

M A N U A L S

Psicologia i Logopèdia



Mariona Portell
Jaume Vives

Investigación en psicología y logopedia

Introducción a los diseños
experimentales, cuasiexperimentales
y *ex post facto*

2.^a
edición

UAB Universitat Autònoma
de Barcelona

Servei de
Publicacions

INVESTIGACIÓN EN PSICOLOGIA Y LOGOPEDIA
Introducción a los diseños experimentales, cuasiexperimentales y *ex post facto*

Mariona Portell Vidal
Jaume Vives Brosa

INVESTIGACIÓN
EN PSICOLOGIA Y LOGOPEDIA
Introducción a los diseños experimentales,
cuasiexperimentales y *ex post facto*

Departamento de Psicobiología y de Metodología
de las Ciencias de la Salud

Universitat Autònoma de Barcelona
Servei de Publicacions
Bellaterra, 2024

DADES CATALOGRÀFIQUES RECOMANADES PEL SERVEI DE BIBLIOTEQUES DE LA
UNIVERSITAT AUTÒNOMA DE BARCELONA

Investigación en psicología y logopedia. Introducción a los diseños experimentales, cuasiexperimentales y *ex post facto* / Mariona Portell Vidal, Jaume Vives Brosa. — Bellaterra : Universitat Autònoma de Barcelona, Servei de Publicacions, 2024. — (Manuals de la Universitat Autònoma de Barcelona ; 60)

ISBN (físic) 978-84-10202-27-6

I. Psicologia
II. Logopedia
III. Psicobiologia
III. Ciències de la Salut
IV. Ciències del Comportament

159.98
Codi IBIC: JMB

© del text:

Mariona Portell Vidal i Jaume Vives Brosa

Primera edició: 2019

Segona edició: 2024

© d'aquesta edició:

Servei de Publicacions de la UAB, 2024

© Imatge de la coberta:

Mirant. Domènec Batalla, 1996. Pintura oli, amb matèria, sobre tàblex

Edició:

Universitat Autònoma de Barcelona
Servei de Publicacions
Plaça de l'Acadèmia. Edifici A
08193 Bellaterra (Cerdanyola del Vallès)
T. 93 581 10 22
sp@uab.cat
www.uab.cat/publicacions

ISBN (físic) 978-84-10202-27-6
ISBN (digital) 978-84-10202-28-3



Aquest llibre està publicat amb una llicència Creative Commons CC-BY-NC-ND.
El titular de l'obra autoritza a utilitzar els continguts sempre que es reconegui l'autoria.
No es permet fer un ús comercial, ni la generació d'obres derivades.

Índice

Presentación	9
1. LA INVESTIGACIÓN CIENTÍFICA EN PSICOLOGÍA Y LOGOPEDIA	11
Presentación	11
Ficha 1.1. Práctica basada en la evidencia en ciencias de la salud	14
Ficha 1.2. Método científico: estrategias de sistematización	18
Ficha 1.3. El proceso de investigación científica	19
Ficha 1.4. Orientación cuantitativa, orientación cualitativa y métodos mixtos	21
Ficha 1.5. Teorías, hipótesis, constructos y variables.	24
Ficha 1.6. La asociación entre variables y las relaciones causales: criterios de causalidad	27
Ficha 1.7. Introducción al concepto de validez	30
Ficha 1.8. Métodos, diseños y técnicas de investigación en psicología	32
Ficha 1.9. Muestreo	34
Ficha 1.10. Guías de lectura crítica	41
Actividades.	44
Ejemplos de confusión	44
Ejercicios	49
Notas	56
2. DISEÑOS EXPERIMENTALES	57
Presentación	57
Ficha 2.1. Clasificación de los diseños experimentales	59
Ficha 2.2. El azar en el diseño experimental	60
Ficha 2.3. Técnicas de control	61
Ficha 2.4. Diseños unifactoriales	63
Ficha 2.5. Estrategias de comparación	65
Ficha 2.6. Ventajas e inconvenientes de la estrategia de comparación intrasujeto	66
Ficha 2.7. Diseños factoriales.	67
Ficha 2.8. Tipos de diseños factoriales.	73
Ficha 2.9. «Interrogando» a un diseño factorial	75
Actividades.	77
Actividad de identificación de diseños	77
Ejercicios	82
Lectura crítica.	93
Soluciones de la actividad de identificación de diseños	95
Notas	98

3. DISEÑOS CUASIEXPERIMENTALES	99
Presentación	99
Ficha 3.1. Paradigma manipulativo: criterios de demarcación	101
Ficha 3.2. Diseños cuasiexperimentales: estructuras básicas de comparación.	103
Ficha 3.3. Validez y amenazas	105
Ficha 3.4. Diseños cuasiexperimentales: ampliación del número de grupos y/o de los momentos de medida	111
Diseño simple de serie temporal interrumpida	112
Diseño de serie temporal interrumpida con grupo control no equivalente	113
Ficha 3.5. Diseños cuasiexperimentales: alternativas para la formación de los grupos	114
Diseño cuasiexperimental de cohortes	114
Diseño de discontinuidad en la regresión	115
Actividades.	118
Actividad de identificación de diseños	118
Ejercicios	122
Lectura crítica.	131
Soluciones de la actividad de identificación de diseños	132
Notas	134
4. DISEÑOS <i>EX POST FACTO</i>	135
Presentación	135
Ficha 4.1. Paradigma manipulativo y no manipulativo	137
Ficha 4.2. Los diseños <i>ex post facto</i>	139
Ficha 4.3. Diseño etiológico de cohortes	141
Ficha 4.4. Diseño de casos y controles	144
Ficha 4.5. Diseño transversal analítico.	146
Actividades.	148
Actividad de identificación de diseños	148
Ejercicios	151
Lectura crítica.	163
Soluciones de la actividad de identificación de diseños	164
Notas	168
Referencias.	169

Presentación

Esta publicación se dirige a personas que se quieren introducir en la lógica de la investigación en psicología y logopedia, y presenta tres tipos de diseño ampliamente utilizados en ciencias de la salud: los experimentales, los cuasiexperimentales y los denominados *ex post facto* que incluyen los diseños etiológicos de cohortes, los de casos y controles y los transversales analíticos. Además de introducir estos diseños, un segundo objetivo de esta publicación es enmarcarlos en una perspectiva general que aboga por la complementariedad metodológica, sobre la cual esperamos que nuestros lectores sigan ampliando su formación.

El texto está formado por cuatro capítulos. En el primero, presentamos los elementos esenciales del procedimiento de investigación científica estableciendo un esquema básico y unos criterios de clasificación dirigidos a favorecer la adquisición de una visión integrada de las alternativas metodológicas que se usan en las ciencias de la salud. En los restantes capítulos exponemos los tres tipos de diseños mencionados, que se enmarcan dentro de lo que se denomina metodología cuantitativa. Nuestro propósito es que el seguimiento de este texto facilite la adquisición de las competencias necesarias para introducirse en la lectura crítica de artículos sobre estudios primarios y, por extensión, en la práctica basada en la evidencia.

Cada capítulo se estructura en dos partes. La primera está organizada en un conjunto de «fichas» que van recorriendo los diferentes bloques conceptuales y que están pensadas para que se puedan consultar en distinto orden; esta estructura modular se interrelaciona en un apartado inicial de presentación. En la segunda parte de cada capítulo hay un conjunto de actividades prácticas con distinto nivel de complejidad. Hay un primer bloque de ejercicios creados para incidir gradual y controladamente en cada concepto. Para alcanzar el nivel de dificultad que pretendemos con estos primeros ejercicios, proponemos casos simulados, si bien adjuntamos algunas referencias que permitirán a los lectores interesados la consulta de las investigaciones reales de las cuales se han adaptado. El segundo grupo de actividades se basa en casos reales y, para analizarlos, incluimos adaptaciones simplificadas de guías de lectura crítica.

Afortunadamente, existen muy buenos manuales que tratan la misma temática que nosotros aquí. Creemos que la originalidad de este material es su estructura modular, que en un breve espacio introduce un conjunto de conceptos metodológicos básicos, unido a un sistema de notas finales que, en aquellos casos en que el lector desee una ampliación, facilita el contacto con algunos de estos buenos manuales. Con esta organización, reservamos las notas a pie de página para aquellas referencias o comentarios que puedan ser de interés mientras se hace la lectura.

Este texto se nutre de un conjunto de trabajos previos (Portell, 1995; Portell, 1997; Portell, Vives y Boixadós, 2003; Vives, 2009¹) y de nuestra interacción con centenares de estudiantes con los que hemos tenido el placer de compartir su iniciación a los métodos de investigación, y que con sus preguntas nos han estimulado a pensar y repensar la manera de explicar y de practicar sobre esta materia.

Queremos finalizar esta presentación agradeciendo a todas las personas que con sus comentarios han contribuido a mejorar este material. Una mención especial para nuestra compañera Mercè Boixadós, que revisó con nosotros las primeras versiones y que nos ha aportado sus sugerencias y su experiencia como profesora de esta asignatura.

Mariona Portell
Jaume Vives
Bellaterra, julio de 2019

1. A lo largo del texto no se harán referencias particulares a las siguientes fuentes:
PORTELL, M. 1995. *Fundamentos de metodología. Cuaderno de prácticas*. Bellaterra: Universitat Autònoma de Barcelona. Manuscrito no publicado.
— 1997. *Fundamentos de metodología. Proyecto docente*. Bellaterra: Universitat Autònoma de Barcelona. Manuscrito no publicado.
PORTELL, M.; VIVES, J.; BOIXADÓS, M. 2003. *Mètodes d'investigació: recursos didàctics*. Bellaterra: Servei de Publicacions de la Universitat Autònoma de Barcelona.
VIVES, J. 2009. *Mètodes d'investigació en psicologia. Projecte docent*. Bellaterra: Universitat Autònoma de Barcelona. Manuscrito no publicado.

1

La investigación científica en psicología y logopedia

PRESENTACIÓN

Un estudio científico muestra la efectividad del brain fitness

«Investigadores de la Universidad de California (UCLA) han descubierto que la gente mayor que regularmente utiliza software de gimnasia mental (*brain fitness*) tiene más memoria y habilidades lingüísticas.»¹

En este estudio científico se estudiaron 59 participantes con una media de edad de 84 años, provenientes de comunidades de jubilados del sur de California. Los voluntarios fueron divididos en dos grupos: el primer grupo utilizó un programa de gimnasia mental durante seis meses, una media de 73,5 sesiones de veinte minutos de duración, mientras que un segundo grupo jugó menos de 45 veces durante el mismo período. Los investigadores hallaron que el primer grupo mostró mejoras estadísticamente significativas en la memoria y habilidades de lenguaje, en comparación con el segundo grupo.

Los resultados del estudio se suman a los de los estudios dirigidos a valorar si estas herramientas de ejercicio cerebral pueden ayudar a mejorar el lenguaje y la memoria y, en último término, si pueden ayudar a proteger a los individuos de la disminución cognitiva asociada al envejecimiento y a la enfermedad de Alzheimer.

La pérdida de memoria vinculada a la edad afecta aproximadamente al 40 % de la gente mayor, y se caracteriza por la autopercepción de pérdida de memoria y la disminución en el rendimiento mnésico. Estudios anteriores han demostrado que la participación en las actividades mentales puede ayudar a mejorar la memoria, pero se ha investigado poco para determinar si los juegos de gimnasia cerebral y los programas de entrenamiento de la memoria que se encuentran en el mercado son eficaces para mejorar la memoria. Este es uno de los primeros estudios dirigidos a evaluar los efectos cognitivos de un software de entrenamiento de la memoria.

La investigación fue realizada por la Dra. Karen Miller y por el Dr. Gary Small, ambos del Semel Institute for Neuroscience and Human Behavior de la UCLA.

¿Qué opinión le merecen estos resultados?

Seguramente muy a menudo ha visto en la prensa titulares del tipo «Un estudio científico...» como el que inicia esta introducción. Parece ser que el adjetivo científico da credibilidad a todo lo que califica, de manera que otorgamos autoridad a todo aquel

1. University of California, Los Angeles (UCLA), Health Sciences (2012, August 3). «Memory improves for older adults using computerized brain fitness program». *ScienceDaily*. <http://www.sciencedaily.com/releases/2012/08/120803193555.htm>. Consulta: 26 de marzo de 2018.

conocimiento al que se atribuye el estatus de científico. Pero, ¿qué es el conocimiento científico?

En primer lugar, hay que recordar que existen varios tipos de conocimientos y uno de ellos es el científico. Así, encontramos, por ejemplo, el conocimiento religioso que estructura creencias y prácticas sobre cuestiones existenciales, morales y sobrenaturales, el conocimiento vulgar que es un conocimiento no especializado, o el conocimiento técnico que se centra en la descripción de instrumentos y de su funcionamiento, sin entrar en los fenómenos subyacentes a estos instrumentos. Por ejemplo, el conocimiento técnico describe el funcionamiento de la suspensión de un coche, pero no explica el funcionamiento de la gravedad o de las fuerzas centrífugas, puesto que estas últimas son objeto de estudio científico.

El conocimiento científico¹ se basa en la contrastación empírica y se distingue de otras formas de obtención de conocimiento por los **objetivos**, que se centran en el establecimiento de leyes de la mayor generalidad posible y por ser obtenido de manera **sistemática** a través de un **método estandarizado** de obtención de conocimiento, que es el **método científico**. Una característica importante es el hecho de ser un conocimiento **empírico**. A su vez, también es un conocimiento **teórico**, en el sentido de que la información que se obtiene no es inconexa, sino que se organiza de manera coherente alrededor de teorías. Otra característica importante del método científico es la **replicabilidad**, ya que, en la medida en que al ser repetida una investigación por parte de diferentes investigadores independientes se llega a las mismas conclusiones, el conocimiento científico que se deriva de ello es más confiable.

El conocimiento científico está acompañado de dos mitos destacables. Uno de los mitos es la infalibilidad, lo que contrasta con el hecho de que uno de los pilares del conocimiento científico sea que este es **provisional**, y de aquí surge la importancia de replicar las investigaciones. El otro mito es el de la **objetividad**. La objetividad que se persigue desde el método científico está, actualmente, desvinculada de la tradición positivista más clásica, según la cual es posible conocer la realidad tal como es, desvinculada de quien la observa. En vez de esto, se asume que la realidad solo se puede captar de manera parcial y distorsionada. En palabras de Bunge (1981, p. 16): «El físico atómico perturba al átomo que desea espiar; el biólogo modifica e incluso puede matar al ser vivo que analiza; el antropólogo obstinado en el estudio de campo de una comunidad provoca en la misma ciertas modificaciones. Ninguno de ellos aprehende su objeto tal como es, sino tal como queda modificado por sus propias operaciones; no obstante, en todos los casos, tales cambios son objetivos, y se presume de que se pueden entender en términos de leyes: no son conjurados arbitrariamente por el experimentador. Más aún, en todos los casos el investigador intenta describir las características y la suma de la perturbación que produce en el acto del experimento; procura, en definitiva, estimar la desviación o “error” producido por su intervención activa. Porque los científicos actúan haciendo tácitamente la suposición de que el mundo existiría incluso en su ausencia, aunque, por supuesto, no exactamente de la misma manera». Retomaremos este concepto más adelante al presentar las orientaciones cuantitativa y cualitativa (véase la ficha 1.4).

La investigación científica persigue tres tipos de objetivo. El objetivo **descriptivo** consiste en caracterizar, ya sea cuantitativa o cualitativamente, una variable o una relación entre variables; el **explicativo** se centra en identificar la causa de un determina-

do fenómeno; el **predictivo** informa sobre el valor que se observará en una variable Y cuando otra variable X toma un valor determinado.

En el momento en que la psicología se escinde de la filosofía para convertirse en una disciplina científica con entidad propia, adopta las características antes descritas del conocimiento científico, de las que queremos destacar, dada su importancia, el hecho de ser un conocimiento obtenido de manera sistemática porque adopta el método científico, y el hecho de ser replicable. Ambas características son igualmente cruciales en la investigación en logopedia que, como disciplina científica, ha sido definida por la Internacional Association of Logopedics and Phoniatrics (IALP) como aquella que estudia la prevención, evaluación, tratamiento y estudio científico de los trastornos de la comunicación humana, que engloba todas las funciones asociadas a la comprensión y expresión del lenguaje oral y escrito, así como a cualquier forma de comunicación no verbal. Iniciaremos el capítulo presentando la orientación de la práctica basada en la evidencia en la medida en que esta orientación justifica la necesidad de que el profesional de la psicología y de la logopedia, tanto si se dedica a la investigación como si no, conozca los conceptos básicos de metodología que se introducen en este texto (véase la ficha 1.1).

A partir del estudio de este capítulo, se espera que el lector sea capaz de:

- Entender el concepto y la importancia de la práctica basada en la evidencia (ficha 1.1)
- Situar la ciencia dentro de las diferentes formas de obtención de conocimiento (véase la presentación).
- Tener una visión general de los elementos que están presentes en la investigación científica (presentación y ficha 1.2).
- Conocer las fases de una investigación científica (ficha 1.3).
- Conocer las características generales de las orientaciones cuantitativa y cualitativa, así como el concepto "métodos mixtos" (ficha 1.4).
- Aprender a identificar una hipótesis y también las variables que contiene (ficha 1.5).
- Entender los conceptos de variable independiente (VI) y variable dependiente (VD) y ser capaz de identificarlas dentro de una hipótesis (ficha 1.5).
- Conocer los criterios para evaluar una relación causal (ficha 1.6).
- Introducir el concepto de validez (ficha 1.7).
- Ser capaz de situar las diferentes estrategias metodológicas (ficha 1.8).
- Entender el objetivo del muestreo, y conocer sus principales elementos y técnicas (ficha 1.9).
- Entender la importancia de las guías de publicación y de las herramientas de valoración de la calidad metodológica, saber diferenciarlas y entender su vinculación con la práctica basada en la evidencia (ficha 1.10).

FICHA 1.1. PRÁCTICA BASADA EN LA EVIDENCIA EN CIENCIAS DE LA SALUD

El acceso a una base de datos de artículos de investigación ayuda a hacerse una idea de la gran cantidad de conocimiento científico acumulado sobre importantes temas de la psicología y la logopedia. Así, por ejemplo, si una profesional busca, a partir de Pubmed, información sobre el tratamiento de la dislexia encontrará 189 artículos publicados en prestigiosas revistas (restringiendo la búsqueda sólo a los publicados entre 2000 y 2017). Si este número de artículos le parece elevado, en la tabla siguiente puede observar los que hallaría si el trastorno de interés fuera la afasia, la esquizofrenia o la depresión.

Ejemplos de productividad en las publicaciones científicas sobre cuatro trastornos.

TRASTORNO (...)	BUSCA: ... <i>Treatment o Therapy o Intervention</i>	2000–2017
Esquizofrenia	<i>Schizophrenia</i>	38.563
Depresión	<i>Depression</i>	124.895
Afasia	<i>Aphasia</i>	271
Dislexia	<i>Dyslexia</i>	189

Fuente: Motor de búsqueda Pubmed. <http://www.ncbi.nlm.nih.gov/pubmed/>.

Consulta: 17 de marzo de 2018.

En gran medida, la **práctica basada en la evidencia** (de ahora en adelante, PBE) se puede presentar como una manera de gestionar la colaboración entre la investigación científica y la práctica psicológica. A continuación, introduciremos los antecedentes de esta corriente y sus principales características.

Antecedentes

La constatación de una enorme variabilidad en la práctica clínica y de una falta de fundamentación científica de muchas decisiones llevó a un grupo de médicos vinculados a la Universidad de McMaster a iniciar un movimiento dentro de la enseñanza y la práctica de la medicina que denominaron *evidence based medicine* (Evidence Based Medicine Working Group, 1992). La traducción que se ha consolidado es «medicina basada en la evidencia» (de ahora en adelante, MBE). Habría sido mejor optar por «medicina basada en pruebas», porque en castellano la primera acepción de «evidencia» no es la de pruebas o indicios, tal como ocurre con el término inglés *evidence*, pero ahora mismo está tan extendida la traducción inicial que será difícil modificarla.

La MBE consiste en el uso consciente, explícito y sensato de las mejores pruebas disponibles en la toma de decisiones sobre la atención integral de cada paciente. La práctica de la MBE significa integrar la experiencia clínica individual con la mejor evidencia externa existente derivada de la investigación sistemática. En toda esta práctica el foco es el paciente, no solo su enfermedad o su proceso, de manera que, desde la realización de un diagnóstico efectivo y eficiente hasta la elección de la mejor

opción terapéutica, hay que identificar y considerar sus derechos, sus principios y sus preferencias (Sackett *et al.*, 1996).

Los conceptos que defiende la MBE nos remiten a otras aportaciones previas que también son antecedentes de la PBE en psicología y logopedia. Mencionaremos dos.

Siguiendo en el ámbito médico, el primer antecedente es Archie Cochrane, que, ya en 1972, publicó el libro *Effectiveness and efficiency: Random reflections on the health services*, de gran relevancia para la práctica de la medicina y la evaluación de las intervenciones médicas. Cochrane insistía en la necesidad de justificar la eficacia de las intervenciones aplicadas y, a partir de aquí, surge la propuesta de utilizar los ensayos clínicos aleatorios (diseños experimentales) como base sobre la que apoyar las decisiones clínicas.

La idea de Cochrane de organizar un resumen crítico, por especialidad, adaptado periódicamente, de todos los ensayos clínicos aleatorizados que sean relevantes fue el origen y la inspiración de la Colaboración Cochrane, organismo internacional que intenta ayudar a los profesionales a tomar decisiones sanitarias bien informadas mediante la preparación, actualización y garantía de la accesibilidad de las revisiones sistemáticas de los efectos de las intervenciones de salud (www.cochrane.org). Las revisiones sistemáticas consisten en una revisión exhaustiva de la literatura que trata un tema claramente definido y que utiliza procedimientos sistemáticos y replicables para identificar, seleccionar y evaluar de manera crítica todos los estudios relevantes, con la finalidad de recoger y analizar la información que provenga de los estudios incluidos (Bryman, 2012; Sánchez-Meca y Botella, 2010).

Un segundo antecedente de la PBE son los trabajos de Gene V. Glass y Mary Lee Smith sobre metanálisis (1976, 1977). El metanálisis implica el análisis estadístico de una amplia colección de resultados procedentes de estudios individuales con el propósito de integrar los hallazgos. Con el paso de los años, el metanálisis se ha convertido en un componente esencial del procedimiento de revisión sistemática, siempre y cuando sea posible llevar a cabo una síntesis cuantitativa de los resultados.ⁱⁱ

Práctica basada en la evidencia en psicología y logopedia

La American Psychological Association (APA) define la PBE como la integración de la mejor investigación disponible con la experiencia clínica en el contexto de las características, la cultura y las preferencias de los pacientes (APA Presidential Task Force on Evidence-Based Practice, 2006). El propósito de la PBE en psicología es promover una práctica psicológica eficaz y contribuir a la mejora de la salud pública aplicándole principios de evaluación e intervención psicológica con una base empírica. De manera análoga, la American Speech and Hearing Association (ASHA) ha definido la PBE en la Logopedia, como un enfoque en el que se integra la alta calidad de la investigación con la experiencia profesional, las preferencias de la persona y los valores en el proceso de toma de decisiones clínicas (ASHA, 2005).

Se puede observar el claro paralelismo entre estos planteamientos y el que se ha expuesto al definir la MBE. Probablemente esto ha favorecido que en las publicaciones sobre este tema sean habituales las referencias al «tratamiento» y al «paciente» como destinatario de la práctica. Tal como destaca la APA *Presidential Task Force on Evidence-Based Practice* (2006), en psicología debe tenerse en cuenta que en muchas

situaciones es preferible sustituir el término «paciente» por «cliente» o «consumidor» al referirse a la persona receptora del servicio. Así mismo, puede ser más conveniente hablar de este servicio en términos de «intervención» y no de «tratamiento».

La PBE se articula alrededor de una serie sucesiva de pasos que enumeramos a continuación.

1. Formular la pregunta de la manera más adecuada posible a partir del problema que se nos presenta. El proceso se inicia con un reconocimiento por parte del profesional, que necesita información para atender la demanda que plantea su paciente/cliente. Esta necesidad de información debe convertirse en una pregunta susceptible de respuesta concreta y empírica, y una buena manera de conseguirlo es especificando la información relativa a cinco elementos básicos presentes en cualquier estudio primario empírico dirigido a evaluar una intervención que se agrupan bajo el acrónimo PICOS: *Participants* (características de los participantes [demográficas —edad, género, etc.—, de salud, etc.], así como del contexto y del momento en que se lleva a cabo la investigación), *Interventions* (o exposiciones a factores de riesgo/protectores), *Comparisons o comparator* (características de las intervenciones que se están comparando), *Outcomes* (medidas de la variable dependiente o variable de respuesta), *Study design* (descripción de los diseños utilizados).
2. Localizar las mejores evidencias con que responder. Para hacerlo, hay que acceder a fuentes de información con garantías de rigor, como bases de datos bibliográficas, revistas científicas, repertorios de revisiones sistemáticas o guías de práctica clínica.
3. Evaluar y valorar críticamente la evidencia, determinando su validez y utilidad para dar respuesta a la pregunta inicial.
4. Aplicar las conclusiones a la práctica, considerando los riesgos y beneficios, las expectativas, las preferencias de los pacientes/clientes y sus necesidades emocionales.
5. Evaluar y valorar el rendimiento de esta aplicación.

Tal como refleja el inicio y el final de este proceso, una característica esencial de la PBE es su orientación al usuario de la información (por ejemplo, el psicólogo clínico) y su interés por cubrir las necesidades de su práctica profesional. El proceso empieza examinando las necesidades del paciente/cliente, pasa por la investigación de evidencia y concluye evaluando en qué grado la práctica realizada ha cubierto las necesidades iniciales.

Los pasos intermedios apelan a conceptos y herramientas de tipo metodológico. Saber cómo acceder a las fuentes de información y cómo gestionar la información obtenida es una competencia imprescindible para la PBE. Afortunadamente, los recursos para disponer de información adecuada han evolucionado mucho en los últimos años y las búsquedas de información son cada vez más rápidas y más satisfactorias.

Siguiendo con el esquema del proceso de PBE, otra competencia esencial es la capacidad para evaluar la calidad de la información científica. Deben conocerse los conceptos básicos sobre diseño y análisis de datos para poder valorar críticamente esta información. De acuerdo con esto, en este libro pretendemos ayudarle a adquirir con-

ceptos esenciales como los de validez interna y externa, y también iniciarle en el análisis metodológico de tres importantes familias de diseños de investigación. Es interesante que estos y otros conceptos que veremos en los diferentes capítulos los vaya situando en el marco más general que ofrece la PBE, porque seguro que así esta introducción a la metodología de investigación cobrará mayor sentido.

Reflexión final

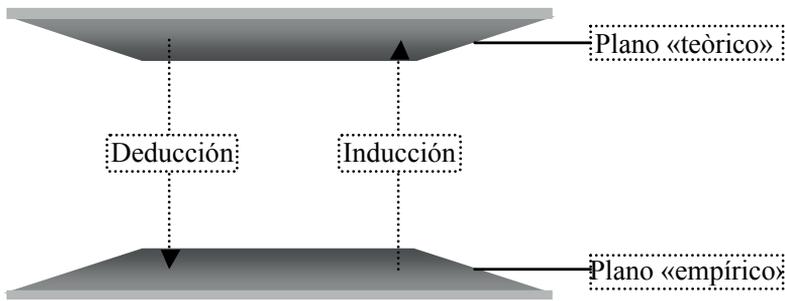
El código deontológico del psicólogo destaca su responsabilidad profesional para ofrecer intervenciones con apoyo empírico que garanticen la validez de sus efectos, y el deber ético de actualización de su competencia profesional. La búsqueda de información válida, su lectura crítica y la incorporación del saber que progresivamente se va adquiriendo con la práctica profesional son tareas que forman parte de la formación constante del especialista en cualquier rama de las ciencias de la salud. La PBE destaca con claridad este vínculo obligatorio entre ciencia y profesión.

La orientación de la PBE goza de gran popularidad y respeto en muchos ámbitos de las ciencias de la salud. A pesar de ello, la PBE no está exenta de controversia. Algunas críticas están relacionadas con el sobre esfuerzo que supone para el profesional sanitario ajustarse a los requisitos de la PBE. En un artículo publicado por Shaneyfelt en el año 2001, se estimaba que un profesional (en su caso, un internista) necesitaría leer, para mantener su conocimiento, veinte artículos diarios ¡todos los días del año! Aunque la disponibilidad de revisiones sistemáticas y guías prácticas reduce la necesidad de lectura de artículos originales, todavía resulta muy difícil mantener el ritmo. Otras críticas son más profundas y están relacionadas con un debate filosófico sobre la valoración de la validez del conocimiento. También existe el temor de que ciertos conflictos de intereses prioricen áreas de investigación en detrimento de otras. Mencionamos estas posturas críticas porque consideramos relevante conocerlas y discutir las, pero este desarrollo escapa a nuestro objetivo en este libro. Aquí solo hemos pretendido presentar la PBE, remarcar su importancia para el profesional y enmarcar el carácter de fundamentación que tienen los métodos, los diseños y las técnicas de investigación para poder comprenderla. Creemos que incluso a los que acabasen siendo críticos ante el enfoque de la PBE les sería útil conocer estos fundamentos.

FICHA 1.2. MÉTODO CIENTÍFICO: ESTRATEGIAS DE SISTEMATIZACIÓN

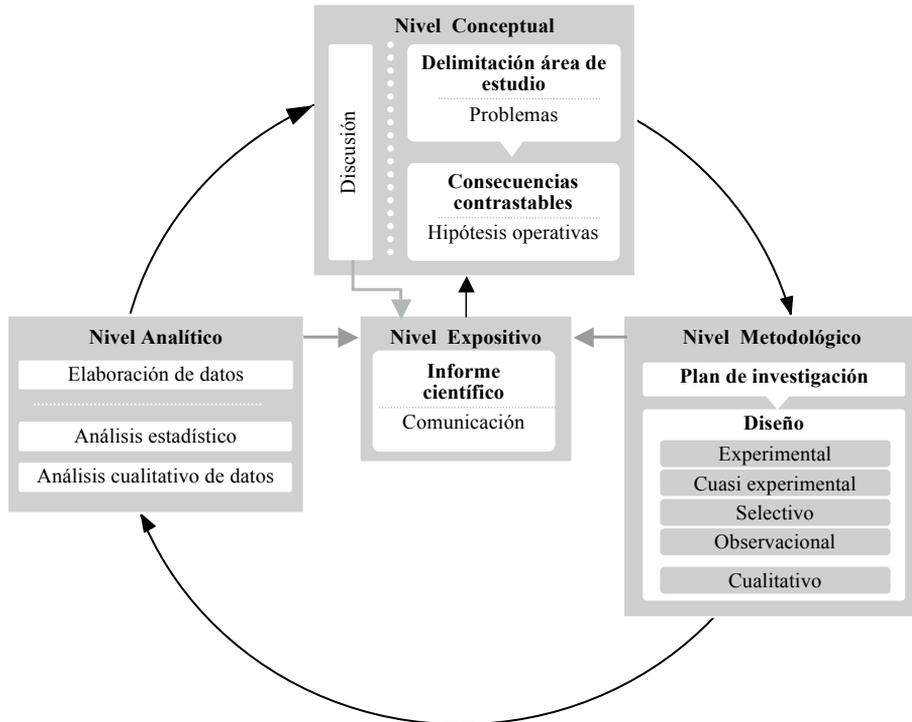
Existen varias estrategias generales para sistematizar la obtención de conocimiento. Por un lado, la estrategia deductiva parte de un cuerpo teórico para deducir consecuencias susceptibles de ser contrastadas empíricamente. Por otro, la inducción sigue el camino contrario: parte de un conjunto de observaciones, esto es datos empíricos, para detectar sus invariancias y, a través de un proceso de generalización, llegar a establecer una ley que se integra en un cuerpo teórico.

A menudo esta combinación se da de manera iterativa: se parte de un proceso deductivo para elaborar hipótesis contrastables y, posteriormente, se comprueba por un proceso de inducción si los datos obtenidos se ajustan al cuerpo teórico; se trata de la estrategia hipotético-deductiva.



FICHA 1.3. EL PROCESO DE INVESTIGACIÓN CIENTÍFICA

El método científico es un proceso **secuencial** a la vez que iterativo. Las fases en que se estructura se pueden organizar en cuatro niveles:² nivel teórico, nivel metodológico, nivel estadístico y nivel expositivo. La relación entre estos niveles y sus elementos básicos se representa en la siguiente figura.



Nivel teórico-conceptual. La investigación científica se inicia a partir de un marco teórico que es, precisamente, donde también se integran las conclusiones extraídas a partir de los resultados de la investigación. Se trata, por tanto, de un proceso circular que otorga a la teoría el papel de ser tanto la generadora de nuevos problemas de investigación como el elemento aglutinador que permite que el conocimiento sea acumulativo.

Si bien es cierto que las fuentes de problemas, en el sentido de generadoras de investigación, son virtualmente inagotables, las fuentes más frecuentes de obtención de problemas son la teoría y la propia investigación: vacíos explicativos o fallos predictivos en una teoría, contradicciones en las conclusiones de diversas investigaciones,

- Inspirados en la propuesta de Arnau (1989) dirigida especialmente a la investigación cuantitativa. Se han modificado varios aspectos para dar cabida a la investigación cualitativa y se ha añadido el nivel expositivo.

desajustes de tipo metodológico, son algunos ejemplos de ello (Bordens y Abbott, 2011).

Una vez identificado y delimitado el problema, y evaluada la viabilidad del estudio que derivará del mismo, hay que pasar del marco teórico general que nos proporciona el problema a su concreción en forma de solución tentativa y contrastable, esto es, la hipótesis operacional (u operativa), que indica explícitamente cómo se medirán las variables objeto de estudio (por ejemplo, la ansiedad se medirá con un instrumento llamado STAI) y/o qué criterios se establecerán para definir determinados valores de variables (por ejemplo, considerar que una persona está ansiosa cuando la puntuación en el STAI sea superior a 65 puntos).

Nivel metodológico. Una vez la hipótesis se ha operativizado, hay que diseñar un plan de investigación para obtener los datos que permitan contrastar esta hipótesis. Uno de los elementos más importantes de este nivel es la elección del método que debe utilizarse y, especialmente, el diseño concreto en que debe ejecutarse para obtener los datos. Cabe señalar que, dentro de la orientación cuantitativa (que incluye los diseños experimentales, cuasiexperimentales, selectivos y observacionales, si bien sobre estos últimos matizaremos su doble adscripción en la ficha 1.5), el diseño de investigación se planifica *a priori* y se convierte en un plan de acción en que cada paso está pautado. En cambio, desde la orientación cualitativa, el plan de investigación es más flexible, y suele distinguirse entre el diseño proyectado, el diseño emergente y el diseño final.

Es aquí, en el nivel metodológico, donde se centra este libro, y especialmente en la orientación cuantitativa.

Nivel analítico. En la orientación cuantitativa, una vez obtenidos los datos y elaborados, hay que analizarlos aplicando los modelos estadísticos más adecuados para comprobar si la hipótesis de partida es estadísticamente verosímil. En la investigación cualitativa, la información que se obtiene se analiza utilizando técnicas no estadísticas.

Nivel expositivo. Coincidimos con Shaughnessy, Zechmeister y Zechmeister (2015) cuando afirman que la investigación científica es una actividad pública y, de hecho, no se convierte en conocimiento al alcance de la comunidad científica, ni puede ser un conocimiento acumulativo, hasta que este no se ha comunicado. El protocolo de comunicación utilizado para hacer pública la investigación es el informe científico, que en psicología se aconseja que siga la normativa de la APA (American Psychological Association).ⁱⁱⁱ El informe científico tiene como principales objetivos facilitar la comunicación entre especialistas y, por extensión, posibilitar la replicabilidad de las investigaciones. Lo que se expone a través del informe científico puede aportar nueva información que, una vez incorporada al marco teórico de partida, puede ser generadora de nuevos problemas que inicien otra iteración en el proceso de investigación.

FICHA 1.4. ORIENTACIÓN CUANTITATIVA, ORIENTACIÓN CUALITATIVA Y MÉTODOS MIXTOS

Si bien hasta hace unos años se consideraba que la psicología científica era sinónimo de orientación cuantitativa, cada vez más se están admitiendo ciertas limitaciones en el conocimiento que aporta esta orientación que se pueden mitigar a través de la orientación cualitativa (Bryman, 2006a; Bryman, 2006b). Dentro de la orientación cualitativa, existe una marcada heterogeneidad en las características de los diseños utilizados, y esto dificulta la caracterización de esta orientación de manera exhaustiva. De todos modos, sí que existen algunas regularidades bastante claras que, además, sirven para poder comparar la orientación cualitativa con la cuantitativa (Creswell, 2009).^{iv}

Uno de los elementos que caracteriza la investigación cualitativa es el hecho de que se basa en procesos de investigación abiertos, es decir, que se modifican a medida que avanza la investigación, iterativos y con un fuerte componente interactivo entre investigador y participante.

La orientación cualitativa parte de la asunción de que no existe una única realidad y de que esta **realidad múltiple** proviene de la **interpretación** que hacen de la misma los **participantes**, y en menor medida los **investigadores**. Así, el principal objetivo del investigador es **comprender los significados** que tienen para los participantes los fenómenos objeto de estudio. Se trata de una comprensión contextualizada que exige una **interacción** intensiva y extensiva **con los participantes**. Estas características desvinculan la investigación cualitativa de la persecución de la objetividad, la alejan de la artificialidad en la situación de investigación y la vinculan a un conocimiento en profundidad pero de alcance limitado en que el investigador deja a un lado la asepsia de la orientación cuantitativa para convertirse en un instrumento de recogida e interpretación de los datos. A continuación se presentan estas características.

- Subjetividad: el hecho de no perseguir la objetividad e, incluso, de admitir la **subjetividad** como herramienta para la comprensión del fenómeno. Por otro lado, y de acuerdo con el concepto de **reflexividad**, el investigador debe poner atención a la forma en que su marco teórico de referencia, sus acciones durante la investigación y sus posteriores interpretaciones afectan a las conclusiones a las que llega y hacerlo explícito en el informe de investigación (Creswell, 2009).
- Realismo (validez ecológica). La importancia de obtener información en el contexto donde de manera natural se produce el fenómeno que se estudiará.
- Conocimiento en profundidad pero de alcance limitado. El **carácter idiográfico** y, por tanto, **focalizado en casos individuales**, del conocimiento. Por otro lado, puesto que no se puede comprender un fenómeno aislado de su contexto, el estudio cualitativo suele tener un carácter **holístico**.
- Protagonismo del investigador: el investigador es quien, en interacción con el fenómeno objeto de estudio y sus actores (los participantes), logra esta comprensión. A menudo esto implica entrar a formar parte del entorno de los participantes y crear un buen *rapport* para **tener acceso a las percepciones, las experiencias, las creencias y los valores** de los participantes. De esta manera, **el investigador se convierte en el instrumento de obtención y análisis de los**

datos más importante: él es quien a menudo crea unos **instrumentos de recogida de datos** que se caracterizan por **no estar estructurados sino abiertos**, quien los adapta si es necesario en el transcurso de la investigación, quien decide explorar otras vías de información u otras hipótesis, y quien tiene un papel clave en la interpretación de la información.

Otro rasgo característico de la investigación cualitativa es su emergencia. La investigación cualitativa puede ser **emergente** en varios aspectos: en las hipótesis, en el propio proceso de investigación y en la generación de la teoría. Así, su voluntad de excluir apriorismos implica utilizar diseños de investigación poco planificados *a priori* y, por tanto, cambiantes y difíciles de catalogar. Se trata de llegar a la comprensión de un fenómeno, y la manera de conseguirla, el plan de investigación, se va construyendo *ad hoc* fruto de la interacción con el fenómeno estudiado.

La tabla que se presenta a continuación resume las principales características que permiten diferenciar entre sí las orientaciones cuantitativa y cualitativa. Hay que tener en cuenta que la comparación que hacemos entre ambas orientaciones se basa en un conjunto de características que no definen exhaustivamente ninguna de las orientaciones por separado. Por otro lado, las diferencias que constatamos no pretenden enfrentar ambas orientaciones, sino poner de manifiesto aquellos aspectos en que se pueden complementar.

<i>ORIENTACIÓN CUANTITATIVA</i>	<i>ORIENTACIÓN CUALITATIVA</i>
Búsqueda de objetividad sin negar la intersubjetividad ³	La objetividad no tiene que ser necesariamente una meta
A menudo, se lleva acabo en entornos estructurados	En general, se lleva a cabo en entornos naturales
Nomotético	Idiográfico
Diseño planificado <i>a priori</i>	Diseño emergente no preestablecido
Generalmente se provoca («se elicit») la producción de información	La información se produce de manera natural
A menudo, la información es recogida por instrumentos estandarizados y validados	En general, el investigador es el elemento principal de obtención de información y no se suelen utilizar instrumentos estandarizados y validados
Vocación confirmatoria	Vocación exploratoria
Análisis estadístico	Análisis conceptual de los resultados

3. Hace referencia a que los fenómenos objeto de estudio son percibidos y elaborados por un determinado grupo de investigadores que tienen un marco teórico común. Este vínculo inevitable entre el fenómeno y quien lo estudia es incompatible con el conocimiento aséptico que caracterizaría el que es puramente objetivo, e implica que el mismo fenómeno es en cierta medida una elaboración compartida por un grupo de personas, esto es, intersubjetividad.

