos adicionales al ANOVA, pero que a pesar de sus similitudes sigue siendo una herramienta mal comprendida (Miller & Chapman, 2001, Milliken & Johnson, 2009).

Una simple extensión del DBA considera cada unidad experimental como un nivel independiente de un factor de bloqueo, en cuyo caso se produce una forma de control experimental denominada bloqueo intrasujeto y el resultado es el diseño de medidas totalmente repetidas (DMTR), donde los tratamientos son administrados en una secuencia aleatorizada a todos los participantes. El DMTR con un factor de tratamiento representa el diseño intrasujetos básico y combina una estructura de tratamiento sencilla con una estructura de control experimental mediante bloqueo intrasujeto. En comparación con el DCA básico de similares condiciones, el diseño intrasujetos suele reducir la varianza de error e incrementar la potencia estadística. Sin embargo, puesto que las diferentes medidas de una misma unidad experimental no se consideran independientes, el análisis estadístico no puede abordarse sin verificar nuevos supuestos (véase Ato y Vallejo, 2007).

Toda una variedad de diseños experimentales más complejos pueden definirse asumiendo una estructura de error única y generalizando los cuatro diseños básicos intersujetos (DCA, DBA, DCL y DVC) y el diseño básico intrasujetos (DMTR) a estructuras de tratamiento factoriales en relación de cruce, permitiendo la utilización de efectos de tratamiento fijos, aleatorios o mixtos (p.e., Pagán, Marín y Perea, 2012). En todos estos casos, la generalización genera mayores dificultades en la interpretación de las interacciones, que en general son muy propensas a errores de interpretación por parte de los investigadores (Meyer, 1991, Aiken & West, 1996, Pardo, Garrido, Ruiz y San Martín, 2007).

El diseño básico DMTR puede generalizarse con facilidad a estructuras de tratamiento factoriales, donde cada unidad experimental representa un bloque que recibe todas las combinaciones factoriales de tratamiento y el conjunto de las unidades constituye un grupo único (p.e. Blanca, Luna, López, Rando y Zalabardo, 2001). Cuando además el conjunto de unidades experimentales se divide en grupos en base a alguna variable de interés para la investigación, el resultado es el diseño de medidas parcialmente repetidas (DMPR) o diseño mixto, así llamado porque consta de un componente intersujetos y un componente intrasujetos, cada uno de los cuales representa un tamaño diferente de unidad experimental y por tanto un término propio de error. Los dos diseños intrasujetos citados (DMTR y DMPR) constituyen de hecho un diseño común, que en adelante llamaremos diseño de medidas repetidas (DMR), el segundo de los cuales es muy similar al que en investigación agrícola y biológica se denomina diseño de parcela dividida ("*split-plot design*", DPD). En el DMR experimental que tratamos aquí se asume que los tratamientos del componente intrasujetos son asignados aleatoriamente o mediante alguna secuencia que garantice la equivalencia de los grupos. Cuando no puedan asignarse aleatoriamente, porque representan una secuencia temporal, el DMR debe ser tratado como cuasiexperimental (véase Arnau, 1995g).

Una estructura factorial diferente se requiere cuando los factores se encuentran en relación de anidamiento, lo que se puede producir tanto con estructuras de tratamiento como con estructuras de control. En ambos casos, unidades experimentales de cierto tamaño se anidan dentro de otras unidades de tamaño mayor y el resultado es el diseño jerárquico (DJE). La generalización de esta situación a cualquier número de variables que incluyan alguna relación de anidamiento conduce a la vasta familia de diseños multinivel (DMN), a la que pertenecen también las dos formas del diseño de medidas repetidas, que emplean dos o más tamaños de unidad experimental y por tanto representan estructuras de error múltiple. Cabe distinguir dos diseños básicos multinivel donde unidades más pequeñas (participantes, pacientes, clientes) se anidan dentro de unidades mayores (clases, clínicas, empresas): en el primero, los sujetos dentro de cada grupo se asignan a azar a los tratamientos; en el segundo, son los grupos en su conjunto los que se asignan a azar a los tratamientos. Referencias generales sobre los diseños multinivel

son Hox (2010) y Snijders & Bosker (2011), pero para el tratamiento de datos multinivel con diseños experimentales la referencia más recomendable es Milliken & Johnson (2002, 2009).

El análisis estadístico asociado a los diseños experimentales es otra cuestión de interés para el investigador aplicado. El tradicional repertorio de pruebas basadas en el contraste de hipótesis ha dejado paso al modelado estadístico, que persigue la especificación, ajuste y comparación de modelos (Ato y Vallejo, 2007; Losilla, Navarro, Palmer, Rodrigo y Ato, 2005). Y dentro del contexto del modelado estadístico, el enfoque univariante mediante los modelos ANOVA y ANCOVA derivados del modelo lineal general se desarrolló para los diseños DCA y DVC donde todos los factores se asumían fijos. La inclusión de factores aleatorios puede ser tolerable en situaciones muy simples y altas condiciones de regularidad, pero para otras situaciones más complejas, como las derivadas de los DMR y en general de los DMN, el enfoque clásico no resulta recomendable por los problemas que genera (vease Ato, Vallejo y Palmer, 2013, Quené y van den Bergh, 2004). La alternativa al enfoque clásico fue en su momento el enfoque multivariante, que aunque sigue siendo hoy por hoy el enfoque más popular, en círculos académicos se considera más apropiada la utilización del enfoque mixto (Ato y Vallejo, 2007, Maxwell & Delaney, 2004, Milliken & Johnson, 2009, Vallejo y Fernández, 1995), un enfoque que afortunadamente se incluye ya en todos los paquetes estadísticos profesionales (GENSTAT, SAS, SPSS, STATA y R).

Diseños de caso único

Los diseños de caso único son diseños experimentales desarrollados en la tradición del control experimental de laboratorio y fueron inicialmente sistematizados por Sidman (1960). Como sucede con los diseños de la estrategia manipulativa, el análisis causal solo será posible cuando se cumplan los dos requisitos básicos (manipulación de variables y control mediante aleatorización), pero puesto que sólo hay una unidad, o a lo sumo un reducido número de unidades, el principio de aleatorización se refiere en la mayor parte de los casos a la asignación de los periodos u ocasiones de observación a los tratamientos. En los casos donde la asignación aleatoria no sea posible, como sucede en muchas aplicaciones clínicas, el investigador debe extremar su precaución a la hora de interpretar los resultados.

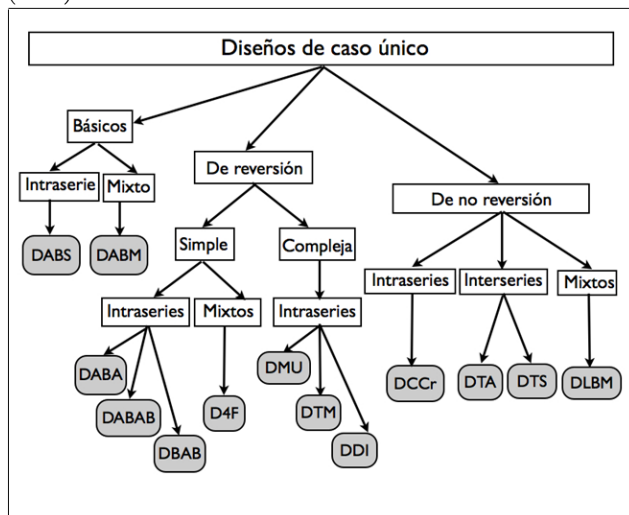
Hay dos dimensiones en la clasificación de los diseños de caso único que, siguiendo a Arnau (1995a,b,c,e), Kennedy (2005) y Ato y Vallejo (2007), nos permite distinguirlos, en una primera dimensión, en función de la reversibilidad de la respuesta a los niveles de la línea base tras retirar o alterar el tratamiento (diseños de reversión y de no reversión) y, en una segunda dimensión, en función de la estrategia de comparación utilizada (diseños interseries, intraseries y mixtos), y se representan en la Figura 4.

El diseño de caso único más básico es el diseño de un solo componente en dos fases A-B, que puede ser intraserie

(diseño básico con $N=1$, DABS) o mixto (diseño básico con N múltiple, DABM).

Los diseños de reversión se clasifican en dos grandes grupos, dependiendo del proceso de reversión, en simples y complejos. En función del número de fases, los diseños de reversión simple pueden ser de dos tipos: intraseries, que incluyen el diseño de tres fases (DABA) y el diseño de cuatro fases (DABAB) y el diseño de retirada (DBAB) y mixtos, que también tienen cuatro fases, y pueden ser de inversión o de generalización (D4F). Por su parte, entre los diseños de reversión compleja cabe citar el diseño multinivel (DMU), el diseño de tratamiento múltiple (DTM) y el diseño de interacción (DDI), que actualmente gozan de gran popularidad en áreas aplicadas.

El más interesante diseño de no reversión intraseries es el diseño de cambio de criterio (DCC), un diseño de fácil aplicación y alta flexibilidad. Entre los diseños de no reversión interseries destacan el diseño de tratamientos alternantes (DTA), y el diseño de tratamientos simultáneos (DTS). Entre los diseños con comparaciones mixtas el más popular es el diseño de línea base múltiple (DLBM), que tiene varias acepciones dependiendo de que los datos de la línea base se registren entre conductas, entre participantes o entre contextos. Más información sobre los diseños de caso único puede consultarse en Arnau (1995c), Ato y Vallejo, 2007 y Kennedy (2005).