Durante muchos años han proliferado planteamientos que confrontan las orientaciones cuantitativa y cualitativa como si se tratara de opciones irreconciliables (Anguera, Blanco-Villaseñor, Losada, Sánchez-Algarra y Onwuegbuzie, 2018; Bryman, 2006b; Johnson y Onwuegbuzie, 2004). En paralelo, sin embargo, también se han consolidado opciones como la metodología observacional, en la que se integran de facto las perspectivas cualitativa y cuantitativa (Anguera, 2010).^v

Poco a poco se ha ido asumiendo que cada orientación tiene puntos fuertes y carencias, y también ha ido ganando fuerza la visión de lo que se ha denominado métodos mixtos que promueven la complementariedad entre la orientación cuantitativa y la cualitativa y su integración.^{vi} De acuerdo con Johnson y Onwuegbuzie (2004, p. 18), el principio rector de la investigación mixta es que «el investigador debería recoger múltiples datos empleando diferentes estrategias, aproximaciones y métodos de tal manera que la mezcla o combinación contenga puntos fuertes complementarios y puntos débiles no solapados».

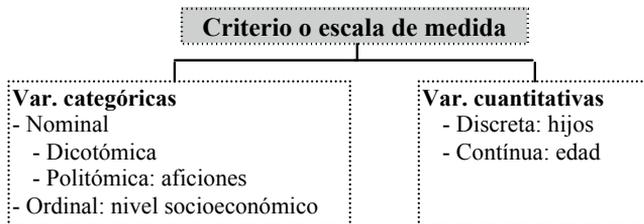
FICHA 1.5. TEORÍAS, HIPÓTESIS, CONSTRUCTOS Y VARIABLES

Decíamos antes que el punto de partida (especialmente en la orientación cuantitativa) y el de destino de la investigación es la teoría. Pero, ¿qué es una teoría?^{vii} Una **teoría** es un conjunto de hipótesis contrastadas y relacionadas que pueden dar una explicación coherente a fenómenos. Por otra parte, una **hipótesis**, que, de acuerdo con la definición anterior, se convierte en la base de las teorías una vez contrastada, es una solución tentativa a un problema. Veamos, por ejemplo, la siguiente hipótesis: «En enfermas de cáncer de mama, la resiliencia aparece más fácilmente en personas con formación superior que en personas con estudios básicos». En esta hipótesis, la resiliencia es una variable muy común en psicología, al igual que lo es la inteligencia, la depresión o la ansiedad: es un constructo. Un **constructo** es una variable no observable de manera directa. Así pues, dado que, a diferencia de otras variables que se pueden observar directamente, como por ejemplo el color del pelo, un constructo no se puede observar y, por tanto, tampoco se puede medir de manera directa; es necesario emplear algún **indicador** que permita medirlo de manera indirecta. En evaluación psicológica, esto se puede hacer con varios instrumentos, y el test es uno de los más populares. Así, por ejemplo, tenemos tests de inteligencia, de depresión o de resiliencia, entre muchos otros.

Para que esta hipótesis sea operativa, hay que establecer cómo se medirá la resiliencia y también qué criterio se establecerá para definir el nivel de estudios. En nuestro ejemplo una hipótesis operativa (operativizada) u operacional sería: «En personas que padecen cáncer de mama, la resiliencia, medida con el CD-RISC, aparece más fácilmente en personas con formación superior (estudios universitarios) que en personas con estudios básicos (ESO)».

Las hipótesis informan sobre unos elementos primordiales en una investigación: nos indican cuáles son las variables que se quieren estudiar. Así pues, definir operativamente una hipótesis implica operativizar (operacionalizar) las variables objeto de estudio, en el ejemplo anterior, resiliencia y formación. De hecho, las variables son un elemento central en la investigación científica y, por tanto, también debemos centrar la atención en ellas. Una **variable** es un atributo susceptible de tomar diferentes valores que pueden ser medidos.

De acuerdo con la escala de medida, las variables se pueden clasificar en los tipos que se indican en la siguiente figura (junto con cada tipo de variable indicamos un ejemplo).



Independientemente de que las variables sean categóricas o cuantitativas, hay que tener en cuenta qué papel desempeñan dentro del diseño de investigación, tal como se muestra en la siguiente figura. Podemos diferenciar las variables en función de si forman parte de lo que en este texto llamamos **contexto** o del **núcleo** de una investigación.

Las variables que forman parte del contexto hacen referencia a factores tanto del medio como del propio sujeto, que, pese a no ser objeto de estudio, pueden tener alguna incidencia sobre lo que se quiere estudiar. Técnicamente, estas variables del contexto son conocidas como **variables extrañas** y podemos identificar dos subtipos: las **variables de confusión**, o enmascaradas, y las **no confundidoras**. Más adelante definiremos el concepto de confusión y hablaremos de un grupo de variables, llamadas terceras variables, dentro de las cuales es posible situar las variables de confusión.

Por otra parte, las variables del núcleo de la investigación son aquellas que son objeto de estudio. Cuando este núcleo incluye más de una variable, hablamos de núcleo relacional. Distinguiamos dos tipos de núcleos relacionales:

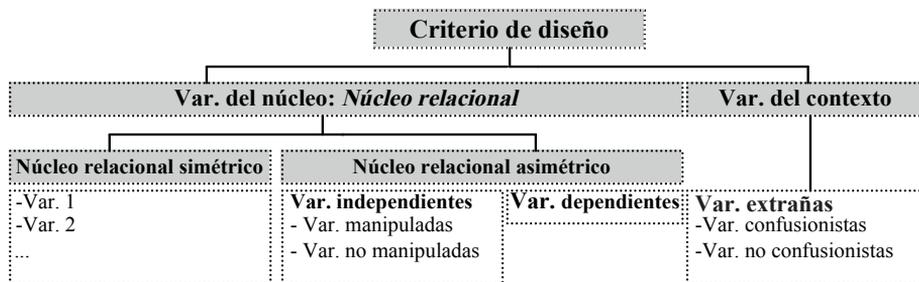
- El **asimétrico**, que supone diferenciar entre **variables independientes (VI)** o variables de exposición, que son la supuesta causa de las variaciones de la **variable dependiente (VD)** o variable de respuesta.⁴ En este caso, se trata de una relación unidireccional $VI \rightarrow VD$.
- El **simétrico**, en el que no se establece una relación causal entre las variables objeto de estudio y lo que interesa analizar simplemente es la covariación que existe entre ellas ($V1 \leftrightarrow V2$).

Así, por ejemplo, la hipótesis «las estrategias de aprendizaje activo mejoran el recuerdo a largo plazo» establece una relación asimétrica o unidireccional entre el tipo de estrategia de aprendizaje (VI) y el recuerdo a largo plazo (VD). En cambio, la hipótesis «hay relación entre la ansiedad y los resultados de los exámenes finales» es de tipo simétrico o bidireccional. Hay dos variables, «ansiedad» y «resultados de los exámenes», que se cree que están relacionadas entre sí, es decir, que cuando varía una también variará la otra, pero en esta hipótesis ninguna de ellas se plantea como la causa hipotética de la otra.

Las VI pueden ser tanto **manipuladas** como **no manipuladas**. El que el investigador manipule una VI significa que este crea las condiciones para que tengan lugar ciertos niveles de la VI, y que estos niveles se aplicarán a los participantes antes de medir la VD. Así pues, manipular una VI implica determinar atributos como la intensidad, la frecuencia o la duración de los niveles de la VI, determinar en qué momento se aplicarán estos niveles a los participantes, y determinar qué participantes recibirán uno u otro nivel de la VI.

Como veremos, la manipulación está presente en los diseños experimentales y cuasiexperimentales, que, a menudo, se agrupan conceptualmente bajo la denominación de «paradigma manipulativo» o «diseños manipulativos».

4. Si bien en los diseños experimentales y cuasiexperimentales se habla de variables independientes y de variables dependientes, en el capítulo 4 veremos que, en el contexto de los diseños *ex post facto*, las variables independientes y las variables dependientes se denominan, respectivamente, variables de exposición y variables de respuesta.



FICHA 1.6. LA ASOCIACIÓN ENTRE VARIABLES Y LAS RELACIONES CAUSALES: CRITERIOS DE CAUSALIDAD

En el ámbito de las ciencias de la salud se acepta que las relaciones causa-efecto tienen lugar en términos de relaciones funcionales probabilísticas. Es decir, asumimos que los valores de una variable que se establece como causa no determinan los resultados específicos de una variable efecto, sino que inciden sobre las distribuciones de probabilidad subyacentes a tales resultados.^{viii} En cualquier caso, la asociación entre variables es una condición necesaria pero no suficiente para establecer una relación causa-efecto. Desde este punto de vista, en ciencias de la salud, la aproximación más clásica para concluir que existe una **relación causal** establece los siguientes **tres criterios**:⁵

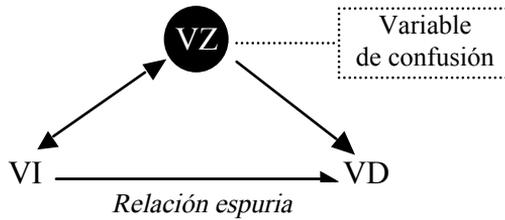
- **Asociación.** Los datos obtenidos en la investigación deben indicar que existe una relación o covariación significativa entre VI y VD (véase la ficha 1.5). Hablamos de covariación entre dos variables cuando, la distribución de los valores de una de ellas difiere en función de los valores de la otra. Para decidir si dos variables se relacionan entre sí debe excluirse que tal relación pueda ser explicada por azar.
- **Temporalidad.** La variable que se establece como causa (VI) debe preceder siempre a su efecto (VD) en el tiempo. Tal precedencia o prioridad temporal implica que toda relación causal es asimétrica o unidireccional.
- **Ausencia de espuriedad.** La tercera condición básica para el establecimiento de relaciones causales implica que no deben existir explicaciones alternativas de la VD diferentes de la variable VI. El esquema de una relación en que no intervienen otras variables (genéricamente denominadas terceras variables), es decir, que solo tenemos VI y VD y una relación unidireccional (causal) entre ellas es tan sencillo como el que se muestra a continuación.

VI -----> VD

Algunos autores, como Wegener y Fabrigar (2000), las llaman relaciones de efectos causales directos para diferenciarlas de las relaciones de efectos causales mediados y de efectos causales moderados, que comentaremos más adelante.

Cuando, además de la VI, existe una variable extraña sobre la que no hay hipótesis respecto a su efecto sobre la VD, que covaría con la VI y que puede explicar el «comportamiento» de la VD, se dice que existe una **relación espuria** que confunde o enmascara la relación entre VI y VD. Estas variables extrañas que provocan la relación espuria son las **variables de confusión**. En la siguiente figura se presenta el patrón de relaciones que caracteriza una relación espuria entre la VI y la VD provocada por la presencia de una variable de confusión (en la figura, VZ).

5. En esta ficha simplificamos la explicación de estos criterios refiriéndolos a los conceptos “variable independiente (VI)” y “variable dependiente (VD)” introducidos en la ficha 1.6. Sin embargo, debemos precisar que los conceptos expuestos en esta ficha 1.7 son aplicables a cualquier investigación (experimental o no) encaminada a estudiar la relación entre una probable causa X y un probable efecto Y.

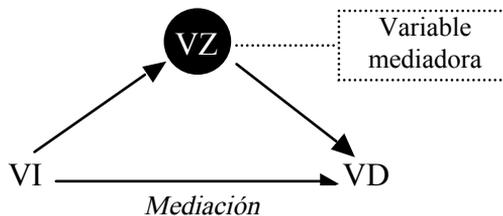


Cuando existe una variable extraña que influye en la VD, covaría (está relacionada o asociada) con la VI y no es una variable moderadora o mediadora entre VI y VD, esta variable se convierte en una variable de confusión y provoca que la relación entre VI y VD sea espuria, ya que en realidad la relación que se está observando es entre VI, VZ y VD.

Evitar que las relaciones espurias confundan los resultados es una de las principales tareas de los investigadores y, para ello, se aplican técnicas de control (véase la ficha 2.3) para evitar que variables extrañas se conviertan en confundidoras. De hecho, la validez interna (un concepto capital en métodos de investigación que se explica en la ficha 1.7) depende de que estas técnicas de control se apliquen de manera eficaz. Resultarán ser eficaces en la medida en que logren romper la covariación entre VZ y VI y eviten, así, la presencia de una relación espuria.

Además de las variables de confusión, las variables **mediadoras** y **moderadoras** también forman parte de las llamadas terceras variables. A diferencia de las anteriores, estas son variables sobre las que se tienen hipótesis (y, por tanto, **son objeto de estudio**, no objeto de control, como es el caso de las confundidoras), pero no en términos de efecto directo sobre la VD, como en el caso de la VI, sino en términos de efecto mediador (variables mediadoras) o moderador (variables moderadoras) entre VI y VD.

Una **variable mediadora** es aquella que se encuentra en medio de la cadena causal entre la VI y la VD. Si en la siguiente figura nos fijamos en la flecha que une VI y VZ, veremos que es una flecha unidireccional, igual que lo es la flecha que une VZ y VD. Las relaciones unidireccionales $VI \rightarrow VZ$, y $VZ \rightarrow VD$ implican que la relación entre VI y VD no es directa sino que está mediada por otra variable (VZ), de tal manera que primero VI causa VZ y, a continuación, VZ causa VD. La relación unidireccional $VI \rightarrow VZ$ es un elemento diferenciador entre mediación y confusión, ya que, en el caso de la confusión, hay una relación bidireccional $VI \leftrightarrow VZ$, lo que implica presencia de covariación, no de causación.

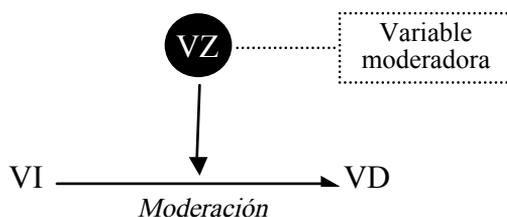


La presencia de una variable mediadora no implica necesariamente que la relación entre VI y VD sea inexistente; de hecho, es frecuente que, a pesar de la presencia de una variable mediadora, la VI explique una parte de la variación de la VD.

Un ejemplo de mediación: la calidad de la educación recibida (VI) afecta a la autoestima (variable mediadora), la cual afecta a la calidad de la educación dada a los hijos (VD).

Tal y como indican Baron y Kenny (1986), la mediación pretende explicar «por qué» o «cómo» se da una relación causal entre VI y VD.

Una variable moderadora es aquella que altera la dirección o la intensidad de la relación entre VI y VD. Al igual que en el caso de las variables mediadoras, las variables moderadoras son objeto de estudio, no de control.



La moderación es muy parecida a la interacción (veremos el concepto de interacción en la ficha 2.7). De hecho, tal como indican Ato y Vallejo (2011), la discusión sobre si ambos efectos son equivalentes continúa vigente, si bien existe un cierto consenso en que la moderación es un caso especial de interacción en que:

- Al igual que en la interacción, el efecto de moderación afecta a una relación causal $VI \rightarrow VD$ (Wu y Zumbo, 2007), de tal manera que los valores que toma la variable moderadora hacen variar la intensidad o la dirección del efecto de la VI sobre la VD. Es decir, el efecto que tiene la VI sobre la VD depende de qué valores tome la VZ (que, en este caso, es la variable moderadora).
- Se diferencia de la interacción en que no existe una hipótesis sobre el efecto directo de la variable moderadora sobre la VD. Así, si bien en una VI puede haber hipótesis tanto sobre su efecto directo sobre la VD como sobre el efecto combinado (interacción) con otras VI sobre la VD, en el caso de la variable moderadora se centra en la hipótesis sobre el efecto que tiene esta en la relación entre VI y VD.

Así, el efecto de moderación pretende explicar «cuándo» o «para a quién» una VI afecta más (o menos) la VD (Baron y Kenny, 1986).

Por ejemplo, en la hipótesis «el apoyo social tiene un efecto sobre los niveles de depresión, si bien este efecto es superior en mujeres que en hombres», la variable apoyo social es la VI, depresión es la VD y género es una variable moderadora. Nótese que aquí género, como variable moderadora, tiene un papel auxiliar en la explicación de la depresión a partir del apoyo social, dado que no se hipotetiza sobre el efecto aislado o directo de género sobre depresión.

FICHA 1.7. INTRODUCCIÓN AL CONCEPTO DE VALIDEZ

Siguiendo la propuesta de clasificación de la validez en investigación cuantitativa de Shadish, Cook y Campbell (2002), podemos decir, de manera genérica, que la **validez** de una investigación es el grado de correspondencia entre lo que se pretende estudiar y lo que en realidad se estudia. Un alto grado de validez implicaría que hay una correspondencia perfecta entre lo que se quiere estudiar y lo que efectivamente se estudia. Por ejemplo, cuando valoramos la validez de un «test» de inteligencia, nos preguntamos si el «test» estudia (mide, en este caso) lo que pretende estudiar, que es la inteligencia.

Hay diferentes tipos de validez: validez interna, validez externa, validez de constructo, etc. En este tema introduciremos los conceptos de validez interna y externa y los ampliaremos en la ficha 3.3.

La **validez interna** es el grado de certeza de que una covariación observada refleja una relación causal. Utilizando los conceptos presentados en las fichas 1.6 y 1.7, podremos decir que habrá un alto grado de validez interna en la medida en que se hayan descartado explicaciones alternativas a la VI de tal manera que la VD se pueda explicar únicamente a partir de la (o las) VI. Estas explicaciones alternativas que hay que evitar tienen un nombre propio que ya conocemos: variables de confusión o enmascaradas.

¿Cómo evitar que las variables extrañas sean variables de confusión? Aplicando **técnicas de control** como, por ejemplo, la aleatorización en el caso de diseños experimentales, o bien la igualación, que es aplicable a cualquier tipo de diseño (véase la ficha 2.3).

Para examinar de manera sistemática las posibles variables confundidoras que pueden estar presentes en un estudio se pueden consultar listas genéricas de amenazas a la validez interna (véase la ficha 3.3).

La **validez externa** se refiere al grado en que una investigación es generalizable a otros individuos, situaciones y momentos. Alcanzar un elevado grado de validez externa, y en especial de realismo, es el principal objetivo de los diseños observacionales (véase la ficha 1.8) en el sentido de velar por la aplicabilidad de los resultados a las situaciones reales sobre las que se quiere incidir. Por su parte, los diseños de encuesta también están centrados en alcanzar un alto grado de validez externa y especialmente en la vertiente que afecta a la generalización a otros individuos. Generalmente, los investigadores trabajan con muestras de participantes. Para que las conclusiones obtenidas con estas muestras se puedan generalizar a la población a la que pertenecen, es necesario que las muestras sean representativas. Empleando técnicas de muestreo que garanticen que toda la población tiene una probabilidad conocida y no nula de formar parte de la muestra y contando con un número suficiente de unidades de observación, se puede alcanzar el criterio de representatividad (véase la ficha 1.9).

La validez externa también es importante en los diseños experimentales, cuasiexperimentales o *ex post facto*; sin embargo, cuando se aplican estos diseños, se centran más esfuerzos en ganar seguridad en la inferencia causal, es decir, en la validez interna.

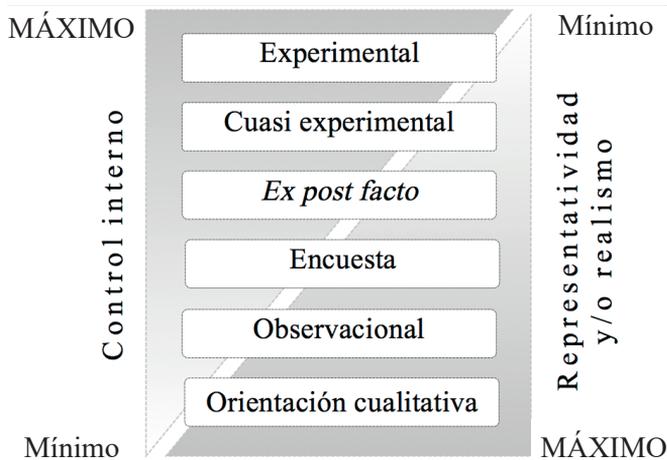
No hay que confundir validez con fiabilidad. La **fiabilidad** tiene que ver con la precisión con que estudiamos algún fenómeno. Esta precisión implica obtener medi-

das estables con independencia de la situación en que se estudie, de quién la estudie y del tiempo transcurrido entre dos medidas. Así, la fiabilidad referida a una investigación es el grado en que, al repetir una investigación empleando el mismo plan de investigación, se obtienen los mismos resultados. Nótese, pues, que la fiabilidad está directamente relacionada con el concepto de replicabilidad.

FICHA 1.8. MÉTODOS, DISEÑOS Y TÉCNICAS DE INVESTIGACIÓN EN PSICOLOGÍA

Los procedimientos empleados en la investigación científica se pueden agrupar en métodos, diseños y técnicas.^x Así, por ejemplo, un diseño de grupos aleatorios o un diseño factorial mixto son diseños experimentales (véase ficha 2.1), un diseño pre-post con control no equivalente es un cuasiexperimento y un diseño de casos y control forma parte del método selectivo (véase ficha 3.4.). Por su parte, hay técnicas específicas de algunos métodos y diseños, mientras que otras son aplicables a todos los diseños. Por ejemplo, la aleatorización (que es una técnica de control) es específica de los diseños experimentales, mientras que el muestreo por conglomerados (técnica de muestreo) o la constancia (técnica de control) son técnicas aplicables a casi todos los diseños.

En la siguiente figura están situados los diferentes tipos de diseños en función del grado de control interno, y de representatividad/realismo.



Revisemos, en primer lugar, los ejes que sirven para ordenar los diferentes grupos de diseños. El eje izquierdo representa el control interno en un continuo que va de menos (parte inferior) a más (parte superior). El control interno se refiere al grado de estructuración impuesta a la situación de estudio y se plantea como el concepto opuesto al grado de naturalidad o realismo de la situación. Un control interno alto supone un nivel de intervención y manipulabilidad también alto y el uso de técnicas de control, por lo que un alto control interno crea las condiciones para obtener conocimiento con una alta validez interna. Por su parte, el eje de la parte derecha de la figura se refiere a los conceptos de representatividad y realismo y nos remite al concepto de validez externa.

Como veremos en los próximos temas, los diseños experimental y cuasiexperimental tienen como principal objetivo contrastar relaciones causales, con la diferencia de que los diseños cuasiexperimentales suelen alcanzar una mejor validez externa, ya que habitualmente se llevan a cabo en entornos más cercanos a la realidad que se quiere estudiar que los experimentales, pero aportan menos validez interna debido a la ausencia de aleatorización.

El método selectivo, que engloba tanto los diseños de encuesta como los *ex post facto*,⁶ es una estrategia particular del método científico que se propone obtener información cuantitativa dirigida a la resolución de problemas de distinta naturaleza (descripción, covariación, predicción, causación, etc.). A tal fin, en el método selectivo se ejerce un «control indirecto» sobre las respuestas objeto de estudio que se centra fundamentalmente en la selección, la estandarización y el control estadístico (Gómez, 1990).

Los diseños *ex post facto* se usan cuando no es posible que el investigador intervenga en la administración de los niveles de la VI, es decir, cuando no es posible manipular la VI. En estos casos, se deben seleccionar sujetos que ya presentan (de ahí el nombre *ex post facto*) determinados valores en la variable de exposición (la variable que juega un papel equivalente a la VI en este tipo de diseños) o bien en la de respuesta (la variable que juega un papel equivalente a la VD en este tipo de diseños). Esta falta de manipulación por parte del investigador le resta validez interna, pero a la vez puede suponer una ventaja en el momento de obtener muestras más representativas y/o trabajar en condiciones más naturales (representatividad/ realismo).

El principal objetivo de los diseños de encuesta es la interrogación general sobre un gran conjunto de aspectos —actuales, pasados o futuros—, con el objetivo de describir tanto variables de manera aislada como relaciones entre variables de una población a partir de una muestra con la mayor representatividad posible (Gómez, 1990; Martínez Arias, 1995). En estos diseños los procedimientos de muestreo adquieren una especial importancia así como los instrumentos empleados para recoger información (el cuestionario y la entrevista).

Los diseños observacionales se presentan como una alternativa capaz de abordar problemas de cualquier naturaleza en los casos en que la prioridad es el registro sistemático y la cuantificación del comportamiento reduciendo a su mínima expresión cualquier alteración del entorno natural de los participantes (Anguera, 2010; Anguera, Blanco y Losada, 2001; Portell, Anguera, Chacón y Sanduvete, 2015). Los diseños desarrollados en el marco de la metodología observacional permiten fundamentar la obtención de evidencia en el seguimiento intensivo de la conducta en su contexto cotidiano, obteniendo descripciones de cambios momento a momento en conductas específicas, con la posibilidad de basar la inferencia causal en la obtención de regularidades apelando a esquemas de causalidad más complejos que el expuesto previamente, como el modelo INUS de John Mackie (cfr. Tacq, 2011). En la ficha 1.5 ya nos hemos referido al carácter de método mixto que se asocia a la metodología observacional y sobre el que puede profundizarse en las lecturas que se referencias en el apartado de notas finales.

Respecto a la orientación cualitativa, su caracterización está resumida en la ficha 1.4.

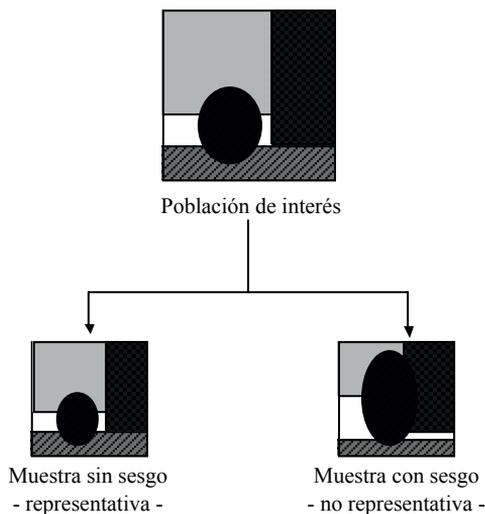
El objetivo del texto es introducir los tres primeros tipos de diseños de la figura anterior (experimental, cuasiexperimental y *ex post facto*) desde una posición que, tal como se evidencia en este capítulo, defiende la conveniencia de una complementariedad metodológica en la investigación en ciencias de la salud.

6. Tal como se indica en el capítulo 4, en inglés estos diseños se suelen agrupar junto con los diseños no manipulativos bajo la denominación *observational design*.

FICHA 1.9. MUESTREO

Los resultados de un análisis de sangre se obtienen a partir de una pequeña cantidad de la sangre que pasa por las venas. Ahora bien, los resultados de la analítica no hacen referencia a este subconjunto de la sangre que tiene una persona, la muestra, sino a los cinco litros de sangre que recorren su cuerpo. Esto es posible porque, previamente, se ha podido comprobar que la composición de esta muestra de sangre es la misma, a escala reducida, que la del resto de la sangre. Es decir, es una muestra representativa.

Salvo el caso de estudios censales, las conclusiones que se obtienen en una investigación con humanos también provienen de muestras, pero, en este caso, las **muestras** se componen de participantes de la investigación que son subconjuntos de una población. Para que las conclusiones obtenidas con estas muestras se puedan generalizar a la población de interés a la que pertenecen, es necesario que las muestras **tengan las mismas características relevantes que la población** de la que forman parte, es decir, que sean representativas ya que de otra forma llevarían a realizar inferencias sesgadas. La siguiente figura simboliza lo que supone que una muestra sea o no representativa; se puede observar que el cuadro izquierdo reproduce en pequeño la composición del cuadro que representa a la población. Este es el objetivo de la aplicación de **técnicas de muestreo**: obtener **muestras representativas**. Hay que tener en cuenta que la representatividad (estadística) de una muestra depende tanto de que la técnica de muestreo se aplique correctamente como de que el tamaño muestral (número de elementos de la muestra) sea suficiente.^{xi}

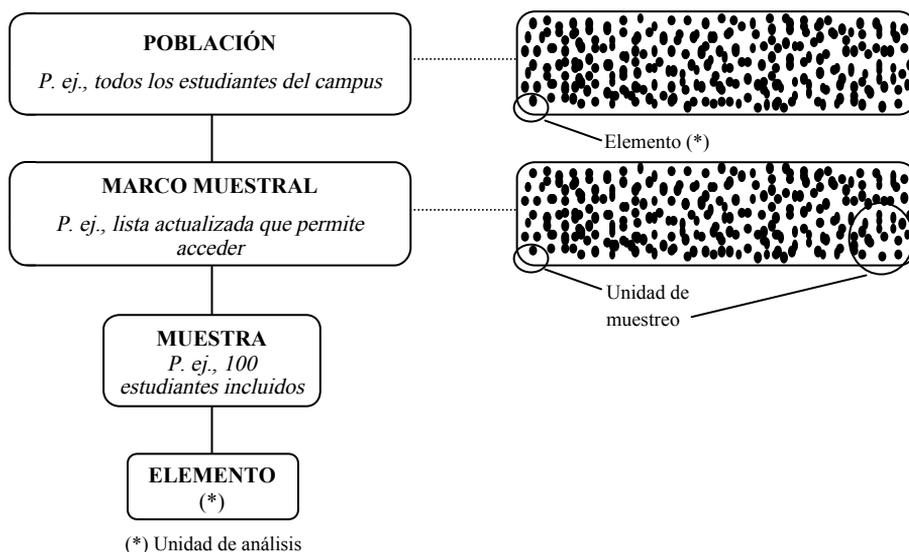


A continuación presentamos los conceptos básicos que intervienen en el muestreo:

- **Población.** Es el conjunto de todos los elementos sobre los que se quiere obtener información. Cuando hablamos de población nos referimos al conjunto de elementos que comparten ciertas características que son objeto de estudio (por

ejemplo, ser estudiantes universitarios, ser trabajadores en activo, etc.), esto es, la población diana. Hay que tener en cuenta que la **población diana** (población objetivo o de interés) a menudo no contiene los mismos elementos que la población sobre la que se extrae la muestra. Por ejemplo, si llevamos a cabo una encuesta telefónica para conocer la intención de voto, si bien la población diana estaría formada por los adultos que pueden votar, la población de la cual se extraería la muestra excluiría aquellos que no tienen teléfono, los que no han podido ser localizados y los que no han querido o no han podido formar parte de la investigación. Es decir, la muestra se obtendría sobre la base de un subconjunto de la población diana que es la **población accesible**.

- **Marco muestral.** Es una lista completa y actualizada de todas las unidades de muestreo pertenecientes a una población (población accesible) que pueden ser incluidas en una muestra. Esta lista debe permitir identificar y localizar las unidades de análisis, y es la base sobre la que se lleva a cabo la selección aleatoria de una muestra; de hecho, sin marco muestral es imposible llevar a cabo un muestreo aleatorio.
- **Unidad de muestreo.** Entidad que puede ser seleccionada para formar parte de una muestra. La unidad de muestreo no tiene por qué coincidir con la unidad de análisis. Por ejemplo, se puede hacer una selección de domicilios (unidad de muestreo), pero la información se obtiene a partir de las personas que viven en estos (unidad de análisis, ver a continuación). Las unidades de muestreo no tienen por qué ser necesariamente elementos simples como personas, también pueden ser agrupaciones tales como hogares, colectivos, comunidades u organizaciones.
- **Unidad de análisis o elemento.** La entidad sobre la que el investigador recoge información y que proporciona datos para el análisis. Al igual que en las unidades de muestreo, las unidades de análisis pueden ser elementos aislados o agrupaciones de elementos como díadas madre-hijo, familias u organizaciones.



Tal como decíamos, las **muestras representativas** se obtienen a través del uso de **técnicas de muestreo** que garantizan que toda la población tiene una probabilidad conocida y no nula de formar parte de la muestra. Si el **marco muestral** está bien definido y el tamaño de la muestra es lo suficiente grande, los **muestreos aleatorios o probabilísticos** son la mejor opción reconocida para lograr el objetivo de representatividad estadística y permiten inferir propiedades de una población a partir de la muestra.

Muestreos aleatorios

Antes de adentrarnos en los diferentes muestreos aleatorios, nos detendremos un momento a aclarar el término «azar». Hay que evitar confundir el significado del término azar que aquí se expone con la definición que asocia la palabra azar con algo fortuito, casual, accidental o sin objetivo. De hecho, cuando en investigación hablamos de hacer algo «al azar», nos referimos a aplicar procedimientos muy sistemáticos.

Un sorteo al azar es una estrategia para hacer equiprobables todos los integrantes de la población en el proceso de selección. Muchas veces, de manera equívoca, se cree que escoger al azar es lo mismo que elegir «a dedo» a los integrantes de la muestra porque el investigador no tiene preferencias por ninguno de ellos, y, por tanto, todo el mundo tiene alguna posibilidad de ser elegido. Se verá que no es así en el siguiente ejemplo. Si hay que elegir una rata de una jaula donde hay muchas, ¿se considera que abrir la jaula y coger una es un procedimiento de selección al azar? Debido a que el investigador no tiene ninguna preferencia por alguna de ellas, se podría pensar que todas tienen la misma probabilidad de salir. Pero, en realidad, coge la que se deja atrapar, ya sea porque no es tan rápida para escaparse respecto a sus compañeras, ya sea porque está enferma... Una manera de asegurar que el azar es el responsable de esta selección consiste en seguir algún procedimiento que garantice que todas las ratas tengan la misma probabilidad de ser escogidas.

Para realizar un sorteo al azar existen diferentes procedimientos. Una manera intuitiva es numerar cada integrante de la población y utilizar un bombo. Nadie, sin embargo, salvo la Lotería Oficial, puede manejar un bombo con miles de números. Un procedimiento más viable sería utilizar un bombo pequeño con diez bolas numeradas de 0 a 9. Se procede a efectuar un sorteo para el número que va en la primera unidad, se devuelve al bombo la bola extraída y se hace una segunda extracción para el número correspondiente a la decena, un tercero para la centena, y así sucesivamente. De esta manera, no importa el tamaño de la población.

Si bien, hace años, la mayoría de investigadores utilizaban una tabla de números aleatorios, actualmente se emplean aplicaciones para generar números aleatorios. Cualquiera que sea el medio, los números deben ser obtenidos a través de un procedimiento probadamente aleatorio, lo que implica que el conjunto de números debe cumplir las condiciones de equiprobabilidad e independencia; esto quiere decir que:

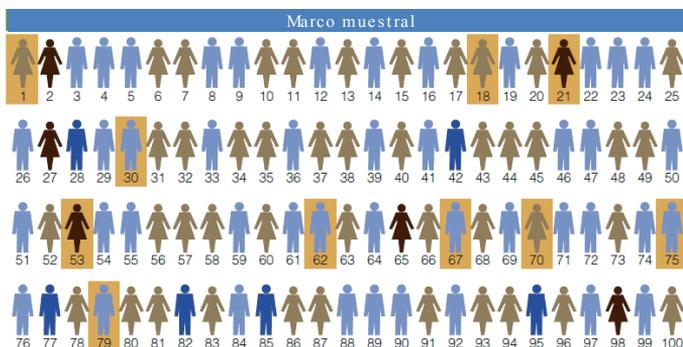
- Todo número tiene la misma probabilidad de salir. Esto significa que cualquier la cifra tiene la probabilidad de aparecer.
- La aparición de una cierta cifra no condiciona para nada la cifra siguiente ni está condicionada por la anterior.

En función del número de etapas en que se obtienen muestras aleatorias, podemos diferenciar entre:

- Muestreo en una etapa: conlleva un único proceso de extracción al azar que se corresponde con un muestreo aleatorio simple, sistemático, estratificado o por conglomerados.
- Muestreo polietápico: conlleva procesos de extracción al azar en diversas fases del muestreo. Por ejemplo, en una primera etapa, se realiza un muestreo por conglomerados, y en una segunda etapa, se aplica un muestreo aleatorio simple dentro de cada uno de los conglomerados elegidos en la primera etapa.

Son **muestreos aleatorios o probabilísticos** el muestreo aleatorio simple, el muestreo aleatorio estratificado, el muestreo sistemático y el muestreo por conglomerados.

- **Muestreo aleatorio simple.** Se caracteriza por seleccionar unidades de muestreo simples de tal manera que cada una de estas unidades de la población tiene la misma probabilidad de formar parte de la muestra. Los siguientes pasos ilustran el proceso que se llevaría a cabo en un muestreo aleatorio simple y son la base para comprender el resto de muestreos.
 1. Se define la población, por ejemplo, los estudiantes de psicología de las universidades catalanas, que, por ejemplo, está formada por 210.000 estudiantes ($N = 210.000$).
 2. Se obtiene la lista exhaustiva y actualizada de estudiantes de las universidades catalanas con suficiente información para poder llevar a cabo el procedimiento de selección, es decir, se construye el marco muestral.
 3. Se asigna un número consecutivo de 1 a N , en nuestro caso, de 1 a 210.000.
 4. Se decide el tamaño de la muestra, por ejemplo, 500. Esta decisión tiene una repercusión directa sobre la precisión con la que se pueden generalizar los resultados a la población, pero eso no lo trataremos en este texto (los textos citados en la nota ^{xi} aportan información detallada sobre este punto).
 5. Utilizando una tabla de números aleatorios o el *software* adecuado para generar números aleatorios, se seleccionan los participantes. La figura que aparece a continuación ilustra el proceso.



Números aleatorios			Muestra				
10480	15011	01536	30	67	70	21	62
22368	46573	25595					
24130	48360	22527					
42167	93093	06243					
37570	39975	81837					
77921	06907	11008					
99562	72905	56420					
96301	91977	05463					
89579	14342	63661					
85475	36857	53342					
28918	69578	88231					
63553	40961	48235					
09429	93969	52636					

Adaptado de: Babbie, E. R. (2011), p. 229.

- **Muestreo aleatorio estratificado.** Procedimiento de muestreo en que la población se divide en grupos o «estratos» de acuerdo con alguna característica relevante relacionada con las variables que se estudian. Se extraen muestras aleatorias independientes de cada estrato con el fin de obtener mayor precisión en la representación de los estratos. Así, si se estratifica por la variable sexo, establecemos el estrato «masculino» y el estrato «femenino». La eficacia de esta técnica depende de la intensidad de relación entre la/s variable/s de estratificación y la/s variable/s objeto de estudio. Cuanto mayor sea esta relación, mayor será la ganancia en precisión. La información sobre cuáles deben ser las variables de estratificación se obtiene de estudios previos o de modelos teóricos. El muestreo aleatorio estratificado se puede hacer con varios tipos de **afijación**,⁷ que es el procedimiento por el que se determina cuántas unidades de cada estrato habrá en la muestra. De acuerdo con el tipo de afijación, tenemos varios tipos de muestreo aleatorio estratificado:
 - Muestreo aleatorio estratificado con afijación simple: se selecciona el mismo número de elementos de cada estrato. Siguiendo con el ejemplo de la estratificación por la variable sexo, se seleccionaría el mismo número de hombres y de mujeres.
 - Muestreo aleatorio estratificado con afijación proporcional: el número de elementos es proporcional al tamaño del estrato en la población. Si la población se compone de un 60 % de hombres y un 40 % de mujeres, la muestra debería reflejar esta proporción de ambos estratos.
 - Muestreo aleatorio estratificado con afijación óptima: para establecer el número de unidades muestrales de cada estrato se tiene en cuenta también la heterogeneidad de la población en las características relacionadas con los estratos, de manera que una mayor heterogeneidad supone mayor número de casos.
- **Muestreo sistemático**
 Consiste en seleccionar de una lista un elemento de cada k elementos. Se parte de un marco muestral (la lista de elementos), se selecciona al azar la primera

7. La afijación también recibe el nombre de «asignación», que es la traducción más directa de *allocation*, el término inglés utilizado habitualmente. En este texto utilizamos «afijación» para no confundir este concepto con el de asignación que veremos en los diseños experimentales y cuasiexperimentales.

unidad seleccionada y, a partir de esta, se selecciona una unidad de cada k elementos. Por ejemplo, a partir del marco muestral de estudiantes matriculados en psicología en la UAB, se selecciona al azar la primera estudiante que formará parte de la muestra y, a partir de esta, se selecciona un estudiante de cada 25.

Un elemento esencial que debe valorarse en este muestreo es el orden en que están listados los elementos del marco muestral. Si el orden de los elementos puede llevar a sesgar la muestra, es preciso, ante todo, reordenar la lista.

– **Muestreo por conglomerados**

Técnica de muestreo en que las unidades de muestreo son agrupaciones ya formadas o grupos naturales (por ejemplo, ciudades, escuelas, facultades, etc.), de tal manera que se seleccionan aleatoriamente estas agrupaciones en lugar de seleccionar elementos simples o aislados como en los anteriores muestreos. En este caso, la unidad de muestreo no coincide con la unidad de análisis, ya que, pese a que se seleccionan grupos, la información que se recoge y analiza es la de los elementos que están dentro de estos grupos.

En contraste con el muestreo estratificado, en el que la composición de los estratos es homogénea, en el muestreo por conglomerados se espera que la composición de los conglomerados sea heterogénea.

Muestras no aleatorias

En una de las obras que les hemos recomendado para ampliar conocimiento sobre el muestreo, el profesor Luís C. Silva (2000) se refiere a los “méritos del azar” para sintetizar las dos grandes ventajas de utilizarlo para la construcción de muestras: (1) los procedimientos probabilísticos satisfacen esa exigencia intuitiva de protegernos ante distorsiones sistemáticas (sean deliberadas o no) en la elección de los elementos de la muestra, y (2) sólo cuando el método es probabilístico se podrá solucionar la tarea de medir el grado de precisión con que se realizan las estimaciones sobre los parámetros poblacionales. A pesar de estas importantes ventajas no siempre es posible aplicar muestreo aleatorio y, en su lugar, se usan procedimientos para la construcción de muestras como los que se indican a continuación:

- **Muestras accidentales o de conveniencia:** aquellas en las que se incluyen los participantes según el grado de accesibilidad y de disponibilidad. Por ejemplo, un profesor que lleva a cabo una investigación con sus alumnos.
- **Muestras a propósito:** aquellas en que la inclusión/exclusión de los participantes depende de que cumplan determinados criterios. Por ejemplo, en una investigación sobre atención con estímulos visuales, se podrían excluir a las personas con problemas de visión y con problemas de atención, así como incluir únicamente a personas con un determinado nivel de formación.
- **Muestras por cuotas:** a partir de la identificación de una variable relevante o más de una para el estudio y del conocimiento de la distribución de estas variables en la población, se determina el número de participantes que tiene que haber en la muestra de cada nivel de estas variables. El punto de partida de las muestras por cuotas es similar al del muestreo estratificado (conocimiento de la

distribución poblacional de características relevantes para la investigación), pero existe una diferencia clave: la selección de unidades no se basa en un procedimiento aleatorio y, por tanto, sigue sin poder contar con los "méritos del azar" a los que nos hemos referido previamente.

Por ejemplo, si la distribución de la variable género es relevante para estudiar el éxito académico en estudiantes de psicología, y suponiendo que la distribución sea del 75 % de mujeres y el 25 % de hombres, a fin de obtener una muestra para cuotas de 100 estudiantes de psicología, cogeríamos a 75 mujeres y 25 hombres.

El muestreo y el error

El uso de un muestreo adecuado y llevado a cabo de manera rigurosa pretende disminuir los errores que pueden ir apareciendo desde que se confecciona el marco muestral hasta que se obtienen los datos. En este transcurso, se pueden dar errores de cobertura, errores de muestreo, errores de no respuesta y errores de medida.

El **error de cobertura** se produce cuando el marco muestral es incompleto o inadecuado. Es decir, cuando no incluye determinados elementos de la población de interés, o cuando incluye pocos (infracobertura) o bien incluye demasiados elementos (sobrecobertura).

El **error de muestreo** es implícito al hecho de trabajar con muestras en lugar de trabajar con poblaciones enteras (como se hace en los censos). La muestra que se extrae de una población es una de las muchas posibles muestras que se habrían podido extraer. De hecho, esta muestra tiene unas características concretas que no serían las mismas si obtuviéramos otra muestra de la misma población. Aunque el proceso de muestreo sea aleatorio y se haya ejecutado rigurosamente, las características de la muestra no serán exactamente iguales que las características de la población. Estas diferencias entre muestra y población se denominan error de muestreo.

El **error de no respuesta** se dan cuando hay un cierto número de unidades seleccionadas para formar parte de la muestra que no participan (total o parcialmente) en la investigación, ya sea porque se niegan o porque no es posible contactar con ellas.

Los **errores de medida** se da cuando las respuestas no reflejan el verdadero valor de las características o atributos de la población. Las fuentes más frecuentes de estos errores son: la influencia del entrevistador, el instrumento de recogida de datos, la actitud del participante y su grado de cooperación, y el modo de recogida de datos. Algunos de estos sesgos se corresponden con amenazas a la validez de constructo que se tratan en la ficha 3.3.

Así pues, hay que tener en cuenta que un muestreo aleatorio planificado y ejecutado adecuadamente permite minimizar los errores de cobertura y de muestreo, pero no minimiza errores igualmente importantes como los de no respuesta o los de medida.

FICHA 1.10. GUÍAS DE LECTURA CRÍTICA

Del capítulo 2 en adelante, se encuentra un recurso de aprendizaje que consideramos esencial para retener y poner en práctica los elementos más básicos de los métodos, diseños y técnicas de investigación: las guías de lectura crítica.

Estas guías de lectura crítica que proponemos se inspiran en las guías de publicación (*reporting guidelines*) y en las herramientas de valoración de la calidad metodológica. Estos dos recursos van de la mano de tres de los elementos más relevantes en ciencia: la replicabilidad, íntimamente relacionada con la fiabilidad, y la validez.

Las **guías de publicación** son directrices dirigidas a los autores sobre lo que deben informar respecto a los métodos empleados y los resultados hallados en una investigación para que el informe de investigación (que a menudo se concreta en un artículo de revista científica) sea transparente, exhaustivo y sin sesgo.

Existen guías de publicación para diseños de investigación concretos. Así, por ejemplo, para los diseños que se exponen en el capítulo 2 el referente es la CONSORT (CONsolidated Standards Of Reporting Trials; cfr. Shulz, Altman y Moher, 2010), para los diseños que se exponen en el capítulo 3 la guía más conocida es la TREND (Transparent Reporting of Evaluations with Nonrandomized Design; cfr. Jarlais, Lyles, Crepez y TREND Group, 2004), y para los diseños que se introducen en el capítulo 4 la guía más conocida es la STROBE (STrengthening the Reporting of Observational Studies in Epidemiology; cfr. Vandembroucke et al., 2007). Otra guía de referencia es la PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses); cfr. Liberati et al., 2009), creada para informar sobre las revisiones sistemáticas y metanálisis que hemos mencionado en la Ficha 1.1. Para otros tipos de diseños existe menos consenso sobre las respectivas guías, pero paulatinamente se van justificando nuevas propuestas. Por ejemplo, para presentar informes basados en la metodología observacional que hemos mencionado en la Ficha 1.9 se ha propuesto la guía GREOM (Guidelines for Reporting Evaluations based on Observational Methodology; cfr. Portell, Anguera, Anguera, Chacón-Moscoso y Sanduvete-Chaves, 2015), para la investigación de orientación cualitativa que hemos mencionado en la Ficha 1.4 se propone la guía COREQ (COnsolidated criteria for REporting Qualitative research; cfr. Tong, Sainsbury y Craig (2007) y para la síntesis de este tipo de estudios la guía ENTREQ (ENhancing Transparency in REporting the synthesis of Qualitative research; cfr. Tong, Flemming, McInnes, Oliver y Craig, 2012). Todas las guías citadas están recogidas en el sitio web de la red EQUATOR (Enhancing the Quality and Transparency of Health Research; <http://www.equator-network.org/>), donde se compilan las principales propuestas para estandarizar la elaboración de informes de investigación. Es importante destacar también las guías de publicación lideradas en los últimos años desde la APA (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008; Appelbaum, Cooper, Kline, Mayo-Wilson, Nezu y Rao, 2018; Levitt, Bamberg, Creswell, Frost, Josselson y Suárez-Orozco, 2018) que complementan la normativa establecida en su famoso manual de publicación (APA, 2010).

Se suele distinguir entre guías de publicación y herramientas de valoración de la calidad metodológica. Estrictamente, las guías de publicación están dirigidas a mejorar la calidad de la publicación de los informes de investigación, es decir, a obtener

informes científicos con una cantidad y calidad de información suficiente para hacer posible la replicabilidad de la investigación, la inclusión de los resultados de varias investigaciones en revisiones sistemáticas y metanálisis, así como la valoración objetiva de la calidad metodológica. Desde este punto de vista, lo cierto es que indirectamente las guías de publicación pueden mejorar la calidad metodológica de la investigación, ya que, por una parte, centran la atención sobre aspectos primordiales de las fases de diseño, ejecución y análisis de datos, y por otro, el hecho de que los investigadores no sean capaces de dar respuesta a una información requerida en estas guías es generalmente indicador de que hay algún aspecto fundamental de la investigación que es inadecuado.

Por su parte, las **herramientas de valoración de la calidad metodológica** se centran en evaluar diversos aspectos considerados clave para poder valorar la validez y la fiabilidad de los estudios. Así, conocer los instrumentos de valoración de la calidad puede convertirse para el investigador en una guía de cómo debe llevar a cabo la investigación en cuanto a cuestiones metodológicas clave; para los editores de revistas puede servir para establecer los criterios mínimos de calidad comunes que debe tener un artículo para poder ser publicado; y para el lector de literatura científica (por ejemplo, un profesional de la salud que actualiza sus conocimientos leyendo los artículos publicados recientemente sobre un tema) puede convertirse en una guía para valorar la calidad de la evidencia que aporta una investigación.

A pesar de que no hay consenso sobre la definición estricta de la calidad metodológica, sí hay varias propuestas de evaluación que tienen muchos puntos en común (véase, por ejemplo, Chacón-Moscoso, Sanduete-Chaves y Sánchez-Martín, 2016). Sin embargo, ninguna de las propuestas que mencionamos a continuación se ha convertido en un referente o *gold standard* para medir la calidad.

Los instrumentos de valoración de la calidad evalúan dimensiones de calidad metodológica relacionadas con la validez externa, la validez interna, la validez de conclusión estadística y diversos aspectos relacionados con la fiabilidad y la validez de la medida de las variables objeto de estudio (véase, por ejemplo, Valentine y Cooper, 2008). Pocas herramientas valoran todos estos aspectos de la calidad, pero casi todas valoran dimensiones relacionadas con la validez interna o con riesgo de sesgo, como la comparabilidad inicial o sesgo de selección (*selection bias*), el mantenimiento de la comparabilidad durante el estudio (*performance bias*), la aplicación de cegamiento (*blinding*) o la presencia de mortalidad selectiva (*attrition* o *loss to follow-up*). Ya hemos introducido algunos de estos conceptos en la ficha 1.8 y, dada su importancia, seguiremos trabajándolos en el resto del libro, especialmente el de validez, sus amenazas y las técnicas de control (véase la ficha 3.3).

Algunos de los instrumentos que cubren estas dimensiones son los de Berra, Elorza-Estrada y Sánchez (2008), Downs y Black (1998) y Jarde, Losilla, Vives y Rodrigo (2013).

Tanto las guías de lectura como los instrumentos de evaluación de la calidad son muy importantes para el desarrollo de una especialidad científica. Hay que tener en cuenta que la gran cantidad de información disponible y el ritmo cada vez más alto al que se publica información científica ponen de relieve la necesidad de la síntesis de investigación (*research synthesis*) a través de revisiones sistemáticas y metanálisis. La calidad de la síntesis de investigación depende de la calidad de los artículos primarios

en los que se basa. Si la calidad metodológica de los artículos primarios es mala, las revisiones sistemáticas y metanálisis resultantes también serán de baja calidad y, por tanto, las conclusiones que se deriven pueden estar gravemente sesgadas. Hay que recordar que la síntesis de investigación es un elemento fundamental de la práctica basada en la evidencia (véase la ficha 1.1), de manera que cualquier decisión que se tome depende principalmente de la calidad metodológica de estas síntesis.

No podemos finalizar este apartado sin hacer mención de la Colaboración Cochrane que, por un lado, contiene la base de datos de revisiones sistemáticas de intervenciones, en especial las basadas en experimentos (lo que en epidemiología se denomina RCT [*randomized controlled trial*]) y, por otro, proporciona un manual (Higgins, Green y Cochrane Collaboration, 2008) sobre cómo llevar a cabo revisiones sistemáticas. Si bien este manual contiene información sobre la valoración de la calidad de los estudios primarios, no se trata de una herramienta para la valoración de la calidad metodológica ni tampoco de una guía de publicación.

En definitiva, tanto las guías de publicación como las herramientas de valoración de la calidad metodológica pretenden contribuir a mejorar la calidad de la ciencia facilitando la replicación, la autocorrección, la síntesis de investigación y, por extensión, la práctica basada en la evidencia.

ACTIVIDADES

Ejemplos de confusión

A continuación, presentamos unos ejemplos basados en un trabajo previo (Domènech, Granero, Losilla y Portell, 1997) en el que se ilustra el concepto de variable de confusión. Estos ejemplos se enmarcan en el contexto de una investigación dirigida a estudiar si la sobreprotección de la madre (variable de exposición o VE) es un factor de riesgo del trastorno por mutismo selectivo (variable de respuesta o VR) en población pediátrica. Los datos (simulados) se han obtenido de acuerdo con un diseño etiológico de cohortes (véase la ficha 4.4). Concretamente, supondremos que se han seleccionado dos grupos de niños (con tamaño $n = 500$ cada uno) según si sus madres presentan o no conductas sobreprotectoras, y que se han seguido hasta la edad escolar registrando cuántos de ellos presentan mutismo selectivo.

Ejemplo 1. Relación entre sobreprotección de la madre y el trastorno por mutismo selectivo (ME)

La figura A1 presenta los datos hipotéticos obtenidos en un estudio para medir la relación entre la sobreprotección de la madre y el mutismo selectivo. En este caso, también se ha registrado la presencia o ausencia de trastornos del habla (por ejemplo, el tartamudeo), porque en la literatura especializada se presenta como un factor de riesgo del mutismo selectivo.

A partir de la primera tabla, se observa que en la muestra expuesta a la sobreprotección materna la proporción de niños con mutismo selectivo es superior ($30/500 = 0,06$) a la proporción observada entre los que no tienen madres sobreprotectoras ($20/500 = 0,04$).

Para medir la intensidad de la relación entre las dos variables que estamos estudiando, se pueden calcular diferentes índices, entre los que está la diferencia de proporciones y la razón de proporciones. Estos índices se conocen como medidas de riesgo. En este punto, únicamente haremos un uso intuitivo de la diferencia de proporciones para ilustrar el concepto de confusión. En la parte inferior de la figura A1 aparecen los valores de las correspondientes diferencias expresadas en términos de porcentajes para facilitar su interpretación.

Los resultados del análisis ponen de manifiesto que en la muestra de niños sobreprotegidos por la madre hay un 2 % más de casos con mutismo selectivo que en la muestra de los que no están sobreprotegidos.

La tabla de la derecha de la figura A1 presenta los resultados de la investigación separando los niños de la muestra en función de si presentan o no presentan trastornos del habla. Obsérvese que los resultados que se obtienen por esta vía coinciden totalmente con los de la tabla global. Así pues, en este caso los trastornos del habla no actúan como variable de confusión.

		SOBREPROTECCIÓN				Sin trastornos del habla		Con trastornos del habla	
		No	Sí			No	Sí	No	Sí
Con ME		20 (4 %)	30 (6 %)	→	12 (4 %)	18 (6 %)	8 (4 %)	12 (6 %)	
Sin ME		480 (96 %)	470 (94 %)		288 (96 %)	282 (94 %)	192 (96 %)	188 (94 %)	
Total		500 (100 %)	500 (100 %)		300 (100 %)	300 (100 %)	200 (100 %)	200 (100 %)	
		↓				↓		↓	
		d = 6 % - 4 % = 2 %				d = 6 % - 4 % = 2 %		d = 6 % - 4 % = 2 %	

Figura A1. La variable «trastornos del habla» no produce confusión.

Ejemplo 2. Confusión que incrementa el efecto

La figura A2 presenta un segundo conjunto de datos hipotéticos. Como se puede ver en la tabla de la izquierda, entre los niños sobreprotegidos por la madre la proporción de los que presentan mutismo electivo ($40/500 = 0,05$) es superior a la proporción observada entre los niños que no están expuestos a la sobreprotección materna ($10/500 = 0,02$).

Supongamos que una diferencia entre ambas proporciones igual al 6 % nos lleva a concluir que existe una relación estadísticamente significativa entre las variables. A partir de estos resultados, ¿podemos suponer que la sobreprotección materna es un factor de riesgo del trastorno por mutismo electivo?

Antes de responder a la pregunta anterior, debe comprobarse la ausencia de variables de confusión que puedan provocar una relación espuria. Como ejemplo, en este caso, estudiaremos el papel que tiene el trastorno del habla como potencial confundidora.

	SOBREPROTECCIÓN		Sin trastornos del habla		Con trastornos del habla	
	No	Sí	No	Sí	No	Sí
Con ME	10 (2 %)	40 (8 %)	8 (1,7 %)	2 (1,7 %)	2 (1 %)	38 (1 %)
Sin ME	490 (98 %)	460 (92 %)	472 (98,3 %)	118 (98,3 %)	18 (90 %)	342 (90 %)
Total	500 (100 %)	500 (100 %)	480 (100 %)	120 (100 %)	20 (100 %)	380 (100 %)
	↓		↓		↓	
	d = 8 % - 2 % = 6 % (medida sesgada)		d = 1,7 % - 1,7 % = 0 %		d = 1 % - 1 % = 0 %	

Figura A2. La variable «trastornos del habla» produce confusión incrementando el efecto.

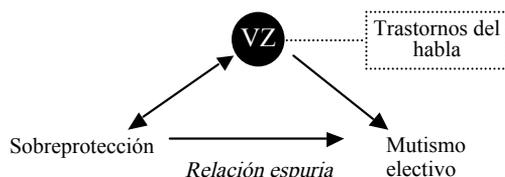
Las tablas situadas a la derecha de la figura A2 presentan los resultados anteriores después de haber fijado los trastornos del habla de los niños de la muestra. Si se examinan estas tablas nos daremos cuenta de que no hay diferencias entre la proporción de niños que presentan mutismo electivo en función de la sobreprotección de la madre.

¿Estos resultados contradicen los obtenidos a partir del análisis de la tabla de la izquierda de la figura A2? ¿Cómo es posible llegar a conclusiones opuestas si el estudio se ha realizado con la misma muestra de sujetos?

Para contestar a estas preguntas, hay que volver a examinar con detalle las tablas de la figura A2. Al observar estas tablas, se comprobará que los grupos de niños sobreprotegidos y no sobreprotegidos están desajustados en cuanto a los trastornos del habla. En efecto, las variables sobreprotección de la madre y trastornos del habla están estrechamente relacionadas: la mayor parte (380) de los 500 niños con madres sobreprotectoras presentan trastornos del habla, mientras que la mayor parte (480) de los 500 sujetos que no tienen madres sobreprotectoras no presentan trastornos del habla.

Además, el trastorno del habla está relacionado con el trastorno por mutismo selectivo, porque en el grupo de los que presentan trastorno del habla hay un 10 % (40/400) de casos que presentan mutismo selectivo, frente al 1,7 % (10/600) que se observa en el grupo sin trastorno del habla.

Así pues, la $d = 6\%$ refleja una relación espuria entre la sobreprotección de la madre y el trastorno por mutismo selectivo a través del trastorno del habla, porque este último está relacionado tanto con la sobreprotección de la madre como con el mutismo selectivo:



Ejemplo 3. Confusión que enmascara el efecto

La figura A3 presenta un tercer conjunto de resultados hipotéticos del estudio sobre la relación entre la sobreprotección materna y el mutismo selectivo; en este caso también se ha registrado si el niño ha sufrido un trauma antes de los tres años, por ser un factor de riesgo conocido del mutismo selectivo.

		Sin trauma antes de los 3 años		Con trauma antes de los 3 años	
SOBREPROTECCIÓN		SOBREPROTECCIÓN		SOBREPROTECCIÓN	
		No	Sí	No	Sí
Con ME	40 (8%)	2 (2,4%)	20 (5,3%)	38 (9,1%)	20 (16,7%)
Sin ME	460 (92%)	80 (97,6%)	360 (94,7%)	360 (90,9%)	100 (83,3%)
Total	500 (100%)	82 (100%)	380 (100%)	418 (100%)	120 (100%)
↓		↓		↓	
$d = 8\% - 8\% = 0\%$ (medida sesgada)		$d = 5,3\% - 2,4\% = 2,9\%$		$d = 16,7\% - 9,1\% = 7,6\%$	

Figura A3. La variable «trauma antes de los tres años» produce confusión enmascarando el efecto.

La primera tabla pone de manifiesto una ausencia de relación entre la sobreprotección de la madre y el mutismo selectivo ($d = 0$). Sin embargo, cuando se estudia la relación dentro de los dos estratos definidos por el hecho de haber sufrido un trauma antes de los tres años, se observa una relación entre la sobreprotección de la madre y el mu-

tismo electivo ($d = 2,9 \%$; $d = 7,6 \%$). En este caso, el factor de riesgo de haber sufrido un trauma antes de los tres años produce una confusión que enmascara la relación.

El efecto de enmascaramiento se produce porque en el grupo de los 500 niños no expuestos a la sobreprotección materna hay más sujetos del grupo que ha sufrido un trauma (418 frente a 82), mientras que en el grupo de los 500 niños que tienen una madre sobreprotectora hay más niños del grupo que no ha sufrido un trauma (380 frente a 120).

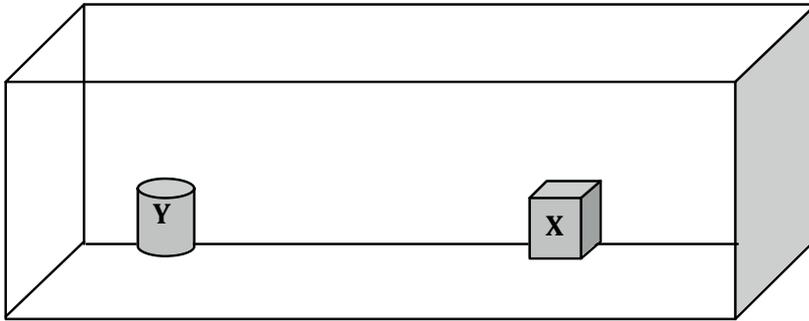
EJERCICIOS

Ejercicio 1.1. La ciencia... y la inferencia

Cuando una rata se sitúa delante del objeto X corre hacia el objeto Y (véase la figura siguiente). El estudiante A dice que la rata «tiene miedo de X»; el estudiante B considera que a la rata «le gusta Y»; el estudiante C considera que en la determinación del comportamiento operan diferentes factores simultáneamente y, en este sentido, considera que la rata «tiene miedo de X y le gusta Y»; finalmente, el estudiante D sugiere que los estudiantes A, B y C hacen valoraciones antropomorfistas y que la única cosa que se puede decir ante esta observación es que la rata «va de X hacia Y».

Asumiendo que X e Y son objetos pequeños y fáciles de mover, ¿cómo podría determinar cuál de los estudiantes tiene «más razón»?

¿Qué inferencias y definiciones operativas están implicadas en la contrastación de las hipótesis de los estudiantes?



Ejercicio 1.2. Clasificación de variables (I)

Operativice cada una de las siguientes variables y clasifíquelas en función del criterio de medida:

<i>Variables</i>	<i>Operativitzación</i>	<i>Clas. C. de medida</i>
Estado civil		
Calidad de la enseñanza		
Tiempo de reacción		
Nacionalidad		
Número de hijos		
Nivel de estudios		
Temperatura		
Tiempo invertido en la discriminación de estímulos auditivos		
Color preferido		
Consumo de tabaco		

<i>Variables</i>	<i>Operativitzación</i>	<i>Clas. C. de medida</i>
Tipo de neurosis Sexo Peso Fumar Nota en un examen CI (cociente de inteligencia) Ansiedad		

Ejercicio 1.3. Clasificación de variables (II)

A continuación, presentamos un conjunto de hipótesis, problemas y resúmenes de investigaciones. En cada caso ha de identificar las variables independientes (VI) y dependientes (VD), y en el caso de las VI, indicar si son manipulables.

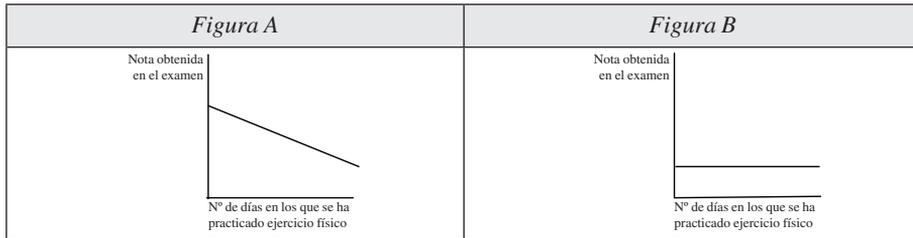
1. La enseñanza asistida por ordenador (EAO) mejora el rendimiento en matemáticas.
2. La mayoría de universitarios catalanes hablan más de dos idiomas.
3. A la semana del nacimiento, los bebés reconocen la voz de su madre.
4. En una situación de examen, el formato de respuesta (abierto o tipo test) influye sobre la ansiedad.
5. La prevalencia del tabaquismo es similar en hombres y en mujeres.
6. La percepción de los riesgos asociados al tabaquismo está relacionada con el consumo de tabaco.
7. En los niños de dos años, el juego en paralelo es más característico que el juego en grupo.
8. La ansiedad influye sobre el rendimiento en una situación de examen.
9. La posición de los productos en las estanterías de los hipermercados influye en la decisión de compra.
10. Un psicólogo clínico asegura que los refuerzos sociales, como la alabanza, tienden a hacer aumentar la autoconfianza del paciente. Así pues, alaba a su paciente cada vez que le dice algo positivo sobre él mismo y no le presta atención cuando dice algo negativo.
11. Un investigador asume que los niños que han recibido un entreno perceptual-motor tendrán una ejecución más buena en tareas de coordinación mano-ojo que los niños que no han recibido este entreno.

Ejercicio 1.4. Ejercicio físico y memoria

Diferentes estudios establecen que practicar ejercicio después de haber estudiado es la mejor táctica para reforzar la memoria. Las figuras siguientes relacionan el número de días en que un grupo de estudiantes realizan ejercicio físico 4 horas después de estudiar para un examen y la nota obtenida en el examen.

Tomando en consideración la información de la Ficha 1.6, interprete los resultados sintetizados en la Figura A.

Tomando en consideración la información de la Ficha 1.6, interprete los resultados sintetizados en la Figura B.



Ejercicio 1.5. Variables independientes manipuladas y no manipuladas

A continuación, presentamos un conjunto de hipótesis algunas de las cuales provienen de Gambara (2002). En cada una de ellas debe: a) identificar la variable independiente (VI); b) identificar la variable dependiente (VD); c) indicar si la VI se puede manipular y ejemplificar una manera de hacerla operativa.

1. Los introvertidos y extrovertidos difieren en su capacidad para resolver problemas espaciales.
2. El refuerzo positivo mejora el aprendizaje de las ratas al recorrer un laberinto.
3. El aprendizaje de conceptos probabilísticos en adolescentes es más fácil con tareas de contenido de la vida cotidiana que con tareas de contenido abstracto.
4. En los carteles publicitarios, la posición de la marca anunciante influye sobre la atención que les prestan los potenciales compradores.
5. Las caras de sujetos de un grupo racial diferente (por ejemplo, caucásico y negro) son más difíciles de recordar que las pertenecientes al mismo grupo.
6. Los sujetos expertos en una materia (por ejemplo, historia) aprenden mejor un material si este pertenece a su dominio de conocimientos que si es ajeno al mismo.
7. Los jugadores de ajedrez expertos (jugadores profesionales) y no expertos (jugadores ocasionales) no difieren en la memorización de aspectos no significativos del juego (fichas de ajedrez distribuidas al azar en el tablero).
8. Los estímulos que están emparejados con estímulos muy diferentes (por ejemplo, «A» y «Z») se detectan más rápidamente que los que lo están con estímulos similares (por ejemplo, «A» y «a»), cuando los pares se presentan en el centro de una pantalla de ordenador.
9. La emocionalidad de los estímulos presentados influye sobre la tasa cardíaca.
10. La prevalencia del tabaquismo es similar en hombres y en mujeres.
11. Se quiere estudiar el efecto que tiene el consumo del cannabis sobre la memoria. Para evaluarlo se seleccionan 22 participantes voluntarios que presentan un consumo elevado de cannabis (documentado en la entrevista clínica y mediante analíticas semanales) y 22 participantes que no consumen. Después de 6 meses de estudio, se evalúa si los sujetos presentan déficits de memoria utilizando el test MIC (Memory Inventory MIC).

12. El ejercicio físico puede aumentar el grado de autoeficacia que tienen las personas de cara a la gestión de su tiempo medida con la prueba TR2-Autoefficacy.
13. En el año 2003 la fundación TIC salud de la Generalitat de Catalunya seleccionó una muestra aleatoria de alumnos de bachillerato de Barcelona para establecer la edad a la que se iniciaron en el hábito de tomar alcohol los fines de semana.
14. La enuresis nocturna, o emisión involuntaria de orina, es frecuente en la infancia. Mayor y colaboradores han puesto de manifiesto que una alta actividad cortical durante el sueño constituye un rasgo fisiológico subyacente a la emisión involuntaria de orina (la enuresis nocturna). Para testar su hipótesis de que una alta actividad cortical induce la emisión involuntaria de orina durante el sueño, seleccionaron un grupo de personas con alto riesgo de sufrir enuresis nocturna y registraron información sobre su actividad cerebral y, al mismo tiempo, se registró si se daba o no un episodio de emisión involuntaria de orina.
15. Unos investigadores quieren comprobar si una intervención basada en las técnicas artísticas del clown es más eficaz que la musicoterapia puede disminuir el malestar psicológico de los niños que reciben tratamiento de quimioterapia.
16. El ciberacoso o ciberbullying se ha definido como aquel tipo de acoso protagonizado por un individuo o grupo que, mediante el uso de las nuevas tecnologías de la información y la comunicación (teléfonos móviles, correo electrónico, redes sociales, blogs, páginas webs, etc.), agrede deliberadamente y reiterada a alguien que no puede defenderse fácilmente por sí mismo. Recientemente se seleccionó una muestra aleatoria de 19.063 alumnos de uno de los mayores distritos escolares de los EEUU que completaron un cuestionario sobre uso y experiencias con Internet. Se observó una relación estadísticamente significativa entre las experiencias de intimidación (ya sea como acosador o como víctima) y la ideación suicida.
17. Se hipotetiza que existe un alto riesgo de sufrir depresión mayor (Major Depressive Disorder, MDD) entre la población con acné.

Ejercicio 1.6. Práctica basada en la evidencia

Responda a las siguientes preguntas a partir de las lecturas que indicamos en cada caso:

1. ¿Qué es y qué relación existe entre lo que la APA Presidential Task Force on Evidence-Based Practice (2006) identifica con los acrónimos EST y EBPP?
Lectura: APA PRESIDENTIAL TASK FORCE ON EVIDENCE-BASED PRACTICE. 2006. Evidence-based practice in psychology. *American Psychologist*, 61, 271–85.
2. ¿Qué diferencias y similitudes hay entre la Colaboración Cochrane y la Colaboración Campbell?
Lectura: SÁNCHEZ-MECA, J.; BORUCH, R. F.; PETROSINO, A.; ROSA A. I. 2002. «La colaboración Campbell y la práctica basada en la evidencia». *Papeles del Psicólogo*, 22, 83, 44–48.
3. ¿Cómo saber si un tratamiento psicológico es eficaz según Labrador *et al.* (2003)?
Lectura: LABRADOR, F.J.; VALLEJO, M. A.; MATELLANES, M.; ECHEBURÚA, E.; BADOS, A.; FERNÁNDEZ-MONTALVO, J. 2003. *La eficacia de los trata-*

mientos psicológicos. Documento de la Sociedad Española para el Avance de la Psicología Clínica y de la Salud. Siglo XXI. Noviembre de 2002. INFO-COP, 17, 25–30.

Ejercicio 1.7. Construcción de muestras

Identifique los procedimientos empleados para seleccionar las muestras en cada uno de los siguientes estudios.

1. Un profesor de la UAB pasa una encuesta a sus alumnos (de diferentes cursos) con el objetivo de conocer la actitud de los jóvenes universitarios en relación con el tabaquismo.
2. Para conocer el comportamiento arriesgado de los adolescentes, se seleccionan al azar treinta institutos del Estado y se entrevista a sus estudiantes.
3. Se seleccionan al azar a cien estudiantes de primer curso de Psicología de la UAB para conocer su opinión sobre una conferencia organizada por la asignatura Fundamentos de Metodología.
4. Un psicólogo industrial pretende estudiar la relación entre el turno de trabajo y la productividad de los operarios en la cadena de montaje. Debido a que algunos estudios previos establecen una relación entre la productividad y el sexo, decide seleccionar al azar a cincuenta hombres y cincuenta mujeres.
5. Una compañía de seguros se propone estudiar una hipótesis sobre la propensión al accidente en el ámbito de la conducción de vehículos y selecciona al azar a cien clientes. Los archivos de la compañía revelan que un 50 % de los conductores no han tenido ningún accidente, un 35 % han tenido uno y un 15 % han tenido más de uno. De acuerdo con estos datos, se decide seleccionar al azar a cincuenta sujetos que no hayan tenido accidentes, a treinta y cinco sujetos que hayan tenido uno y a quince que hayan tenido más de uno.
6. ¿Cómo se llama el muestreo empleado en la investigación del ejercicio 2.4 («Comportamiento del consumidor»)?
7. ¿Cómo se llama el tipo de muestreo utilizado en la investigación del ejercicio 2.5 («¿Cuál es el eslogan más apropiado»)?
8. ¿Cómo se llama el tipo de muestreo utilizado en la investigación del ejercicio 2.11 («Tratamiento computarizado de la anomia»)?
9. ¿Cómo se llama el tipo de muestreo utilizado en la investigación del ejercicio 3.5 («El tratamiento del trastorno por estrés posttraumático en víctimas de agresiones sexuales»)? ¿Qué función cumplen los criterios de admisión?

Ejercicio 1.8. Referenciación (I)

Las referencias bibliográficas están formadas por un conjunto de campos cuyos valores permiten identificarlas y diferenciarlas entre sí. La American Psychological Association (APA) ha establecido una normativa que especifica cómo se deben citar estas referencias bibliográficas. Así, el estilo de citación de la APA determina, para cada tipo de referencia bibliográfica: los campos que deben aparecer, en qué orden y su

contenido, los delimitadores entre campos, los literales que identifican los subtipos de referencias y el aspecto en que debe imprimirse cada campo. De entre las ventajas que supone esta normativa, podemos destacar el hecho de que estandariza la cita bibliográfica y facilita la comunicación entre investigadores, ya que permite identificar fácilmente el tipo de material bibliográfico que se está citando.

Para poder hacer este ejercicio, debe buscar los criterios de referenciación de la APA para libros, compilaciones, capítulos de compilación y artículos. Puede recurrir a la fuente original, es decir, al manual de publicación de la APA (APA, 2010), a otros que sintetizan esta información, como Vives, Portell y Boixadós (2009), o bien a la Purdue Online Writing Lab.

A continuación presentamos una serie de referencias que contienen la información necesaria para que puedan ser citadas. Lleve a cabo la cita bibliográfica de estas referencias aplicando la normativa de la APA.

1. Libro:
Autores (nombre y apellido): Alan Bryman / *Título:* Social Research Methods / *Año de publicación:* 2012 / *Editorial:* OUP / *Ciudad editorial:* Oxford.
2. Artículo:
Autores: L. Jack, S. Hayes, J. Scharalda, B. Stetson, N. H., Jones-Jack, M. Valliere, W. R. Kirchain, C. LeBlanc / *Título:* Appraising Quantitative Research in Health Education: Guidelines for Public Health Educators / *Año de publicación:* 2010 / *Revista:* Health Promotion Practice/Volum: 11 / *Número:* 2 / *Páginas:* 161–165 / DOI: 10.1177/1524839909353023
3. Capítulo de libro o de compilación:
Autores del capítulo: D.M. Haybron / *Título del capítulo:* Philosophy and the science of subjective well-being / *Autores de la compilación:* M. Eid, R. J. Larsen / *Título de la compilación:* The science of subjective well-being / *Año de publicación:* 2008 / *Editorial:* Guilford Press / *Ciudad editorial:* Nueva York, NY / *Páginas:* 17–43.
4. Compilación:
Autores: Harris Cooper, Larry V. Hedges, Jeffrey C. Valentine / *Título:* The Handbook of Research Synthesis and Meta-Analysis / *Año de publicación:* 2009 / *Editorial:* Russell Sage Foundation / *Ciudad editorial:* Nueva York.

Ejercicio 1.9: Referenciación (II)

Identifique a qué tipo de referencia corresponde cada una de las siguientes citas bibliográficas:

1. Mazzucchelli, T. G., Kane, R. T. i Rees, C. S. (2010). Behavioral activation interventions for well-being: A meta-analysis. *The Journal of Positive Psychology*, 5(2), 105–121. doi:10.1080/17439760903569154.
2. American Psychological Association (2010). *Publication manual of the American Psychological Association* (6a ed.). Washington: Autor.
3. Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2012). *Research methods in psychology*. Nueva York, NY: McGraw-Hill.

4. Sailor, W., Dunlap, G., Sugai, G., & Horner, R. (Eds.). (2008). *Handbook of Positive Behavior Support*. Nueva York: Springer.
5. Cozby, P. C. i Bates, S. (2011). *Methods in behavioral research*. Nueva York, NY: McGraw-Hill.
6. Bunge, M. (1995). La filosofía es pertinente a la investigación científica. A F. Mora (Ed.). El problema cerebro-mente (pp. 55-71). Madrid: Alianza Editorial.

Notas

- i Dado el carácter introductorio de este texto, presentamos aquí algunos de los elementos que caracterizan el método científico, pero dejamos fuera muchos otros. Para una ampliación didáctica de los temas tratados en esta ficha recomendamos la obra de León y Montero (2015).
- ii Para ampliar información sobre revisiones sistemáticas y metanálisis, le recomendamos el libro de Cooper, Hedges y Valentine (2009).
- iii Puede profundizar en todos los aspectos de la normativa APA con su manual de publicación (APA, 2010).
- iv Para una introducción didáctica a la metodología cualitativa en el estudio del comportamiento puede consultar el manual de Riba (2007) y el de León y Montero (2015).
- v Para una introducción a la metodología observacional recomendamos Anguera (2003, 2010) y para una síntesis puede consultar Anguera, Blanco, Losada y Portell (2018). Para profundizar en el tipo de combinación cuantitativa y cualitativa que se produce en la metodología observacional puede consultar Sánchez-Algarra y Anguera (2013) y Anguera, Camerino, Castañer, Sánchez-Algarra & Onwuegbuzie (2017).
- vi Para profundizar en los métodos mixtos puede consultar Creswell (2009), Creswell y Plano-Clark (2007) y Teddlie y Tashakkori (2008, 2010). Desde el punto de vista de la integración entre lo cualitativo y lo cuantitativo vinculada a la observación sistemática, recomendamos los trabajos sobre métodos mixtos de Camerino, Castañer y Anguera (2012), Castañer, Camerino y Anguera (2013).
- vii Puede profundizar en estos conceptos en los textos de Anguera (1989) y de Mayor (1989).
- viii Puede ampliar este concepto de causalidad en Ato (1995) y Moreno (1989).
- ix La noción de asociación entre variables se introduce de manera clara en el manual de estadística descriptiva de Solanas, Salafranca, Fauquet y Nuñez (2005). Para una introducción al concepto de significación estadística puede consultar Pardo y San Martín (2010).
- x Para una ampliación de los criterios de clasificación de los métodos y diseños que aquí se proponen se puede consultar Moreno, Martínez y Chacón (2000) y Chacón, Sanduvete, Portell y Anguera (2013). Otros autores proponen otros criterios para clasificar los diseños. Véase, por ejemplo, Ato, López y Benavente (2013) para una propuesta de clasificación basada en los objetivos que persigue la investigación.
- xi Para profundizar en los conceptos que se exponen en esta ficha puede consultar Lohr (2010), Martínez-Arias (1995) y Silva (2000).

2 Diseños experimentales

PRESENTACIÓN

Si le preguntamos qué es un «experimento científico», muy probablemente nos dará una respuesta, dado que «experimento» es un término de uso común. Lo cierto es, sin embargo, que la popularización de algunos términos científicos conlleva, a menudo, una excesiva simplificación o, incluso, una tergiversación de su significado. El término experimento es un ejemplo de ello cuando se utiliza como sinónimo de investigación científica o bien cuando se confunde con una investigación llevada a cabo en un laboratorio. En este capítulo, veremos que la experimentación es un tipo de investigación y que el hecho de que se lleve a cabo en un laboratorio no es una característica definitoria ni distintiva de un experimento. Así, por ejemplo, se puede estudiar experimentalmente el efecto que tiene una estimulación magnética eléctrica transcraneal de la corteza temporal sobre el recuerdo, y también se puede usar un diseño experimental para evaluar el efecto de una formación en gestión del estrés laboral entre los trabajadores de una empresa.