DABS: diseño AB simple. DABM: diseño AB con N múltiple.
 DABA: diseño intraseries ABA. DABAB: diseño intraseries ABAB.
 DBAB: diseño intraseries BAB. D4F: diseño en 4 fases.
 DMU: Diseño multinivel. DTM: Diseño de tratamiento múltiple.
 DDI: diseño de interacción. DCCr: diseño de cambio de criterio.
 DTA: diseño de tratamientos alternantes.
 DTS: diseño de tratamientos simultáneos
 DLBM: diseño de línea base múltiple.

Figura 4. Diseños de caso único.

Se han propuesto muchos y muy variados procedimientos estadísticos para el análisis estadístico de diseños de caso único, desde la simple exploración visual hasta complejos procedimientos paramétricos lineales y no lineales y las

pruebas de aleatorización (véase Arnau, 1995d, Ato y Vallejo, 2007).

Diseños cuasiexperimentales

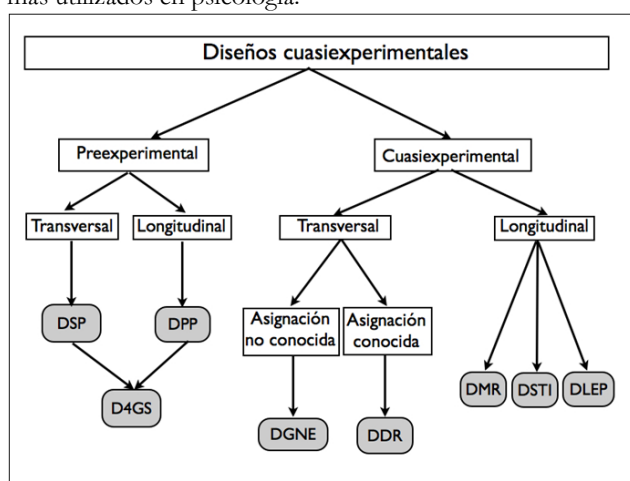
Los diseños cuasiexperimentales persiguen el mismo objetivo que los diseños experimentales, o sea, el establecimiento de relaciones causa-efecto, y cumplen con el requisito de la manipulación de al menos una VI, pero no es posible (o no es ético) cumplir con el requisito de la asignación aleatoria para asegurar que no haya diferencias entre grupos antes de asignar un tratamiento o programa. Para compensar esta ausencia se recurre al empleo de grupos de tratamiento, grupos de control, medidas pretest y postest y técnicas de control experimental (p.e., métodos de emparejamiento, véase Stuart & Rubin, 2008) con el objeto de controlar las diferencias preexistentes entre grupos. Sin embargo, en determinadas áreas de la psicología aplicada los diseños cuasiexperimentales son más utilizados que otras alternativas de diseño. Obviamente, como consecuencia de la no equivalencia inicial de los grupos, el análisis causal presenta muchas más dificultades en los diseños cuasiexperimentales que en los experimentales, aunque se han propuesto algunas técnicas de control estadístico tales como las puntuaciones de propensión (Luellen, Shadish & Clark, 2005) para mejorar el análisis de datos.

Hay muchas variedades posibles de diseños cuasiexperimentales. Shadish, Cook y Campbell (2002) presentan una detallada clasificación de los diseños cuasiexperimentales. Una presentación más breve puede consultarse en Ato (1995a), Vallejo (1995a) y Ato y Vallejo (2007). Siguiendo a Judd & Kenny (1981), la clasificación que utilizamos aquí requiere que el investigador decida en primer lugar entre dos formas de comparación, transversal (comparación intersujetos o entre grupos no equivalentes) y longitudinal (comparación intra sujetos o entre medidas múltiples) y en segundo lugar en la forma de emplear la asignación no aleatoria (conocida o no conocida).

Cabe distinguir dos módulos básicos, que suelen denominarse genéricamente diseños básicos o preexperimentales, sobre los cuales se construyen los diseños cuasiexperimentales auténticos. El primer módulo es el diseño solamente postest (DSP), que tiene un grupo experimental al que se le administra el programa, intervención o tratamiento, y otro grupo de control, pero carece de medidas pretest, y por tanto sólo utiliza comparaciones intersujetos. El segundo módulo es el diseño pretest-postest (DPP), que tiene un único grupo con medidas antes (pretest) y después (postest) del tratamiento, y por tanto sólo emplea comparaciones intrasujetos. Una combinación peculiar de estos dos diseños básicos conduce al diseño de cuatro grupos de Solomon (D4GS), que cuando emplea aleatorización se convierte en un diseño experimental auténtico (p.e., Whitman, Van Rooy, Viswesvaran & Alonso, 2008), pero en cualquiera de sus versiones, experimental o cuasiexperimental, su principal ventaja es la capa-

cidad de probar y controlar la reactividad del instrumento de medida (García, Frías y Pascual, 1999).

Entre los diseños cuasiexperimentales transversales, los dos casos prototípicos son el diseño con grupos no equivalentes (DGNE), que representa también una combinación especial de los dos módulos preexperimentales citados y donde la asignación no es aleatoria ni tampoco conocida (p.e., Pons, González y Serrano, 2008), y el diseño de discontinuidad de la regresión (DDR), que es similar al diseño DGNE pero donde la asignación a tratamiento sí es al menos conocida, basada en la adopción de un punto de corte. Ambos diseños pueden mejorarse incluyendo técnicas de control estadístico como el uso de covariantes. Entre los diseños cuasiexperimentales longitudinales destacamos el diseño de medidas repetidas, donde los tratamientos del factor intrasujetos no han sido aleatorizados, el diseño de series temporales interrumpidas (DSTI), que es una extrapolación del módulo DPP con un grupo y múltiples medidas pretest y postest y puede generalizarse también a múltiples grupos, y el diseño longitudinal en panel (DLEP), que en su forma más simple requiere un grupo con dos variables medidas en al menos dos momentos temporales, y puede ser generalizado a múltiples variables y múltiples momentos. La Figura 5 resume los diseños preexperimentales y cuasiexperimentales más utilizados en psicología.



DSP: diseño solamente postest. DPP: diseño pretest-postest.
 D4GS: diseño cuasiexperimental de 4 grupos de Solomon.
 DGNE: diseño con grupos no equivalentes.
 DDR: diseño de discontinuidad de la regresión.
 DMR: diseño cuasiexperimental de medidas repetidas.
 DSTI: diseño de series temporales interrumpidas.
 DLEP: diseño longitudinal en panel.

Figura 5. Diseños cuasiexperimentales.

Teniendo en cuenta que todos los diseños cuasiexperimentales pueden ser analizados con múltiples procedimientos estadísticos, las opciones analíticas más comunes son, para los diseños transversales (D4GS, DGNE y DDR), alguna de las alternativas del modelo lineal generalizado en cualquiera de sus modalidades: regresión, ANOVA, ANCOVA y análisis de las puntuaciones de cambio (Ato, 1995b,c, Ato y

Vallejo, 2007, Ato, Losilla, Navarro, Palmer y Rodrigo, 2005). Respecto a los diseños longitudinales, para el diseño de medidas repetidas pueden consultarse Arnau, 1995g,h y Vallejo y Fernández, 1995. Para el diseño DSTI lo más apropiado es el análisis de series temporales si existe un elevado número de medidas pretest y postest (Arnau, 1995e,f, Vallejo, 1995a,b, 1996, Vallejo, Arnau, Bono, Cuesta, Fernández y Herrero, 2002) o bien, si existe un reducido número de medidas, algunos de los procedimientos alternativos que se exponen con detalle en Arnau (1995c) y en Arnau y Bono (2004). Para el diseño DLEP son apropiados los procedimientos analíticos tratados en Arnau y Gómez (1995). Los métodos intensivos longitudinales constituyen hoy por hoy una recomendable clase de modelos propuestos para analizar datos masivos recogidos a lo largo del tiempo (Bolger & Laurenceau, 2013).

Estrategia asociativa

Estudios comparativos

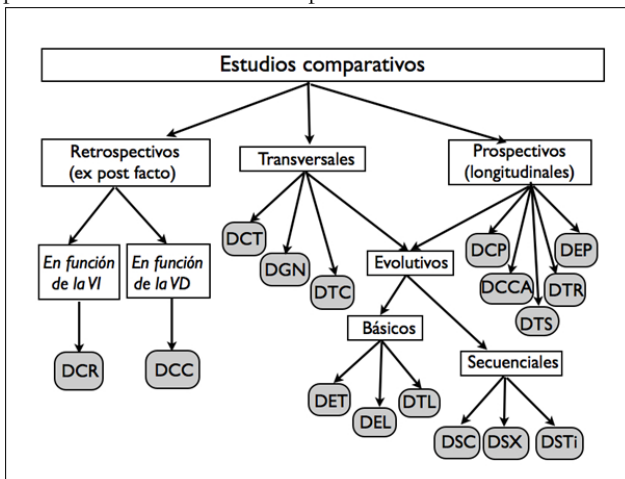
Los estudios comparativos son estudios que analizan la relación entre variables examinando las diferencias que existen entre dos o más grupos de individuos, aprovechando las situaciones diferenciales creadas por la naturaleza o la sociedad. Puesto que son en esencia estudios no experimentales, no se utilizan variables manipuladas (o tratándose de una variable susceptible de manipulación, que no sea posible manipularla por razones éticas o administrativas), ni tampoco asignación aleatoria de los participantes. En su forma más característica estos estudios emplean variables atributivas (también llamadas de selección o clasificación) como VIs, en contraposición a variables manipuladas, que son más características de los estudios experimentales y cuasiexperimentales, y a variables medidas, que suelen ser comunes en los estudios predictivos y explicativos.