Presentamos los elementos de un experimento idealⁱ a partir de un ejemplo. Suponga que se diseña un experimento para estudiar el efecto de un nuevo fármaco sobre la fijación de la atención en niños con TDAH (trastorno por déficit de atención con hiperactividad). La definición de la variable independiente (VI) requiere delimitar, como mínimo, dos condiciones experimentales. En este ejemplo, en una se administra el fármaco y en la otra un placebo (una sustancia muy similar pero sin el principio activo). Una muestra de niños con TDAH se asigna al azar a dos grupos: un grupo recibe el fármaco y el otro, el placebo (en este segundo caso, hablamos de un grupo control por placebo). Una vez efectuada esta manipulación, a todos los niños se les mide la fijación de la atención (la variable dependiente, VD) mediante un procedimiento válido y fiable. Si este fuera un experimento ideal, se esperaría que durante todo el estudio la única diferencia entre los dos grupos fuera el hecho de recibir o no el principio activo; así, si se mantiene la comparabilidad durante todo el transcurso de la investigación, en el caso de observar cambios en la fijación de la atención, es válido atribuirlos al efecto del nuevo fármaco. La asignación al azar y el control por placebo forman parte de un conjunto de técnicas de control dirigidas a que los dos grupos sean lo más equivalentes u homogéneos posible en cuanto a variables extrañas (véase la ficha 2.3). Como veremos a continuación, la asignación al azar juega un papel fundamental en los diseños experimentales.

Las características que delimitan los diseños experimentales y permiten diferenciarlos de otras familias de diseños, como los cuasiexperimentales y los *ex post facto*, son las que se indican a continuación:

- Manipulación del VI: el investigador administra los valores de la VI a cada participante (véase la ficha 1.5 para una presentación de las variables manipuladas y no manipuladas).
- Aleatorización: asignación aleatoria (asignación al azar) de los participantes en las diferentes condiciones experimentales —si la estrategia de comparación es intersujeto— o en los distintos órdenes de presentación —si se opta por una estrategia de comparación intrasujeto. El hecho de emplear un mecanismo probadamente aleatorio para asignar los participantes a una u otra condición experimental hará posible, si la muestra es grande, que las variables extrañas queden bien equilibradas —en cuanto a variables extrañas conocidas y desconocidas— en las diferentes condiciones experimentales. Las fichas 2.2, 2.4 y 2.5 amplían los conceptos de asignación aleatoria y de estrategia de comparación.

La manipulación de la VI y la aleatorización, junto con la aplicación de técnicas de control, son procedimientos dirigidos a incrementar la validez interna. Alcanzar un alto grado de validez interna es fundamental en la contrastación de hipótesis causales, que son el objetivo principal cuando se aplican diseños experimentales.

A partir del estudio de las siguientes fichas se espera que el lector sea capaz de:

- Saber clasificar los diseños experimentales (ficha 2.1).
- Entender y saber diferenciar entre selección y asignación de la muestra (ficha 2.2).
- Conocer las técnicas de control dirigidas a mejorar la validez interna (ficha 2.3).
- Ser capaz de representar un diseño experimental esquemáticamente (fichas 2.4 y 2.7).
- Conocer las principales características de las dos estrategias de comparación (fichas 2.5 y 2.6).
- Entender el concepto de interacción (ficha 2.7).
- Conocer los diseños factoriales y sus variantes más importantes (fichas 2.7, 2.8 y 2.9).

FICHA 2.1. CLASIFICACIÓN DE LOS DISEÑOS EXPERIMENTALES

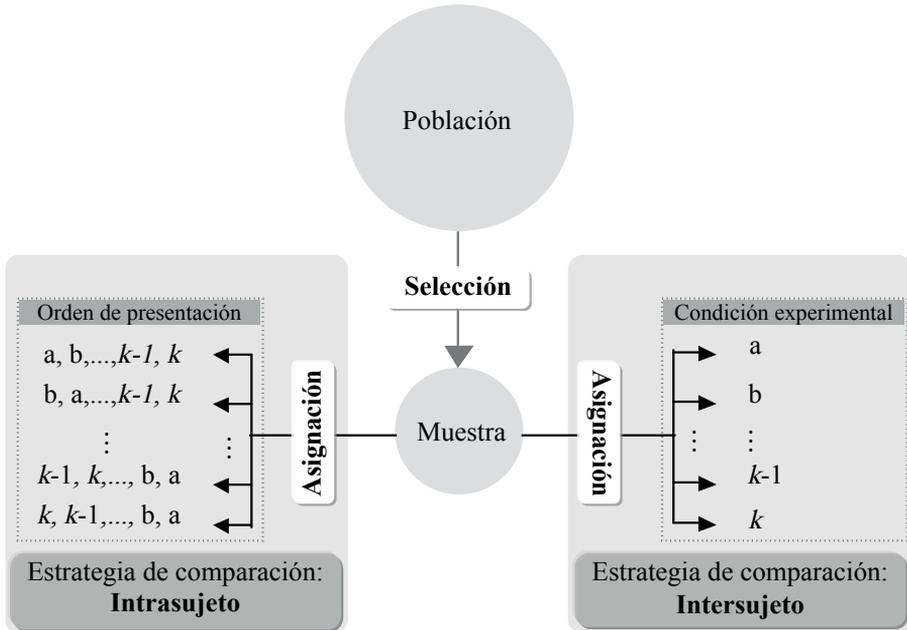
Los diseños experimentales se pueden clasificar en función del número de VD que miden (univariable y multivariable), el número de VI que incluyen (unifactoriales y factoriales, véanse las fichas 2.4 y 2.7), la estrategia de comparación que utilizan para cada VI (intersujeto, intrasujeto, mixto; véase la ficha 2.5), y el tipo de técnica de control basada en la aleatorización que aplican (asignación aleatoria sin restricciones, bloqueo y contrabalanceo; véase la ficha 2.3). La combinación de estos criterios permite identificar los diseños experimentales de uso más habitual en la investigación en ciencias de la salud.

<i>Núm. VD</i>	<i>Núm. VI</i>	<i>Estrategia de comparación</i>	<i>Tipo de aleatorización (con o sin restricciones)</i>	<i>Diseño</i>
Univariable	Unifactorial	Intersujeto	Asignación al azar (aleatorización sin restricciones)	Grupos aleatorios
			Bloqueo	Grupos aleatorios con bloques
		Intrasujeto	Sujeto como control propio · Contrabalanceo	Intrasujeto con contrabalanceo
	Factorial	Intersujeto	(...)	Factorial intersujeto (...)
		Intrasujeto	(...)	Factorial intrasujeto (...)
		Mixto	(...)	Mixto (...)
	Multivariable	(...)	(...)	(...)

(...) Puede replicar las estructuras anteriores.

FICHA 2.2. EL AZAR EN EL DISEÑO EXPERIMENTAL

Cuando se analiza una investigación experimental, es importante comprender la diferencia entre el procedimiento para construir la muestra de estudio y el procedimiento usado para distribuir los elementos que forman la muestra en las condiciones experimentales.

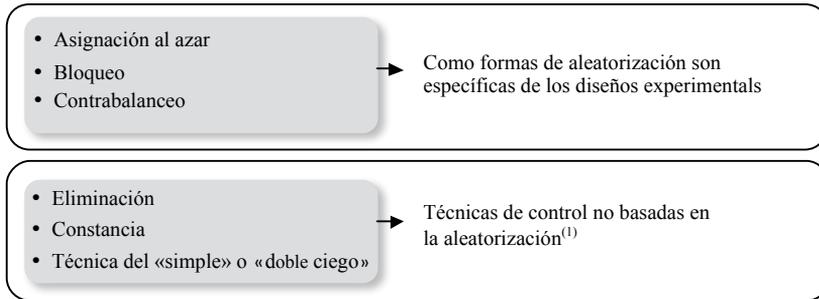


Con el término «**selección**» identificamos el proceso por el que se escoge (se selecciona) una muestra de individuos de una población. Hablamos de selección aleatoria cuando se emplea un mecanismo aleatorio para elegir los elementos de la población que formarán parte de la muestra de estudio. El hecho de que la selección sea aleatoria contribuye de forma positiva a la consecución de validez externa. En la ficha 1.9 se ha presentado el concepto «azar» en el proceso de selección de una muestra y ya se ha insistido sobre los peligros de confundir el significado del término azar, que aquí se expone, con la noción de azar como algo fortuito, casual, accidental o sin objetivo.

En los diseños experimentales intersujeto, la «**asignación**» hace referencia al mecanismo por el que los sujetos de una muestra (o participantes en el estudio) son repartidos en los diferentes grupos a los que se aplican las diferentes condiciones experimentales. En el diseño experimental intrasujeto, todos los sujetos de la muestra pasan por todas las condiciones experimentales, lo que nos podría llevar a pensar que la asignación aleatoria no tiene cabida. En este caso, la aleatorización no se aplica a las condiciones experimentales, sino a los diferentes órdenes de presentación de los niveles de la VI (en la ficha 2.5 se explica el motivo).

FICHA 2.3. TÉCNICAS DE CONTROL

En una investigación experimental se pueden distinguir dos tipos de técnicas de control: 1) las basadas en alguna forma de aleatorización y que son específicas del diseño experimental, 2) y las no basadas en aleatorización y que son aplicables tanto a los diseños experimentales como a los no experimentales.



Nota.⁽¹⁾ Dentro de este grupo de técnicas también hay que tener en cuenta las aplicaciones del apareamiento que no se combinan con asignación aleatoria.

La **asignación al azar** (en inglés *randomization* o *random assignment*) es un procedimiento de distribución aleatoria de una muestra entre los grupos que se comparan en un experimento (grupos experimentales). Se espera que, de este procedimiento, resulten grupos equivalentes respecto a fuentes de confusión tanto conocidas como desconocidas. Hay que tener en cuenta que esta expectativa de equivalencia puede verse comprometida debido a factores como un reducido tamaño de muestra en relación con el número de condiciones experimentales o una alta heterogeneidad de la población de origen. En estos casos, es recomendable imponer alguna restricción previa al proceso de aleatorización, como sucede al aplicar la técnica del bloqueo. Otro aspecto que hay que tener en cuenta es que la asignación al azar se enfoca a la creación de grupos equivalentes al inicio, pero que la validez de los resultados depende de que esta equivalencia entre los grupos se mantenga durante todo el experimento (excepto en cuanto al nivel de la VI que reciben), y esto puede suponer la aplicación complementaria de otras técnicas de control, como por ejemplo la eliminación o la constancia.

El **bloqueo** (en inglés *blocking*) es un procedimiento que consiste en formar grupos homogéneos (que denominamos bloques) de individuos en función de los valores obtenidos en una o más variables relacionadas con la VD. La aleatorización por bloques (en inglés *block randomization*) supone que, una vez creados los bloques, los sujetos de cada bloque son asignados aleatoriamente a las diferentes condiciones experimentales. El control basado en la formación de grupos homogéneos a partir del **apareamiento** de participantes (en inglés *matching*) es una técnica que también se utiliza sin aplicar asignación al azar. Una versión sencilla de esta técnica de apareamiento consiste en: 1) determinar y medir una variable o un conjunto de variables que pueden influir de manera decisiva en la VD; 2) establecer parejas de participantes que presentan valores similares a la variable o variables medidas previamente, y 3) asignar un miembro de cada pareja a cada grupo de estudio.ⁱⁱ

El **contrabalanceo** (en inglés *counterbalancing*) se aplica en los diseños intrasujeto. En estos diseños los participantes pasan por todas las condiciones experimentales. Así, podemos considerar que los diseños intrasujeto extienden la noción de bloque homogéneo a su punto más extremo: el **mismo sujeto como procedimiento de control**. En un diseño intrasujeto la asignación aleatoria se aplica a los diferentes órdenes de presentación según la técnica del contrabalanceo, para controlar los efectos de periodo (véase la ficha 2.5). El contrabalanceo es completo cuando los sujetos se asignan a todos los posibles órdenes de presentación; en caso contrario, se habla de contrabalanceo incompleto (un ejemplo es el diseño en cuadrado latino).ⁱⁱⁱ

La **eliminación** (en inglés *elimination*) es una técnica de control que supone mantener la variable que se quiere controlar en su valor nulo. Nos podemos plantear la aplicación de esta técnica con variables como, por ejemplo, el ruido (en este caso el valor nulo se correspondería con el silencio absoluto) o la luz (en este caso, la oscuridad); sin embargo, hay muchas otras para las que no tiene sentido, dado que no podemos definir un valor nulo para ellas (por ejemplo, la inteligencia o el sexo de los participantes).

Controlar una variable mediante la técnica de la **constancia** (en inglés *constancy* o *restriction*) supone que esta variable tenga el mismo nivel para todos los participantes. La constancia es una buena alternativa cuando no es posible eliminar una variable de la situación experimental.

El **simple ciego** y el **doble ciego** (en inglés *blinding*, *single-blind* y *double-blind*) consisten en ocultar información sobre características del «tratamiento» administrado para controlar fuentes de confusión, como las expectativas del participante y del experimentador (estas amenazas a la validez se definen en la ficha 3.3). Se considera simple o doble ciego en función de si el desconocimiento es solo del participante o lo es también de la persona que le está administrando el «tratamiento». Vinculada a la asignación aleatoria, existe una técnica denominada **ocultación de la asignación** (*concealment of allocation* o *allocation concealment*), que no debe confundirse con la técnica del doble ciego. La ocultación de la asignación supone que la persona que distribuye los participantes a las diferentes condiciones experimentales no conoce hasta el último momento cuál es el grupo de comparación al que se asignará cada participante, para evitar que tenga influencia en las asignaciones de los participantes en los grupos. En este caso, el sesgo que se quiere controlar es lo que en la ficha 3.3. definimos como la amenaza de selección. Así, por ejemplo, si partimos de una muestra y la persona encargada de planificar la asignación no conoce a qué grupo es asignado cada participante y, además, la persona encargada de aplicar la intervención no sabe si está aplicando la intervención activa o un placebo y el participante tampoco sabe si está recibiendo o no la intervención activa, en este caso estarían presentes tanto la ocultación de la asignación como el doble ciego. Si bien el ejemplo ilustra un caso de aplicación conjunta, interesa que se note que se pueden aplicar por separado.

FICHA 2.4. DISEÑOS UNIFACTORIALES

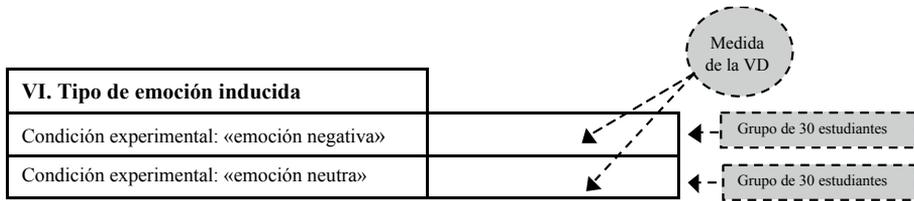
Los diseños unifactoriales intersujeto son diseños experimentales caracterizados por la presencia de **una sola VI** y por el uso de un procedimiento probadamente aleatorio (asignación al azar sin restricciones o con restricciones, como ocurre con el bloqueo) para asignar los sujetos solo a una de las posibles condiciones experimentales. Así, se comparan los valores de la VD obtenidos en grupos de sujetos diferentes, y por eso decimos que la estrategia de comparación es **intersujeto** (entre diferentes sujetos, en inglés, *between-subject*).

Los diseños experimentales unifactoriales **intrasujeto** (en inglés, *within-subject*) se caracterizan por la presencia de una sola VI y por el uso de un procedimiento probadamente aleatorio (por ejemplo, contrabalanceo completo) para asignar los sujetos a uno de los posibles órdenes de presentación de las condiciones experimentales. Así, todos los participantes pasan por todas las condiciones experimentales, y se comparan los valores de la VD obtenidos en el mismo grupo de sujetos en momentos diferentes; por eso decimos que la estrategia de comparación es intrasujeto.

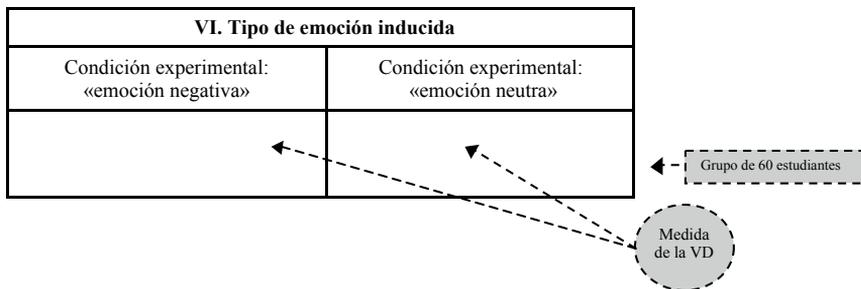
Ilustraremos la estructura de los dos diseños a partir de un ejemplo enmarcado en las investigaciones sobre el efecto de las emociones inducidas en el recuerdo.¹ Suponga que se pretende contrastar la hipótesis de que una emoción negativa sesga el recuerdo hacia las palabras negativas. Para contrastar experimentalmente esta hipótesis, se planifica una manipulación de inducción emocional basada en cortos de películas validados en un estudio previo. Para inducir la emoción negativa se usa *Cry Freedom* [«Grita libertad»] y para inducir una emoción neutra usa *Sticks*. Como material de aprendizaje y posterior recuerdo, se utilizan presentaciones en PowerPoint en las que aparecen 22 palabras de tres sílabas, de las cuales 6 son positivas (por ejemplo, lealtad), 6 negativas (por ejemplo, injuria) y el resto neutros (por ejemplo, materia). El procedimiento experimental implica tres fases: 1) visualización de la presentación en PowerPoint; 2) visualización del corto de película, y 3) prueba de recuerdo de las palabras de la presentación en PowerPoint. Supondremos que la muestra de este estudio está formada por 60 estudiantes.

La siguiente tabla representa la estructura de un diseño unifactorial **intersujeto**. En este caso, la muestra de 60 estudiantes se asigna aleatoriamente a dos grupos (30 en cada uno). Los dos grupos de participantes los representamos en filas diferentes de la tabla; un grupo se expone a la condición experimental «emoción negativa» y el otro, a la condición «emoción neutra». La columna de la derecha la reservamos para representar al final del estudio los valores obtenidos en la VD (por ejemplo, media y desviación estándar de palabras recordadas). Para medir el efecto de la emoción negativa, se puede comparar el recuerdo medio de los participantes en ambos grupos. La comparación puede basarse también en otros índices estadísticos alternativos a la media,^{iv} pero, en cualquier caso, lo que caracteriza a este diseño es que la medida de efecto se basa en comparar respuestas de grupos de sujetos diferentes (estrategia de comparación intersujeto).

1. Puede encontrar una investigación real sobre el efecto de las emociones en el recuerdo en el siguiente artículo: BENEYTO, V. B.; GARCÍA, E. 2012. ¿Es la memoria del optimista menos influenciada por las emociones negativas? *Psicothema*, 24 (2), 199–204.



La siguiente tabla representa la estructura de un diseño unifactorial **intrasujeto**. En este caso, los 60 participantes pasan por las dos condiciones experimentales, en dos sesiones separadas por una semana, y al acabar la proyección de cada corto se mide la VD, esto es, realizan la prueba de recuerdo (con listas de palabras no idénticas pero equivalentes).



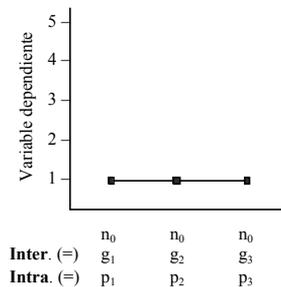
El hecho de que toda la muestra pasa por las dos condiciones experimentales lo representamos ubicando las categorías de la VI en columnas diferentes dentro de la misma fila. Aquí también se aplica una forma de aleatorización, denominada contrabalanceo, de modo que cada uno de los 60 estudiantes es asignado al azar a uno de los dos posibles órdenes de presentación de las condiciones experimentales. Así, la mitad de la muestra visualiza en primer lugar el corto *Cry Freedom* y en segundo lugar *Sticks*, y la otra mitad visualiza los cortos en orden inverso. En este diseño, la comparación para medir el efecto de la emoción negativa involucrará la respuesta de los propios participantes en dos momentos distintos (estrategia de comparación intrasujeto). Es importante darse cuenta de que aquí disponemos de la respuesta de 60 personas en cada condición experimental, mientras que, a igual tamaño de muestra, con el diseño intersujeto presentado previamente solo disponemos de la respuesta de 30 personas en cada condición experimental.

Por último, queremos comentar un aspecto que es aplicable a ambos diseños. Es habitual que, aparte de aplicar alguna forma de aleatorización como las expuestas, también se apliquen otras técnicas de control mostradas en la ficha 2.3. Por ejemplo, en este caso se ha mantenido constante el número de sílabas de las palabras y también podemos suponer que se han mantenido constantes variables como las condiciones de la sala o la edad de los participantes, que se sabe que pueden influir en el recuerdo. El control de todas aquellas variables que pueden confundir el efecto de la manipulación experimental, es decir, el efecto de la VI, es un aspecto fundamental en la investigación experimental, dado que su principal objetivo es aportar conclusiones válidas sobre la relación causal entre la VI y la VD.

FICHA 2.5. ESTRATEGIAS DE COMPARACIÓN

Supongamos que tenemos una VI con tres niveles (n_0, n_1, n_2) y se quiere valorar la eficacia de los niveles n_1 y n_2 respecto a la condición control (nivel 0: n_0) que supone la ausencia de tratamiento.

Si el estudio se hace a partir de un diseño **intersujeto**, dispondremos de tres grupos de sujetos diferentes (uno que recibirá el nivel 0, otro el nivel 1 y otro el nivel 2 de la VI). El objetivo de las técnicas de control (ficha 2.3) será que los tres grupos sean equivalentes; es decir, que sean lo más homogéneos posible en relación con las potenciales variables de confusión, tanto conocidas como desconocidas. Si todas las variables extrañas se encuentran perfectamente equilibradas en los tres grupos y les administramos el mismo tratamiento (por ejemplo, el nivel 0), ¿qué esperamos encontrar? Pues los resultados esperados serían los mismos valores de la VD en los tres grupos (es decir, una línea paralela al eje de abscisas tal como muestra la figura).



- Intersujeto** ▾ Niveles de la VI (n_0, n_1, n_2)
Grupos de sujetos «equivalentes» (g_1, g_2, g_3)
- Intrasujeto** ▾ Niveles de la VI (n_0, n_1, n_2)
Periodos «equivalentes» (p_1, p_2, p_3)

Si el estudio se hace a partir de un diseño **intrasujeto**, entonces la homogeneidad entre los sujetos a quienes se aplica los diferentes tratamientos está garantizada, ya que son los mismos. En este sentido, decimos que en un diseño intrasujeto el mismo sujeto actúa como procedimiento de control. Aquí la problemática de la comparabilidad se desplaza hacia otro punto: la homogeneidad entre los períodos u ocasiones de presentación y de respuesta. El objetivo de las técnicas de control será incrementar la comparabilidad entre las ocasiones en que se mide el efecto de cada nivel de la VI. Si a un mismo grupo de sujetos le administramos el mismo tratamiento en tres períodos diferentes (por ejemplo, el nivel 0) y estos tres momentos son equivalentes (el estado de estos individuos es comparable, la situación es comparable, etc.), ¿qué esperamos encontrar? Pues los resultados deberían mostrar los mismos valores de la VD en los tres períodos (es decir, una línea paralela al eje de abscisas tal como muestra la figura).

Así, para analizar la calidad de un diseño intersujeto, hay que prestar atención a todo lo que puede amenazar la equivalencia (la comparabilidad) de los grupos de sujetos que se comparan. En cambio, cuando se analiza un diseño intrasujeto, hay que examinar los factores que pueden comprometer la equivalencia entre los períodos.

FICHA 2.6. VENTAJAS E INCONVENIENTES DE LA ESTRATEGIA DE COMPARACIÓN INTRASUJETO

La comparación a nivel intrasujeto supone ventajas considerables en relación con la comparación intersujeto, pero también comporta importantes inconvenientes.

<p>Ventajas:</p> <ul style="list-style-type: none"> □ Sujeto como control propio □ Economía (sujetos y procedimiento) 	<p>Inconvenientes:</p> <ul style="list-style-type: none"> □ Efecto de persistencia □ Efecto de período (práctica, cansancio, motivación, etc.)
--	---

En cuanto a las ventajas, cabe destacar que el diseño intrasujeto es un excelente instrumento para sustraer de la comparación cualquier confusor vinculado a las diferencias individuales. Podemos considerar que este diseño lleva al extremo la noción de restricciones a la aleatorización que hemos expuesto al referirnos al bloqueo (ficha 2.3), utilizando el mismo sujeto como procedimiento de control. Sobre esta base, el diseño intrasujeto se convierte en una estrategia que requiere una muestra mucho menor que un diseño intersujeto para conseguir la misma potencia en la detección de diferencias entre tratamientos. Además, también se considera un diseño más económico desde el punto de vista de la misma situación experimental.

Los inconvenientes de la estrategia de comparación intrasujeto se pueden agrupar en efectos de período y de persistencia. El **efecto de período** se refiere al cambio en la respuesta atribuible a factores ajenos a la manipulación experimental que covarían con la secuencia en que se mide esta respuesta. Entre estos factores, podemos mencionar el aprendizaje de la tarea asociada a la medida de la VD, la fatiga que supone repetir esta tarea, la motivación con la que una persona participa en diferentes sesiones de un experimento o el efecto que tiene la simple práctica o repetición de una tarea. Así, si el aprendizaje, la fatiga, la motivación y/o la práctica varían a lo largo de los diferentes periodos en los que se aplican los diferentes niveles de la VI y/o entre las diferentes ocasiones en que se mide la VD, entonces los diferentes periodos no son equivalentes, y se puede confundir el efecto. Para controlar el efecto de período se usa la técnica del contrabalanceo (véase la ficha 2.3).

El **efecto de persistencia** (también denominado residual o *carry-over*) se refiere a la prolongación del efecto de un nivel de la VI más allá del periodo de su presentación, por lo que aún no se ha extinguido cuando se aplica el siguiente nivel de la VI, con lo cual se confunden sus efectos. La estrategia más adecuada para evitar estos problemas exige la introducción de periodos de reposo entre la aplicación de los distintos niveles de la VI; sin embargo, hay ocasiones en que esto es insuficiente y, en estos casos, se puede recurrir a otros diseños más avanzados que permiten valorar la hipótesis de igualdad de los efectos de persistencia.^v En los casos en que el efecto de un nivel de la VI no se pueda extinguir (por ejemplo, ciertos aprendizajes o el efecto de una cirugía radical), entonces no será posible estudiarlo de manera intrasujeto.

FICHA 2.7. DISEÑOS FACTORIALES

Los diseños factoriales incluyen más de una VI o factor, y al menos una de estas variables o factores es manipulada y aleatorizada.^{vi} Los diseños factoriales presentan diferentes ventajas respecto a los unifactoriales.^{vii} La ventaja más importante es que, además de permitir estudiar el efecto que ejerce cada factor por separado sobre la VD, también permite estudiar el efecto que tiene sobre la VD la interacción entre estos factores. Una segunda ventaja es de ahorro de tiempo y de participantes, ya que para conseguir la misma información que se obtiene con un diseño factorial hay que aplicar varios diseños unifactoriales. Una tercera ventaja tiene que ver con que la conducta raramente está influenciada por una sola variable; así, la característica del diseño factorial de incluir simultáneamente varias VI lo hace más capaz de abordar modelos más completos y, en consecuencia, podemos suponer que permitirá alcanzar más validez.^{viii}

Ilustraremos los principales elementos del diseño factorial a partir de un ejemplo del campo de la geriatría, ámbito en el que se va incrementando el interés por establecer la eficacia de tratamientos de tipo no farmacológico.² Hay estudios que muestran la eficacia del entrenamiento cognitivo para mejorar la ejecución de diferentes tareas en adultos sanos. Asimismo, se ha demostrado que el ejercicio aeróbico puede mejorar la memoria, la atención y el tiempo de reacción. Partiendo de esta base, suponga que queremos contrastar experimentalmente las siguientes hipótesis planteadas respecto a adultos de más de 69 años que no presentan deterioro cognitivo: 1) el entrenamiento cognitivo incrementa su función cognitiva; 2) un ejercicio físico aeróbico basado en entrenamiento de resistencia progresiva (ERP) incrementa la función cognitiva, y 3) si el entrenamiento cognitivo se acompaña del ejercicio físico, entonces el incremento en la función cognitiva será superior que si se aplica solo el entrenamiento cognitivo.

Es posible contrastar las hipótesis 1 y 2 usando sendos diseños unifactoriales. En el primer diseño la VI sería el «entrenamiento cognitivo (activo / placebo)», y en el segundo diseño la VI sería el «ejercicio físico» (ERP / placebo). Sin embargo, si se aplican estos dos diseños nos faltará información para valorar la tercera hipótesis. A continuación veremos cómo un único diseño de tipo factorial permite contrastar las tres hipótesis.

Un diseño factorial se define combinando los niveles de las VI que lo integran. En este caso, el diseño involucra dos VI con dos niveles cada una (diseño factorial simbolizado con «2 x 2») tal como se expone en la ficha 2.9) y su combinación dará lugar a las cuatro condiciones experimentales que ilustra la figura F2.7–A.

2. Este ejemplo presenta datos simulados y se basa en una simplificación de los siguientes trabajos, donde puede encontrar información ampliada sobre esta temática:

GATES, N.J. et al. 2011. «Study of Mental Activity and Regular Training (SMART) in at risk individuals: A randomised double blind, sham controlled, longitudinal trial». *BMC Geriatrics*, 11: 19. Doi: 10.1186/1471-2318-11-19.

LEGAULT, C. et al. 2011. «Designing clinical trials for assessing the effects of cognitive training and physical activity interventions on cognitive outcomes: The Seniors Health and Activity Research Program Pilot (SHARP-P) Study, a randomized controlled trial». *BMC Geriatrics*, 11: 27. Doi: 10.1186/1471-2318-11-27.

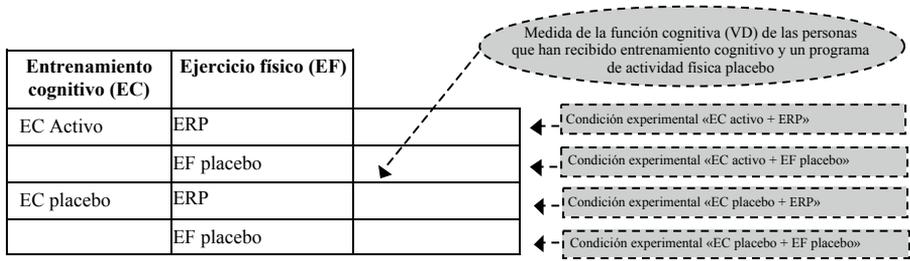


Figura F2.7-A. Condiciones experimentales

Suponga que partimos de una muestra de 100 adultos de más de 69 años que cumplen un conjunto de condiciones de inclusión en el estudio que los hacen similares respecto a potenciales variables de confusión (por ejemplo, funciones cognitivas similares, ausencia de patología orgánica o psiquiátrica conocida y probada capacidad para seguir el entrenamiento cognitivo y físico planificado). Cada participante se asigna al azar a una de las cuatro condiciones experimentales que hemos definido, de forma que en cada condición experimental hay 25 participantes. Así, por ejemplo, la casilla de la VD identificada con la flecha corresponde a los 25 participantes que durante 6 meses hacen dos sesiones semanales de 45 minutos de entrenamiento cognitivo y 45 minutos de ejercicio físico simulado (un programa que tiene la misma estructura que el programa de ERP, pero que no tiene carácter progresivo ni implica actividades que aumentan notablemente la frecuencia cardíaca, la fuerza o el equilibrio). Durante el período de estudio, todos los participantes tienen contactos de la misma frecuencia y duración con los entrenadores. El diseño es doble ciego; así, al ingresar al estudio, los participantes saben que serán asignados al azar a una de las cuatro condiciones mencionadas, pero ni ellos ni los responsables de aplicar las intervenciones no saben cuál es la intervención que se hipotetiza como más favorable.

A continuación, presentamos tres grupos de gráficos (figuras F2.7-B, F2.7-C y F2.7-D) que ilustran tres posibles resultados (simulados) del diseño factorial que estamos usando como ejemplo. Las cuatro marcas (dos cuadrados y dos círculos) de la primera gráfica de cada bloque representan las medias de la mejora de la función cognitiva de los cuatro grupos de personas expuestas a cada condición experimental. Así, por ejemplo, en la primera gráfica de la figura F2.7-B observamos que la mejora media en la función cognitiva de las 25 personas expuestas a «EC Placebo + EF Placebo» (media: 2 puntos) respecto a las expuestas a «EC Activo + EF Placebo» (media: 1 punto) es 1 punto superior (diferencia entre medias).

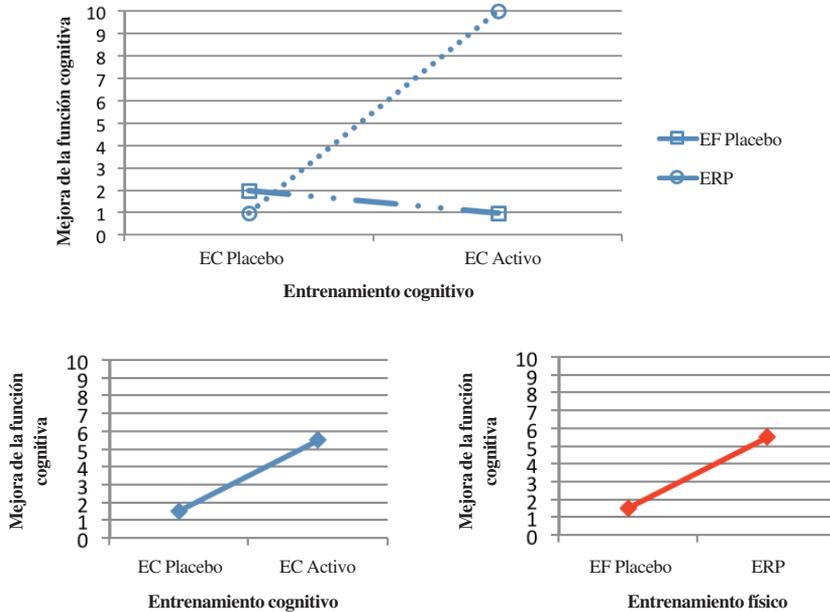


Figura F2.7-B. Interacción

Cuando se analiza un diseño factorial surgen dos preguntas: 1) ¿qué información se puede extraer del diseño?, y 2) de esta información, ¿cuál es útil para formular conclusiones del efecto de las VI sobre la VD?

1) ¿Qué información se puede extraer del diseño?

Entre la información que podemos extraer del diseño factorial debemos distinguir entre:

- Los **efectos simples**, que son los efectos de una VI (sobre la VD) fijando los niveles de las restantes VI del diseño. El primero de los tres gráficos (el que está en la parte superior) de las figuras F2.7-B, F2.7-C y F2.7-D representa los efectos simples del entrenamiento cognitivo para cada uno de los niveles de la otra VI (ejercicio físico placebo y ERP). Gráficamente, la pendiente de las dos rectas representa los dos efectos simples.^{ix} Por ejemplo, si nos fijamos en la figura F2.7-B, la línea punteada con una clara pendiente positiva representa el efecto simple del entrenamiento cognitivo cuando se administra ERP. Así, observamos que la mejora media de los participantes que han recibido ERP y que han recibido entrenamiento cognitivo es 9 puntos superior a la de los que también han recibido ERP pero combinado con EC-Placebo; la conclusión es que el entrenamiento cognitivo produce una mejora de la función cognitiva cuando se combina con ERP. Sin embargo, si examinamos el segundo efecto simple de este gráfico, las conclusiones varían, ya que la mejora media de los participantes que reciben entrenamiento cognitivo es 1 punto más baja

que la de los participantes que han recibido EC-Placebo (que es de 2 puntos), por lo que la conclusión de este segundo efecto simple es que el entrenamiento cognitivo no mejora la función cognitiva cuando no se combina con ERP. Así pues, el efecto que tiene el entrenamiento cognitivo sobre la mejora en la función cognitiva cambia según si se realiza (o no) un tipo de ejercicio físico aeróbico.

- Los **efectos principales**, que se refieren al efecto global de una VI (sobre la VD) sin tener en cuenta los niveles de las restantes VI del diseño. Los dos últimos gráficos de las figuras F2.7-B, F2.7-C y F2.7-D representan los principales efectos de las dos VI del ejemplo: el efecto principal del entrenamiento cognitivo y el efecto principal del ejercicio físico. Así, si nos fijamos en el segundo gráfico de la figura F2.7-B, observamos que los dos puntos que definen la recta son la media de la mejora en la función cognitiva del grupo EC-Placebo y del grupo EC-Activo, sin separarlos en función del tipo de programa de actividad al que han sido expuestos. Podemos calcular estos dos valores a partir de la media de las medias representadas en el gráfico previo de los efectos simples ($[(2 + 1) / 2]$ y $[(10 + 1) / 2]$). Sugerimos que dibuje esta recta sobre el gráfico de los efectos simples y observará que la recta correspondiente al efecto principal se encuentra en la media entre los puntos que definen la recta de los dos efectos simples. Una segunda actividad que le sugerimos hacer es relacionar las 8 medias que se representan en los tres gráficos de la figura F2.7-B con las 4 casillas de la tabla del diseño.

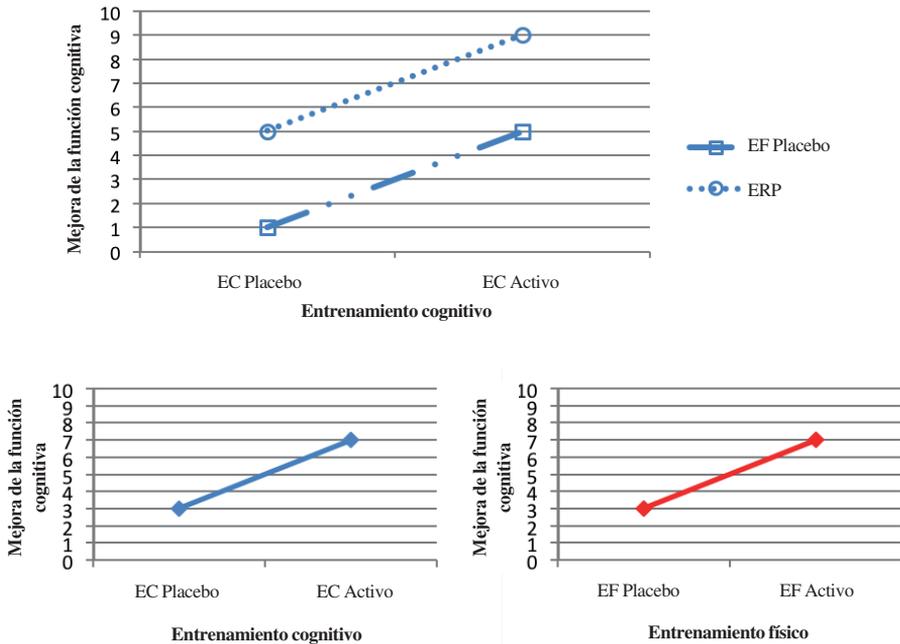


Figura F2.7-C. Ausencia de interacción.

La comparación de los efectos simples de las figuras F2.7-B y F2.7-C facilita la definición del concepto «interacción». A diferencia de lo que ocurre en la figura F2.7-B, en la F2.7-C observamos que el efecto simple del entrenamiento cognitivo cuando hay EF-Placebo es equivalente al efecto simple del entrenamiento cognitivo cuando se recibe el ERP. Puede observarse que en ambos casos el entrenamiento cognitivo produce una mejora media de la función cognitiva de 4 puntos. En el gráfico esto queda representado por dos rectas paralelas (la misma pendiente). Cuando las líneas que representan los efectos simples son paralelas, decimos que no hay interacción entre las VI. En cambio, cuando no lo son, sugiere la existencia de interacción entre las VI al actuar sobre la VD.

2) De la información que se puede extraer de un diseño factorial, ¿en qué nos basaremos para formular conclusiones?

En un diseño factorial habrá que empezar siempre analizando la interacción y, en función de si la detectamos o no, basaremos la interpretación de los resultados en los efectos simples o en los efectos principales. Tal como se puede observar en los gráficos de la figura F2.7-C, en ausencia de interacción los efectos principales son suficientes para resumir los resultados obtenidos, dado que los efectos simples no aportan información adicional. En cambio, si examina la figura F2.7-B, observará que las conclusiones basadas en los efectos principales serían falsas, dado que el efecto del entrenamiento cognitivo sobre la función cognitiva solo supone una ganancia cuando se acompaña de ERP; hay que concluir, pues, en función de los efectos simples. Finalmente, si examina la figura F2.7-D, se encontrará con un caso en el que también hay interacción (obsérvese que las líneas que representan los efectos simples no son paralelas) y que, por tanto, hay que basar la interpretación los resultados en los efectos simples; en casos como este en que las pendientes de las rectas que representan los efectos simples son del mismo signo cabe señalar que aunque el efecto principal es interpretable como un promedio, las conclusiones basadas únicamente en los efectos principales serían poco precisas.

Bloque-c.

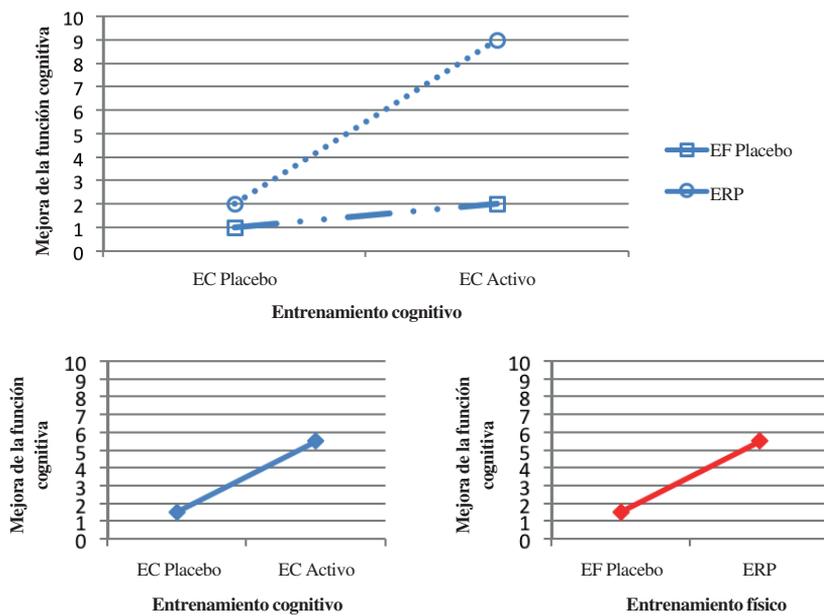


Figura F2.7-D. Interacción.

FICHA 2.8. TIPOS DE DISEÑOS FACTORIALES

En función de la estrategia de comparación, podemos distinguir tres tipos de diseños factoriales:

- Diseño factorial intersujeto: en todas las VI se aplica una estrategia de comparación intersujeto (el ejemplo presentado en la ficha 2.7 es un diseño factorial de este tipo).
- Diseño factorial intrasujeto: en todas las VI se aplica una estrategia de comparación intrasujeto.
- Diseño factorial mixto: combina una o más VI que aplican estrategia de comparación intersujeto con otras VI que aplican la estrategia intrasujeto.

Para ilustrar este punto supongamos una investigación sobre los factores que influyen en las preferencias de los hombres por productos cosméticos. Se plantea un diseño factorial para estudiar cómo influyen dos factores sobre la preferencia por una crema facial: la presentación (tubo o bote) de la crema y el aroma (inodora, cítrica o madera). Se trata de un diseño factorial 2×3 , dado que combina una VI con dos niveles con una segunda VI con tres niveles, por lo que plantea seis condiciones experimentales (véase la ficha 2.9). A continuación, mostramos cómo este mismo diseño factorial 2×3 requiere un número distinto de grupos en función de la estrategia de comparación que se usa para cada VI.

Diseño factorial intersujeto 2×3

VI: Presentación	VI: Aroma	
Tubo	Inodora	
	Madera	
	Cítrica	
Bote	Inodora	
	Fusta	
	Cítrica	

Interviene un grupo de participantes diferente por cada condición experimental. En total seis grupos de participantes diferentes.
Cada participante solo recibe un tipo de presentación combinada con un tipo de aroma.

Diseño factorial intrasujeto 2×3

Tubo			Bote		
Inodora	Madera	Cítrica	Inodora	Madera	Cítrica

Interviene un único grupo de participantes que pasa por las 6 condiciones experimentales en periodos diferentes.
Cada participante recibe 6 combinaciones de presentación y aroma.

Diseño factorial mixto 2×3 (intrasujeto en el primer factor)

VI: aroma	VI: Presentación	
	Tubo	Bote
Inodora		
Madera		
Cítrica		

Interviene un grupo de participantes diferente por cada nivel de la VI intersujeto.
Si el factor intrasujeto es el primero tal como está esquematizado en la tabla, los mismos participantes asignados, por ejemplo, al aroma «inodora» recibirán la crema en los dos formatos de presentación. En total intervendrán tres grupos de participantes diferentes. Aquí cada participante recibe los dos tipos de presentación pero siempre combinada con el mismo aroma.
Nótese que si el factor intrasujeto fuera el aroma intervendrían dos grupos de participantes diferentes. Aquí cada participante recibiría los tres tipos de aroma pero siempre combinadas con la misma presentación.

Existe una extensión del diseño factorial mixto que queremos comentar explícitamente, ya que es de uso frecuente en ciencias de la salud. Se trata del caso en que una de las variables que conforman el diseño factorial son las ocasiones en que se mide la VD. Podemos ejemplificar esta situación a partir del ejemplo que hemos usado para explicar la interacción (ficha 2.7). Suponga que al ejemplo expuesto le añadimos tres ocasiones de medida de la función cognitiva (VD), tal como esquematiza la siguiente tabla. A diferencia de la comparación intrasujeto que hemos explicado hasta ahora (fichas 2.4, 2.5, 2.6), aquí no se observa la respuesta bajo diferentes niveles de una VI o condiciones experimentales, sino que se observa únicamente el registro de la VD en diferentes ocasiones con el fin de estudiar su evolución temporal. Observe que son dos variables intersujeto las que determinan la condición experimental que recibirá el participante, y que en las tres ocasiones de medida la condición experimental es la misma. Así, en este caso especial del diseño mixto, no tenemos una VI intrasujeto manipulada. Por último, añadiremos que se suele usar el término «medidas repetidas» (en inglés, *repeated measures*) como genérico para englobar los diseños que miden la VD en más de una ocasión.

Entrenamiento cognitivo (EC)	Ejercicio físico (EF)	Ocasión de medida de la función cognitiva (VD)		
		Inicio	6 meses	18 meses
EC Activo	Ejercicio aeróbic			
	EF placebo			
EC placebo	Ejercicio aeróbic			
	EF placebo			

La ocasión de medida de la función cognitiva (VD) se analiza como una variable intrasujeto del diseño.

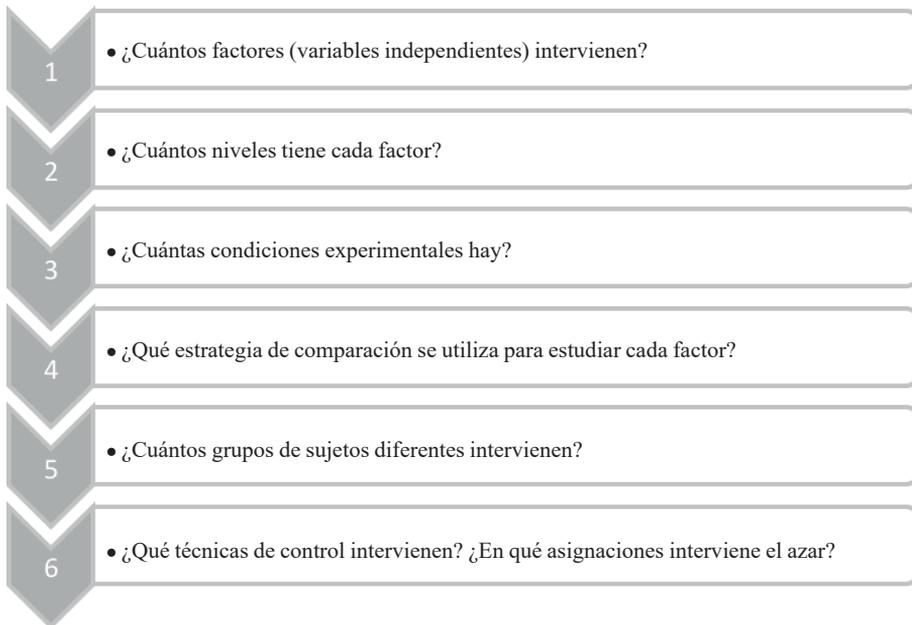
FICHA 2.9. «INTERROGANDO» A UN DISEÑO FACTORIAL

Existen diseños factoriales mucho más complejos que los presentados hasta ahora, ya que pueden incorporar más VI y más niveles por cada una de ellas. La siguiente notación permite simbolizar la estructura del diseño factorial:

$$\text{N}^{\circ} \text{ de niveles}_{\text{VI-1}} \times \text{N}^{\circ} \text{ de niveles}_{\text{VI-2}} \times \dots \times \text{N}^{\circ} \text{ de niveles}_{\text{VI-j}}$$

Así, en el caso del ejemplo de la ficha 2.7, indicaremos que el número de niveles de cada VI se combina formando una estructura factorial 2×2 (dos VI con dos niveles cada una). Con la misma lógica, si expresamos que tenemos un diseño factorial $2 \times 3 \times 5$ debemos entender que se trata de un diseño con tres factores o VI, la primera con dos niveles, la segunda con tres y la tercera con cinco.

En la siguiente figura proponemos una secuencia de cuestiones que facilitan la lectura de una investigación basada en un diseño experimental de tipo factorial.



En primer lugar, se trata de identificar las VI implicadas y los niveles que tiene cada una. A partir de aquí, se pueden establecer las condiciones experimentales del diseño, que son el resultado de combinar los niveles de cada VI o factor. Así, en un diseño 2×2 tenemos cuatro condiciones experimentales y en un diseño $2 \times 3 \times 5$ tenemos treinta. A continuación, hay que establecer la estrategia de comparación que se usa para comparar los niveles de cada VI y determinar si el diseño factorial es intersujeto, intrasujeto o mixto; esto permitirá determinar el número de grupos diferentes de participantes que intervienen en el diseño. Finalmente, hay que examinar las técnicas

de control aplicadas. Recuerde que para definir el diseño como experimental requeríamos que, como mínimo, se aplique la aleatorización a una de las VI (ficha 2.7). Evidentemente, cuantas más comparaciones del diseño se controlen con técnicas basadas en la aleatorización, más se incrementa la posibilidad de obtener conclusiones con mayor validez interna, si contamos con un número suficiente de participantes..

ACTIVIDADES

Actividad de identificación de diseños

Identifique y represente esquemáticamente los diseños utilizados en cada uno de los siguientes estudios.

1. Estudio de los efectos de la aplicación de un programa de entrenamiento en habilidades sociales dirigido a madres y sus hijos³

En el presente trabajo se evalúa la eficacia diferencial del entrenamiento conductual en habilidades sociales en dos situaciones: 1) entrenamiento en habilidades sociales solo a madres; 2) entrenamiento en habilidades sociales a madres e hijos/as. La muestra está formada por 33 parejas madre-hijo/a. Estas parejas son asignadas de forma aleatoria a tres grupos (dos condiciones experimentales y una condición control). Los niños y niñas presentan una edad media de 7,8 años, de los cuales 16 son del sexo masculino y 17 del femenino. Todos ellos están escolarizados y pertenecen a familias de un nivel socioeconómico medio-bajo. No se encuentran diferencias estadísticamente significativas entre el resultado de las dos situaciones de entrenamiento; en cambio, sí se observan diferencias entre cada uno de los grupos que han recibido entrenamiento con respecto al grupo control.

2. Efectos de la visión de escenas violentas en la conducta agresiva infantil⁴

El objetivo del presente estudio es determinar la influencia de la exposición a escenas violentas en la conducta agresiva infantil. La hipótesis de los autores es que hay diferencias en la conducta verbal y física infantil en función del tipo de imágenes violentas visualizadas. Estas pueden ser imágenes violentas justificadas, injustificadas o neutras. La muestra está formada por 60 participantes (30 niñas y 30 niños) de 12 años. Los participantes son asignados de manera aleatoria a una de las tres condiciones experimentales; la variable sexo se controla mediante la técnica de bloqueo. De este modo, cada tipo de escena será visto por 10 chicos y 10 chicas. Todos ellos ven una película, de 15 minutos de duración, correspondiente a cada condición experimental (escenas de violencia ficticia justificada, escenas de violencia ficticia injustificada o escenas ficticias sin violencia o neutras). Una vez vista la película, todos los niños y niñas pasan a participar en un juego cooperativo que requiere la colaboración de todos los integrantes del grupo para su resolución. Se graban las conductas y luego se codifican los actos agresivos, verbales y físicos, de cada uno de los participantes en una plantilla de registro observacional.

3. Adaptado de: OLIVARES, J.; MARTÍNEZ, M.; LOZANO, M. 1997. «Estudio de los efectos de la aplicación de un programa de entrenamiento en habilidades sociales dirigido a madres e hijos». *Psicología Conductual*, 5 (2), 277-93. <http://www.funveca.org/revista/PDFespanol/1997/art08.2.05.pdf>. Consulta: 18 de noviembre de 2013.
4. Adaptado de: PEÑA, M. E.; ANDREU, J. M.; MUÑOZ, M. J. 1999. «Efectos de la visión de escenas violentas en la conducta agresiva infantil». *Psicothema*, 11, nº 1, 27-36.

3. Calidad percibida del vino

Se desea estudiar el efecto de la información que aparece en la etiqueta de las botellas de vino sobre la calidad percibida por el consumidor. Con este objetivo se diseña una investigación con cuatro condiciones. En la primera condición, la etiqueta indica que el vino es catalán, en la segunda, que el vino es español, en la tercera, que el vino es italiano y, en la cuarta, que el vino es francés. A partir de los registros facilitados por entidades que realizan cursos de cata de vino en la comarca del Penedès, se seleccionan al azar 48 personas adultas. Todos los participantes han realizado un curso de cata durante un fin de semana, tienen edades comprendidas entre los 25 y los 65 años, y un 4% son fumadores. La tarea de los participantes es estimar la calidad del vino, con una escala de 0 a 10, en las cuatro condiciones experimentales. Se aplica la técnica del contrabalanceo.

4. Análisis de la influencia de la conducta en la ilusión de control⁵

Diferentes investigaciones establecen que la percepción de control es un predictor importante del bienestar físico y psicológico. En general, cuando las personas pueden elegir entre ejercer control o no, muestran una clara preferencia por ejercerlo. Pero no siempre que creemos controlar algo es realmente así. En algunas ocasiones, se presenta lo que se denomina «ilusión de control», es decir, se actúa como si se pudiera influir en ciertos hechos que en realidad son aleatorios o dependen de factores ajenos a la persona implicada.

En un trabajo pionero sobre la ilusión de control Langer (1975) empleaba diferentes tareas de juegos de azar. Este investigador identificó cuatro elementos clave como factores que favorecen la ilusión de control: familiaridad, implicación, competición y conocimiento previo. A partir de esta investigación, surgieron un conjunto de trabajos como el de Alloy y Abramson (1979), basados en los juicios de contingencia, que ponían de manifiesto la importancia de la presencia de una conducta en la generación de ilusión de control. En los juicios de contingencia se presentan, de forma secuencial, una serie de ensayos en los que los participantes deben optar por realizar una de dos conductas posibles (concretamente, deben escoger entre realizar una determinada acción, como pulsar una palanca, o no hacer nada) y observar con posterioridad si la conducta elegida va seguida o no de la presencia de una luz, con el objetivo final de que el participante establezca un juicio sobre el grado de control que su conducta ejerce sobre la aparición de la luz.

El objetivo de esta investigación es dar respuesta a la pregunta sobre si en una situación en la que la persona puede realizar una respuesta voluntaria (control libre) genera la misma ilusión de control que en una situación equivalente sin presencia de conducta (predicción).

Los participantes son un grupo de 30 estudiantes de primer curso de psicología (20 mujeres y 10 hombres) con edades comprendidas entre 18 y 23 años.

Siguiendo el paradigma de los juicios de contingencia, se diseñan dos pruebas experimentales de 40 ensayos cada una con un programa de ordenador:

5. Adaptado de: ROVIRA, T.; FERNÁNDEZ, J.; EDO, S. 2000. «Análisis de la influencia de la conducta en la ilusión de control». *Anales de Psicología*, 16 (1), 13–21.

- «Prueba de control libre»: los estímulos empleados son un ruido de 400 hz, 90 dB y 400 ms de duración (indica el inicio de los ensayos) y un cuadrado verde de 10 x 10 cm como estímulo a controlar. La prueba está programada de forma que, con posterioridad a la conducta del participante (pulsar una tecla o no), el cuadrado aparece en el 50 % de los ensayos.
- «Prueba de predicción»: tras el mismo ruido de inicio que en la prueba anterior, y durante el intervalo de tiempo en que anteriormente el participante ejecuta la conducta, se presenta uno de los dos estímulos de los que son posibles predicciones del cuadrado verde: un círculo blanco vacío o un círculo blanco lleno. Cada uno de ellos aparece en la mitad de los ensayos y seguido a su vez por el cuadrado en el 50 % de las ocasiones.

Antes de iniciar las pruebas se explica que el objetivo general del estudio consiste en observar la capacidad de controlar acontecimientos del entorno y relaciones entre estímulos. Todas las instrucciones aparecen en la pantalla de un ordenador y al finalizar cada prueba se entrega un cuaderno donde deben anotar su juicio de contingencia. El juicio de contingencia consiste en indicar su juicio de control o de predicción de la aparición del cuadrado verde. La emisión de este juicio se realiza mediante una escala de 10 cm de longitud, con los extremos marcados del 0 al 100, y fragmentada con intervalos de 1 cm. El 0 significa no tener ningún control (en la prueba de control libre) o predicción (en la prueba de predicción) sobre la aparición del cuadrado, y el 100 significa tener un control/predicción total sobre la aparición del cuadrado.

Todos los participantes pasan por las dos pruebas, pero no todos las realizan con el mismo orden: la mitad hace primero la prueba de control libre y después la de predicción, y la otra mitad sigue el orden inverso. Los participantes son asignados aleatoriamente a uno de los órdenes de presentación (control libre-predicción o predicción-control libre). Entre cada una de las pruebas se hace una tarea distractora con diferentes estímulos que van apareciendo en el ordenador.

5. Tratamiento del dolor postoperatorio⁶

En un hospital se realiza un estudio sobre el efecto de dos técnicas para el tratamiento del dolor postoperatorio tras una intervención quirúrgica superficial, y el efecto del momento de inicio de aplicación de estas técnicas. Las técnicas son la acupuntura y la terapia musical; además, se define una tercera condición de control por placebo que incluye sesiones de duración y estructura similar a las de aplicación de las técnicas mencionadas. Las tres intervenciones (Placebo —P—, Acupuntura —A— y Terapia Musical —TM—) se aplican durante 15 días antes (15-d), 7 días antes (7-d) o inmediatamente antes de la intervención quirúrgica (0-d). Se dispone de un total 45 pacientes de un centro hospitalario, los cuales son asignados al azar a una de las nueve condiciones experimentales. Cada grupo está formado por cinco pacientes. Para evaluar el efecto conjunto de ambas VI, se mide el tiempo de remisión del dolor postoperatorio.

6. Puede encontrar más información sobre terapia musical en: WHITAKER, M. H. 2010. «Sounds soothing: Music therapy for postoperative pain». *Nursing*, 40 (12), 53–54.

6. Tiempo de reacción

En el contexto de una línea de investigación sobre los factores que influyen en el tiempo de reacción, un investigador se plantea estudiar el efecto del tipo de estímulo empleado y del tipo de ruido del ambiente en el que se presenta el estímulo. Con este objetivo, selecciona dos tipos de estímulos: una luz blanca (LB) y un ruido grave (RG). Se combinan estos dos tipos de estímulos con tres tipos de situaciones ambientales: silencio (S), música clásica (MC) y música rock (MR). Entre los estudiantes de logopedia de primer curso se seleccionan al azar 72 sujetos y se asignan, de manera equilibrada y al azar, a una condición experimental. Los estudiantes pasan por la correspondiente situación experimental en el mismo laboratorio y a la misma hora, y a todos ellos se les mide el tiempo de reacción.

7. El entorno psicosocial y los neurotransmisores⁷

En el ámbito de la clásica discusión entre genética y ambiente, se diseña un estudio para contrastar el supuesto de que se produce un efecto de interacción entre los neurotransmisores y el ambiente, de manera que un mismo neurotransmisor puede tener diferentes efectos en función del entorno psicosocial.

Una muestra de 50 monos rhesus se divide al azar en dos grupos. Uno de los grupos puede controlar las cosas que hay en su jaula y el otro no. En concreto, uno de los grupos tiene libre acceso a juguetes y alimentos, y el otro puede tener juguetes y comida solo después de que los miembros del primer grupo hayan terminado. De este modo, se espera conseguir que un grupo crezca con la sensación de control sobre el entorno y el otro grupo no. Posteriormente, a todos los sujetos se les administra dos dosis de un antagonista de la benzodiazepina⁸ con una semana de intervalo entre dosis. La primera dosis es lo suficientemente baja para que no tenga efecto en el sistema nervioso. La segunda dosis sí es suficiente para producir efecto. En la situación de dosis baja no hay diferencias entre grupos, mientras que en la situación de dosis alta los monos criados con poco control muestran evidentes signos de ansiedad y pánico, y corren a esconderse en algún rincón de la jaula. Los miembros del grupo con control no estaban angustiados, sino que eran agresivos y atacaban a cualquiera que se les acercara.

8. La influencia normativa⁹

Cuando las personas deben manifestar un juicio sobre algún aspecto de la realidad en presencia de otros, suelen tener dos preocupaciones principales: tener razón y dar bue-

7. Adaptado de: BARLOW, D. H.; DURAND, V. M. 2001. *Psicología anormal. Un enfoque integral*. Madrid: Thomson Learning.

8. Los antagonistas de la benzodiazepina producen una ansiedad extrema. Las pocas veces que se han administrado a personas, estas aseguran haber vivido la sensación más horrible de su vida, a pesar de que su efecto es muy breve.

9. Adaptado de: VAN AVERMAET, E. 1992. «Influencia social en los grupos pequeños». En: M. Hewstone, W. Stroebe, J. P. Codol y G. M. Setephenson (eds.). *Introducción a la psicología social*. Barcelona: Ariel. P. 340–68.

na impresión a los demás. Centrémonos en la segunda de las preocupaciones. Si nos sentimos dependientes de los demás para satisfacer una gran variedad de necesidades, consideraremos importante poder maximizar nuestro atractivo para con ellos. En la medida que el desacuerdo con los demás nos puede hacer sentir menos atractivos y, al contrario, el acuerdo provoca evaluaciones positivas, las personas tendemos, especialmente en situaciones de ambigüedad, a adaptar nuestro juicio al de los demás por razones normativas. La conformidad provocada por el deseo de ser querido o por la aversión a ser rechazado es debida a lo que en psicología denominamos *influencia normativa*.

Para poder evidenciar los efectos de la influencia normativa en situaciones de ambigüedad se selecciona al azar una muestra de 50 estudiantes universitarios a los que se encomienda la tarea de emitir un juicio. Se trata de identificar el color de 20 figuras geométricas que aparecen en pantalla, 10 de las cuales están pintadas con un color ambiguo y las otras 10 con un color no ambiguo. Para controlar el efecto del orden de presentación se aplica la técnica del contrabalanceo. Los participantes deben hacer el reconocimiento de las figuras con color ambiguo y de las figuras con color no ambiguo.

Por otra parte, con respecto a la situación de identificación, se definen tres condiciones: individual, en pareja y grupal. Los participantes son asignados aleatoriamente a una de estas condiciones. En la situación de pareja, el participante emite el juicio en primer lugar y en presencia de un cómplice del experimentador. En la situación grupal, el juicio se emite en presencia de seis personas, cuatro de las cuales son cómplices de los experimentadores y emiten su juicio en primer lugar; el juicio de los cómplices siempre está alejado de una respuesta «razonablemente»¹⁰ correcta. Los investigadores evalúan las discrepancias entre el juicio individual y en pareja en relación con el que se emite en la situación grupal.

10. Decimos razonablemente correcta porque los sujetos no son expertos en la identificación de colores. Esto hace que no sea posible comparar este juicio con el que emitirían expertos en el tema. Por este motivo, los investigadores establecen unos parámetros dentro de los que se considera que la respuesta es adecuada.

EJERCICIOS

Ejercicio 2.1: ¿Se trata de un experimento?

McGuigan (1983, p. 164–165) explica la siguiente anécdota que le sucedió con un estudiante que estaba realizando su trabajo de máster:

El estudiante utilizó la velocidad para recorrer un laberinto como indicador del constructo que quería contar. El objetivo era evaluar el efecto de dos condiciones de entrenamiento sobre dicho constructo. Su razonamiento fue que el grupo 1, que recibía el entrenamiento que se consideraba más adecuado, debía mostrar más velocidad. Pero el experimento nunca fue más allá. Para asignar las ratas a los grupos, «el experimentador» fue a las jaulas y las primeras ratas que fueron a parar a sus manos fueron asignadas al grupo 1; las restantes, al grupo 2. Sin duda, los animales más activos se asomaron primero y el experimentador las cogió, mientras que los menos activos se recogieron en la parte posterior. Al grupo 1 van a parar muy probablemente las ratas más rápidas. Por lo tanto, se confundieron los tratamientos experimentales y la habilidad inicial (innata) de las ratas para correr.

Posiblemente, después de hacer esto, este estudiante tuvo más de un quebradero de cabeza para finalizar la máster (¡esperemos que lo acabara consiguiendo!). En todo caso, nos plantea un excelente ejemplo para reflexionar sobre una serie de conceptos básicos del diseño experimental.

1. Identifique la(s) variable(s) independiente(s) y las dependiente(s) que quería relacionar este estudiante.
2. En esta investigación, ¿qué tipo de variable es «la habilidad inicial de las ratas para correr»?
3. ¿Cómo se llama la relación que se establece entre las tres variables indicadas en las preguntas precedentes?
4. ¿Qué técnica de control debería haber utilizado el estudiante?

Ejercicio 2.2: Psicofármacos

Una fobia es un trastorno de ansiedad que se caracteriza por un miedo intenso e irracional a un estímulo o una situación. La respuesta ante este estímulo o situación es la evitación o la fuga.

Estudios previos indican que un nuevo fármaco ISRS (inhibidor selectivo de la recaptación de serotonina) tiene un efecto ansiolítico que provoca una disminución de la intensidad y frecuencia de aparición de la sintomatología característica de los trastornos de ansiedad, pudiendo llegar hasta la remisión total del trastorno. Dado que las fobias son un trastorno de ansiedad, un investigador mantiene la hipótesis de que estas también serán sensibles a los efectos terapéuticos de este ISRS. Para contrastar esta hipótesis, se establecen dos grupos: un grupo (grupo 1) al que se administra una píldora diaria con una dosis del ISRS y otro grupo (grupo 0) al que se administra una píldora diaria de características idénticas a la administrada al grupo anterior, pero que no contiene ISRS.

Debido a que existen diferentes tipos de fobias y que cada una de ellas conlleva diferencias en relación con la frecuencia, intensidad y duración de la sintomatología, el investigador decide incluir en el estudio únicamente a individuos con fobia social. Por otra parte, es sabido que, a mayor tiempo transcurrido desde el inicio del trastorno hasta el inicio del tratamiento, más se agravan los síntomas y más difícil es la remisión, sobre todo si ha pasado más de un año desde el inicio del trastorno. Este aspecto también se tuvo en cuenta al planificar la recogida de datos.

Se selecciona una muestra de 20 personas con fobia social con una duración del trastorno de entre 8 y 12 meses. Posteriormente, se asignan los sujetos al azar a dos grupos. Pasado el periodo de tratamiento, se mide la ansiedad de los pacientes mediante el test T (los valores que puede tomar este test van de 0 a 100 puntos: a más puntuación más ansiedad). En la tabla E2.2 se presentan los resultados de la investigación.

Tabla E2.2. Resultados de la investigación

<i>Caso</i>	<i>Grupo</i>	<i>Ansiedad</i>	<i>Caso</i>	<i>Grupo</i>	<i>Ansiedad</i>
3	1	13	6	0	35
4	1	18	10	0	39
19	1	22	9	0	29
7	1	34	18	0	39
20	1	27	14	0	11
8	1	36	1	0	25
16	1	38	12	0	20
17	1	8	13	0	15
5	1	20	15	0	10
11	1	9	2	0	30

1. Formule la hipótesis de esta investigación. Tal como la ha formulado, ¿se trata de una hipótesis operativa?
2. ¿Cuál es la VD?
3. Identifique la VI e indique sus niveles.
4. ¿Cómo se llama el efecto que se pretende controlar al administrar una píldora de características idénticas a la que contiene el ISRS, pero que no contiene el principio activo?
5. En el enunciado se mencionan dos posibles variables de confusión. Identifíquelas e indique cómo se han controlado.
6. ¿Qué función cumple la asignación aleatoria de los sujetos a los grupos de tratamiento?
7. ¿Cómo se denomina el diseño utilizado?
8. Calcule la media y la desviación estándar de la ansiedad en cada uno de los grupos.
9. ¿Cómo construiría un indicador numérico del efecto del nuevo ISRS?

10. A partir de los resultados obtenidos, ¿considera que el tratamiento farmacológico ha sido eficaz?
11. Suponga que se considera que una reducción de la ansiedad medida con el test T empieza a ser relevante a partir de 10 puntos. A partir de los resultados de este estudio, ¿qué recomendaría para el tratamiento de las fobias?
12. ¿El diseño de esta investigación permite medir el efecto placebo?
13. Desde un punto de vista ético, ¿cómo valora la actuación de este investigador?

Ejercicio 2.3: Asignación aleatoria y bloqueo

En la tabla E2.3 se presentan una serie de características de una muestra de alumnos de ESO seleccionados para estudiar el efecto de un nuevo método de enseñanza sobre el rendimiento. Para realizar la investigación se necesita distribuir la muestra en dos grupos que sean comparables. A uno de los grupos se le aplicará el nuevo método de enseñanza y al otro, el método tradicional.

1. Asigne aleatoriamente los alumnos de la muestra a los dos grupos (m. nuevo / m. tradicional). Para cada grupo indique la distribución porcentual de la variable sexo y la media de CI (tabla E2.3b).
2. Forme dos bloques a partir de la variable CI, en función de si tienen las puntuaciones más altas o más bajas en esta variable. Asigne aleatoriamente los sujetos de cada bloque a los dos niveles de la VI. Para cada grupo indique la media de CI (tabla E2.3c).
3. Asigne los sujetos a los dos niveles de la VI usando la técnica del apareamiento con posterior asignación aleatoria. Para cada grupo indique la media de CI (tabla E2.3d).
4. ¿Hay alguna diferencia en el diseño implícito en las tres preguntas anteriores? ¿De qué manera puede valorar cuál es el mejor?

Tabla E2.3a. Sexo y CI de una muestra de 24 alumnos

<i>Caso</i>	<i>Sexo</i>	<i>CI</i>	<i>Caso</i>	<i>Sexo</i>	<i>CI</i>
1	M	90	13	F	118
2	M	87	14	F	89
3	M	106	15	M	100
4	F	86	16	M	128
5	M	92	17	F	120
6	F	107	18	M	100
7	F	93	19	F	121
8	F	90	20	F	101
9	M	108	21	F	90
10	F	97	22	M	85
11	M	113	23	F	127
12	M	97	24	F	85

Tabla E2.3b. Asignación aleatoria sin restricción

<i>Caso</i>	<i>Sexo</i>	<i>CI</i>	<i>Método</i>	<i>Caso</i>	<i>Sexo</i>	<i>CI</i>	<i>Método</i>
1	M	90		13	F	118	
2	M	87		14	F	89	
3	M	106		15	M	100	
4	F	86		16	M	128	
5	M	92		17	F	120	
6	F	107		18	M	100	
7	F	93		19	F	121	
8	F	90		20	F	101	
9	M	108		21	F	90	
10	F	97		22	M	85	
11	M	113		23	F	127	
12	M	97		24	F	85	

	<i>Método de enseñanza</i>	
<i>Sexo</i>	Tradicional (T)	Nuevo (N)
Masculino (M)		
Femenino (F)		
XQI		

Tabla E2.3c. Asignación aleatoria con bloqueo

	<i>Caso</i>	<i>CI</i>	<i>Método</i>
BAJO	1	90	
	2	87	
	4	86	
	5	92	
	7	93	
	8	90	
	10	97	
	12	97	
	14	89	
	21	90	
	22	85	
	24	85	

	<i>Caso</i>	<i>CI</i>	<i>Método</i>
ALTO	3	106	
	6	107	
	9	108	
	11	113	
	13	118	
	15	100	
	16	128	
	17	120	
	18	100	
	19	121	
	20	101	
	23	127	

Tabla E2.3d. Asignación aleatoria con apareamiento

<i>Bloque</i>	<i>Caso</i>	<i>CI</i>	<i>Caso</i>	<i>CI</i>	<i>M. TRAD.</i>	<i>M. NUEVO</i>
1	22	85	24	85		
2	4	86	2	87		
3	14	89	1	90		
4	8	90	21	90		
5	5	92	7	93		
6	10	97	12	97		
7	15	100	18	100		
8	20	101	3	106		
9	6	107	9	108		
10	11	113	13	118		
11	17	120	19	121		
12	16	128	23	127		

Ejercicio 2.4: Comportamiento del consumidor

El comportamiento del consumidor se define como el proceso de decisión y la actividad física que los individuos realizan cuando evalúan, adquieren, usan o consumen bienes o servicios. En el marco de esta especialidad se realiza una investigación para evaluar si los adolescentes, chicos y chicas de entre 16 y 17 años, prefieren alimentos (tipo snacks y derivados lácteos) que al anunciarlos se asocian a un personaje famoso. Con este objetivo, la responsable de la investigación selecciona al azar diez institutos de clase media del Barcelonés, de modo que la muestra con la que se lleva a cabo la investigación está formada por 616 hombres y mujeres de entre 16 y 17 años que van a

los mencionados institutos. Estos chicos y chicas se asignan al azar a dos grupos (A y B). A los que forman parte del grupo A se les convoca a la proyección de una película en la que diferentes alimentos aparecen asociados a personajes famosos. A los hombres y mujeres que forman parte del grupo B se les pasa otra película en la que los mismos alimentos no aparecen asociados a ningún personaje famoso. A continuación, se invita a todos los participantes a una merienda en la que están los alimentos que aparecían en las películas, mezclados con otros que no aparecían, y se registra el tipo de alimento escogido por los chicos y chicas.

1. Identifique y clasifique la VI.
2. Identifique y clasifique la variable dependiente.
3. ¿De qué tipo de diseño se trata?
4. Si se asigna el mismo número de sujetos en cada tratamiento, ¿cuántos sujetos recibirán cada uno de los niveles de la VI?
5. Si los elementos de la muestra se hubieran dividido en dos grupos en función del sexo, y los sujetos de cada grupo se hubieran asignado al azar a cada condición experimental, ¿cómo se denominaría el diseño utilizado?
6. Si los pacientes de la muestra hubieran sido todos chicas, ¿cómo se llamaría el diseño utilizado?
7. Indique, al menos, una variable que debería haber controlado y que en el resumen no consta como tal. ¿Con qué técnica la controlaríamos?

Ejercicio 2.5: ¿Cuál es el eslogan más apropiado?

Un psicólogo que trabaja en una empresa de publicidad desea comprobar el impacto de dos eslóganes nuevos de un producto de limpieza en comparación con el que se está utilizando. A partir de la información facilitada por un grupo de comerciantes, se seleccionan 30 consumidores de otros productos de limpieza. A cada uno de los consumidores se le presentan los tres eslóganes empleando la técnica del contrabalanceo. Al día siguiente de examinar cada eslogan se mide el grado de recuerdo del mensaje y la actitud hacia la compra del producto.

1. Identifique y clasifique la(s) variable(s) independiente(s).
2. Identifique y clasifique la(s) variable(s) dependiente(s).
3. ¿De qué tipo de diseño se trata?
4. En esta investigación, describa los pasos que implica la utilización de la técnica del contrabalanceo.
5. ¿A cuántos sujetos se aplicará cada condición experimental?

Ejercicio 2.6: Tamaño muestral en los diseños factoriales

Partiendo de un ejercicio propuesto por Gambará (2002), a continuación le planteamos de manera abstracta cuatro diseños factoriales. En cada caso se trata de que represente esquemáticamente cada uno de los diseños y que responda a las preguntas que formulamos a continuación.

1. Se realiza un experimento con un diseño factorial 3×3 . El número de observaciones en cada condición experimental es de 9. Si las 2 VI son intersujeto, ¿cuál será su tamaño muestral?
2. En un diseño factorial 2×3 la primera VI es intersujeto y la segunda, intrasujeto. Si en cada condición experimental queremos disponer de 10 datos, ¿con cuántos sujetos tendremos que contar?
3. Partimos de un diseño factorial 2×2 , con ambas VI intersujeto. Contamos con 80 sujetos; ¿cuántos sujetos tendremos en cada condición experimental?
4. Dos investigadores realizan un experimento y ambos utilizan un diseño factorial 4×2 ; pero el investigador A considera ambas variables intrasujeto, y el investigador B, intersujeto. En cada condición queremos obtener 10 puntuaciones. ¿Deberán contar los dos investigadores con el mismo número de sujetos? ¿Por qué?

Ejercicio 2.7: El riesgo y el control percibido por los conductores jóvenes

En el ámbito de la seguridad vial, uno de los factores que se ha relacionado con la alta incidencia de accidentes graves entre los conductores jóvenes es el riesgo percibido. Los resultados obtenidos en diferentes investigaciones han llevado a formular la siguiente hipótesis: en situaciones en las que hay una alta probabilidad de accidente, los conductores jóvenes perciben menos riesgo que los adultos, si bien esta relación está modulada por el grado de control que perciben sobre la situación. Para contrastar esta hipótesis, se realiza un estudio de laboratorio usando un simulador de conducción que permite emular diferentes situaciones entre las que varía la probabilidad de accidente y el grado de control que los participantes ejercen sobre la situación.

Todos los participantes visualizan un coche que se mueve a velocidad constante (80 km/h), en tres situaciones diferentes: 1) situación = DS día soleado y con perfecta visibilidad; 2) situación = NL noche con lluvia, y 3) situación = NLN noche con lluvia y niebla. Los participantes son asignados aleatoriamente a distintos órdenes de presentación de las situaciones mencionadas. En cuanto al grado de control, se definen dos condiciones: 1) condición = conductor, en la que el participante manipula el volante del simulador como si fuera el conductor del vehículo; 2) condición = acompañante, en la que el participante visualiza la situación en calidad de acompañante de un cómplice del investigador que manipula el volante del simulador. El simulador está programado para que en la «condición = conductor» no se produzca ningún accidente, independientemente del manejo del volante que hace el participante.

Se seleccionan 30 jóvenes de 20 años, y 30 adultos de 35 años. Las condiciones de admisión en el estudio son las siguientes: 1) tener permiso de conducir, 2) no haber estado nunca implicado en un accidente grave, 3) haber realizado, en promedio, un mínimo de 30 km semanales durante los 12 meses previos, 4) los participantes adultos deben tener un mínimo de diez años de experiencia como conductores, y los jóvenes un mínimo de un año, 5) no dedicarse profesionalmente al transporte de pasajeros o mercancías.

Los participantes de cada grupo de edad son asignados aleatoriamente a una de las dos condiciones definidas previamente (conductor-acompañante). Después de presenciar la evolución del coche en cada una de las situaciones (DS, NL, NLN), el participante indica el riesgo percibido en una escala de 0 a 100.

Los resultados muestran que el riesgo percibido por los adultos es superior al percibido por los jóvenes en todas las condiciones, e independientemente del tipo de control. Por otra parte, en la condición en la que la probabilidad de accidente es más alta, el riesgo percibido es mayor entre los acompañantes (control bajo), que entre los conductores (control alto).

1. Escriba la hipótesis operativa de esta investigación.
2. Identifique la(s) variable(s) dependiente(s) y clasifíquela(s) en función del criterio de medida.
3. Identifique la(s) variable(s) independiente(s), clasifíquela(s) en función del criterio de medida e indique si está(n) manipulada(s) o no.
4. Indique la(s) técnica(s) de control empleada(s) y la función que cumplen en este estudio.
5. Clasifique la investigación en función de los dos criterios que se indican a continuación. Además de marcar la alternativa, en la línea de puntos debe escribir el aspecto que caracteriza el criterio de clasificación:
 - Criterio 1: en función de
 - a) diseño intersujeto
 - b) diseño intrasujeto
 - c) diseño mixto
 - Criterio 2: en función de
 - a) diseño unifactorial
 - b) diseño factorial
6. Esquematice el diseño utilizado.
7. Elabore una representación gráfica que encaje con la información incluida en el resumen sobre los resultados obtenidos.
8. ¿Estos resultados apoyan la hipótesis planteada?
9. Indique dos tipos de validez que se consideran en la evaluación global de la metodología de una investigación. Haga un comentario sobre cada una de ellas aplicado a la investigación que está analizando.