Un estudio comparativo suele adoptar uno de varios enfoques temporales posibles. Se denomina retrospectivo (o *ex post facto*) cuando la VI y la VD han ocurrido antes del comienzo del estudio, transversal cuando la definición de VI y VD se realiza concomitantemente en el tiempo, y prospectivo (o longitudinal) cuando se observan al comenzar el estudio y se prolongan a lo largo del tiempo. Dentro de esta estructura general pueden definirse varios diseños de investigación en psicología que se muestran en la Figura 6.

Estudios retrospectivos o *ex post facto*

Los estudios retrospectivos utilizan información histórica para remontarse hacia atrás en el tiempo examinando eventos que han tenido lugar con anterioridad (de ahí su denominación alternativa de estudios *ex-post-facto*; véase Suchman, 1967). Aunque el periodo temporal objeto del estudio puede abarcar muchos años, una ventaja de los estudios retrospectivos es el escaso tiempo que se tarda en completarlo, puesto que solo requiere registrar y analizar los datos. Por esta razón

son particularmente útiles para estudiar la asociación entre variables que no ocurren frecuentemente o son difíciles de predecir o cuando existe un largo intervalo temporal entre una supuesta causa y su efecto. Son típicos de la investigación en ciencias de la salud (en particular, la salud mental y la epidemiología; véase Mann, 2003), de donde se han importado a disciplinas afines tales como la psicopatología y psicología clínica. En el contexto clínico, los estudios retrospectivos se utilizan usualmente en lugar de un experimento para plantear hipótesis causa-efecto en situaciones donde no es posible utilizar variables manipuladas.



DCR: diseño de cohortes retrospectivo. DCC: diseño caso-control.
 DCT: diseño de cohortes transversal. DGN: diseño de grupos naturales.
 DTC: diseño transcultural. DCP: diseño de cohortes prospectivo.
 DCCA: diseño caso-control anidado. DTS: diseño transversal simultáneo.
 DTR: diseño transversal repetido. DEP: diseño (longitudinal) en panel.
 DET: diseño evolutivo transversal. DEL: diseño evolutivo longitudinal.
 DTL: diseño evolutivo "time-lag" o de retardo temporal.
 DSC: diseño secuencial de cohorte. DSX: diseño secuencial transverso.
 DSTi: diseño secuencial de tiempo.

Figura 6. Estudios comparativos.

Pueden diferenciarse dos tipos de estudios retrospectivos, dependiendo de si la definición de los grupos de participantes se establece en base a la variable independiente o la variable dependiente (Cohen, Manion & Morrison, 2012, Kirk, 2013). En el diseño de cohorte retrospectivo (DCR, también llamado de cohorte histórico), se utilizan los registros para identificar dos grupos de sujetos en función de si han sido o no expuestos a la variable independiente, y después se comparan en las variables dependientes (o sea, su característica diferencial es que hay una sola VI y una o más VDs, p.e., Ríos, Godoy y Sánchez, 2011). Una cohorte representa en este contexto a un grupo de personas que han experimentado un evento significativo (p.e., nacimiento, divorcio o enfermedad) durante un determinado intervalo temporal. Por el contrario, en el diseño caso-control (DCC) los registros se utilizan para identificar dos grupos en función de si muestran evidencia de la variable dependiente (los casos) o no la muestran (los controles), y después se com-

paran en términos de su exposición previa a las variables (véase Figura 6).

Por ejemplo, supongamos que un investigador se interesa por estudiar la asociación entre asistencia a centros de educación infantil (VI) y rendimiento académico en primaria (VD). Si se empleara un diseño de cohorte retrospectivo, el investigador podría utilizar dos grupos de niños que difieren por haber asistido o no a centros de educación infantil y después comparar el rendimiento académico medio obtenido en primaria. Es decir, se define primero la VI de forma retrospectiva y se mide después la VD, pero ambas variables ocurren antes de que el estudio comience. Si se empleara el diseño caso-control, el investigador podría utilizar dos grupos de niños que muestran diferencias en rendimiento académico en primaria (casos y controles) y después examinar retrospectivamente para cada caso y control la asistencia a cursos de educación infantil en sus primeras etapas. Es decir, se define primero la VD y se evalúa después retrospectivamente la VI, pero ambas ocurren también antes de que el estudio comience.

Un rasgo característico de los estudios retrospectivos es su interés explícito por formular hipótesis del tipo causa-efecto con el objeto de reproducir las características de la metodología experimental en un contexto no experimental y con variables que no son susceptibles de manipulación. De hecho en la investigación educativa los diseños retrospectivos o ex post facto se denominan también diseños causal-comparativos (véase Brewer & Kuhn, 2010, Gay, Mills & Airasian, 2012) mientras que en la literatura estadística se corresponden en cierta medida con los diseños de experimentos inversos (véase Loy, Goh & Xie, 2002). Sin embargo, es importante destacar la existencia de numerosas amenazas potenciales contra la validez interna en los estudios retrospectivos. Shadish, Cook y Campbell (2002, p.131-133) discuten hasta 53 amenazas generales contra la validez en el diseño caso-control, lo que hace prácticamente inviable el establecimiento de relaciones causales en estos estudios (véase no obstante Farrant, 1977). Con todo, los estudios retrospectivos/ex post facto suelen encontrar relaciones interesantes que pueden ser ulteriormente objeto de investigación experimental y en ocasiones es posible utilizar procedimientos de control sofisticados para conseguir condiciones en cierta medida similares a la obtenida mediante aleatorización que permitan el análisis de relaciones causa-efecto (véase Vélez, Egurrola y Barragán, 2013).

Estudios transversales

A diferencia de los estudios retrospectivos, los estudios transversales se definen en un momento temporal determinado y siguen una tradición eminentemente asociativa donde el interés en el establecimiento de relaciones causa-efecto es secundario. Mientras los estudios retrospectivos y prospectivos de cohorte son el mejor medio para evaluar cuestiones de incidencia (o sea, se proponen determinar el número de nuevos casos que se desarrollan durante un intervalo tempo-

ral específico), los estudios transversales se utilizan primordialmente para evaluar cuestiones de prevalencia (o sea, se proponen determinar el número de casos que existen en una determinada población en un momento temporal específico). Los estudios transversales son además idóneos para el estudio de VDs que permanecen estables en el tiempo, es decir, que no son susceptibles de cambio.

En epidemiología y psicopatología, los estudios transversales utilizan uno o más grupos de participantes que son evaluados en un momento temporal determinado en una o más VDs. El más simple es el diseño de cohortes transversal (Hudson, Pope & Glynn, 2005), que utiliza muestras de cohortes de forma transversal y después evalúa retrospectivamente la historia de exposiciones (VIs) y resultados (VDs) en los miembros de la cohorte en un periodo de tiempo específico (p.e., Stefani y Feldman, 2006). Es obvio que, como consecuencia del proceso de muestreo, algunos participantes de la muestra no habrán tenido exposición (VI) ni habrán experimentado resultado (VD) alguno. Esta es una diferencia esencial con los estudios de cohorte (retrospectivo y prospectivo), donde el historial de exposiciones y/o resultados afecta a todos los sujetos de la cohorte (Mann, 2003).

En algunas áreas de la psicología aplicada es frecuente el diseño de grupos naturales (DGN, véase Shaughnessy, Zechmeister & Zechmeister, 2012), cuyo objetivo es la comparación en una o más VDs de grupos preexistentes, donde los grupos se seleccionan usando participantes que pertenecen a niveles de variables que son fuente de diferencias individuales (p.e., sexo, inteligencia, grupo étnico, trastorno psicopatológico, etc), y pertenecen a una misma cultura. Cuando los grupos naturales pertenecen a culturas diferentes el resultado es el diseño transcultural (DTC, véase Matsumoto & van Vijver, 2011). El desarrollo de influyentes encuestas transculturales, tales como la *European Social Survey* o el *Program for International Assessment of Students' Achievements* (PISA), ha popularizado este tipo de investigaciones y ha permitido la explotación de nuevas opciones analíticas que no habían sido previamente consideradas (véase Davidov, Schmidt & Billiet, 2011).

Los estudios transversales requieren esfuerzos mínimos en economía y tiempo y son bastante eficientes cuando se proponen hipótesis asociativas, pero plantean serios problemas de validez si se persiguen hipótesis causa-efecto.

Estudios prospectivos (o longitudinales)

En los estudios prospectivos o longitudinales, las VIs y VDs se observan después de haberse iniciado la investigación (aunque la VI puede en ocasiones referirse a una situación anterior). De ahí que el coste económico y el tiempo requerido para completar un estudio prospectivo sea mucho mayor que el requerido por un estudio retrospectivo, aunque como contrapartida mejora sensiblemente su fiabilidad. Y debido a su condición de longitudinal, algunas de las potenciales amenazas contra la validez interna de los estudios retrospectivos no se consideran en los estudios prospectivos,

aunque los inconvenientes para el establecimiento de relaciones causa-efecto siguen siendo patentes, principalmente los problemas de sesgo y de confundido, que además se agravan como consecuencia del abandono de sujetos (“*attrition*”) que se produce en tanto mayor grado cuanto más se prolonga en el tiempo.

Uno de los diseños prospectivos más comunes es el diseño de cohorte prospectivo (DCP), cuya metodología es similar al diseño de cohorte retrospectivo, pero la cohorte es evaluada hacia adelante, en varios momentos a lo largo del tiempo, utilizando una o más VDs relevantes (por ejemplo, una escala de depresión). Usualmente este diseño incluye una única cohorte y se prolonga en el tiempo hasta determinar el resultado de interés (por ejemplo, la incidencia de la depresión), en cuyo caso los miembros de la cohorte que no desarrollan depresión sirven como controles internos. Pero también puede incluir más de una cohorte, un grupo expuesto y otro no expuesto, en cuyo caso este último grupo sirve como control externo (Mann, 2003). Un diseño con estructura similar al diseño caso-control es el diseño caso-control anidado (DCCA), en el que casos y controles se seleccionan desde dentro de un estudio prospectivo, permitiendo reducir sensiblemente el tiempo y el coste económico en comparación con el diseño caso-control (véase Ernster, 1994).

Hay varios diseños prospectivos o longitudinales de gran interés en varias áreas aplicadas de la psicología (Arnaú, 1995d, Bijleveld et al., 1998, Taris, 2000). Aunque no tiene una orientación estrictamente longitudinal, el diseño más básico es el diseño transversal simultáneo (DTS), en el que varios grupos se miden en una misma ocasión y su generalización constituye el diseño transversal repetido (DTR), que consta de al menos un DTS administrado en dos o más momentos temporales. En ambos casos, el interés de los investigadores es estudiar el cambio a nivel individual en alguna variable de interés, pero no el cambio debido a la edad, que es objeto de los estudios evolutivos. En el diseño en panel (DEP) uno o más grupos de sujetos son evaluados en un número limitado de momentos temporales (usualmente, entre 3 y 5), mientras que en el diseño longitudinal en panel se amplía el número de momentos temporales de evaluación, y es una generalización del diseño DCP (porque emplea múltiples cohortes), del DTS (porque cada una de las diferentes cohortes constituyen grupos de edad) y del DEP (porque la muestra de participantes que constituye la cohorte es evaluada en diferentes momentos temporales). Arnaú (1995c,e,f) y Gómez (1995) tratan en detalle algunas cuestiones analíticas de muchos diseños longitudinales. Bijleveld et al. (1998), Fitzmaurice, Laird & Ware (2009), Menard (2008) y Singer & Willett (2003) cubren en profundidad todas las cuestiones relevantes del diseño y el análisis de datos longitudinales.

Estudios evolutivos

En el contexto de la psicología del desarrollo, los estudios comparativos se han interesado tradicionalmente por analizar las diferencias o los cambios que se producen en

comportamientos o habilidades de interés evolutivo en función de la edad o del desarrollo. Es muy importante determinar si el interés del investigador son las diferencias de edad o los cambios debidos al tiempo para decidir si un estudio es evolutivo (Schaie & Caskie, 2006).

Singer & Willet (2003) han propuesto un interesante modelo multinivel general que distingue dos etapas en el análisis del cambio en respuesta a dos objetivos diferentes y permite un análisis estadístico separado del cambio inter e intraindividual. En la primera etapa (nivel-1) el objetivo es el cambio intraindividual a lo largo del tiempo, o sea, la trayectoria individual de cada persona. En la segunda etapa (nivel-2) el objetivo son las diferencias individuales en el cambio intraindividual y la detección de las variables que explican tales diferencias.

Se han definido tres diseños evolutivos básicos dentro de esta tradición (Baltes, Reese & Nesselroade, 1977, Schaie, 1965, Rosel, 1995). En primer lugar, el diseño evolutivo transversal (DET), que utiliza muestras de diferentes grupos de edad evaluados a nivel intersujetos en un mismo momento temporal y persigue como objeto encontrar diferencias en función de la edad en la medida de una o más VDs. A pesar de su nombre, la dimensión longitudinal en este diseño se debe a la comparación entre diferentes grupos de edad. La característica diferencial que justifica la inclusión de este diseño en los estudios evolutivos es que la variable clave es la edad, mientras que en los estudios transversales y en el diseño transversal simultáneo la edad es una variable de control (Taris, 2000). El diseño es bastante eficiente, porque permite aproximar un estudio longitudinal desde una perspectiva transversal, pero no es posible evaluar el cambio intraindividual y los efectos de edad se confunden con los efectos evolutivos. En segundo lugar, el diseño evolutivo longitudinal (DEL), que se interesa por el análisis del cambio que se produce en un grupo de individuos observado en repetidas ocasiones a lo largo del tiempo. En tercer lugar el diseño evolutivo de retardo temporal (DTL, "*time-lag design*"), donde se registran observaciones en varios momentos temporales, como en el diseño longitudinal, pero se utiliza una muestra de individuos del mismo grupo de edad que es diferente para cada momento temporal.

Estos tres diseños son casos especiales del modelo evolutivo general de Schaie (1965), quien postuló que cualquier cambio evolutivo puede ser potencialmente descompuesto en una (o más) de tres dimensiones independientes de edad cronológica (referida al número de años desde el nacimiento al momento cronológico en que se evalúa el comportamiento), cohorte o generación (referida al grupo de individuos que participan de un contexto común en un mismo momento cronológico) y periodo (referida a la ocasión temporal en que se realiza la medida). Las tres dimensiones se confunden de forma que una vez especificadas dos de ellas, la tercera viene determinada. Por ejemplo, un individuo de una cohorte nacida en 1950 y evaluada en el periodo 2010 se sabe que tiene una edad de 60 años en el momento de la medida. Así, en el diseño evolutivo transversal, edad y cohorte se confun-

den y por tanto no pueden observarse efectos debidos al periodo, ya que los grupos de edad estudiados deben pertenecer a diferentes cohortes. El diseño evolutivo longitudinal es menos asequible en términos de tiempo y coste económico que el anterior, pero se confunden edad y periodo y por tanto no pueden observarse efectos debidos a la cohorte. Por su parte, en el diseño evolutivo de retardo temporal se confunden los efectos de periodo y cohorte (Schaie, 1994, Schaie & Caskie, 2006).

Sin embargo, imponiendo restricciones apropiadas sobre una de las tres dimensiones del modelo, es posible analizar los otros dos componentes y su interacción (Arnau, 2005, Bijleveld et al., 1999, Menard, 2008). Tres estrategias se han propuesto para analizar todas las combinaciones de dos de los tres componentes, produciendo tres formas de diseño alternativas a los diseños evolutivos clásicos llamados genéricamente diseños secuenciales (véase Schaie, 1994), a saber: el diseño secuencial de cohorte (DSC), el diseño secuencial transverso (DSX) y el diseño secuencial de tiempo (DSTi). En general, estos diseños son más eficientes que los diseños evolutivos básicos. Por ejemplo, un diseño longitudinal con niños de 5 a 13 años observados cada 2 años requiere para ser completado 8 años, mientras que un diseño secuencial equivalente reduce a la mitad el tiempo requerido para realizar el estudio, utilizando un grupo de niños de 5 años (que se observa a los 5, 7 y 9 años) y otro grupo de niños 9 años (que se observa a los 9, 11 y 13 años). Más información sobre métodos de investigación en psicología evolutiva puede encontrarse en Martínez-Arias (1983) y sobre los diseños evolutivos en general en Vega (1983).

Estudios predictivos y explicativos

Otro tipo de investigaciones no experimentales que derivan de la estrategia asociativa son los estudios cuyo propósito principal es explorar las relaciones entre variables con el objeto de pronosticar o explicar su comportamiento. Durante mucho tiempo se han conocido también con el nombre genérico de estudios correlacionales, pero esta denominación se considera hoy incorrecta porque el tipo de análisis estadístico no es precisamente el tema crucial en este tipo de estudios (Cook & Campbell, 1979). Sin embargo, para los estudios que únicamente exploran relaciones entre variables con coeficientes de correlación simple seguiremos empleando el término de diseños correlacionales.

Tres características distintivas permiten definir operativamente este tipo de estudios respecto de los demás estudios de naturaleza no experimental:

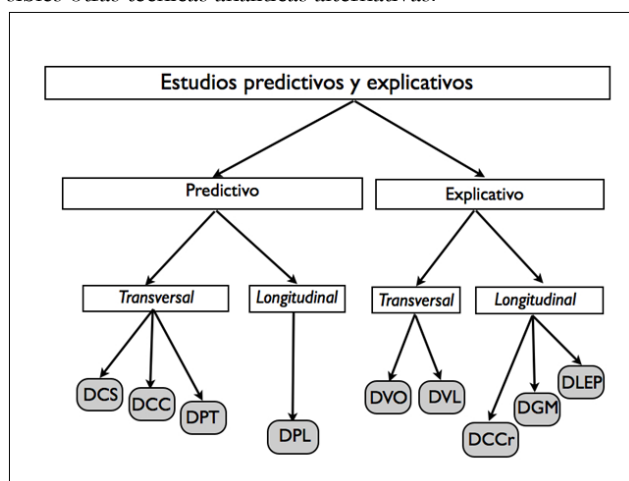
- a) la existencia de una muestra única de participantes que no es usualmente seleccionada aleatoriamente;
- b) la medida de cada participante de la muestra en dos o más variables que suelen ser de naturaleza cuantitativa (o sea, variables medidas, no manipuladas) pero ocasionalmente también de naturaleza categórica, y

c) la disponibilidad, como punto de partida para abordar el análisis estadístico, de una matriz de correlaciones (o covarianzas) entre variables.

Siguiendo a Pedhazur & Smelkin (1990), hemos clasificado los trabajos que siguen esta línea de investigación, dependiendo de la complejidad del objetivo que se persigue, en estudios predictivos y explicativos, y se representan en la Figura 7.

Estudios predictivos

La forma más básica que suele adoptar este tipo de estudios ocurre cuando el objetivo de la investigación es simplemente explorar una relación funcional entre dos o más variables, sin distinción alguna entre ellas. Puesto que no se utiliza ninguna forma de control de variables extrañas sobre la relación funcional, el diseño resultante se denomina diseño correlacional simple (DCS). Para el análisis estadístico de la asociación entre variables suele utilizarse algún coeficiente de correlación (o una matriz de coeficientes de correlación) apropiado a la naturaleza métrica de las variables (p.e., Sicilia, Aguila, Muyor, Orta y Moreno, 2009), pero también son posibles otras técnicas analíticas alternativas.