Ejercicio 2.8: La percepción del tiempo

Un equipo de investigación está interesado en el estudio de los factores que influyen sobre la percepción del tiempo. Los trabajos previos sugieren que las personas infravaloran el tiempo cuando se les presenta información novedosa. Por otra parte, indican que se sobrevalora el tiempo a medida que se incrementa la magnitud del intervalo. Para contrastar estas hipótesis, se selecciona a 90 estudiantes con puntuaciones similares en el cuestionario de «Matutinidad-vespertinidad» de Horne y Östberg. En todos ellos se presentan 50 palabras, la mitad de las cuales son de uso común (información no novedosa) y las otras, de uso no común (información novedosa). Cada palabra se proyecta en una pantalla y al desaparecer el participante debe indicar cuánto tiempo ha permanecido visible. El orden de presentación de los dos tipos de palabras (novedosas y no novedosas) se aleatoriza y cada participante recibe todas las palabras, pero una sola vez cada una. Respecto a la magnitud del intervalo de tiempo en que cada palabra se mantiene visible en la pantalla se definen tres condiciones: 2 segundos, 6 segundos

y 10 segundos. Los sujetos son asignados al azar a una de las tres magnitudes mencionadas y realizan los 50 juicios temporales el mismo día y a la misma hora. Los resultados indican una infravaloración del tiempo cuando la información es novedosa y una sobrevaloración cuando no lo es, si bien el efecto de esta variable está modulado por la magnitud del intervalo: cuando la información es novedosa el mayor sesgo se produce con la magnitud baja, mientras que cuando es no novedosa el mayor sesgo se produce con la magnitud alta.

1. Identifique las VI y VD.
2. Indique si hay alguna VI que no ha sido manipulada.
3. Identifique y esquematice el diseño utilizado.
4. Para poder tener 30 participantes en cada condición experimental, ¿cuántos participantes se necesitarían?
5. En este estudio, ¿qué papel tiene la variable matutinidad-vespertina?

Ejercicio 2.9: Sobre la importancia del refuerzo verbal

Un equipo de investigación se propone estudiar la relación entre la extraversión y el refuerzo propuesta por la teoría de Gray (1970, 1982, 1987) en la población adolescente. De acuerdo con lo que se desprende de esta teoría, los extrovertidos serán más susceptibles a los premios que los introvertidos, mientras que los introvertidos serán más susceptibles a los castigos.

Se diseña un estudio con una muestra de adolescentes de sexo masculino. La captación de los participantes se realiza mediante un anuncio en la prensa local en el que se ofrece una compensación a cambio de la participación en el estudio. Se usa el EPQ (Eysenck y Eysenck, 1975) para clasificar a los voluntarios en la dimensión extraversión/intraversión. La etapa de reclutamiento finaliza cuando se dispone de 30 adolescentes clasificados en cada condición.

A la mitad de los introvertidos se les asigna al azar a la condición de premio y a la otra mitad a la condición de castigo; de manera similar, se asignan los sujetos extrovertidos. A cada participante se le presenta un conjunto de cien tarjetas; en cada una de ellas hay escrito un verbo (por ejemplo, caminar) y la lista de pronombres personales (yo, tú, él, ella, nosotros, etc.). La tarea que deben hacer los chicos consiste en construir, para cada tarjeta, una oración que contenga el verbo y empiece por un pronombre personal. En función de la condición a la que ha sido asignado el participante, cuando usa los pronombres «yo, nosotros» el investigador lo refuerza verbalmente («bien») o lo castiga también verbalmente («pobre») durante una serie de ensayos. Al finalizar la recogida de datos, para cada participante se genera una puntuación que se usa como indicativo de la susceptibilidad al condicionamiento.

Los resultados indican que no hay diferencias entre la susceptibilidad al castigo de introvertidos y extrovertidos. En cuanto a los premios, se observan diferencias en el sentido de que los introvertidos son más susceptibles al condicionamiento.

1. ¿Cuál es la hipótesis de esta investigación?
2. Identifique y clasifique las variables de este estudio.
3. Esquematice y clasifique el diseño utilizado.

4. Indique cuántos grupos son necesarios para llevar a cabo el estudio.
5. Si el número de participantes por grupo es el mismo, cuántos participantes habrá en cada nivel o combinación de niveles de la(s) variable(s) independiente(s)?
6. ¿De qué manera puede afectar a la validez de la investigación el que la muestra esté formada por adolescentes que reciben una compensación por el hecho de participar en la misma?
7. ¿Los resultados del estudio sugieren la presencia de una interacción?
8. ¿Los resultados del estudio están planteados como efectos simples y/o como efectos principales?

Ejercicio 2.10: La aplicación del Programa Lidcombe a través de webcam para el tartamudeo temprano¹¹

Objetivo. El tratamiento con cámara web es potencialmente útil para el tratamiento de casos de tartamudeo temprano en el que los clientes están aislados de servicios de tratamiento especializados por razones geográficas y de otro tipo. El objetivo del presente ensayo fue comparar los resultados de la aplicación presencial y a través de webcam del tratamiento del programa Lidcombe (Packman et al., 2015) para el tartamudeo temprano.

Método. El diseño aplicado fue un ensayo controlado aleatorizado de no inferioridad. Los participantes fueron 50 niñas con una media de edad de 5 años al inicio del tratamiento. Los participantes se asignaron aleatoriamente al tratamiento estándar del programa Lidcombe o al tratamiento con el programa webcam Lidcombe. Los resultados se evaluaron a través del porcentaje de sílabas con tartamudeo a los 9 meses después de la aplicación del programa y de la gravedad de la tartamudez típica informada por los padres.

Resultados. No hubo pruebas suficientes de una diferencia postratamiento del porcentaje de sílabas tartamudeadas entre los tratamientos del programa Lidcombe estándar y de webcam. Tampoco se encontraron diferencias entre grupos en cuanto a la gravedad de la tartamudez típica reportada por los padres.

Conclusiones: Este ensayo confirmó la viabilidad de la intervención del programa webcam Lidcombe. Parece ser tan eficaz y económicamente viable como el tratamiento estándar (presencial) del programa Lidcombe.

1. Indique cuál/es son las VI.
2. Indique cuál/es son las VD.
3. Identifique las variables extrañas.
4. ¿Se ha controlado la variable Edad?
5. ¿Qué técnicas de control se han aplicado?
6. Haga el esquema del diseño.
7. ¿Cuál es el diseño utilizado?
8. Si se asignan el mismo número de participantes a cada condición experimental, ¿cuántos participantes habrá en cada condición experimental?

11. Adaptado de: BRIDGMAN, K., ONSLOW, M., O'BRIAN, S., JONES, M. y BLOCK, S. (2016). Lidcombe Program Webcam Treatment for Early Stuttering: A Randomized Controlled Trial. *Journal of Speech, Language, and Hearing Research*, 59(5), 932-939. https://doi.org/10.1044/2016_JSLHR-S-15-0011

Ejercicio 2.11: Tratamiento computarizado de la anomia

El tratamiento asistido por ordenador de la anomia es cada vez más frecuente dada la facilidad que aporta a la hora de aumentar la intensidad y la duración de los tratamientos. Los tratamientos asistidos por ordenador para la anomia se pueden clasificar en supervisados y no supervisados. Por otro lado, el hecho de ofrecer pistas a la hora de nombrar palabras puede ayudar a mejorar la anomia. Las pistas pueden ser de diversos tipos, pero comúnmente se utilizan pistas fonológicas, semánticas o múltiples pistas (*multicue*).

Unos investigadores mantienen la hipótesis que el aprendizaje supervisado combinado con múltiples pistas es el tratamiento más eficaz para la anomia. Para contrastar esta hipótesis, combinaron tanto el tratamiento computarizado supervisado como el no supervisado con tres tipos de pistas: fonológicas, semánticas y *multicue*.

Seleccionaron una muestra de 120 personas afásicas, con un grado de afectación similar de anomia, de edades comprendidas entre 35 y 40 años. Cada participante fue asignado aleatoriamente a una de las condiciones experimentales que resultan de combinar la supervisión (presente / ausente) con las diferentes pistas. El grado de anomia se evaluó a través de las puntuaciones obtenidas con el *Boston Naming Test* (Kaplan, Goodglass y Weintraub 2001).

Los resultados indicaron que el tratamiento computarizado supervisado es más eficaz que el no supervisado sólo cuando se combina con *multicue*. Cuando las pistas eran fonológicas o semánticas no se observaron diferencias entre ambos tipos de tratamiento.

1. Indique cuál/es son las VI.
2. Indique cuál/es son las VD.
3. Identifique las variables extrañas.
4. ¿Qué técnicas de control se han aplicado?
5. Haga el esquema del diseño.
6. ¿Cuál es el diseño utilizado?
7. Si se asignan el mismo número de participantes a cada condición experimental, ¿cuántos participantes estarán asignados a la condición de tratamiento supervisado con pistas fonológicas?
8. Si todos los participantes hubieran recibido ambos tipos de tratamiento computarizado
 - a. ¿Cuál sería el diseño utilizado?
 - b. Haga el esquema del diseño de la pregunta anterior.
 - c. En este caso, ¿cuántos participantes habría en cada condición experimental?
9. Si el sexo es una variable extraña, ¿podríamos controlarla a través de bloqueo? En caso afirmativo, ¿cómo se llevaría a cabo ese bloqueo?
10. ¿Se podría controlar el sexo aplicando alguna otra técnica de control?
11. ¿Se puede estudiar la interacción en esta investigación?
12. ¿Los resultados están formulados en términos de efectos principales o de efectos simples?
13. ¿Considera que hubo interacción?

LECTURA CRÍTICA

Busque los artículos del siguiente cuadro. Para hacer la investigación le sugerimos que utilice el Google Scholar y/o la base de datos Pubmed.

A partir del resumen y de la información del apartado «Método», haga un análisis metodológico del artículo y complete la ficha de lectura crítica que encontrará a continuación. Excluya de este análisis todas las cuestiones relativas al análisis de datos y a la validez de conclusión estadística, ya que escapan a los contenidos expuestos en este texto introductorio.

Artículo 1

Título: «Eficacia del entrenamiento cognitivo basado en nuevas tecnologías en pacientes con demencia tipo Alzheimer»

Autores: Bernardino Fernández-Calvo, Roberto Rodríguez-Pérez, Israel Contador, Alicia Rubio-Santorum y Francisco Ramos

Revista: *Psicothema*

Año de publicación: 2011

Artículo 2

Título: «Emoción y memoria de reconocimiento: la discriminación de la información negativa como un proceso adaptativo»

Autores: Fernando Gordillo León, José María Arana Martínez, Lilia Mestas Hernández, Judith Salvador Cruz, Juan José García Meilán, Juan Carro Ramos y Enrique Pérez Sáez

Revista: *Psicothema*

Año de publicación: 2010

Artículo 3

Título: «Integrated motivational interviewing and cognitive behavioural therapy for people with psychosis and comorbid substance misuse: randomised controlled trial»

Autores: Christine Barrowclough, Gillian Haddock, Til Wykes, Ruth Beardmore, Patricia

Conrod, Tom Craig, Linda Davies, Graham Dunn, Emily Eisner, Shôn Lewis, Jan Moring, Craig Steel i Nicholas Tarrier

Revista: *BMJ*

Artículo 4

Título: «Group therapy for adolescents with repeated self harm: randomised controlled trial with economic evaluation»

Autores: J. M. Green, A. J. Wood, M. J. Kerfoot, G. Trainor, C. Roberts, J. Rothwell, A. Woodham, E. Ayodeji, B. Barrett, S. Byford y R. Harrington

Revista: *BMJ*

Año de publicación: 2011

CRITICAL READING CARD: ARTICLE
(FICHA DE LECTURA CRÍTICA: ARTÍCULO)



Remember to thoroughly read the article once found and extract the following information. If you obtain a summary which does not provide enough information you may mark «lack of information» in the corresponding box.

-1- Write the article's **bibliographic reference** according to the APA's referencing style

--

-2- Make a **list of the methodological terms** that appear in the method section not related with the study's subject. The first column includes the term in English and the other two the translation in Catalan and Spanish. Include the new terms in your English/Catalan/Spanish Vocabulary.

English	Catalan	Spanish

-3- Indicate which are the **objectives** and **hypothesis** of the research.

--

-4- Does the study address a clearly focused issue? An issue can be 'focused' in terms of the population studied, the intervention given, the comparator given or the outcomes considered.

--

-5- Identify which are the **independent and dependent variables (IV and DV)**.

IV	DV

-6- Is exposure to IV levels measured in a standard, valid and reliable way?

--

-7- Are DV measured in a standard, valid and reliable way?

--

-8- Is it possible to identify the **sampling procedure used**? If so, just copy the information presented about it.

--

-9- What percentage of individuals refused to participate?

--

-10- Classify the design used according to the criteria presented in this chapter (number of IV and DV, comparison strategy and randomization type).

--

-11- What factors other than the IV could affect the DV? Include potential confounding factors like differences in baseline characteristics between groups.

--

-12- Are individuals within groups blind to which intervention group they belong AND are those delivering the intervention blind to the intervention group?

--

-13- Aside from the experimental intervention, were the groups treated equally?

--

-14- If there are any within-subject IV, aside from the experimental intervention, were the periods homogeneous?

--

-15- Identify the **control techniques** used and which variables have been controlled with.

--

-16- Link the previous answers with the concepts of internal and external validity applied to the conclusions of this study.

Internal Validity	
External validity	

SOLUCIONES DE LA ACTIVIDAD DE IDENTIFICACIÓN DE DISEÑOS

1. Estudio de los efectos de la aplicación de un programa de entreno en habilidades sociales dirigido a madres y a sus hijos
Diseño experimental intersujeto de tres grupos

<i>VI: tipo de entreno</i>	
G. exp1: Solo a madres	
G. exp2: A madres e hijos/as	
G. control	
VD: Habilidades sociales	

2. Efectos de la visión de escenas violentas en la conducta agresiva infantil
Diseño experimental intersujeto de grupos aleatorios con bloques

<i>Bloque: sexo</i>	<i>VI: grado de violencia</i>	
[Chicos]	ficticia justificada	
	ficticia injustificada	
	neutras	
[Chicas]	ficticia justificada	
	ficticia injustificada	
	neutras	
VD: agresividad		

3. Calidad percibida del vino
Diseño experimental intrasujeto

<i>VI: información de la etiqueta de las botellas</i>			
«vino catalán»	«vino español»	«vino italiano»	«vino francés»
VD: sobre la calidad percibida del vino			

4. Análisis de la influencia de la conducta en la ilusión de control
Diseño experimental intrasujeto

<i>VI: tipo de prueba</i>	
control libre	predicción
VD: juicio ilusorio	

5. Tratamiento del dolor postoperatorio
Diseño experimental factorial intersujeto 3 x 3

<i>VI₁: técnicas tratamiento</i>	<i>VI₂: días antelación de la aplicación</i>	
placebo	15 días	
	7 días	
	0 días	
acupuntura	15 días	
	7 días	
	0 días	
terapia musical	15 días	
	7 días	
	0 días	
VD: dolor postoperatorio		

6. Tiempo de reacción
Diseño experimental factorial intersujeto 2 x 3

<i>VI₁: tipo de estímulos</i>	<i>VI₂: tipo de situaciones ambientales</i>	
luz blanca	silencio	
	música clásica	
	música rock	
ruido grave	silencio	
	música clásica	
	música rock	
VD: tiempo de reacción		

7. El entorno psicosocial y los neurotransmisores

Diseño experimental factorial mixto 2×2

	<i>VI₂: dosis antagonista de la benzodiazepina</i>	
VI ₁ : control ambiente	baja	alta
sí		
no		
VD: signos de ansiedad		

8. La influencia normativa

Diseño experimental factorial mixto 2×3

	<i>VI₁: situación de ambigüedad</i>	
VI ₂ : situación de identificación	sí	no
individual		
en pareja		
grupala		
VD: influencia normativa		

Notas

- i La forma en que delimitamos la experimentación en este texto no es la única posible y se ajusta a lo que Cook y Campbell (1986) denominaban *tradición del control estadístico*. A lo largo del siglo xx esta tradición se ha ido consolidando como el enfoque normativo de la investigación experimental para las ciencias sociales y las ciencias de la salud. Así, este enfoque se presenta en muchos textos de psicología bajo el epígrafe de *enfoque experimental clásico o tradicional* (por ejemplo, Arnau, 1995; Viader, 1996). Además, bajo esta tradición se ha diferenciado el experimento verdadero del cuasiexperimento (Shadish, Cook y Campbell, 2002) que desarrollaremos en el capítulo 3.
- ii Para una completa revisión de las diferentes formas de apareamiento, se puede consultar Ato (1991). Dentro de esta revisión se destaca la diferencia que supone el hecho de incorporar la variable de bloqueo dentro del diseño experimental para sistematizar su estudio, respecto a otras formas de apareamiento, y también se exponen sus repercusiones en el modelado posterior de los datos.
- iii Para una introducción a los diferentes tipos de contrabalanceo, puede consultar el capítulo 8 de León y Montero (2015).
- iv Se pueden usar índices estadísticos distintos a la media, pero, dado el carácter introductorio del texto, simplificaremos el cálculo de efectos y lo presentaremos únicamente en términos de diferencia entre medias. Encontrará una introducción a las pruebas estadísticas de comparación intersujeto e intrasujeto en Domènech, Granero, Losilla y Portell (1999) y en Pardo y San Martín (2010).
- v Encontrará una introducción a estos diseños en Portell y Domènech (1997).
- vi En este texto asumimos una definición muy genérica de diseño factorial, pero en sentido estricto hay que diferenciar entre los diseños totalmente aleatorios de aquellos que solo presentan aleatorización parcial, como el diseño *split-plot* o de muestra dividida; se puede profundizar en esta distinción en Balluerka y Vergara (2002).
- vii La lista de ventajas no es exhaustiva, dado que omitimos aquellas que para comprenderlas es necesario tener conocimientos sobre estadística inferencial. Para una exposición de las ventajas del diseño factorial desde el punto de vista de la reducción de la varianza de error y la potencia de la prueba, se puede consultar Balluerka y Vergara (2002).
- viii El concepto de interacción que se desarrollará en esta ficha es aplicable al efecto de moderación descrito en la ficha 1.6, dado que, como está explicado, la moderación es un caso particular de la interacción.
- ix Para simplificar la explicación se hace una presentación del concepto de simple y principal basada únicamente en el análisis gráfico y descriptivo de la diferencia entre medias. Esta aproximación a la didáctica de la interacción está basada en los trabajos de León y Montero (2001, 2003), que recomendamos para ampliar los conceptos que aquí se presentan. Este abordaje debe completarse con el análisis estadístico inferencial, que permite tomar decisiones sobre la significación estadística de estas diferencias, y con la definición de los correspondientes criterios clínicos para valorar la relevancia de las diferencias encontradas. Encontrará información sobre el análisis estadístico inferencial del diseño factorial en Balluerka y Vergara (2002) y en Pardo y San Martín (2010).

3. Diseños cuasiexperimentales

PRESENTACIÓN

La cuasiexperimentación se ha definido como una investigación que tiene todos los elementos de un experimento, con una notable excepción: la aleatorización. En la investigación aplicada son muchas las ocasiones en las que se hace difícil distribuir aleatoriamente a los participantes a las diferentes condiciones experimentales. ¿Ponemos un ejemplo?

Se desea evaluar una nueva intervención formativa higiénico-sanitaria dirigida a trabajadores de empresas de restauración comunitaria. La intervención combina sesiones teóricas con prácticas y discusión en pequeños grupos de trabajo, y se distribuye en cinco sesiones a lo largo de cuatro semanas. Para evaluar la eficacia de la intervención, se quiere disponer de dos grupos: uno que recibirá la intervención y uno que actuará como control. Debido a la duración de la intervención, la empresa que ha accedido a participar en el estudio plantea restricciones en cuanto a los centros de trabajo y los turnos en los que se puede aplicar. Finalmente, la intervención se aplica a los trabajadores involucrados en la manipulación de alimentos que hacen el turno de tarde en uno de los centros de trabajo más grandes de la empresa. A los trabajadores del turno de la mañana de ese mismo centro no se les hace la intervención formativa. Antes de comenzar la intervención, se miden los conocimientos higiénico-sanitarios de los trabajadores que manipulan alimentos en los dos centros. Una semana después de finalizar la intervención, se mide nuevamente el conocimiento de todos los trabajadores.

Fíjese que la intervención (variable independiente, en adelante VI) es manipulable, es decir, que el investigador tiene la «potestad» de crear y aplicar los niveles de la VI. Haciendo uso de esta «potestad», establece que son dos los niveles que puede tomar la VI (formación / ausencia de formación) y estos son administrados en las condiciones que ha definido. Si revisa la ficha 1.5, en la que se presenta el concepto de VI manipulada, comprobará que este ejemplo se ajusta al concepto de manipulación. Ahora bien, ¿qué implicaría hacer aquí asignación al azar? Pues que trabajadores ubicados en diferentes centros y/o turnos puedan recibir la intervención formativa, mientras que otros trabajadores en los mismos centros y turnos no reciban la intervención. Los problemas logísticos asociados son evidentes, hasta el punto de que en algunas situaciones aplicadas puede ser inviable la asignación al azar. Hay que tener en cuenta que la falta de aleatorización obliga al investigador a ser más cauto a la hora de emitir conclusiones respecto a una posible relación causa-efecto, puesto que sin la aleato-

rización difícilmente se puede garantizar que los grupos que estamos comparando (ej., formación / ausencia de formación) estén equilibrados en cuanto a variables extrañas conocidas o desconocidas.

Entre las razones que limitan el uso del control basado en la aleatorización, se encuentran las siguientes:ⁱ

- Ética: los estudios no aleatorizados a veces son la única alternativa ética para hacer investigación. Considere, por ejemplo, una situación en salud pública en la que la urgencia de una intervención formativa hace difícil dejar potenciales usuarios sin que la reciban. Casos como este limitan el uso no solo de la aleatorización, sino también de todas aquellas alternativas que se basan en la comparación con un grupo control sin tratamiento.
- Coste: los estudios no aleatorizados son habitualmente más baratos que sus homólogos experimentales.
- Naturalidad: los estudios no aleatorizados suponen una menor alteración de la situación cotidiana.

En las fichas 3.2, 3.4 y 3.5 se introducen un conjunto de diseños propuestos para contrastar hipótesis causales en situaciones en las que no es posible aplicar aleatorización. Previamente, en la ficha 3.1 se presentan los criterios para establecer los límites entre estos diseños cuasiexperimentales y los diseños experimentales explicados en el capítulo previo.

Los diseños que presentaremos en este tema son muy relevantes para comprender mucha de la investigación aplicada que se publica en psicología, logopedia y en ciencias afines. Así, por ejemplo, proliferan en el ámbito de la evaluación de programas y de manera especial en la evaluación de intervenciones en contexto educativo, en salud comunitaria o en el ámbito de las políticas de prevención.ⁱⁱ

Los autores vinculados al desarrollo de los diseños cuasiexperimentales han realizado también valiosas aportaciones a la teoría de la validez de la inferencia causal (por ejemplo, Campbell y Stanley, 1966; Cook y Campbell, 1979; Shadish, Cook y Campbell, 2002). Por este motivo, hemos escogido este capítulo para ampliar la noción de validez introducida en el capítulo 1. En la ficha 3.3 se presenta esta ampliación y se revisan las principales amenazas a los distintos tipos de validez.

A partir del estudio de las fichas siguientes se espera que el lector sea capaz de:

- Entender y saber diferenciar entre experimento y cuasiexperimento (ficha 3.1).
- Saber clasificar y representar los diseños cuasiexperimentales (fichas 3.2, 3.4 y 3.5).
- Entender y saber diferenciar entre cuatro tipos de validez (ficha 3.3).
- Conocer las principales amenazas a la validez (ficha 3.3).

FICHA 3.1. PARADIGMA MANIPULATIVO: CRITERIOS DE DEMARCACIÓN

Los diseños experimentales y cuasiexperimentales se aplican en estudios dirigidos a contrastar hipótesis causales. Estas dos familias de diseños, a parte del objetivo, también comparten el hecho que implican la manipulación de la VI cuyo efecto pretenden establecer. Tal como ilustra la siguiente tabla, la diferencia entre el diseño experimental y el cuasiexperimental está en la presencia o no de aleatorización.

Diseño	Criterios	
	Manipulación	Aleatorización
Experimental	Sí	Sí
Cuasiexperimental	Sí	No

En la investigación aplicada que tiene por objetivo cuantificar el efecto de una intervención, encontramos muchos casos en los que el investigador puede mantener algunos aspectos similares a un diseño experimental (por ejemplo, crear y aplicar los niveles de la VI o mantener constantes ciertas variables), pero, en cambio, resulta muy dificultoso usar técnicas de aleatorización para determinar «el que» o «cuando» recibe la intervención. A modo de ejemplo, considere un estudio para valorar el efecto del aprendizaje basado en problemas (ABP) sobre el rendimiento académico y compare el impacto diferencial sobre la vida cotidiana de los participantes que pueden tener los dos planteamientos siguientes:

1. En una asignatura que se imparte en grupos de mañana y de tarde, al segundo se le aplica metodología de ABP y al de la mañana, la técnica expositiva tradicional.
2. En una asignatura que se imparte en grupos de mañana y de tarde, los estudiantes son repartidos al azar en dos grupos, uno de los cuales recibe la asignatura siguiendo metodología de ABP y el otro, la técnica expositiva tradicional.

Son obvios los inconvenientes que puede suponer el segundo planteamiento, que requeriría que algunos o todos los estudiantes cambiaran su horario habitual para seguir la asignatura en el grupo al que han sido asignados, con el efecto que esto tendría sobre el horario del resto de asignaturas de su grupo de matrícula. El primer planteamiento, que es el cuasiexperimental, conlleva claras ventajas en casos como estos, si bien conviene profundizar en sus limitaciones para interpretar y valorar correctamente la validez de los datos que se obtienen.

La falta de aleatorización de los participantes a los grupos de tratamiento supone una reducción en el grado de control interno. La importancia atribuida a la aleatorización (ya sea asignación al azar, bloqueo o contrabalanceo) se explica por el control que ejercen estas técnicas sobre las potenciales variables confundidoras desconocidas por el investigador, lo que revierte en un incremento de la validez interna. Sin embargo, el hecho de que una investigación sea experimental porque se cumplen los criterios de manipulación y aleatorización no garantiza *per se* que obtengamos un alto grado de validez interna.

Ejercer un alto grado de control interno implica diseñar la investigación de tal manera que las explicaciones alternativas a la propuesta por la hipótesis de investigación sean lo menos plausibles posible. El control interno no se puede delegar únicamente en la manipulación y la asignación aleatoria, sino que, habitualmente, será necesario conocer las potenciales amenazas a la validez interna y ejercer un control específico sobre las mismas.

Así, el hecho de obtener un alto grado de validez interna termina siendo la resultante de la correcta aplicación tanto de técnicas de control propias del diseño experimental —como la asignación al azar o el contrabalanceo—, como de técnicas de control de uso más general —como, por ejemplo, la constancia. Así pues, se puede deducir que la falta de asignación aleatoria no implica ausencia de validez interna, sino una potencial disminución de la misma. Y es que hay que recordar que la validez interna no es una propiedad dicotómica, sino que es una cuestión de grado.

FICHA 3.2. ESTRUCTURAS BÁSICAS DE COMPARACIÓN

Los diseños cuasiexperimentales han sido sistematizados y popularizados por Donald T. Campbell, Thomas D. Cook, William R. Shadish y Julian C. Stanley (por ejemplo, Campbell y Stanley, 1966; Cook y Campbell, 1979; Shadish, Cook y Campbell, 2002).ⁱⁱⁱ Presentamos, en primer lugar, las estructuras básicas de comparación a partir de las cuales se configuran los diseños cuasiexperimentales que presentamos en fichas posteriores.

Para la esquematización de los diseños seguiremos la notación clásica de Campbell y Stanley (1966), siendo:

- O_i observación o medida de la variable dependiente (en adelante, VD) en el momento «i»;
- X tratamiento o intervención; contraviniendo ligeramente la notación original, en los diseños que involucran a más de un grupo usaremos el guión bajo («_») para clarificar el espacio en blanco que se utiliza para simbolizar la ausencia de intervención;
- línea discontinua que indica que los grupos que separa son no equivalentes.

Diseño pretest-postest de un solo grupo	$O_1 X O_2$
Diseños solo postest con grupo no equivalente	$\begin{array}{c} X O_1 \\ \text{---} \text{---} \text{---} \\ _ O_1 \end{array}$

La primera estructura básica de comparación está incluida en el **diseño pretest-postest** de un solo grupo. En este diseño se mide la VD antes y después de aplicar una intervención. De este modo, la O_1 simboliza la línea base u observación (medida de la VD) preintervención, la X la aplicación de la intervención y la O_2 la observación (medida de la VD) postintervención. El fundamento inferencial de este diseño es la comparación entre el pretest (O_1) y el postest (O_2). Se espera que, como consecuencia de la intervención, los resultados en O_2 sean diferentes a los observados en O_1 .

Ejemplo: Siguiendo con el ejemplo presentado en la introducción, este diseño supondría usar un único grupo de trabajadores y seguir el siguiente procedimiento: 1) medir los conocimientos higiénico-sanitarios de los trabajadores de la muestra, 2) aplicarles la intervención formativa y 3) volver a medir sus conocimientos higiénico-sanitarios con el mismo instrumento usado para obtener la medida pretest. El resultado esperado es que los conocimientos observados después de la intervención sean superiores a los que mostraban al inicio.

La segunda estructura básica de comparación está incluida en los **diseños solo postest con grupo no equivalente**. En este diseño, un grupo que ha experimentado la intervención X se compara con un grupo que no lo ha hecho con el fin de establecer el efecto de X.

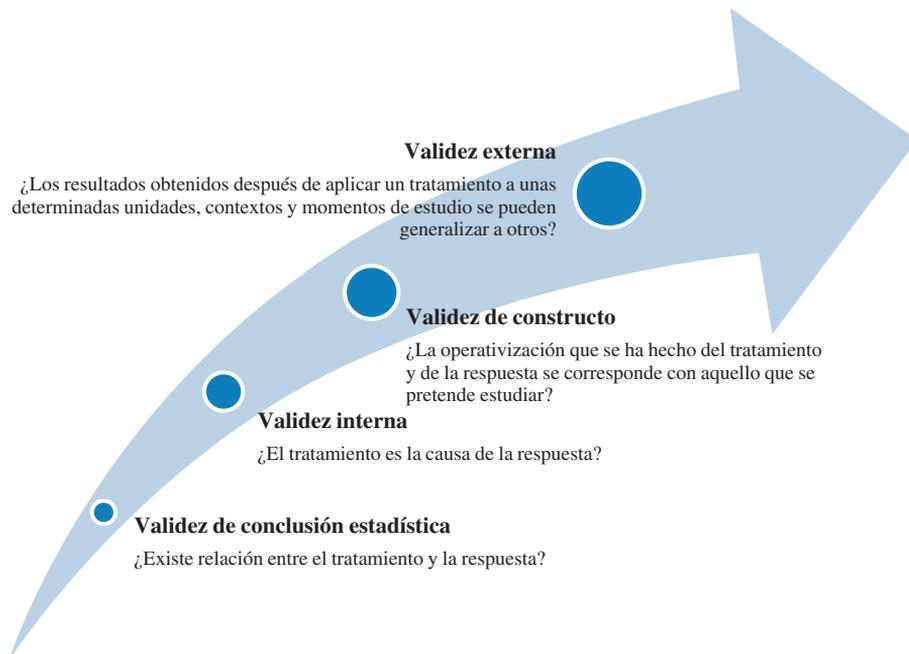
Ejemplo: Siguiendo con el ejemplo sobre la evaluación de la intervención para incrementar los conocimientos higiénico-sanitarios, este diseño supone definir dos grupos de trabajadores, uno de tratamiento y uno de control, procurando que sean lo más similares posible. Tal y como refleja el esquema de este diseño, la medida de los conocimientos higiénico-sanitarios (variable dependiente) se realiza en el mismo momento en los dos grupos (O_1). Se espera que, como consecuencia de la intervención, los resultados en O_1 del grupo de tratamiento sean diferentes a los observados en O_1 en el grupo sin intervención. Este grupo se denomina grupo control «no equivalente» para destacar que la asignación de los participantes en los grupos no se ha hecho al azar y que, por tanto, incrementa el riesgo de que los dos grupos no estén equilibrados respecto a las variables extrañas (en la notación propuesta, este punto queda simbolizado con la línea discontinua), de tal manera que aumenta el riesgo de falta de comparabilidad entre los grupos.

Los dos diseños presentados, que denominaremos **preexperimentales**, difícilmente permiten realizar inferencias causa-efecto válidas, ya que están expuestos a múltiples amenazas a la validez interna. En la ficha 3.3 se presenta una sistematización de las principales amenazas que hay que tener en cuenta.

FICHA 3.3. VALIDEZ Y AMENAZAS

En el tema 1 ya se ha introducido el concepto de validez (ficha 1.7) indicando, de forma genérica, que la validez de una investigación hace referencia al grado en que estudia lo que pretende estudiar. Del mismo modo, decimos que un instrumento es válido si mide realmente esa variable o aquel constructo que pretende medir. Por tanto, la validez está relacionada con la idea de «verdad» o «certeza»; concretamente, con el grado de confianza respecto de la veracidad de los datos obtenidos. Con este comentario queremos destacar, también, que la validez debe asociarse al conocimiento obtenido a partir de una investigación y evitar juicios apriorísticos sobre la validez de determinados métodos y diseños. Es necesario, también, tener presente que nunca se puede tener la certeza absoluta de que las inferencias que se hacen a partir de los resultados de una investigación sean correctas o ciertas. Así, no hablamos de la validez de un estudio en términos absolutos, sino que nos referimos a la misma en términos de grado.

La siguiente figura sintetiza los principales conceptos utilizados en la valoración de la validez de una investigación dirigida a contrastar una hipótesis causal.^{iv}



Antes de definir con más detalle cada tipo de validez, nos gustaría que prestase atención a dos aspectos. En esta figura se ponen en relación cinco elementos relevantes para hacer una valoración de validez que se suelen identificar con el acrónimo UTOST, siendo:

- u: unidades en las que se estudia la relación entre el tratamiento y la respuesta (normalmente personas)
- t: tratamiento o intervención cuyos efectos se quieren estudiar
- o: observaciones (*Observations* o *Outcomes*) sobre el resultado del tratamiento (respuestas)
- s: *setting* o contexto de estudio
- t_i: tiempo o momento de estudio

Un segundo aspecto que queremos destacar a partir de esta figura es el carácter acumulativo de las cuestiones relacionadas con la validez, por lo que la figura debe leerse en sentido ascendente. Así, por ejemplo, tiene poco sentido preguntarse sobre la existencia de una relación causal si previamente no se aporta evidencia de la validez de conclusión estadística sobre tal relación. Del mismo modo, tiene poco sentido preguntarse sobre la generalización de una relación causal en contextos diferentes si previamente no se aportan evidencias sobre la existencia de tal relación en el contexto y entre los participantes en el estudio.

Pasamos a continuación a profundizar en cada tipo de validez y a presentar los principales factores que pueden amenazarla, teniendo en cuenta que se consideran amenazas de la validez las razones específicas por las que podemos estar parcial o completamente equivocados a la hora de inferir, por ejemplo, una covariación, relación causal, representatividad de los constructos, o a la hora de generalizar estas relaciones a otras poblaciones, situaciones o momentos históricos.

- La **validez de conclusión estadística** se refiere a la validez de la inferencia sobre: 1) si existe relación (covariación) entre el tratamiento y el resultado, y 2) cuán fuerte es esta relación.

Son factores que amenazan la validez de conclusión estadística:^v la baja potencia estadística, la violación de los supuestos de las pruebas estadísticas, el problema de la tasa de error, la falta de fiabilidad de las medidas, las restricciones de rango de los valores de las variables, la falta de fiabilidad de la administración de los tratamientos, la sobrevariabilidad en el contexto de investigación, la heterogeneidad aleatoria de las unidades de respuesta y la estimación imprecisa del tamaño del efecto.

- La **validez interna** se refiere a la validez de la inferencia sobre si la covariación observada entre el tratamiento (t) y el resultado (o) refleja una relación causal en el sentido de que fue hipotetizada (el tratamiento como causa del resultado: $t \rightarrow o$). Para fundamentar esta inferencia, hay que aportar pruebas que permitan valorar en qué grado los resultados se habrían producido en ausencia del tratamiento. Así, la validez interna de una investigación aumenta en la medida en que se aporta evidencia que ayude a descartar las hipótesis rivales o explicaciones del resultado alternativas al efecto de la propia VI.

A continuación, definimos las principales amenazas a la validez interna indicando, primero, el término que se suele usar para designarlas:^{vi}

1. *Precedencia temporal ambigua*: falta de claridad en el orden en que actúan las variables. En el esquema causal que estamos usando, para poder concluir que el

tratamiento es la causa del resultado observado, es necesario que el tratamiento se haya producido antes de la medida del resultado (es la condición de «temporalidad» presentada en la ficha 1.6). Esta condición queda garantizada por cuanto el tratamiento es una variable manipulada por el investigador y esta manipulación se realiza adecuadamente. Volviendo al ejemplo de la introducción, hablaríamos de precedencia temporal ambigua si en un centro de trabajo no hubiera garantía de que la medida del conocimiento postratamiento se hace después de la formación.

2. *Selección*: diferencias sistemáticas de los participantes, anteriores a la aplicación de una intervención (esto es, aplicación de los niveles de la VI) que pueden causar el efecto (el resultado) observado. Por ejemplo, suponga que se usa un diseño solo postest con un grupo control no equivalente para estudiar el efecto de un nuevo método de entrenamiento en natación para niños y que el grupo de tratamiento está formado por voluntarios, mientras que el grupo control lo forman niños que acuden a un centro de verano. Es posible que los del primer grupo estén mucho más motivados por el aprendizaje de la natación y que la eventual mejora en el rendimiento esté más relacionada con esta motivación que con la intervención administrada.
3. *Historia*: todos los acontecimientos externos (excepto el tratamiento) que ocurren entre el inicio de la implementación del tratamiento y la medida del resultado; el hecho de descartar la amenaza de la historia supone descartar que estos acontecimientos externos son la causa de los resultados observados. Un ejemplo de esta amenaza lo encontramos en el siguiente caso. Suponga que se mide la dieta restrictiva de un grupo de chicas adolescentes con normopeso antes y después de recibir una intervención para mejorar su autoestima, y que durante el periodo de intervención un reportaje sobre famosas gravemente afectadas por trastornos del comportamiento alimentario tiene un fuerte impacto mediático. En caso de que haya una reducción en la dieta restrictiva de las adolescentes, el efecto de la intervención está confundido con el efecto del reportaje.
4. *Maduración*: cambios naturales que experimentan los participantes en un estudio y que se darían en ausencia del tratamiento, como, por ejemplo, envejecer, tener hambre o adquirir experiencia. Estos cambios suponen una amenaza cuando se confunden con el efecto del tratamiento. Así, por ejemplo, considere el caso en que la evaluación del efecto de un programa de acogida en niños inmigrantes se basa solo en la comparación de resultados entre el rendimiento inicial y el observado al final del curso. Hay muchos aspectos madurativos que pueden explicar el cambio observado en parte de la introducción del programa de acogida.
5. *Regresión estadística (en la media)*: cuando los participantes presentan puntuaciones extremas en alguna variable, estas puntuaciones tienden a ser menos extremas cuando las volvemos a medir y este cambio se puede confundir con el efecto del tratamiento. Así, por ejemplo, suponga que se aplica un entrenamiento específico para mejorar los aciertos en los tiros libres en el baloncesto. Se empieza por aquellos jugadores que tienen el peor porcentaje de aciertos en los últimos dos partidos (< 10 %) y se observa que este porcentaje mejora. Esta mejora se podría atribuir a la regresión en la media de las puntuaciones extremas.

6. *Pérdida no aleatoria de sujetos (mortalidad selectiva)*: cuando diferentes tipos de participantes abandonan selectivamente diferentes condiciones que se comparan en el estudio. Ponemos un ejemplo. Suponga que para evaluar una dieta para niños con obesidad se comparan los resultados obtenidos en una escuela en la que se aplica la dieta con los obtenidos en niños de características similares que acuden a otra escuela en la que no se aplica la dieta. Hablaríamos de mortalidad selectiva si, una vez iniciada la intervención, un número importante de los niños obesos de la escuela que aplicaba la dieta especial hubieran dejado de usar el servicio de comedor, mientras que los niños obesos de la otra escuela hubieran seguido usando el servicio.
7. *Administración de pruebas*: el hecho de realizar una medida previa de una variable puede influir en las siguientes medidas hechas con la misma prueba, por lo que la práctica, la familiarización o factores relacionados se pueden confundir con el efecto de la intervención que se evalúa. Así, por ejemplo, suponga que se usa el rendimiento en una tarea de memoria hecha por ordenador como VD en un diseño pretest-postest de un solo grupo, dirigido a evaluar el efecto de un taller de memoria para gente mayor. La práctica en la tarea puede hacer que en la segunda ocasión de medida se obtengan mejores resultados solo por el hecho de repetirla.
8. *Instrumentación*: cuando se produce un cambio en el instrumento de medida, ya sea humano o mecánico, a lo largo del período de estudio, lo que provoca cambios en los resultados observados que se pueden confundir con el efecto del tratamiento. Así, por ejemplo, hablaremos de sesgo de instrumentación si en un curso sobre mejora de la redacción y estilo de escritura académica la profesora que corrige las redacciones modifica el criterio de corrección al evaluar las que se entregan al final con respecto a las que se entregaron al inicio del curso.

Estas amenazas no tienen por qué darse de manera aislada e independiente, y suelen presentarse de forma combinada. Por tanto, estos efectos pueden interactuar e influir así en los resultados de la investigación. Para minimizar estas amenazas, hay que tenerlas presentes al planificar el diseño incorporando técnicas de control y también, una vez recogidos los datos, mediante técnicas de control estadístico.

- **Validez de constructo**: grado de correspondencia entre las instancias particulares de un constructo y el concepto representado por el constructo. El término constructo hace referencia a un concepto teórico inobservable (por ejemplo, ansiedad, depresión, inteligencia, personalidad, aptitudes, actitudes, etc.). La definición operativa de estos constructos presenta considerables dificultades. Hay que tener en cuenta que todos los elementos antes mencionados al referirnos al concepto «utost;» suponen instancias particulares de conceptos genéricos que se simbolizan con el acrónimo en mayúsculas (UTOSTi). Así, por ejemplo, nos podemos referir a una formación sobre lavado de manos (T) que en el estudio se instancia en unos rótulos específicos con un mensaje motivacional y unas instrucciones (t).

La validez de constructo se ve amenazada por aquellos factores que comprometen la correspondencia entre lo que ha sucedido durante el estudio y los constructos utili-

zados para describir los resultados obtenidos. A continuación, presentamos una relación de amenazas a la validez de constructo:^{vii}

1. *Inadecuada explicación de los constructos*: no explicar de manera precisa un constructo puede llevar a inferencias incorrectas sobre las relaciones entre la operativización y el constructo.
2. *Confusión de constructos*: al operativizar se pueden ver implicados más de un constructo, y la falta de descripción de todos ellos puede dar lugar a inferencias incompletas del constructo. Por ejemplo, si la autoestima se ve afectada por la autoconfianza, entonces el efecto de estos constructos se puede confundir.
3. *Sesgos derivados del uso de una operación única*: en este sentido, se aconseja que los constructos se operativicen de múltiples maneras. Por ejemplo, diferentes indicadores de una actitud.
4. *Sesgos derivados del uso de un solo método*: de forma similar que en el caso anterior, el consejo es diversificar las técnicas empleadas. Por ejemplo, usar medidas fisiológicas y autoinformes para medir la ansiedad.
5. *Confusión entre constructo y niveles de constructo*: esta confusión puede llevar a concluir sobre un aspecto general (por ejemplo, la higiene postural) cuando en realidad solo se ha incidido sobre un aspecto (por ejemplo, postura de un segmento corporal) .
6. *Sensibilidad del tratamiento a factores estructurales*: la estructura factorial de una medida puede cambiar como resultado de un tratamiento, y este cambio puede permanecer oculto si siempre se usa la misma medida.
7. *Cambios diferenciales en la reactividad a los autorregistros*: así, un participante puede modificar su interés en hacer un autoinforme cuando se da cuenta de que no ha sido asignado al grupo de tratamiento.
8. *Reactividad ante la situación experimental*: las respuestas de los participantes pueden reflejar no solo los efectos del tratamiento, sino también sus percepciones ante la situación experimental (o de tratamiento).
9. *Expectativas del experimentador*: el experimentador puede influir en las respuestas de los participantes transmitiendo sus expectativas sobre las respuestas deseadas.
10. *Efectos de novedad y interrupción*: los participantes pueden responder inusualmente bien ante una innovación o inusualmente mal por el hecho de interrumpir su rutina.
11. *Igualación compensatoria*: cuando una intervención provoca unos efectos deseables, los encargados de implementar las diferentes condiciones se pueden resistir a hacer algo muy diferente, lo que iguala las condiciones.
12. *Rivalidad compensatoria*: cuando los participantes asignados a la condición control se muestran motivados para mostrar que pueden obtener resultados similares a los que reciben en el tratamiento.
13. *Desmoralización o resentimiento*: supone que los participantes asignados a la condición más desfavorable pueden cambiar sus respuestas o incluso abandonar el estudio.
14. *Difusión de la intervención*: hace que los participantes asignados a una condición se expongan a estímulos que son propios de otras condiciones.

- **Validez externa:** como ya se ha comentado en el primer tema, la validez externa tiene que ver con el juicio sobre la generalización. De manera más específica, la podemos definir como la validez de la inferencia sobre si la relación causal se mantiene al variar las unidades (participantes), los contextos, las variables de tratamiento, las variables de resultado (*outcomes*) y los momentos de medida. Así, por ejemplo, si en un estudio sobre formación en higiene hecho con trabajadores estadounidenses del sector de la restauración se demuestra un incremento en las buenas prácticas, la valoración de la validez externa supone aportar pruebas para determinar el alcance de la generalización de esta relación a otros perfiles de trabajador, a adaptaciones de la intervención formativa en otras lenguas, en otros contextos de aplicación y a otros momentos.

Las amenazas a la validez externa se presentan en términos de interacciones. Esto es así porque la posibilidad de generalizar quedará limitada mientras el efecto observado no se mantenga al variar las unidades, los contextos, los momentos, las variables de tratamiento y las de resultados. A continuación, presentamos algunos casos ilustrativos de estas interacciones entre la relación causal y los siguientes factores:^{viii}

1. *Interacción con las unidades:* cuando el efecto hallado en ciertas unidades (participantes) no se mantiene en otras unidades.
2. *Interacción con las variaciones de tratamiento:* cuando el efecto hallado en una variación del tratamiento no se mantiene en otras variaciones del tratamiento, o cuando un tratamiento se combina con otros, o cuando solo se emplea parte de un tratamiento.
3. *Interacción con los resultados (outcomes):* cuando el efecto hallado sobre ciertos resultados (outcomes) no se mantiene en otros; por ejemplo, las conclusiones sobre la eficacia de una formación higiénico-sanitaria pueden variar si se mide el conocimiento de las buenas prácticas o la realización de buenas prácticas.
4. *Interacciones con el contexto:* el efecto hallado en un determinado lugar podría no mantenerse en otros contextos o entornos.
5. *Mediación dependiente de contexto:* cuando el efecto de una variable mediadora (véase la ficha 1.6) de una relación causal en un contexto no se mantiene en otros contextos.

FICHA 3.4. DISEÑOS CUASIEXPERIMENTALES: AMPLIACIÓN DEL NÚMERO DE GRUPOS Y/O DE LOS MOMENTOS DE MEDIDA

En la ficha 3.2 se han esquematizado dos diseños que están altamente expuestos a las amenazas a la validez interna presentadas en la ficha 3.3. Para mejorar estos diseños con vista a tener más garantías para inferir relaciones causales, existen dos opciones no excluyentes (ni exhaustivas): 1) ampliar el número de grupos mediante la incorporación de uno o más grupos de control, y (2) ampliar el número de observaciones o medidas de la VD. A continuación presentamos tres diseños representativos de ambos tipos de ampliación.^{ix}

Diseño pretest-postest con grupo control no equivalente

$$\begin{array}{c} O_1 X O_2 \\ \hline O_1 - O_2 \end{array}$$

Diseño simple de serie temporal interrumpida

$$O_1 O_2 O_3 X O_4 O_5 O_6$$

Diseño de serie temporal interrumpido con grupo control no equivalente

$$\begin{array}{c} O_1 O_2 O_3 X O_4 O_5 O_6 \\ \hline O_1 O_2 O_3 - O_4 O_5 O_6 \end{array}$$

El **diseño pretest-postest con grupo control no equivalente** es el diseño cuasiexperimental más popular. En su forma básica, consta de un grupo al que se administra la intervención X y un grupo de comparación sin tratamiento. Como es característico de los cuasiexperimentos, entre estos grupos que reciben los dos niveles de la VI del diseño no hay asignación aleatoria. En cuanto a la VD, se toman las observaciones en el mismo momento temporal y con el mismo instrumento de medida en dos ocasiones: un pretest, antes de administrar el tratamiento, y un postest, después de su administración. Aquí, la medida pretest contribuye a controlar el grado de no equivalencia inicial (sesgo de selección).

El ejemplo presentado en la introducción del tema se ajusta a este tipo de diseño. Podemos analizar sus resultados comparando el incremento del conocimiento que se observa en el grupo de la mañana que ha recibido la formación X con respecto al incremento de conocimiento que se observa en el grupo de la tarde que ha actuado como control. Al analizar estos resultados, la principal amenaza a la validez interna radica en la selección, es decir, en la falta de homogeneidad entre los dos grupos, que puede interactuar con otros factores como la maduración, la instrumentación, la regresión estadística y la historia. Así, por ejemplo, puede suceder que en el grupo de la tarde se concentren los trabajadores con una formación inferior o más inmigrantes con dificult-

tades para comprender la lengua o con cualquier otro rasgo distintivo que altere el cambio que se produce como consecuencia del tratamiento. También podría suceder que en el grupo de la tarde el supervisor encargado de recoger los cuestionarios que evalúan el conocimiento sea menos insistente con el retorno de los cuestionarios, con lo que se acaben recogiendo solo las respuestas de los trabajadores más motivados y que tienen más conocimiento.

Cualquier acción que contribuya a mejorar la comparabilidad de los grupos o a controlar estadísticamente las diferencias que puedan existir entre ellos supone una mejora a este diseño. Una vía de mejora sería el uso de la técnica de control del apareamiento (véanse las fichas 2.3 y 3.5). El diseño pretest-postest con un grupo control no equivalente también puede mejorar añadiendo medidas pretest, como hace el diseño con doble pretest que esquematizamos a continuación:

$$\begin{array}{c} O_1 O_2 X O_3 \\ \hline O_1 O_2 - O_3 \end{array}$$

Diseño simple de serie temporal interrumpida

El diseño simple de serie temporal interrumpida supone una ampliación del diseño pretest-postest de un solo grupo añadiendo series de observaciones de la misma variable a lo largo del tiempo (serie temporal) antes y después de aplicar X. El hecho de disponer de series de medidas pretratamiento y postratamiento permite valorar si el cambio observado en una medida (VD) se puede atribuir al efecto del tratamiento, controlando factores como la tendencia previa de respuesta, el comportamiento cíclico de la VD o la regresión a la media.

Como ejemplo de la aplicación de este diseño, suponga que en una empresa dedicada a la restauración comunitaria se realizan una serie de medidas durante un mes sobre las buenas prácticas en el lavado de manos. Una vez verificada la estabilidad y la regularidad de estas conductas, el servicio de prevención de la empresa implementa una intervención sobre las pautas para una adecuada higiene de las manos. A continuación, se toman nuevamente medidas sobre la práctica del lavado de manos de los trabajadores durante treinta días. Los efectos de la implementación del programa (X) se pueden apreciar en las diferencias entre las series obtenidas antes y después de X mostrando diferentes pautas^x que ilustramos en las siguientes figuras:

- cambio de nivel sin evidenciar tendencia (figura a)
- cambio de nivel sin cambio de tendencia (figura b)
- cambio de tendencia pero no en el nivel medio (figura c)
- cambio de nivel y de tendencia (figura d)

Figura a

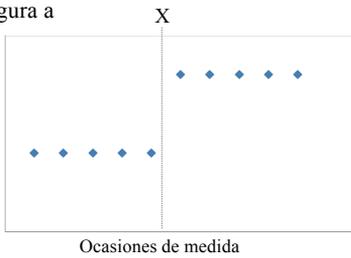


Figura b

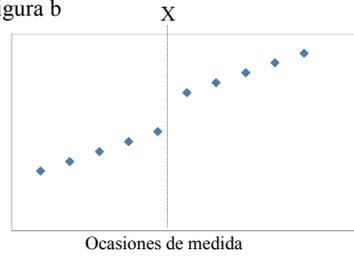


Figura c

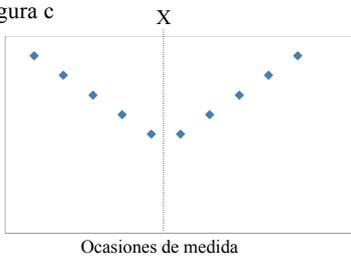
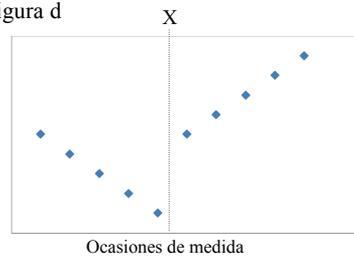


Figura d



Cabe señalar que los principales problemas que presenta este tipo de diseños están relacionados con el estudio de la estabilidad de la serie obtenida.

Diseño de serie temporal interrumpida con grupo control no equivalente

A partir del diseño anterior, se han desarrollado variantes. Entre ellas destacamos la que añade un grupo control que permitirá obtener más evidencia de cara a descartar amenazas como, por ejemplo, la de historia. Los diseños de serie temporal interrumpida con grupo control no equivalente constituyen uno de los cuasiexperimentos más efectivos, pero también de los más costosos.

FICHA 3.5. DISEÑOS CUASIEXPERIMENTALES: ALTERNATIVAS PARA LA FORMACIÓN DE LOS GRUPOS

En cualquier diseño de comparación de grupos, la principal condición que debe cumplir el grupo control es que sea lo más similar posible al grupo que recibe el tratamiento, y el sesgo de selección se presenta cuando esta homogeneidad queda comprometida. Cuando no es posible usar el azar para construir los grupos, se usan diferentes procedimientos con garantías desiguales con respecto a la amenaza de selección:

- El más sencillo, pero a la vez el menos garantista, es el uso de grupos naturales ya formados antes de iniciar la investigación. Un ejemplo lo encontramos en el caso presentado en la introducción, en el que un grupo lo formaban los trabajadores del turno de la mañana y el otro los trabajadores del turno de la tarde.
- Un segundo procedimiento, más costoso pero también más recomendable, es la aplicación de la técnica del apareamiento (*matching*) ya presentada en la ficha 2.3. Así, por ejemplo, en el estudio sobre el efecto de un entrenamiento en relajación en actores profesionales, la técnica del apareamiento supondría buscar para cada actor que ha accedido a recibir la intervención un actor que comparta características con respecto a variables que pueden afectar el efecto de la relajación e ir formando así el grupo control.

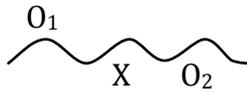
Los dos diseños que presentamos a continuación suponen nuevas alternativas para velar por la homogeneidad de los grupos de comparación en situaciones en las que no es posible realizar asignación aleatoria.

Diseño cuasiexperimental de cohortes¹

El término «cohorte» hace referencia a un grupo cuyos miembros comparten alguna característica o experiencia significativa en un período de tiempo determinado. Por ejemplo, las personas nacidas en el mismo año son la cohorte (generación) de ese año. Las escuelas y muchas otras organizaciones formales e informales se caracterizan por tener una afluencia regular de cohortes. El hecho de que estas cohortes tiendan a presentar características similares plantea oportunidades de cara a la formación de los grupos en un diseño cuasiexperimental.

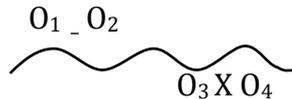
En su estructura más simple, el **diseño de cohortes básico** consiste en la comparación entre medidas tomadas en dos cohortes que se suceden en una organización, correspondiendo la primera a la cohorte previa a la intervención, y la segunda, a la cohorte que ha recibido la intervención. Se utiliza una línea ondulada para simbolizar que los grupos que se comparan son dos cohortes de la misma organización. Así, el diseño que acabamos de comentar quedaría representado de la siguiente forma:

1. En los textos sobre cuasiexperimentación, este diseño se denomina únicamente «diseño de cohortes», pero en este texto añadimos el término «cuasiexperimental» para evitar la confusión con otro diseño también denominado de cohortes pero que no implica la manipulación de la VI. Este diseño será presentado en el capítulo 4 con la denominación de «diseño etiológico de cohortes».



El clásico ejemplo de este diseño es el trabajo de Judith H. Minton (1975, cf. Shadish, Cook y Campbell, 2002), dirigido a evaluar el efecto de la primera temporada de emisión de la serie infantil *Sesame Street* (*Barrio Sésamo*) sobre las puntuaciones en el MRT (Metropolitan Readiness Test) en niños de primer año de guardería. Esta autora localizó una guardería en la que aplicaban el MRT al final del primer año y como grupo control usó los hermanos mayores que habían ido a la misma guardería y para los que también se disponía de las puntuaciones de MRT antes de que se iniciara la emisión de *Sesame Street*. De esta manera, se disponía de un grupo control equivalente en edad y con el mismo estatus madurativo y que por la proximidad respecto al entorno educativo se esperaba que redujera el sesgo de selección en referencia al diseño análogo que no se basa en cohortes (véase la ficha 3.2). Sin embargo, la selección sigue siendo la amenaza más importante de este diseño junto con la historia.

El diseño anterior se puede mejorar añadiendo un pretest. Este nuevo diseño se denomina **diseño de cohortes con pretest** o **también diseño de ciclos institucionales**, y se representa de la siguiente manera:



La lógica del diseño es similar a la del diseño con grupo de control no equivalente con la diferencia de que los grupos que se comparan son cohortes y que ambos grupos no se miden simultáneamente, sino en momentos diferentes. Esto supone ventajas con respecto a la similitud de los grupos, pero también supone que las observaciones de la cohorte anterior se hacen en un periodo temporal previo, lo que incrementa la amenaza de la historia.

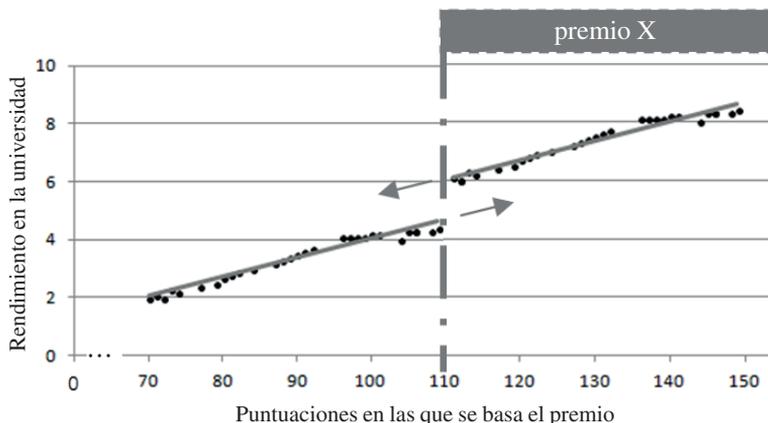
Diseño de discontinuidad en la regresión

Se trata de un diseño con dos grupos iguales a muchos de los anteriores, pero presenta una característica que lo hace único dentro de la cuasiexperimentación, y es el hecho de que el criterio de asignación de los participantes a los grupos es conocido. Ilustraremos esta característica especial a partir de un ejemplo adaptado del original de Campbell y Stanley (1966).

Suponga que se quiere evaluar el efecto que tiene la concesión de un premio de fin de estudios secundarios sobre el rendimiento posterior del estudiante durante el primer año de grado. Suponga también que para el otorgamiento del premio se tienen en cuenta diferentes indicadores de seguimiento del bachillerato que se integran en una medida que va de 0 a 150 puntos, y que establece que el premio implica haber conseguido más de 110 puntos. En general, un premio es una recompensa por el trabajo

realizado, pero ¿en qué grado se convierte también en un aliciente que provoca nuevos éxitos? Si mira de responder a esta pregunta con los diseños presentados hasta ahora se dará cuenta de la dificultad, y por eso resulta tan atractiva la aportación que hace el diseño de discontinuidad en la regresión.

La siguiente figura nos sirve para ilustrar la lógica del diseño. En ella representamos la relación entre la puntuación en la que se basa el premio (eje de abscisas) y el rendimiento al terminar el primer curso en la universidad (eje de ordenadas). Las dos líneas inclinadas representan dos rectas de regresión, una ajustada a la nube de puntos de los que no obtienen el premio y la otra ajustada a la nube de puntos de los que obtienen el premio.^{xi} La línea vertical representa el punto de corte a partir del cual se otorga el premio.



El diseño basa la inferencia en la evaluación de las predicciones (que hemos simbolizado con dos flechas) de las rectas de regresión en torno al punto de corte que determina el otorgamiento del premio. Suponga que se obtiene un resultado, como el de la figura anterior, en el que observamos claras diferencias en el rendimiento de los que han recibido el premio respecto al grupo que no lo ha recibido. Si prolongamos la recta de los que no han recibido el premio por encima del punto de corte (tendencia indicada con la flecha que tiene la punta hacia la derecha), observamos que el valor de rendimiento predicho estaría por debajo de lo observado en el grupo premiado. Un razonamiento similar, pero en sentido contrario, se puede aplicar para el grupo que no recibió el premio. Así pues, este diseño nos aporta una valiosa evidencia para concluir sobre la eficacia del premio académico en el bachillerato sobre el rendimiento en enseñanza superior.

Podemos representar la estructura de este diseño de la siguiente manera:

$$\begin{array}{cccc} O_A & C & X & O_2 \\ O_A & C & _ & O_2 \end{array}$$

O_A es la medida en una variable cuantitativa que se usa para asignar los participantes a los grupos; C indica que las unidades son asignadas a las condiciones sobre la

base del valor en un punto de corte; el resto de símbolos tienen el mismo significado que en los diseños precedentes.

El aspecto más relevante de este diseño es que cada participante puede ser clasificado según las puntuaciones en la variable O_A y que hay un punto de corte en esta variable en función del cual se hace la asignación. La asignación a un grupo u otro depende de este punto de corte y, por este motivo, decimos que, en este caso, el criterio de asignación es conocido, lo que hace que este diseño cuente con una consideración superior a la del resto de diseños cuasiexperimentales en cuanto a la validez interna que puede aportar.

Como conclusión de esta ficha, queremos destacar que los dos tipos de diseños presentados solo pueden ser aplicados si se cumplen requisitos que en algunos ámbitos de investigación son poco frecuentes (por ejemplo, tener la muestra organizada en cohortes o conocer las puntuaciones previamente en una variable sobre la que, además, haya un punto de corte claro para el propósito del estudio). Cuando se dan estas condiciones, ambos grupos de diseños presentan las ventajas que hemos ido indicando, pero no siempre será factible aplicarlos.

ACTIVIDADES

Actividad de identificación de diseños

Identifique y represente esquemáticamente los diseños utilizados en cada uno de los siguientes estudios.

1. Ritmos circadianos y trastorno afectivo emocional

La fototerapia es una técnica de tratamiento que emplea radiaciones electromagnéticas de origen natural o artificial para el tratamiento de enfermedades y se ha aplicado al reajuste de los ritmos circadianos.

Los resultados de varias investigaciones indican que algunos trastornos afectivos están relacionados con los ritmos circadianos. De hecho, muchos pacientes con trastorno afectivo emocional muestran una cierta falta de orden en los ritmos biológicos circadianos, como insomnio primario, hipersomnia, etc. Para contrastar la hipótesis de que la fototerapia puede tener efectos positivos en el trastorno afectivo emocional, se selecciona una muestra de 30 pacientes diagnosticados de trastorno afectivo emocional y se les aplica fototerapia. Tanto antes como después de aplicar el tratamiento, se mide el grado de depresión de los pacientes mediante un cuestionario.

2. Impacto de la práctica de ejercicio en la salud psicológica de las mujeres²

El presente estudio se centra en la evaluación de los efectos del ejercicio físico regular sobre el estado de ansiedad en mujeres de mediana edad. Se parte de la hipótesis de que el ejercicio continuado reduce el nivel de ansiedad. Para contrastar esta hipótesis, se obtiene una muestra de 39 pacientes femeninas de entre 35 y 50 años procedentes de una consulta psicológica, que presentan elevados niveles de ansiedad. En primer lugar, se miden los niveles de ansiedad estado con el SAI (State Anxiety Inventory). Después de excluir a las mujeres con puntuaciones extremas, la muestra queda reducida a 32 pacientes. Una semana después, y al cabo de un mes, se vuelve a pasar el SAI para comprobar la estabilidad de las puntuaciones. A continuación, se aplica un programa de ejercicio físico durante diez semanas. El programa incluye sesiones de ejercicios aeróbicos y de resistencia, y las pacientes deben asistir a las mismas dos veces por semana. La duración de cada sesión es de una hora. El nivel de ansiedad se vuelve a evaluar con el mismo instrumento (SAI) al cabo de un mes, a los seis meses y a los doce meses de haber finalizado el programa.

3. Efecto de la relajación sobre la productividad

En una entidad bancaria, la dirección pretende aumentar la productividad de los empleados que se encargan de la entrada de datos. El equipo de psicólogos del departa-

2. Adaptado de: MEYER, B. B. et al. 2001. «The impact of an exercise program on the psychosocial function of older adult women: a pilot study». En: A. Papaionnaou, M. Goudas y Y. Theodorakis (eds.) In the dawn of the new millennium. Proceedings of the Xth World Congress of Sport Psychology. Vol. 2. Tesalónica: Christodoulidis. P. 48–50.

mento de Recursos Humanos recomienda una intervención basada en la aplicación diaria de dos sesiones de relajación de diez minutos cada una. El objetivo es restablecer los recursos atencionales que se van agotando a medida que transcurre la sesión de trabajo. Para evaluar la eficacia de la intervención, se desea disponer de dos grupos: uno que recibirá la intervención basada en relajación y un grupo control. Finalmente, se decide hacer las sesiones de relajación a los trabajadores del turno de la mañana y mantener a los del turno de la tarde en las mismas condiciones en que estaban. Antes de comenzar la intervención, se mide la productividad de todos los trabajadores durante el último mes mediante un índice que tiene en cuenta tanto el número de datos entrados como el número de errores cometidos por unidad de tiempo. Un mes después del inicio de la intervención, se mide de nuevo la productividad tanto en el turno de la mañana como en el de la tarde.

4. Prevención de trastornos del comportamiento alimentario³

En el ámbito de la prevención de trastornos del comportamiento alimentario (TCA) se desea evaluar el efecto de un programa basado en alfabetización en medios y conocimientos de nutrición sobre las actitudes alimentarias de las adolescentes. A todas las niñas de primero de ESO de una escuela pública se les mide la actitud hacia la alimentación con la escala EAT (Eating Attitudes Test); posteriormente reciben la intervención y al terminar se les vuelve a medir las actitudes con el EAT. Un grupo de niñas de una segunda escuela pública de la misma población actúa como grupo control y responde a la EAT en los mismos momentos que las niñas de la otra escuela.

5. Fomento de la práctica deportiva de los niños⁴

El equipo de dirección de un centro escolar quiere comparar la efectividad relativa de los profesores de educación física del mismo centro con la de monitores externos contratados para estimular la práctica deportiva de los niños durante las sesiones de iniciación al deporte. Para ello se compara la práctica deportiva acumulada al finalizar el curso de los niños entrenados por un monitor externo, con la misma medida obtenida en los niños del curso anterior que fueron entrenados por un monitor del centro.

6. Autoeficacia y afrontamiento de la ansiedad en el deporte⁵

El entrenador de un equipo de baloncesto, después de un año de gran cantidad de partidos perdidos en los últimos minutos, pide asesoramiento a un psicólogo del deporte.

3. Adaptado de: RAICH, R. M.; SÁNCHEZ-CARRACEDO, D.; LÓPEZ-GUIMERÀ, G.; PORTELL, M.; MONCADA, A.; FAUQUET, J. (2008). «A Controlled Assessment of School-Based Preventive Programs for Reducing Eating Disorder Risk Factors in Adolescent Spanish Girls», *Eating Disorders*, 16 (3), 255–272.
4. Adaptado de: HERNÁNDEZ-MENDO, A. (2002). «Investigando con la realidad en psicología del deporte: el uso de diseños cuasiexperimentales». *Lecturas: Educación Física y Deportes (EFDeportes.com)*, 46. Recuperado de: <http://www.efdeportes.com/efd46/invest.htm>
5. Adaptado de: HERNÁNDEZ-MENDO, A. (2002). «Investigando con la realidad en psicología del deporte: el uso de diseños cuasiexperimentales». *Lecturas: Educación Física y Deportes (EFDeportes.com)*, 46. Recuperado de: <http://www.efdeportes.com/efd46/invest.htm>

El psicólogo diseña una intervención destinada a generar estrategias de afrontamiento de ansiedad y a incrementar la autoeficacia percibida. La evaluación de esta intervención se ajusta al siguiente procedimiento: *a)* se dispone de las grabaciones de doce partidos pretratamiento a partir de las cuales se obtienen medidas del rendimiento en los últimos minutos; *b)* durante un periodo sin competición aplica las sesiones presenciales de intervención psicológica; *c)* se mide el rendimiento en los últimos minutos en doce partidos de la temporada siguiente.

7. Efecto de un videojuego sobre el reconocimiento de palabras en niños disléxicos

En su trabajo de final de grado una estudiante de logopedia desea estudiar el efecto del programa Tradislexia⁶ sobre el reconocimiento de palabras en niños disléxicos. Con este objetivo selecciona una muestra de 22 niños (8 niñas y 14 niños) de segundo y tercer ciclo de Educación Primaria con un rango de edad entre 9 y 12 años. Con un CI ≥ 75 según el test que mide el factor G de Cattell, escala 2 forma A. Se tomó como criterio de selección la identificación por parte del profesor de niños con problemas específicos en lectura (DAL), es decir, que presentaran un retraso de uno o dos años en lectura o escritura, y que sus dificultades no estuvieran asociadas a problemas sensoriales, físicos, motores o intelectuales. También tuvo en cuenta que presentaran valores similares en tiempo invertido en lectura de pseudopalabras y en memoria de trabajo. A todos los niños les aplicó la *Batería de Evaluación de los procesos lectores de los niños de Educación Primaria PROLEC*. A continuación, llevó a cabo la aplicación del videojuego *Tradislexia*. La duración fue de 30 minutos diarios aproximadamente, y se realizaba todos los días de la semana (aproximadamente, entre 16 y 20 sesiones durante un mes). Todos los niños realizaban las mismas tareas en el mismo orden establecido previamente por el programa informático. Una semana después de finalizado el tratamiento la estudiante administraron de nuevo las pruebas de lectura de palabras y pseudopalabras del Prolec a todos los participantes.

8. Intervenciones basadas en sistemas aumentativos y alternativos de comunicación⁷

Se desea contrastar la hipótesis según la cual una intervención basada en el uso de sistemas aumentativos y alternativos de comunicación (Augmentative and Alternative Communication, AAC), usando dibujos, esquemas, imágenes, etc. reduce la ansiedad de los padres de niños con dificultades de comunicación cuando a los niños se les aplica una cirugía menor. La muestra está formada por 60 padres o madres de niños con dificultades de comunicación que acceden a participar el día que a su hijo/a se le practica una cirugía menor (por ejemplo, una extracción dental compleja). A los participantes se les pide que vengan solos con su hijo/a (es decir, todos los padres o madres carecen de un acompañante adulto). El estudio se realiza en dos clínicas (A y B). Los niños

6. Adaptado de: JIMÉNEZ, J.F. y ROJAS, E. (2008). Efectos del videojuego Tradislexia en la conciencia fonológica y reconocimiento de palabras en niños disléxicos. *Psicothema*, 20(3), 347-353.

7. Adaptado de: THUNBERG, G., TÖRNHAGE, C. J., & NILSSON, S. (2016). Evaluating the Impact of AAC Interventions in Reducing Hospitalization-related Stress: Challenges and Possibilities. *Augmentative and Alternative Communication*, 32(2), 143-150. <http://doi.org/10.3109/07434618.2016.1157703>

tienen edades comprendidas entre 3 y 15 años, y los diagnósticos son variados (autismo, hiperactividad, déficit de atención y discapacidad intelectual). Los 40 progenitores que acuden a la clínica A forman el grupo control y los 20 progenitores restantes que acuden a la clínica B forman parte del grupo a los que se orienta en el uso de AAC durante el día de la cirugía. En ambos grupos se mide el nivel de ansiedad antes y después de la cirugía.

EJERCICIOS

Ejercicio 3.1: Amenazas a la validez en la valoración de programas

A continuación encontrará unos resúmenes de investigaciones dirigidas a evaluar el efecto de distintas intervenciones. Identifique, como mínimo, una posible amenaza a la validez de cada una de ellas.

1. Una televisión comarcal quiere evaluar el efecto, en jóvenes fumadores, de una serie de tres documentales sobre los peligros del tabaquismo. Se selecciona una muestra de fumadores(as) y se les pide que realicen un autoinforme de su consumo durante un año. La emisión de la serie documental se realiza a los seis meses de iniciarse el estudio (se puede garantizar que todos los participantes ven los tres programas de una hora de duración). Los resultados indican una disminución estable del consumo a partir del momento de emisión de los documentales. En una tertulia radiofónica, el director del programa expone el éxito obtenido y defiende, con contundencia, la necesidad de que se produzca un nuevo incremento del precio del tabaco, igual o superior al que se produjo durante el período de emisión de los documentales.
2. El psicólogo de una institución tutelar de menores detecta un grupo de niños que presentan rasgos de privación cultural marcada y una velocidad de aprendizaje excesivamente lenta para su edad. Tras valorar la inteligencia de los niños, establece un contexto educativo altamente controlado y los somete a un programa de aprendizaje especial durante dos años. Al finalizar este periodo se detecta una ganancia de 1,75 de edad mental, lo que lleva a concluir que el programa de aprendizaje especial favorece el desarrollo cognitivo.
3. Los programadores de un nuevo procesador de textos quieren evaluar si el módulo de autoaprendizaje que han diseñado es eficaz. Con este objetivo, seleccionan una muestra de potenciales usuarios y les hacen realizar una prueba de escritura con el procesador de texto. A continuación, los instruyen para que realicen la sesión de autoaprendizaje (que dura una hora). Al terminar esta sesión vuelven a realizar la prueba de escritura y se observa que el número de errores ha disminuido sensiblemente respecto a los cometidos al inicio de la sesión.
4. Se diseña un programa para incrementar el rendimiento de los trabajadores de una cadena de montaje. Para evaluar su eficacia se realiza una medida antes y otra después de aplicar el programa. La medida de rendimiento se basa en la contabilización de las piezas producidas en la cadena que superan el control de calidad (esta valoración la realiza automáticamente un robot). Los responsables del estudio se dan cuenta de que los criterios de calidad se han modificado durante el período de estudio y de que son más estrictos durante la segunda fase.
5. Después de una operación retorno de vacaciones que supera, con creces, la media de accidentes de tráfico registrados durante los últimos veinte años, el gobierno decide poner en marcha una campaña de prevención de accidentes. Al año siguiente un políptico se congratula de los efectos de la campaña dada la disminución del número de accidentes en relación con el año anterior.

6. Se pretende estudiar un nuevo diseño de las pantallas que utilizan los vigilantes de una entidad bancaria para favorecer la detección de movimientos atípicos. Con este objetivo, se seleccionan dos grupos de vigilantes. Un grupo realiza las pruebas con el nuevo diseño de pantalla, y el otro, con el diseño tradicional. No se observan diferencias ni en la eficacia ni en la eficiencia de los dos grupos. Además, cuando se compara la ejecución del grupo control con datos previos sobre el rendimiento de los vigilantes se observa una mejora significativa. El responsable de esta investigación se entera de que durante la recogida de datos circulaba el rumor de que la empresa estaba preparando una reestructuración de plantilla.
7. Un psicólogo propone un nueva técnica para incrementar la asertividad. Para apoyar su propuesta aporta los resultados obtenidos en diez pacientes a los que aplicó la técnica durante cinco semanas. Antes y después de aplicarla, midió su asertividad observando el comportamiento de los pacientes durante las sesiones de terapia de grupo.

Ejercicio 3.2: La percepción de la seguridad

Los responsables de la seguridad de una empresa de aparcamiento están planificando la instalación de cámaras de videovigilancia en todos los puntos conflictivos de un local que ha sufrido diferentes atracos durante los últimos meses. El equipo de psicólogos de la empresa aprovecha esta circunstancia para valorar el efecto de este tipo de aparatos sobre la percepción de la seguridad de las instalaciones. Diseñan un cuestionario que mide la seguridad percibida y la aplican a las personas que tienen plazas de aparcamiento alquiladas antes de que se haga pública la decisión de instalar cámaras de vigilancia. Posteriormente los usuarios reciben una carta informándoles de que se están instalando las cámaras de vigilancia. Cuando las cámaras ya hace dos meses que funcionan aplica nuevamente el cuestionario a los mismos usuarios.

1. ¿Cuál es el objetivo de esta investigación?
2. Identifique y clasifique las variables implicadas en esta investigación.
3. Esquematice y clasifique el diseño utilizado.
4. Si disponemos de cien sujetos para realizar esta investigación y el número de sujetos que recibe cada tipo de tratamiento es el mismo, ¿cuántos sujetos recibirán el tratamiento cuyo efecto se pretende estudiar?
5. Indique, al menos, una amenaza a la validez interna de este estudio.
6. ¿Cómo se podría mejorar la validez interna de esta investigación?
7. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez de constructo de esta investigación? Indique, al menos, dos aspectos.
8. ¿Dispone de suficiente información para identificar la técnica de muestreo empleada? Justifique la respuesta.
9. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez externa de esta investigación? Indique, al menos, dos aspectos.

Ejercicio 3.3: Prevención de riesgos laborales

El servicio de prevención de una empresa del sector servicios quiere estudiar el efecto de una campaña de prevención de accidentes laborales. Se dispone de recursos para realizar diversas medidas de los accidentes que se producen antes y después de la campaña; asimismo, existe la posibilidad de realizar una aplicación gradual de la campaña.

1. Detalle las características del estudio si este se desarrollara de acuerdo con un diseño pre-post sin grupo control. Especifique, como mínimo:
 - a. ¿Cuántos grupos de trabajadores participarían en el estudio? ¿Cómo seleccionarías a los participantes? ¿Pediría que los trabajadores participantes cumplieran algún requisito especial?
 - b. ¿Cuáles serían las variables y cómo las operativizaría?
 - c. El esquema del diseño.
 - d. Las principales amenazas a la validez interna.

2. Detalle las características del estudio si éste se desarrollara de acuerdo con un diseño pre-post con grupo control no equivalente. Especifique, como mínimo:
 - a. ¿Cuántos grupos de trabajadores participarían en el estudio? ¿Cómo seleccionarías a los participantes? ¿Pediría que los trabajadores participantes cumplieran algún requisito especial?
 - b. ¿Cuáles serían las variables y cómo las operativizaría?
 - c. El esquema del diseño.
 - d. Las principales amenazas a la validez interna.

Ejercicio 3.4: Las aulas de informática de la Facultad

Suponga que, gracias a una ampliación del presupuesto asignado a la Facultad de Psicología de la UAB, se consigue disponer de cuatro aulas más de informática con cuarenta ordenadores cada una. La Facultad quiere comprobar hasta qué punto es cierto lo que se da por supuesto, es decir, la influencia positiva de esta mejora en la dotación de cara al rendimiento académico en las asignaturas en que se realizan prácticas con ordenador. Con este objetivo, al finalizar el curso académico se hace un cálculo del porcentaje de aprobados en las asignaturas mencionadas y se compara con los datos extraídos de las actas del curso anterior, cuando aún no se había incrementado el número de ordenadores (hay decir que en los dos cursos estaba vigente el mismo plan de estudios).

1. ¿Cuál es la hipótesis que se pretende contrastar? Identifique y clasifique la variable independiente. Defina operativamente la variable dependiente.
2. ¿Cómo se llama el diseño utilizado?
3. Indique tres amenazas a la validez interna de esta investigación. Deberá justificar la respuesta.
4. Realice una representación gráfica del diseño empleado siguiendo la notación propuesta por Cook y Campbell (1979).

5. ¿Cuál sería el procedimiento a seguir para contrastar experimentalmente esta hipótesis mediante un diseño de grupos aleatorios?
6. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez de constructo de esta investigación? Indique, al menos, dos aspectos.
7. ¿Dispone de suficiente información para identificar la técnica de muestreo empleada? Justifique la respuesta.
8. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez externa de esta investigación? Indique, al menos, dos aspectos.

Ejercicio 3.5: Intervención en violencia de género

Este enunciado se enmarca en la investigación sobre el tratamiento a la comunidad de los condenados por delitos de violencia de género. En este ámbito, uno de los objetivos es disponer de programas de tratamiento para disminuir la hostilidad (agresividad, ira, rabia, irritabilidad, cólera, resentimiento), las ideas persecutorias (susplicacia, miedo a la pérdida de autonomía, necesidad de control, dificultades en la expresión de la hostilidad) y la sintomatología depresiva. A fin de conocer la eficacia de un programa cognitivo de tratamiento sobre la reducción de las ideas persecutorias, un equipo de investigación⁸ aplica este programa a 130 penados. Previamente habían evaluado sus ideas persecutorias. Ninguno de los participantes en el estudio tenía trastornos neurológicos o mentales que desaconsejasen su inclusión en un programa de rehabilitación de delincuentes estándar. Con posterioridad a la aplicación del programa, se evaluaron nuevamente las ideas persecutorias de los 130 participantes y los resultados mostraron una reducción de las mismas.