DCS: diseño correlacional simple. DCC: diseño correlacional controlado. DPT: diseño predictivo transversal. DPL: diseño predictivo longitudinal. DVO: diseño explicativo con variables observables. DVL: diseño explicativo con variables latentes. DCCr: diseño explicativo de curvas de crecimiento. DGM: diseño explicativo de grupos múltiples. DLEP: diseño explicativo longitudinal en panel.

Figura 7. Diseños predictivos y explicativos.

Resulta obvio que en el DCS el grado de control ejercido sobre las terceras variables que potencialmente pueden afectar a la relación analizada es nulo y que todas las variables tienen el mismo estatus metodológico (en realidad, todas se consideran variables criterio), por lo que este diseño es altamente propenso a padecer muchas de las amenazas contra la validez interna y externa. Puede mejorarse sensiblemente cuando la relación funcional entre las variables se explora

controlando una (o más) tercera(s) variable(s). El procedimiento de control más apropiado en este caso es el control estadístico mediante residualización o parcialización, en cuyo caso el diseño resultante se denomina diseño correlacional controlado (DCC). El análisis estadístico utiliza coeficientes de correlación parcial o semiparcial de orden k , donde k se refiere al número de variables controladas, o bien una matriz de coeficientes de correlación parcial o semiparcial (p.e., Piamontesi et al., 2012).

Cuando el objetivo de la investigación es explorar una relación funcional mediante el pronóstico de alguna variable criterio a partir de uno o más predictores, se aplica un diseño predictivo transversal (DPT), donde es común la utilización de los términos predictor, en sustitución del de variable independiente, y criterio, en sustitución del de variable dependiente. Con el DPT no suele emplearse ningún procedimiento de control de variables extrañas y el procedimiento estadístico más común consiste en especificar un modelo de regresión apropiado, lineal o no lineal, dependiendo de la naturaleza y la distribución de la variable criterio, y puede también tratarse desde una perspectiva multinivel. Los modelos lineales generalizados (Ato et al., 2005, Dobson, 2002, Hardin & Hilbe, 2007) y los modelos multinivel (Hox, 2010) son especialmente recomendables en este contexto. Cuando la variable criterio es numérica y distribuida normalmente, suele especificarse un modelo de regresión normal (p.e., Jiménez, Martínez, Miró y Sánchez, 2012), cuando la variable es binaria o dicotómica un modelo de regresión logística (p.e., Planes et al., 2012) y cuando presenta una estructura jerárquica un modelo multinivel (p.e. Núñez, Vallejo, Rosario, Tuero y Valle, en prensa). En muchos de estos casos es también posible, e incluso en ocasiones conveniente, evaluar la eficacia de la predicción mediante una clasificación de los participantes. Es útil en este contexto, además de la regresión logística, la aplicación de otras técnicas multivariantes tales como el análisis discriminante, la clasificación y segmentación jerárquicas y las redes neurales (véase Levy y Varela, 2003).

Si el investigador dispone de muchos predictores, hay dos medios alternativos de introducirlos en una ecuación de regresión: o bien todos de una vez (introducción simultánea) o bien mediante una serie de pasos (introducción secuencial). En este último caso, hay dos opciones generales para determinar un orden lógico de introducción de los predictores: o bien mediante un procedimiento automático, dirigido por el computador ("stepwise selection"), o mediante un procedimiento sustantivo, dirigido por el investigador en función del objetivo de la investigación ("hierarchical selection"). Se han esgrimido muchas razones para preferir la selección jerárquica a la selección paso a paso (Snyder, 1991, Thompson, 1995, 2001), la principal de las cuales es que con la selección jerárquica es posible incorporar en las primeras etapas del proceso las variables extrañas que el investigador considere pertinentes, para analizar los predictores objeto de la investigación liberados de la influencia de las variables extrañas en las etapas finales.

Tanto los diseños correlacionales como el diseño predictivo transversal son en esencia diseños de medida única. Una generalización del DPT al caso longitudinal conduce al diseño predictivo longitudinal (DPL), que se caracteriza por registrar una medida intensiva de las variables criterio y por intentar la eliminación de muchos artefactos metodológicos (Mindick & Oskamp, 1979), aunque adolece también de los problemas asociados a la investigación longitudinal en términos de economía y tiempo requerido.

Estudios explicativos

Otro tipo de investigaciones que derivan de la estrategia asociativa son los estudios explicativos, que se proponen con el objetivo de probar modelos acerca de las relaciones existentes entre un conjunto de variables, tal y como se derivan de una teoría subyacente. En un modelo de regresión, el modelo estadístico básico de los diseños predictivos, se definen con claridad los papeles de predictor y criterio, pero postulan estructuras de modelado muy simples (por lo general, un criterio y varios predictores) y consideran todas las variables como manifiestas u observables y medidas sin error, lo que limita considerablemente su aplicabilidad. Su generalización a modelos más avanzados, incluyendo estructuras de modelado simultáneas y variables no observables o latentes, permiten definir un sistema de ecuaciones de regresión donde las variables pueden intercambiar sus papeles de predictor y criterio. Para aclarar en mayor medida su papel en el contexto de cada una de las ecuaciones de regresión del sistema se utilizan alternativamente los términos de variables exógenas (ajenas al sistema) y variables endógenas (internas al sistema), en lugar de los términos de predictor y criterio.

Dos situaciones de investigación se contemplan aquí (véase Figura 7 anterior). En primer lugar, el diseño explicativo con variables manifiestas (DVM), que se caracteriza por definir una red estructural de relaciones entre variables que puede representarse mediante un sistema de ecuaciones de regresión, asumiendo que todas las variables son manifiestas. El procedimiento estadístico más apropiado para analizar los datos de este tipo de diseño parte de una matriz de correlaciones (o covarianzas) y aplica después el análisis de senderos (“*path analysis*”). El ajuste de los datos empíricos al modelo propuesto se evalúa utilizando medidas de ajuste apropiadas para modelos estructurales (véase Kline, 2011). Sin embargo, como se apuntó más arriba, los investigadores deben tener siempre presente que este análisis asume que las variables manifiestas son manifestaciones fiables de sus correspondientes constructos, lo que no siempre se puede justificar (véase Cole & Preacher, 2013).

Dada una relación funcional entre un predictor y un criterio, una de las más interesantes propiedades del diseño DVO es su potencial para investigar los efectos de una o más terceras variables (Ato y Vallejo, 2011), lo que justifica el incremento sensible de investigaciones que someten a prueba modelos de mediación y moderación (Jose, 2013). En su forma más simple, un modelo de mediación se interesa por

los procesos intervinientes que producen una relación funcional o un efecto de tratamiento (p.e., Gartzia et al., 2012), mientras que un modelo de moderación se interesa por los procesos que afectan a la magnitud de la relación o efecto (p.e., Rodrigo et al., 2012).

En segundo lugar, el diseño explicativo con variables latentes (DVL) distingue en un modelo una parte estructural (que representa un modelo estructural de relaciones entre variables, como en el diseño DVO) y una parte de medida (que incluye los diferentes indicadores que definen un constructo o variable latente) y se representa también mediante un sistema de ecuaciones estructurales, donde algunas variables son observables y otras son latentes. Se conocen como modelos SEM o modelos de ecuaciones estructurales (“*structural equation models*”). Hay al menos dos enfoques estadísticos apropiados para estimar los parámetros de los modelos SEM, un procedimiento basado en las covarianzas (“*covariance-based SEM*”), que es el más extendido (Kline, 2011, Levy y Varela, 2006) y para el que se han diseñado programas de análisis específicos (AMOS, CALIS, EQS, LISREL, MPLUS), y un procedimiento basado en las varianzas (“*variance-based SEM*”), denominado “*partial least squares path modeling*” (PLS-PM) que suele aplicarse en ocasiones en que no se cumplen las severas condiciones que exige el primero relativas al tamaño muestral, normalidad multivariante e independencia de las observaciones (Esposito Vinci et al., 2010, Hair, Hult, Ringle & Sarstedt, 2013), y para el que también existen programas de análisis apropiados (p.e., el programa PLS-PM incluido en el paquete R).

Los DVL permiten controlar cualquier variable extraña simplemente incluyéndola en el modelo estructural como variables observables y especificando sus relaciones con el resto de variables, observables y latentes (p.e., Rosario et al., 2012). Además también es posible probar efectos de mediación y moderación, usualmente comparando el ajuste de dos modelos alternativos que difieren entre sí por incluir o no efectos de las terceras variables (p.e., Buelga, Cava y Musitu, 2012).

Una interesante generalización de los DVL que puede utilizarse para modelar datos longitudinales en los que se observan medidas repetidas de al menos una variable en un número de ocasiones conduce al diseño de curvas de crecimiento (DCCr). Una asequible introducción a este tipo de diseño puede consultarse en Arnau y Balluerka (2004) y Preacher, Wichman, McCallum & Briggs (2008). Otros diseños explicativos longitudinales que emplean modelos SEM, tales como el diseño explicativo con grupos múltiples (DGM) y el diseño explicativo longitudinal en panel (DLEP), pueden consultarse en Little (2013b).

Estrategia descriptiva

La estrategia descriptiva representa, junto a la estrategia asociativa, una de las dos formas características de los estudios no experimentales porque no cumple ninguno de los dos criterios básicos de la investigación experimental (manipulación

de variables y control mediante asignación aleatoria). Pero mientras que en la estrategia asociativa los objetivos de investigación se traducen en hipótesis que persiguen las comparaciones entre grupos y la predicción o explicación de comportamientos o procesos, en la estrategia descriptiva el objetivo de la investigación es la definición, clasificación y/o categorización de eventos para describir procesos mentales y conductas manifiestas, que no suele requerir el uso de hipótesis.

Hay dos grandes tipos de estudios incluidos dentro de la estrategia descriptiva, que representan también formas alternativas de recogida de datos que se utilizan en la mayoría de los estudios no experimentales: los estudios observacionales, donde se registran comportamientos que son objeto de observación y se clasifican de acuerdo con códigos arbitrarios, y los estudios selectivos, donde se registran opiniones o actitudes en una escala de respuesta usualmente mediante cuestionario. Siguiendo a Kish (1987), la característica esencial de un estudio observacional es el *realismo* con que se investigan los comportamientos que son objeto de observación, mientras que la característica básica de un estudio selectivo es la *representatividad* de la muestra seleccionada respecto de la población objetivo.

Es importante advertir que el análisis de datos que se espera en un estudio descriptivo no es necesariamente un análisis descriptivo; en muchas ocasiones se requiere un análisis más complejo utilizando un amplio repertorio de técnicas analíticas univariantes y multivariantes (por ejemplo, técnicas de clasificación, de reducción de la dimensionalidad o de escalamiento). El tipo de estudio no depende del análisis estadístico realizado, sino del objetivo que se persigue. Cuando el objetivo de la investigación es descriptivo, el investigador no suele plantear hipótesis concretas para someter a prueba empírica y presentará resultados que describan apropiadamente comportamientos o procesos mentales, ya sea con estadísticos descriptivos o con otras técnicas analíticas más sofisticadas.

Estudios observacionales

Un estudio observacional es un plan de investigación de gran versatilidad procedimental para registrar la conducta espontánea de una unidad (participante, díada, equipo, etc.) mediante técnicas de observación específicas y siguiendo un plan de muestreo de conductas en contextos naturales (Rabadán y Ato, 2003).

Siguiendo a Anguera (1995) y a Shaughnessy, Zechmeister y Zechmeister (2012), los estudios observacionales se distinguen de forma general en función del método de observación utilizado, o bien mediante la observación indirecta (no reactiva) de comportamientos que ocurrieron previamente, como sucede cuando se analiza la escena de un crimen en busca de indicios probatorios o se consultan registros de archivo sobre comportamientos de interés, o bien mediante observación directa. En este último caso, el investigador debe crear un contexto de observación que puede variar a lo

largo de un continuo que discurre desde la utilización de métodos sin intervención (observación naturalista) en el contexto de observación al empleo de métodos con intervención, entre los cuales se encuentra la observación participante (donde el observador juega un papel dual, observando conductas y participando en la situación observada) y la observación estructurada (donde el observador interviene en la situación con la finalidad de facilitar el registro de comportamientos que pueden ser difíciles de registrar mediante observación naturalista como sucede en los estudios de Piaget). En todas estas situaciones el investigador debe tener en cuenta las específicas limitaciones que presenta esta metodología entre las cuales destacan: 1) el escaso control que usualmente se obtiene de las variables extrañas (de hecho no suele emplearse el rótulo de “variable” en este contexto); 2) la posible dependencia de los juicios subjetivos y 3) los problemas de reactividad derivados de la situación de observación.

Los estudios observacionales se adaptan bien a muchas situaciones que interesan a la investigación cualitativa (Anguera, 1986, 1995, Marshall & Rossman, 2006), tales como los estudios narrativos, que exploran a un solo participante, los estudios fenomenológicos, que exploran a un grupo reducido de participantes y los estudios etnográficos, que se interesan por una cultura. Especialmente apropiada es la metodología observacional en el estudio de casos, donde el investigador estudia intensivamente aspectos específicos de la conducta de una entidad (que puede incluir uno o más sujeto/s, díada/s, escuela/s, hospital/es, equipo/s o empresa/s) durante un periodo de tiempo (véase Woodside, 2010) y también en los estudios generadores de teoría (basados en la “grounded theory” propuesta por Strauss & Corbin, 1998) que se proponen generar o descubrir una teoría o patrón analítico de un proceso, acción o interacción a partir de los datos recopilados de los participantes que lo han experimentado (p.e., Fernández-Alcántara et al., 2013).

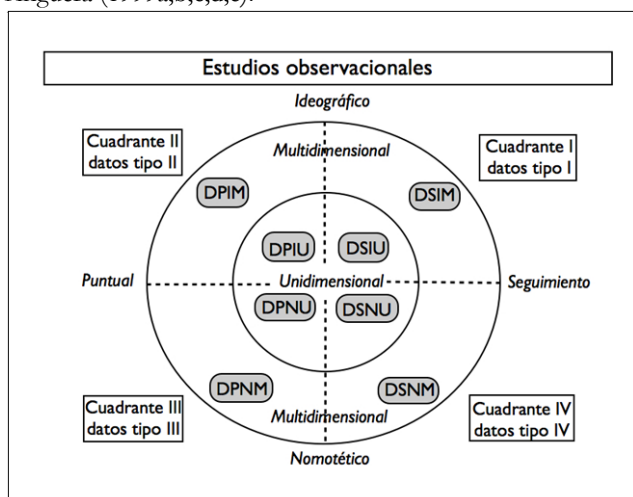
A pesar de la diversidad de estudios posibles, la metodología que se aplica en los estudios observacionales es única y “... se caracteriza básicamente por la perceptividad del comportamiento, la habitualidad en el contexto, la espontaneidad de la conducta observada y la elaboración a medida de instrumentos de observación” (Anguera, Blanco, Hernández y Losada, 2011, p. 64). En este sentido, puede definirse el diseño observacional como una guía flexible que permite tomar decisiones acerca de la recogida de datos (incluyendo la construcción de instrumentos de observación), así como de la gestión y el análisis de los datos, destacando tres criterios clave para distinguirlos (Anguera, Blanco y Losada, 2001, Anguera, Blanco, Hernández y Losada, 2011):

- a) las unidades de estudio observadas se distinguen entre sí en función de su nivel de integración en ideográfico (una única unidad de observación, individual o agregada) y nomotético (una pluralidad de unidades de observación);
- b) la temporalidad del estudio permite distinguir entre observaciones realizadas en un corte transversal (puntual) o longitudinal (seguimiento), y

c) la dimensionalidad del estudio, referida a los niveles de respuesta registrados, facilita la distinción entre investigaciones que requieren un único nivel de respuesta (unidimensional) e investigaciones que requieren varios (multidimensional).

El cruce de las dos opciones posibles para cada uno de los tres criterios citados produce ocho tipos de diseños observacionales que se han representado en la Figura 8. Cada uno de los cuatro cuadrantes de la figura se corresponde también con las categorías de una taxonomía propuesta por Sakett (1978) relativa a la naturaleza de los datos obtenidos a partir de los diseños observacionales, en los que se diferencia entre datos tipo I (secuenciales y evento-base), datos tipo II (concurrentes y evento-base), datos tipo III (secuenciales y tiempo-base) y datos tipo IV (concurrentes y tiempo-base).

Anguera, Blanco y Losada (2001) y Blanco, Losada y Anguera (2003) ilustran la naturaleza de cada uno de los diseños observacionales y presentan algunos ejemplos de uso en numerosas aplicaciones de la psicología clínica, la psicología escolar, la psicología del deporte, la etología (animal y humana) y la observación en contextos naturales, que se han recogido en una serie de publicaciones coordinadas por la profesora Anguera (1999a,b,c,d,e).



DPIM: diseño puntual ideográfico multidimensional.
 DSIM: diseño de seguimiento ideográfico multidimensional.
 DPNM: diseño puntual nomotético multidimensional.
 DSNM: diseño de seguimiento nomotético multidimensional.
 DPIU: diseño puntual ideográfico unidimensional.
 DSIU: diseño de seguimiento ideográfico unidimensional.
 DPNU: diseño puntual nomotético unidimensional.
 DSNU: diseño de seguimiento nomotético unidimensional.

Figura 8. Estudios observacionales.

Los diseños observacionales que se describen aquí se clasificarán como descriptivos cuando su objetivo encaje dentro de una estrategia de investigación descriptiva. Sin embargo, muchos estudios representativos de lo que hoy se conoce como *metodología observacional* trascienden en gran medida el marco descriptivo y no tienen fácil encaje en esta categoría.

Utilizando los diseños observacionales tratados aquí, se proponen en ocasiones analizar relaciones funcionales entre comportamientos, utilizando técnicas muy similares a los diseños de caso único (pero en general primando la validez externa sobre la validez interna) y empleando procedimientos de análisis de datos que dependen de la naturaleza de los datos registrados, mediante una transformación sutil del registro cualitativo en datos cuantitativos, y por tanto deberán ser ubicados dentro de la estrategia asociativa, aunque empleando los mismos diseños observacionales que tratamos aquí. La característica común de todos los diseños observacionales, se encajen o no dentro de la estrategia descriptiva, es el uso de la observación como técnica de recogida de datos. Más información sobre la moderna metodología observacional puede consultarse en muchas publicaciones de la profesora Anguera y de otros miembros de su dilatado equipo de investigación.

Estudios selectivos

Los estudios selectivos constituyen una de las formas más comunes de investigación en psicología basada en el uso del método de encuesta por muestreo, y su característica distintiva fundamental es la utilización de la técnica del autoinforme para recabar información empírica (en particular, mediante entrevistas y cuestionarios) sobre una muestra de participantes, que se asume representativa de una población, y cuyos elementos se determinan mediante algún plan de muestreo, con el objeto de investigar alguna característica de la población (Martínez-Arias, 1995a). Dentro de esta definición cabe incluir muchas encuestas nacionales (por ejemplo, la *Encuesta Nacional de Salud*, de periodicidad quinquenal, para conocer aspectos de la salud de los ciudadanos) e internacionales (por ejemplo, la *Encuesta Social Europea*, de periodicidad bianual, para recoger información sobre las opiniones y actitudes de los ciudadanos europeos). No se incluyen aquí los estudios selectivos que utilizan encuestas con objetivos ajenos a la investigación del comportamiento.

Una cuestión esencial en la aplicación de un estudio selectivo es la representatividad de la muestra respecto de la población, una cuestión que concierne a la validez externa de la investigación. La representatividad de la muestra debe establecerse utilizando un conjunto de variables sociodemográficas marcadoras no relacionadas con la temática de la encuesta y comprobando que las estimaciones muestrales no difieren de los parámetros de la población dentro de un margen de error preestablecido. El mayor grado de representatividad se alcanza cuando se emplea muestreo probabilístico, pero es muy común en psicología aplicada el uso de muestreo no probabilístico, y en particular los métodos subjetivos de muestreo, que tienen el menor grado de representatividad.

Además del objetivo de la investigación, el diseño de un estudio selectivo debe tener en cuenta el procedimiento de recogida de la información y adoptar una serie de decisiones esenciales previas al muestreo, durante el muestreo y posteriores al muestreo (Martínez-Arias, 1995c). Los diseños se-

lectivos pueden ser clasificados en función de varias perspectivas. Desde una perspectiva metodológica pueden ser descriptivos (cuando el objetivo del estudio es obtener información puntual de poblaciones poniendo el acento en la precisión en las estimaciones de parámetros) y analíticos (cuando el objetivo es la comparación de grupos, la predicción o la explicación de comportamientos). Desde una perspectiva temporal, un estudio selectivo puede ser transversal (DPT y DNPT), donde se emplea una o más muestras de una misma población, manteniendo constante el tiempo, de serie temporal (DPST y DNPST), que son encuestas transversales con una nueva muestra utilizando el mismo instrumento para cada oleada, usualmente manteniendo constante la edad), longitudinal (DPL y DNPL), donde se emplea una muestra en varios momentos temporales para evaluar el cambio, manteniendo constante la cohorte) y secuencial de cohorte longitudinal (DPCS y DNPCS), que combina elementos de los tres diseños anteriores estudiando cohortes longitudinalmente mientras nuevas cohortes se van añadiendo secuencialmente.

Dentro de la estrategia descriptiva, los estudios selectivos que interesan son obviamente los descriptivos, pero no los analíticos, que como sucede con los estudios observacionales deben ubicarse dentro de otras alternativas no experimentales de la estrategia asociativa dependiendo del objetivo que se persiga (estudios comparativos, predictivos o explicativos), pero su característica común es el uso de la encuesta como instrumento de recogida de datos. Los diseños selectivos, ya sean descriptivos o analíticos, constituyen todo un arsenal de procedimientos de la denominada *metodología de encuesta* que se diferencian en función de su grado de representatividad (probabilístico vs. no probabilístico) y de la perspectiva temporal de la recogida de datos y se representan en la Figura 9.

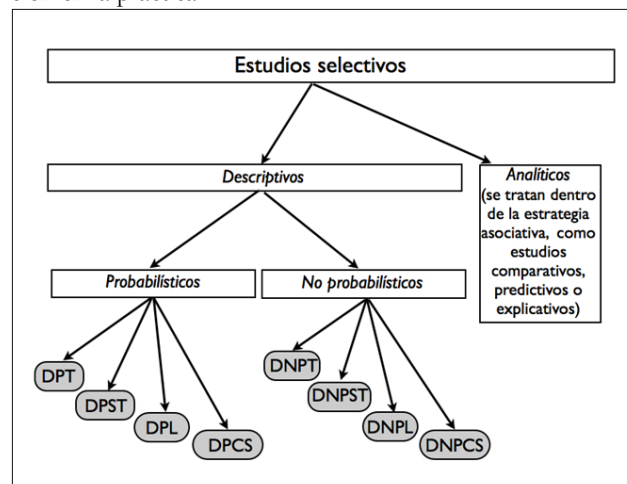
Todo el proceso de aplicación y estimación de una encuesta puede consultarse en Gómez (1990) y Martínez Arias (1995a,b,c). Información más detallada relativa a la investigación mediante encuesta puede encontrarse en una enciclopedia en dos volúmenes editada por Lavrakas (2008).

Conclusiones

Partiendo de un marco contextual y de unos principios básicos para fundamentar los aspectos metodológicos del desarrollo, la articulación y la valoración de un informe de investigación, en este trabajo se aborda una clasificación de los diseños de investigación más usuales en psicología donde se unifican criterios metodológicos y criterios analíticos para facilitar su identificación.

Es obvio que cualquier clasificación será por su propia naturaleza ambigua, y que en algunos casos nos encontraremos con diseños que no tienen una fácil ubicación. Este trabajo presenta un total de 71 diseños básicos, que pueden ser fácilmente generalizados para abarcar un gran espectro de

todas las opciones de diseño que un investigador tiene disponibles. La intención futura de sus autores es ampliar en sucesivas reediciones esta presentación preliminar, incluyendo nuevos diseños básicos que probablemente no se habrán tenido en cuenta, eliminando diseños que pueden resultar obsoletos y desarrollar una presentación más detallada donde se traten cada uno de los diseños con sus posibles extensiones, sus opciones analíticas y algunos ejemplos de su aplicación en la práctica.



DPT: diseño selectivo probabilístico transversal;
 DPST: diseño selectivo probabilístico de serie temporal;
 DPL: diseño selectivo probabilístico longitudinal;
 DPCS: diseño selectivo probabilístico de cohorte secuencial;
 DNPT: diseño selectivo no probabilístico transversal;
 DNPST: diseño selectivo no probabilístico de serie temporal;
 DNPL: diseño selectivo no probabilístico longitudinal;
 DNPCS: diseño selectivo no probabilístico de cohorte secuencial unidimensional.

Figura 9. Estudios selectivos.

Dada la problemática situación creada en nuestro país por la necesidad de publicar para conseguir nuevas metas académicas y profesionales, que está produciendo como consecuencia una mayor profundización en aspectos teóricos en los expertos “sustantivos” y de los aspectos metodológicos y analíticos en los expertos “metodólogos”, nuestra firme recomendación es fomentar una colaboración más estrecha entre los dos colectivos y potenciar el desarrollo de trabajos de investigación más competitivos.

Agradecimientos.- Nuestro especial agradecimiento al equipo directivo de la revista *Anales de Psicología*, por darnos la oportunidad de desarrollar este trabajo, y a quince expertos, sustantivos y metodólogos, que han respondido a nuestra llamada para revisar un primer borrador y nos han aportado sugerencias muy valiosas para mejorarlo. La versión final contiene muchas de las sugerencias recibidas. Los errores que todavía persistan en el trabajo sólo deben ser imputables a sus autores.

Referencias

- Ader, H. J., & Mellenbergh, G. (1999). *Research methodology in the life, behavioral and social sciences*. London, UK: Sage Publications.
- Aiken, L. S., & West, S. G. (1996). *Multiple regression: testing and interpreting interactions*. Thousand Oaks, CA: Sage Publications.
- 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. (1986). La investigación cualitativa. *Educar*, 10, 23-50.
- Anguera, M. T. (1991). *Metodología observacional en la investigación psicológica* (vol. I). Barcelona: PPU.
- Anguera, M. T. (1993). *Metodología observacional en la investigación psicológica* (vol. II). Barcelona: PPU.
- Anguera, M. T. (1995). La metodología cualitativa. En M. T. Anguera, J. Arnau, M. Ato, M. R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de 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.
- Anguera, M. T., Chacón, S. y Blanco, A. (Eds.) (2008). *Evaluación de programas sociales y sanitarios: Un abordaje metodológico*. Madrid: Síntesis.
- American Psychological Association. (2010). *Publication Manual of the American Psychological Association* (6th Ed.) Washington, DC: Author.
- Arnau, J. (1995a). Metodología de la investigación psicológica. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez-Arias, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 23-44). Madrid: Síntesis.
- Arnau, J. (1995b). Fundamentos metodológicos de los diseños de sujeto único. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez-Arias, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 163-178). Madrid: Síntesis.
- Arnau, J. (1995c). Diseños experimentales de sujeto único. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez-Arias, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 179-194). Madrid: Síntesis.
- Arnau, J. (1995d). Análisis estadístico de datos para los diseños de sujeto único. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez-Arias, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 195-243). Madrid: Síntesis.
- Arnau, J. (1995e). Diseños de investigación longitudinal. En J. Arnau (Ed.), *Diseños longitudinales aplicados a las ciencias sociales y del comportamiento* (pp. 35-53). México: Limusa.
- Arnau, J. (1995f). 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). México: Limusa.
- Arnau, J. (1995g). 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). México: Limusa.
- Arnau, J. (1995h). 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). México: 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). México: Limusa.
- Arnau, J., Anguera, M.T. y Gómez, J. (1990). *Metodología de la investigación en Ciencias del Comportamiento*. Murcia: Secretariado de Publicaciones de la Universidad de Murcia.
- Ato, M. (1991). *Investigación en Ciencias del Comportamiento: Fundamentos*. Barcelona: PPU/DM.
- Ato, M. (1995a). Tipología de los diseños cuasiexperimentales. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez-Arias, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 245-270). Madrid: Síntesis.
- Ato, M. (1995b). Análisis estadístico I: Diseños con variable de asignación no conocida. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez-Arias, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 271-304). Madrid: Síntesis.
- Ato, M. (1995c). Análisis estadístico II: Diseños con variable de asignación conocida. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez-Arias, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en psicología* (pp. 305-320). Madrid: Síntesis.
- Ato, M. y Vallejo, G. (2007). *Diseños experimentales 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., Losilla, J. M., Navarro, B., Palmer, A. y Rodrigo, M. F. (2005). *El modelo lineal generalizado*. Girona: Documenta Universitaria - EAP, S.L.
- Bailey, R. A. (2008). *Design of comparative experiments*. Cambridge, UK: Cambridge University Press.
- 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. e Hidalgo, M. D. (2005). The controversy over null hypothesis significance. *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.
- 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.
- Breiman, L., Friedman, J., Olshen, R. y Stone, C. (1984). *Classification and regression trees*. New York, NY: Wadsworth Publishing.
- Brewer, E. W., & Kuhn, J. (2010). Casual-comparative design. In N.J. Salkind (Ed.), *Encyclopedia of research design* (pp. 124-131). Thousand Oaks, CA: Sage Publication.
- Buelga, S., Cava, M. J. y Musitu, G. (2012). Reputación social, ajuste psicológico y victimización entre adolescentes en el contexto escolar. *Anales de Psicología*, 28, 180-187.
- Byrne, D. W. (2000). Common reasons for reject manuscripts at medical journals: a survey of editors and peer reviewers. *Science Editor*, 23,39-44.
- Cajal, B., Gervilla, E. y Palmer, A. (2012). When the mean fails, use an M-estimator. *Anales de Psicología*, 28, 281-288.
- Cohen, L., Manion, L., & Morrison, K. (2012). *Research methods in Education* (6th Ed.). New York, NY: Routledge.

- Cole, D.A. & Preacher, K.J. (2013). Manifest variable path analysis: potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, 18. (Recuperado de http://www.quantpsy.org/pubs/cole_preacher_%28in.press%29.pdf).
- Cook, T. D. y Campbell, D. T. (1979). *Quasiexperimentation: Design and analysis issues for field settings*. Chicago, IL: Rand-McNally.
- Cuesta, M., Fonseca, E., Vallejo, G. y Muñoz, 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 G. R. Hancock & R. O. Mueller, (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.
- Dobson, A. J. (2002). *Introduction to generalized linear models* (2nd Ed.). Boca Raton, FL: Chapman & Hall.
- 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.
- 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.
- Fernandes-Taylor, S., Hyun, J. K., Reeder, R. N., & Harris, A. (2011). Common statistical and research design problems in manuscripts submitted to high-impact medical journals. *BMC Research Notes*, 4, 304, doi:10.1186/1756-0500-4-304
- 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.
- Fitzmaurice, G., Laird, N., & Ware, J. (2009). *Applied longitudinal analysis*. New York, NY: Wiley.
- 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.
- 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.
- Gay, L. R., Mills, G. E., & Airasian, P. W. (2012). *Educational research: competencies for analysis and applications* (10th Ed.). NJ: Pearson.
- Gennettian 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.
- 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.
- Gómez, J. (2005). 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). México: Limusa.
- Graham, J. W. (2012). *Missing data: Analysis and design*. New York, NY: Springer.
- Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: A broad practical approach*, (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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 Ed.). 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.
- Harris, A., Reeder, R., & Hyun, J. (2009a). Common statistical and research design problems in manuscripts submitted to high-impact psychiatry journals: what editors and reviewers want authors to know. *Journal of Psychiatric Research*, 43, 1231-1234.
- Harris, A., Reeder, R., & Hyun, J. (2009b). Common statistical and research design problems in manuscripts submitted to high-impact public health journals. *The Open Public Health Journal*, 2, 44-48.
- Harris, A., Reeder, R., & Hyun, J. (2011). Survey of editors and reviewers of high-impact psychology journals: statistical and research design problems in submitted manuscripts. *The Journal of Psychology*, 145, 195-209.
- Hoening, 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.
- Hox, J. (2010). *Multilevel analysis: Techniques and applications*. 2nd Edition. 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.
- 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 negativo: 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.
- 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.
- Kirk, R. E. (2005). The importance of effect magnitude. En S.F. Davis (Ed.), *Handbook of research methods in Experimental Psychology* (pp. 83-105). Oxford, UK: Blackwell.
- Kirk, R.E. (2013). *Experimental design: procedures for the behavioral sciences* (3rd Ed.). 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 Guilford Press.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd Ed.). New York, NY: The Guilford Press.
- Knapp, T. R., & Mueller, R. O. (2010). Reliability 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.
- Lavrakas, P. J. (Ed.) (2008). *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage Publications.
- 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 n° 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. (2001). Post hoc power analysis: an idea whose time has passed?. *Pharmacotherapy*, 21, 405-409.
- Levy, J. P. y Varela, J. (Eds.) (2003). *Análisis multivariante 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*. 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 Ed.). Thousand Oaks, CA: Sage.
- Martínez-Arias, M. R. (1983). Métodos de investigación en psicología evolutiva. En A. Marchesi, M. Carretero y J. Palacios (Eds.), *Psicología evolutiva I: Teoría y métodos* (pp. 323-372). Madrid: Alianza Universidad.
- Martínez-Arias, M. R. (1995a). El método de las encuestas por muestreo: conceptos básicos. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en Psicología* (pp. 385-421). Madrid: Síntesis.
- Martínez-Arias, M. R. (1995b). Diseños muestrales probabilísticos. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de 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, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en Psicología* (pp. 485-510). Madrid: Síntesis.
- Martínez-Arias, M. R. (1995d). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- 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 Ed.). 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.
- 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, D. L. (1991). Misinterpretation of interaction 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: Analysis of covariance*. New York, NY: Van Nostrand.
- Milliken G. A., & Johnson, D. E. (2009). *Analysis of messy data: Designed experiments* (2nd Ed.). 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., Elosua, 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. (en prensa). 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. Recuperado de <http://www.ehu.es/ojs/index.php/psicodidactica/article/view/7127>.
- 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.
- Pagán, A., Marín, J. y Perea, M. (2012). El papel de la sílaba en la codificación posicional de las representaciones ortográficas. *Anales de Psicología*, 28, 954-962.
- 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.
- 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.
- Payne, R. W., Harding, D. A., Murray, D. M. et al. (2009). *A guide to Anova and Design in Genstat release 12*. H.Hempsted, UK: VSN International.
- Pearl, J. (2009). *Causality: models, reasoning and inference*. (2nd Ed.). Cambridge, MA: Cambridge University Press.
- Pedhazur, E. J., & Smelkin, L. P. (1991). *Measurement, Design and Analysis: an integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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.
- 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, H. (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.
- Ramos, M. y Catena, A. (2004). Normas para la elaboración y revisión de artículos originales experimentales en Ciencias del Comportamiento. *International Journal of Clinical and Health Psychology*, 4, 173-189.
- 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.
- 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.
- Rosario, 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). México: Limusa.
- Rosnow, R. L., & Rosenthal, R. (2009). Effect sizes: why, when and how to use them. *Zeitschrift für Psychologie*, 217, 6-14.
- 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., & Caskie, G. I. L. (2006). Methodological issues in aging research. In D.M. Teti (Ed.), *Handbook of Research Methods in Developmental Science* (pp. 21-39). Oxford, UK: Blackwell Publishing.
- 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.
- 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, D. M. (2006). Issues in the use of longitudinal and cross-sectional designs. In D.M. Teti (Ed.), *Handbook of Research Methods in Developmental Science* (pp. 3-20). Oxford, UK: Blackwell Publishing.
- Shadish, W. & Clark, M. H. (2004). An introduction to propensity scores. *Metodología de las Ciencias del Comportamiento*, 4, 291-300.
- Shadish, W., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.
- Shaughnessy, J., Zechmeister, E., & Zechmeister, J. (2012). *Research methods in psychology* (9th Ed.). New York, NY: McGraw-Hill.
- 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.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: modeling change and event occurrence*. New York, NY: Oxford University Press.
- Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis. An introduction to basic and advanced multilevel modeling* (2nd Ed.). Newbury Park, CA: Sage.
- Snyder, P. (1991). Three reasons why stepwise regression methods should not be used by researchers. En B. Thompson (Ed.), *Advances in social science methodology* (Vol.1, pp. 99-105). Greenwich, CT: JAI Press.
- 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.
- 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.
- Taris, T. W. (2000). *A primer in longitudinal data analysis*. Newbury Park, CA: Sage.
- 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.
- Vallejo, G. (1995a). Diseños de series temporales interrumpidas. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *Métodos de investigación en Psicología* (pp. 321-334). Madrid: Síntesis.
- Vallejo, G. (1995b). Análisis de los diseños de series temporales interrumpidas. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (Eds.), *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). México: Limusa.
- Vallejo, G., Ato, M., Fernández, P., & Livacic-Rojas, P. E., (en prensa). Multilevel bootstrap analysis with assumptions violated. *Psicotema*, 25.
- 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., Gotzsche, 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.
- Vega, J. L. (1983). Metodología longitudinal. En A. Marchesi, M. Carretero y J. Palacios (Eds.), *Psicología evolutiva: Teoría y métodos* (pp. 323-372). Madrid: Alianza Universidad.
- 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.
- 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.

(Artículo recibido: 3-6-2013; revisado: 3-7-2013; aceptado: 4-7-2013)