1. ¿Cuál es la hipótesis que se pretende contrastar? Identifique y clasifique la variable independiente. Defina operativamente la variable dependiente.
 2. ¿Cómo se llama el diseño utilizado?
 3. Indique tres amenazas a la validez interna de esta investigación. Justifique la respuesta.
 4. Con vistas a incrementar la validez interna, indique qué técnica o técnicas de control se han empleado.
 5. Haga una representación gráfica del diseño empleado siguiendo la notación propuesta por Cook y Campbell (1979).
 6. Indique un diseño cuasiexperimental que mejore la validez interna del diseño empleado en esta investigación.
 7. ¿Cuál sería el procedimiento a seguir para contrastar experimentalmente esta hipótesis mediante un diseño de grupos aleatorios?
 8. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez de constructo de esta investigación? Indique, al menos, dos aspectos.
8. Adaptado de: NOVO, M.; FARIÑA, F.; SEJO, M. D.; ARCE, R. 2012. «Assessment of a community rehabilitation programme in convicted male intimate-partner violence offenders». *International Journal of Clinical and Health Psychology*, 12 (2), 219–34.

9. ¿Dispone de suficiente información para identificar la técnica de muestreo empleada? Justifique la respuesta.
10. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez externa de esta investigación? Indique, al menos, dos aspectos.

Ejercicio 3.6: Enfrentamiento activo y adaptativo a la crisis del envejecimiento femenino

Diferentes investigaciones⁹ han mostrado que las personas de edad avanzada con más recursos sociales tienden a usar estrategias de afrontamiento activo (*active coping-style*). En cambio, las formas pasivas (*passive coping-style*) consistentes en la ausencia de enfrentamiento o conductas de evitación y negación resultan menos exitosas. A partir de estos datos, un equipo de investigación plantea la hipótesis de que la actividad social, más específicamente el aprendizaje social, cambia el uso de estrategias de afrontamiento cognitivo, conductual y emocional hacia un comportamiento adaptativo. Para contrastar esta hipótesis, se constituyen dos grupos: un grupo de intervención formado por 30 mujeres con edades comprendidas entre 60 y 90 años, y un grupo de control formado por otras 30 mujeres. Las participantes de ambos grupos fueron emparejadas por sexo, edad, nivel socioeconómico y educativo. Se excluyeron de la muestra las mujeres con incapacidad física o cognitiva y que se hubieran quedado viudas. A las participantes de ambos grupos se les aplicó la escala multidimensional de evaluación de los estilos de afrontamiento (BRIEF-COPE) antes y después del período de intervención. La intervención consiste en un ciclo de sesiones psicoeducativas en torno a diferentes materias (filosofía y cine, historia de la Grecia antigua, comunicación humana, sexualidad humana, dibujo y pintura, folclore y narración de cuentos). En las sesiones se usa una metodología basada en la interacción y discusión grupal, la influencia recíproca de experiencias e información y la toma de decisiones compartida, lo cual, de acuerdo con la literatura, implica el aprendizaje social.

1. Identifique y clasifique la variable independiente. Defina operativamente la variable dependiente.
2. ¿Cómo se denomina el diseño utilizado?
3. Indique tres amenazas a la validez interna de esta investigación. Justifique la respuesta.
4. Con vistas a incrementar la validez interna, indique qué técnica o técnicas de control se han empleado.
5. Realice una representación gráfica del diseño empleado siguiendo la notación propuesta por Cook y Campbell (1979).
6. Indique un diseño cuasiexperimental que mejore la validez interna del diseño empleado en esta investigación.

9. Adaptado de: KRZEMIEN, D.; URQUIJO, S.; MONCHIETTI, A. 2004. «Aprendizaje social y estrategias de afrontamiento a los sucesos críticos del envejecimiento femenino». *Psicothema*, 16 (3), 350–6.

7. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez de constructo de esta investigación? Indique, al menos, dos aspectos.
8. ¿Dispone de suficiente información para identificar la técnica de muestreo empleada? Justifique la respuesta.
9. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez externa de esta investigación? Indique, al menos, dos aspectos.

Ejercicio 3.7: Eficacia del entrenamiento de la higiene vocal y los ejercicios de la función vocal en la prevención de los trastornos de la voz en los maestros de escuela primaria¹⁰

Los trastornos de la voz en los maestros tienen un impacto significativo en su funcionamiento y bienestar ocupacional. Se cree que los maestros tienen una alta prevalencia de problemas de voz debido a los entornos acústicos desfavorables en los que trabajan y las altas demandas vocales y los niveles de estrés asociados con la enseñanza. Aunque los problemas de voz que experimentan los docentes deben evitarse porque son causados por factores que los docentes pueden modificar, existe información limitada disponible sobre la efectividad de las diferentes estrategias preventivas. El objetivo de este estudio fue investigar la eficacia del entrenamiento de higiene vocal (HV) y de los ejercicios de función vocal (EFV) para reducir los síntomas vocales, el uso vocal indebido, aumentar el conocimiento del cuidado de la voz, el tiempo de fonación máximo y el rango de frecuencia fonacional máximo en maestros de escuela. Treinta y siete maestros de tres escuelas en Melbourne, Australia, participaron en el estudio. Las escuelas se asignaron a uno de tres grupos: HV, EFV y control sin tratamiento. Los participantes de HV y EFV informaron mejores características vocales y conocimiento de la voz después del entrenamiento, mientras que el grupo de control mostró deterioro en la mayoría de las variables. Los participantes de HV mostraron mayores mejoras que los participantes de EFV. Estos hallazgos indican que es probable que la capacitación preventiva de voz para docentes sea efectiva.

1. Indique cuál/es son las VI.
2. Indique cuál/es son las VD.
3. Identifique las variables extrañas.
4. ¿Qué técnicas de control se han aplicado?
5. Haga el esquema del diseño.
6. ¿Cuál es el diseño utilizado?
7. Compare el diseño de este ejercicio con el de la investigación original.

10. Adaptado de: Pasa, G., Oates, J., & Dacakis, G. (2007). The relative effectiveness of vocal hygiene training and vocal function exercises in preventing voice disorders in primary school teachers. *Logopedics Phoniatrics Vocology*, 32(3), 128-140. <https://doi.org/10.10>

Ejercicio 3.8: Evolución de los errores de denominación en afasia tras un tratamiento fonomotor¹¹

Antecedentes: Aunque los tratamientos para la anomia en la afasia generalmente usan la precisión de los objetivos entrenados y no entrenados para determinar la eficacia del tratamiento, los investigadores han comenzado a emplear el análisis de errores para delinear con mayor concreción los mecanismos implicados en la recuperación de la anomia.

Objetivos: El objetivo de este estudio fue examinar los cambios en los errores de denominación de 24 individuos con afasia después del tratamiento fonomotor.

Métodos y procedimiento: Veinticuatro personas con afasia recibieron tratamiento fonomotor. La denominación por confrontación se evaluó antes del tratamiento, inmediatamente después del tratamiento y 3 meses después del tratamiento. Las respuestas se calificaron según su precisión y tipo de error (e.g., fonológicos, semánticos, omisiones, etc.). Se realizaron pruebas t de medidas repetidas que comparaban la precisión y las proporciones del tipo de error entre el pretratamiento y posttratamiento inmediato, y a los de 3 meses.

Resultados: La precisión en la denominación mejoró de forma significativa inmediatamente después del tratamiento, y las mejoras se mantuvieron 3 meses después del tratamiento. No se observaron cambios significativos en las proporciones de tipo error; sin embargo, se observaron tendencias hacia la disminución de las proporciones de omisiones inmediatamente después del tratamiento y 3 meses después del tratamiento.

Conclusiones: Los resultados son indicativos de un cambio en el procesamiento lingüístico en el que la red lingüística se alteró holísticamente como resultado del tratamiento fonomotor.

1. Indique cuál/es son las VI.
2. Indique cuál/es son las VD.
3. Identifique las variables extrañas.
4. ¿Qué técnicas de control se han aplicado?
5. Haga el esquema del diseño.
6. En comparación con los diseños incluidos en las fichas 3.2 y 3.4, ¿qué ventajas y limitaciones tiene el diseño utilizado?

Ejercicio 3.9: Evaluación de un programa dirigido a niños con trastorno fonológico¹²

El desarrollo del lenguaje es un proceso de esencial importancia en la infancia. Sin embargo, los trastornos del lenguaje constituyen un problema frecuente durante el

11. Adaptado de: MINKINA, I., OELKE, M., BISLICK, L. P., BROOKSHIRE, C. E., POMPON, R. H., SILKES, J. P., & KENDALL, D. L. (2016). An investigation of aphasic naming error evolution following phonomotor treatment. *Aphasiology*, 30(8), 962-980. <https://doi.org/10.1080/02687038.2015.1081139>
12. Adaptado de: GALLEGO, J. L., GÓMEZ, I. A. y AYLLÓN, M. F. (2018). Evaluación de un programa dirigido a niños con trastorno fonológico. *Revista de Investigación en Logopedia*, 5(2), 135-166. <https://doi.org/10.5209/RLOG.58623>

desarrollo de los niños, especialmente los denominados trastornos fonológicos. En este sentido, el estudio que se presenta muestra los resultados de una investigación realizada con escolares de 5, 6 y 7 años de edad con trastorno fonológico. Su objetivo fue comprobar los efectos que tiene un programa de intervención en el desarrollo lingüístico de niños de educación infantil y primaria, en el nivel articulatorio. Participaron en el estudio 50 escolares: 25 controles (13 niños y 12 niñas) y 25 en el grupo de intervención (19 niños y 6 niñas). Se midieron las habilidades fonológicas antes y después de la intervención en ambos grupos. Los resultados mostraron diferencias significativas entre el grupo de tratamiento y el grupo de control, tras la aplicación del programa. Se ha demostrado, por tanto, la eficacia de este programa para mejorar las habilidades fonológicas.

1. Indique cuál/es son las VI.
2. Indique cuál/es son las VD.
3. Identifique las variables extrañas.
4. ¿Qué técnicas de control se han aplicado?
5. Haga el esquema del diseño.
6. ¿Cuál es el diseño utilizado?
7. Si se hubiera mantenido constante la edad de los participantes, ¿cómo se denominaría el diseño aplicado?
8. ¿Cómo valora la última frase del resumen (“Se ha demostrado, por tanto, la eficacia de este programa para mejorar las habilidades fonológicas”)?

LECTURA CRÍTICA

Busque los artículos del siguiente cuadro. Para hacer la búsqueda le sugerimos que utilice el Google Scholar y/o la base de datos Pubmed.

A partir del resumen y de la información del apartado «Método», haga un análisis metodológico del artículo y complete la ficha de lectura crítica que encontrará a continuación. Excluya de este análisis todas las cuestiones relativas al análisis de datos y a la validez de conclusión estadística, ya que escapan a los contenidos expuestos en este texto introductorio.

Artículo 1

Título: «Intervención en estrategias de revisión del mensaje escrito»

Autores: Jesús-Nicasio García; y Olga Arias-Gundín

Revista: *Psicothema*

Año de publicación: 2004

Artículo 2

Título: «Intervención psicológica en la conducta agresiva y antisocial con niños»

Autores: Maite Garaigordobil Landazabal

Revista: *Psicothema*

Año de publicación: 2004

Artículo 3

Título: «Community partnership to affect substance abuse among Native American adolescents»

Autores: Lowe, J.; Liang, H.; Riggs, C.; Henson, J.; Elder, T.

Revista: *The American journal of drug and alcohol abuse*

Año de publicación: 2012

Artículo 4

Título: «Effects of a humor-centered activity on disruptive behavior in patients in a general hospital psychiatric ward»

Autores: Antonio Higuera; Hugo Carretero-Dios; José P. Muñoz; Esther Idini; Ana Ortiz; Francisco Rincón; David Prieto-Merino; María M. Rodríguez del Águila

Revista: *International Journal of Clinical and Health Psychology*

Año de publicación: 2006

CRITICAL READING CARD: ARTICLE
(FICHA DE LECTURA CRÍTICA: ARTÍCULO)



Remember to thoroughly read the article once found and extract the following information. If you obtain a summary which does not provide enough information you may mark «lack of information» in the corresponding box.

-1- Write the article's **bibliographic reference** according to the APA's referencing style.

--

-2- Make a **list of the methodological terms** that appear in the method section not related with the study's subject. The first column includes the term in English and the other two the translation in Catalan and Spanish. Include the new terms in your English/Catalan/Spanish Vocabulary.

English	Catalan	Spanish

-3- Indicate which are the **objectives** and **hypothesis** of the research.

--

-4- Does the study address a clearly focused issue? An issue can be 'focused' in terms of the population studied, the intervention given, the comparator given or the outcomes considered.

--

-5- Identify which are the **independent and dependent variables (IV and DV)**.

IV	DV

-6- Is exposure to IV levels measured in a standard, valid and reliable way?

--

-7- Are DV measured in a standard, valid and reliable way?

--

-8- Is it possible to identify the **sampling procedure used**? If so, just copy the information presented about it.

--

-9- What percentage of individuals refused to participate?

--

-10- Classify the design used according to the criteria presented in this chapter.

--

-11- What factors other than the IV could affect the DV? Include potential confounding factors like differences in baseline characteristics between groups.

--

-12- Aside from the quasi-experimental intervention, were the groups treated equally?

--

-13- Identify the **control techniques** used and which variables have been controlled with.

--

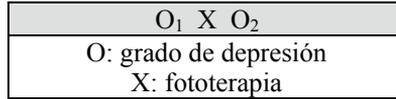
-14- Link the previous answers with the concepts of internal and external validity applied to the conclusions of this study.

Internal Validity	
External validity	

SOLUCIONES DE LA ACTIVIDAD DE IDENTIFICACIÓN DE DISEÑOS

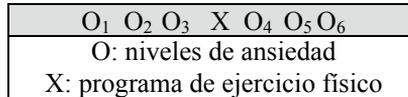
1. Ritmos circadianos y trastorno afectivo emocional

Diseño pretest-postest de un solo grupo



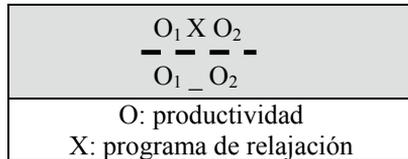
2. Impacto de la práctica de ejercicio en la salud psicológica de las mujeres

Diseño simple de serie temporal interrumpida



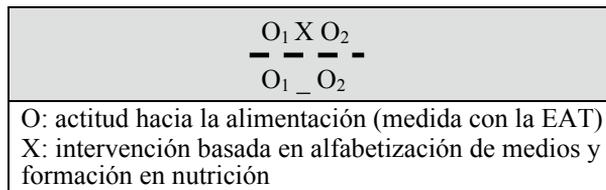
3. Efecto de la relajación sobre la productividad

Diseño pretest-postest con grupo control no equivalente



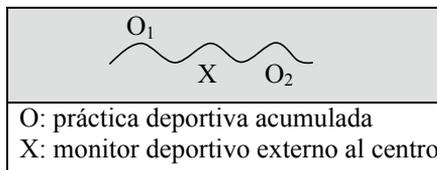
4. Prevención de trastornos del comportamiento alimentario

Diseño pretest-postest con grupo control no equivalente



5. Fomento de la práctica deportiva de los niños

Diseño de cohortes básico



6. Autoeficacia y afrontamiento de la ansiedad en el deporte
Diseño simple de serie temporal interrumpida

$O_1 O_2 O_3 O_4 O_5 O_6 O_7 O_8 O_9 O_{10} O_{11} O_{12} X O_{13} O_{14} O_{15} O_{16} O_{17} O_{18} O_{19} O_{20} O_{21} O_{22} O_{23} O_{24}$
O: medida de rendimiento en los últimos minutos de partido X: intervención psicológica

7. Efecto de un videojuego sobre el reconocimiento de palabras en niños disléxicos
Diseño pretest-postest de un solo grupo

$O_1 X O_2$
O: aciertos en la identificación de palabras X: videojuego

8. Intervenciones basadas en sistemas aumentativos y alternativos de comunicación
Diseño pretest-postest con grupo control no equivalente

$\frac{O_1 X O_2}{O_1 _ O_2}$
O: nivel de ansiedad X: intervención AAC

Notas

- i Para profundizar en la contraposición de ventajas y limitaciones de la aleatorización, puede consultar el texto de Ato (1995a).
- ii Para una panorámica del resto de diseños empleados en evaluación de programas organizados desde una perspectiva que prioriza la complementariedad metodológica, puede consultar Chacón, Sanduverte, Portell y Anguera (2013).
- iii El término cuasiexperimento es de uso común en psicología y educación, pero hay que saber que no se emplea en todas partes. En otros ámbitos de las ciencias de la salud, se suele hablar de «estudio comparativo» o «estudio controlado no aleatorizado» para referirse a estructuras que en este texto presentamos dentro de los diseños cuasiexperimentales. En inglés, es común referirse a los cuasiexperimentos como *quasiexperiments* o *controlled clinical trial* (CCT)
- iv Siguiendo la delimitación entre tipos de validez propuesta por los autores citados, en este texto mantenemos la denominación "validez de constructo"; sin embargo el lector puede consultar estándares actuales (E.g., Standards for Educational and Psychological Measurement, 2014) que no incluyen esta denominación. Puede consultar el texto de Chacón y Shadish (2008) para una exhaustiva revisión en español de la evolución de la tipología de validez y de los principales autores implicados en la propuesta y revisión de estos conceptos.
- v La exposición detallada de las amenazas a la validez de conclusión estadística supera el propósito de este texto y aquí únicamente se citan. Para obtener las correspondientes definiciones y ejemplos de cada caso, recomendamos la lectura del capítulo 2 del texto de Shadish, Cook y Campbell (2002).
- vi Para una exposición detallada de estas amenazas, recomendamos el capítulo 2 del texto de Shadish, Cook y Campbell (2002).
- vii Para una exposición detallada de las amenazas a la validez de constructo, recomendamos el capítulo 3 del texto de Shadish, Cook y Campbell (2002).
- viii Para una exposición detallada de las amenazas a la validez externa, recomendamos el capítulo 3 del texto de Shadish, Cook y Campbell (2002).
- ix Para una relación exhaustiva de los diseños cuasiexperimentales, se puede consultar el texto de Shadish, Cook y Campbell (2002). En lengua española se puede consultar una completa exposición en Ato (1995b) y Chacón, Shadish y Cook (2008).
- x Para una presentación en español del diseño y el análisis de series de tiempo puede consultar el texto de Vallejo (1995).
- xi Encontrará una explicación didáctica de estos conceptos en el capítulo 9 del libro de León y Montero (2003) y en Gambara (2002). Hay que tener presente, también, las limitaciones de este diseño cuando la relación entre las variables no es lineal; para ampliar información sobre este tema, recomendamos el capítulo 7 del texto de Shadish, Cook y Campbell (2002).

4. Diseños *ex post facto*

PRESENTACIÓN

Los diseños que hemos visto hasta ahora —experimentales y cuasiexperimentales— forman parte de lo que podríamos denominar «paradigma manipulativo», dado que todos ellos comparten una característica: la manipulación de las variables independientes (VI). Cabe tener en cuenta que en psicología y logopedia se plantean importantes hipótesis referidas a factores de riesgo difíciles o imposibles de manipular. A modo de ejemplo podemos considerar investigaciones sobre el efecto del abuso infantil en la psicopatología en la edad adulta, el efecto de la dislexia en la autoestima, el efecto de ciertas características de personalidad en la percepción del dolor en la fibromialgia, o el efecto del estrés laboral en las alteraciones de la conducta alimentaria, entre muchas otras. Para comprender cómo se realizan este tipo de estudios presentaremos tres diseños ampliamente utilizados en la investigación en ciencias de la salud: el **diseño etiológico de cohortes**, el **diseño de casos y controles** y el **diseño transversal analítico**.

Existe cierta controversia sobre la denominación genérica de estos tres diseños. En este texto optamos por denominarlos **diseños *ex post facto***ⁱ. Asimismo, consideramos de interés integrar estos diseños junto con los diseños de encuesta —que no se desarrollan en este documento— dentro del **método selectivo**, y diferenciar este último del método observacional. Con esta clasificación, que ya ha sido introducida en la ficha 1.9, pretendemos facilitar el análisis de tres tipos de estudios que, más allá de compartir el hecho de ser “no experimentales”, difieren en un conjunto de aspectos importantesⁱⁱ.

El método selectivo se define como una estrategia particular del método científico encaminada a obtener información cuantitativa para dar respuesta a problemas de diferente naturaleza (descripción, covariación, predicción, causación, etc.). Con esta finalidad, el uso del método selectivo se caracteriza por ejercer un control indirecto sobre las respuestas objeto de estudio basado fundamentalmente en: (1) la selección de participantes en función de su pertenencia a un determinado nivel de las variables relevantes, (2) la estandarización, y (3) el uso de técnicas de control estadístico.

De los dos grupos de diseños incluidos dentro del método selectivo, los diseños *ex post facto*, a diferencia de los diseños de encuesta, se centran en la contrastación de hipótesis causales y se utilizan cuando la naturaleza de la variable que se plantea como posible causa desaconseja su manipulación y el control directo sobre los posibles factores extraños que pueden intervenir. Esta característica fundamental justifica un cam-

bio en la nomenclatura de las variables objeto de estudio (véase ficha 1.6). Así, dado que hay menos garantías de que la dependencia de la variable dependiente (VD) sea únicamente hacia la VI, y también que las VI sean independientes las unas de las otras, la VD se denomina **variable de respuesta** (VR) y el factor explicativo se denomina **variable de exposición** (VE).

Es importante destacar que en estos diseños la investigación comienza cuando ya están presentes los valores de la VE, o de la VE y de la VR —de ahí el nombre *ex post facto*. Según si estos valores que ya se han dado correspondan a la VE o bien tanto a la VE como a la VR, y según cual sea la variable empleada para seleccionar y agrupar a los participantes al inicio de la investigación, podemos diferenciar entre el diseño etiológico de cohortes, el diseño de casos y controles y el diseño transversal analítico.

La aproximación al estudio de la causalidad a partir de los diseños *ex post facto* es altamente compleja y problemática, pero al mismo tiempo resulta un recurso imprescindible en ciencias de la salud. La diferencia entre el acercamiento a la causalidad que nos aportan estos diseños respecto a un diseño experimental estrictamente controlado, se traduce en una reducción del grado de confianza con el que se pueden descartar explicaciones alternativas (véase ficha 1.7).

A partir del estudio de las siguientes fichas, se espera que el lector sea capaz de:

- Conocer las diferencias entre los diseños manipulativos y los no manipulativos (ficha 4.1).
- Diferenciar entre los tres tipos de diseños *ex post facto* (ficha 4.2).
- Entender y saber describir las fases de un diseño etiológico de cohortes (ficha 4.3).
- Entender y saber describir las fases de un diseño de casos y control (ficha 4.4).
- Entender y saber describir las fases de un diseño transversal analítico (ficha 4.5).

FICHA 4.1. PARADIGMA MANIPULATIVO Y NO MANIPULATIVO

La siguiente figura permite ubicar y diferenciar los diseños *ex post facto* de los diseños experimental y cuasiexperimental con respecto a diversos aspectos relacionados con la VE.ⁱⁱⁱ Imagine una investigación en la que se pretende estudiar los efectos de la meditación (con/sin meditación) sobre el rendimiento académico. Si esta investigación se lleva a cabo aplicando un diseño experimental, los participantes se asignan al azar a una de las condiciones experimentales (suponiendo un diseño intersujeto), esto es, o bien a la condición de meditación o bien en la de no meditación; de esta manera, si la muestra es suficientemente grande, hay una alta probabilidad de que los dos grupos sean comparables. El investigador manipula la VE determinando para todos los participantes varios aspectos relacionados con el tratamiento que pueden influir en la VR, tales como tipo de meditación, horario, duración y frecuencia de las sesiones, entorno donde se llevará a cabo, quién conducirá las sesiones, actividad que realizará el grupo asignado a no meditación, etc. De este modo, el investigador provoca o elicit los valores que tomará la VE.

Si la investigación se lleva a cabo con un diseño cuasiexperimental, la formación de los grupos no será aleatoria. De hecho, a menudo el investigador trabajará empleando grupos ya formados y asignando los grupos a los niveles de la VE. Así, por ejemplo, contará con el grupo de la mañana y con el grupo de la tarde. Ante la imposibilidad de asignar aleatoriamente a los estudiantes al grupo de la mañana o de la tarde (sencillamente porque muchos estudiantes se negarían a pasar del grupo de la mañana a la tarde o viceversa), el investigador partiría de estos grupos ya formados y asignaría a uno de ellos la condición de meditación y al otro grupo la condición de no meditación. Al igual que en los experimentos, en los cuasiexperimentos se manipulan los valores de la VE.

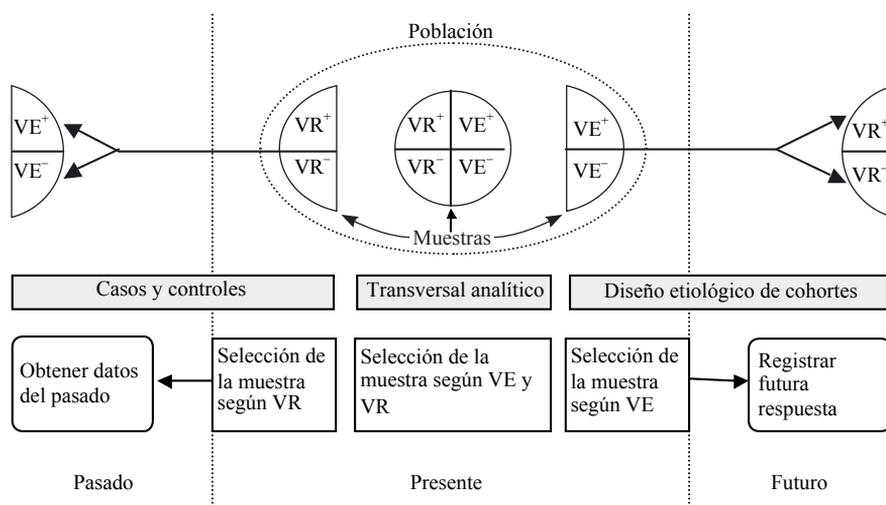
En caso de que el diseño a aplicar sea *ex post facto*, el investigador no solo trabajará con grupos ya formados, sino que, dada la imposibilidad de manipular, simplemente registrará o medirá los valores de la VE que ya presentan los participantes al inicio de la investigación. Es decir, registrará para cada participante si medita regularmente o no, y clasificará a los participantes en estos dos grupos. En esta situación, el investigador no determina el valor que toma la VE por cada participante; simplemente pregunta a cada participante si medita habitualmente o no. Nótese que, por tanto, el investigador no genera ni administra las condiciones en que se lleva a cabo la meditación ni la actividad alternativa de los participantes que no meditan y que, por tanto, se hace mucho más complicado garantizar la homogeneidad de las características tanto de la actividad de meditación como de la de no meditación y, aún más relevante, es mucho más improbable que los dos grupos sean comparables.

Finalmente, obsérvese que la medida de la VR no introduce ninguna diferencia entre los tres diseños.

	<i>VE (VI)</i>			<i>VR (VD)</i>
	<i>Formación de los grupos de comparación</i>	<i>Administración de los valores de la VI/VE</i>	<i>Mecanismo generador de los valores de VE</i>	
Experimental	Aleatorización	Manipulación de la VI	El investigador elicit los valores de la VE	Medida de la VR/VD
Cuasi-experimental	Sin aleatorización. A menudo empleando grupos ya formados y asignando los grupos a los niveles de la VI	Manipulación de la VI	El investigador elicit los valores de la VE	
<i>Ex post facto</i>	Grupos ya formados sin intervención del investigador	No manipulación. Simplemente se mide la VE	Desconocido / ajeno a la acción del investigador	

FICHA 4.2. LOS DISEÑOS *EX POST FACTO*

La figura siguiente (adaptada de Margets, Vorster y Venter, 2002) proporciona una visión general de los principales elementos que diferencian los tres diseños *ex post facto*,^{iv} especialmente en cuanto a las variables que se emplean para seleccionar y clasificar los participantes en grupos: VR en el diseño de casos y controles, VE en el diseño etiológico de cohortes, o ambas simultáneamente en el caso del diseño transversal analítico. La figura representa la situación más sencilla, en la que tanto la VE como la VR tienen únicamente dos niveles. En el caso de la VE, los dos niveles se corresponden a la presencia (VE^+) o ausencia (VE^-) de exposición a un factor de riesgo o bien a un factor protector. En cuanto a la VR, los dos niveles se refieren a la presencia del desenlace de interés (VR^+) o bien a su ausencia (VR^-).



En el **diseño etiológico de cohortes** (véase la ficha 4.3), la selección de participantes se hace en función de la exposición a factores de riesgo, es decir, según los valores que presenten en la VE. Así, la muestra se obtiene identificando unidades de muestreo que estén expuestas (VE^+) o bien que no lo estén (VE^-) a un factor de riesgo (por ejemplo, presencia [VE^+] / ausencia [VE^-] de apoyo social) y se sigue a los participantes para registrar cuáles presentan (VR^+) y cuáles no presentan (VR^-) el acontecimiento o desenlace de interés.

En el otro extremo, en el **diseño de casos y controles** (véase la ficha 4.4), la selección parte del otro extremo de la cadena causal. Concretamente, la muestra se obtiene identificando unidades de muestreo que presenten (VR^+) o bien que no presenten (VR^-) determinados valores de la VR (por ejemplo, presencia/ausencia de un trastorno de ansiedad) y luego se buscan de forma retrospectiva los valores que presentaba la VE en cada participante, es decir, se registra si los participantes estuvieron expuestos o no al factor de riesgo.

Finalmente, en medio tenemos los **diseños transversales analíticos** (véase la ficha 4.5), que hacen la selección de los participantes partiendo de la base de que presenten o no los diferentes valores tanto de la VE como de la VR. En este caso, pues, en el momento de la selección se registran tanto la VE como la VR, y se clasifican los participantes según las cuatro combinaciones de niveles de las variables objeto de estudio: VE⁺ VR⁺ | VE⁻ VR⁺ | VE⁺ VR⁻ | VE⁻ VR⁻.

Cabe destacar que, tanto en el diseño de casos y controles como en el transversal analítico, tanto la VE como la VR ya han tomado sus valores, es decir, ya han pasado. En cambio, en el diseño etiológico de cohortes, solo la VE ha tomado ya sus valores para cada participante, la VR (o más concretamente el valor de la VR que corresponde al desenlace de interés) todavía no se ha dado.

FICHA 4.3. DISEÑO ETIOLÓGICO DE COHORTES

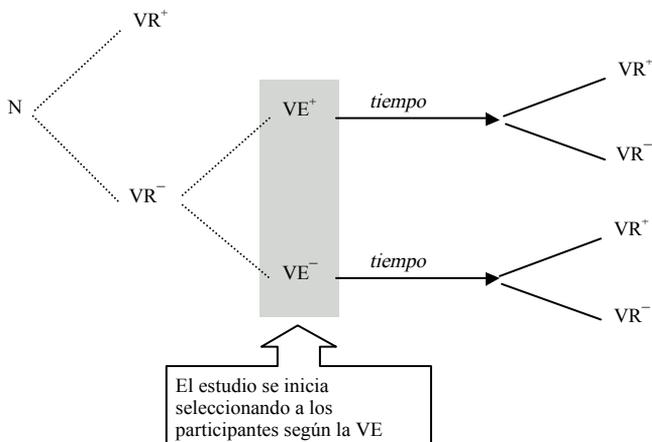
El diseño etiológico de cohortes,^v que en la literatura inglesa generalmente se denomina genéricamente *cohort study* o *cohort design*, es un diseño *ex post facto* en que la VE ya ha tomado sus valores, pero no se ha dado aún el desenlace objeto de estudio (VR). Al iniciarse el estudio, se miden tanto la VE, para clasificar a los participantes, como la VR, para comprobar que no esté presente el desenlace objeto de estudio o acontecimiento de interés. El acontecimiento de interés hace referencia a uno de los valores de la variable de respuesta, como, por ejemplo, la presencia de una enfermedad determinada. En caso de que algún participante presente el desenlace, es excluido de la muestra. De este modo, se garantiza el criterio de temporalidad a que hacíamos referencia en el tema 1 al hablar de relaciones causales (ficha 1.6).

Los participantes son seleccionados y agrupados sobre la base de los valores de la variable de exposición. De este modo se seleccionan participantes que presentan el nivel de interés de la variable de exposición, esto es, participantes expuestos, y otros que no presentan este nivel, los participantes no expuestos. A continuación, se hace un seguimiento durante el período de tiempo que se considera suficiente para que aparezca el acontecimiento de interés.

Este diseño es propenso a la amenaza de selección y resulta poco eficiente y caro cuando el desenlace es de baja prevalencia.

El procedimiento de aplicación de este diseño implica las siguientes fases:

1. Seleccionar una muestra a partir de la población (N). Es un requisito del diseño que ninguno de los participantes presente la categoría o nivel «crítico» de la VR (también llamado desenlace o acontecimiento de interés). Por ejemplo, en una investigación dirigida a contrastar la hipótesis «el locus de control externo es un factor de riesgo para el trastorno de angustia», entenderemos como nivel crítico de la VR el diagnóstico de trastorno de angustia (según criterio DSM: un mínimo de tres crisis en un período de tres semanas no provocadas por estímulos fóbicos, etc.). Simbolizaremos este nivel crítico de la VR como VR⁺ y el resto como VR⁻.
2. Medir la VE y agrupar a los participantes en función de las categorías o los niveles de esta VE. En el caso más sencillo, la VE tendrá dos categorías que simbolizaremos como VE⁺ y VE⁻. La primera categoría se refiere a la exposición al factor (llamado también «factor de riesgo» o, según en qué contexto, factor protector), que, según la hipótesis, puede provocar el desenlace. Siguiendo con el ejemplo del trastorno de angustia, VE⁺ es un locus de control externo. **Nota:** También se podría seleccionar a los participantes de cada grupo a partir de poblaciones separadas, pero hay que tener en cuenta que la fuerza de la evidencia causal dependerá de la homogeneidad de los grupos en todas aquellas variables que pueden producir confusión.
3. Efectuar un seguimiento de las cohortes.
4. Medir la VR. En el caso más sencillo, esto supone clasificar a los participantes en función de la presencia o no del desenlace (VR⁺ || VR⁻).

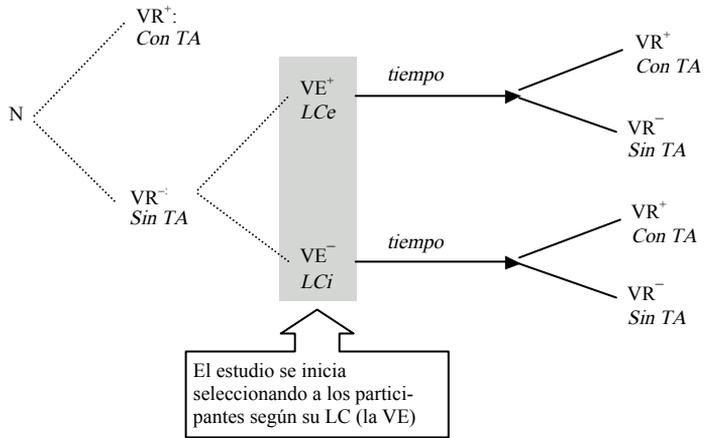


Por ejemplo,¹ en una investigación dirigida a contrastar la hipótesis «el locus de control externo es un factor de riesgo para el trastorno de angustia», entenderemos como nivel crítico de la VR el diagnóstico de trastorno de angustia (según criterio DSM: un mínimo de tres crisis en un período de tres semanas no provocadas por estímulos fóbicos, etc.).

El procedimiento de aplicación de este diseño implicaría las siguientes fases:

1. Seleccionar una muestra que no presente el acontecimiento de interés, en este caso, el trastorno de angustia a partir de la población. Quedan excluidos los participantes que presenten trastorno de angustia al inicio de la investigación.
2. Medir y agrupar a los participantes según tengan un locus de control externo (VE⁺) o interno (VE⁻).
3. Efectuar un seguimiento de las dos cohortes (la del locus de control externo y la del interno (VE⁻).
4. Medir la presencia / ausencia del desenlace (VR⁺ || VR⁻), en este ejemplo, la presencia / ausencia de trastorno de angustia.

1. Otros ejemplos publicados de aplicación del diseño etiológico de cohortes se pueden encontrar en: AUTENRIETH, C. S.; BAUMERT, J.; BAUMEISTER, S. E.; FISCHER, B.; PETERS, A.; DÖRING, A.; THORAND, B. 2010. «Association between domains of physical activity and all-cause, cardiovascular and cancer mortality». *European Journal of Epidemiology*, 26 (2), 91–99. Doi: 10.1007/s10654-010-9517-6. HANSEN, B. M.; NILSSON, O. G.; ANDERSON, H.; NORRVING, B.; SÄVELAND, H.; LINDGREN, A. 2013. «Long term (13 years) prognosis after primary intracerebral haemorrhage: A prospective population based study of long term mortality, prognostic factors and causes of death». *Journal of Neurology, Neurosurgery & Psychiatry*, 84 (10), 1150–5. Doi:10.1136/jnnp-2013-305200.



TA: trastorno de angustia | LCe: locus de control externo | LCi: locus de control interno

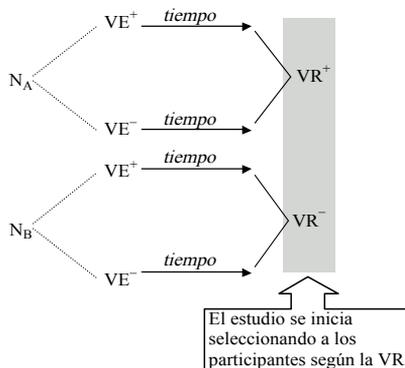
FICHA 4.4. DISEÑO DE CASOS Y CONTROLES

En el diseño de casos y controles^{vi} (en inglés, *case-control*), tanto la VE como la VR ya han tomado sus valores al inicio de la investigación. Los participantes se seleccionan según los valores que presentan en la VR, de tal manera que hay una parte de la muestra formada por participantes que ya presentan el acontecimiento de interés, que llamamos casos, y otra muestra de individuos que no lo presentan, los controles. A continuación, se evalúa cuáles son los factores de riesgo (o los factores protectores) a que han sido expuestos los individuos, así como la intensidad y la duración de la exposición.

Se desprende de esta estructura que el diseño de casos y controles, a diferencia del diseño etiológico de cohortes, no garantiza el criterio de temporalidad (dado que no se puede garantizar que la causa preceda al efecto), y que permite superar la limitación de lo anterior con respecto al estudio de fenómenos de baja prevalencia. Este diseño es propenso a los sesgos retrospectivos y de selección. En cuanto al primero, se recomienda hacer uso de fuentes de información independientes del individuo. Para hacer frente al segundo, hay que explicitar claramente los criterios de selección de los casos y, por otra parte, el grupo de control debería provenir de la subpoblación de individuos que está en riesgo de presentar el desenlace y debería ser tan similar al grupo de casos como para considerar que la única diferencia relevante entre ambos grupos es el hecho de presentar o no el acontecimiento de interés. A menudo se utiliza el apareamiento para obtener un grupo control lo más comparable posible en cuanto a las variables extrañas.

El procedimiento de aplicación de este diseño implica las siguientes fases:

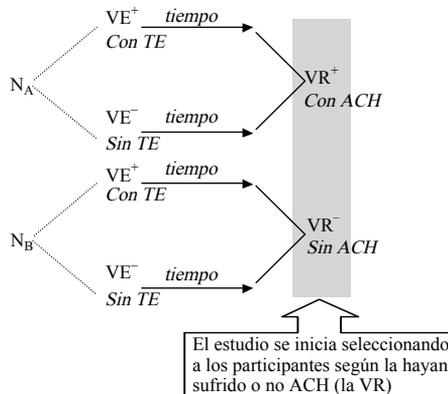
1. Seleccionar una muestra de individuos que presentan la categoría o nivel «crítico» de la VR (también denominado desenlace o evento de interés). Por nivel crítico entendemos el nivel de la VR que se quiere explicar. Simbolizaremos este nivel crítico de la VR como VR⁺ y la población correspondiente con NA. Este es el grupo de «casos».
2. Seleccionar una muestra de individuos que no presentan la categoría o nivel «crítico» de la VR. Simbolizaremos esta categoría como VR⁻ y la población correspondiente con NB. Este es el grupo de «controles».
3. Medir la VE. En general, la medida de estas variables lleva a buscar información relativa al pasado de los participantes.



Por ejemplo,² en una investigación dirigida a contrastar la hipótesis «el exceso de trabajo es un factor de riesgo de accidente cerebrovascular hemorrágico», el desenlace o nivel crítico de la VR es la presencia de un accidente cerebrovascular hemorrágico.

El procedimiento de aplicación de este diseño implicaría las siguientes fases:

1. Seleccionar una muestra de individuos que presentan el desenlace de interés, en este caso, el accidente cerebrovascular hemorrágico. Este grupo corresponde a los casos.
2. Seleccionar una muestra de participantes que no hayan sufrido ningún accidente cerebrovascular hemorrágico y que sean lo más parecidos posible en relación con las variables extrañas relevantes para este estudio. Por ejemplo, se podrían emparejar controles y casos en edad, género y tipo de trabajo.
3. Medir la exposición a horas de trabajo extenuante por semana. De esta manera, se pueden establecer dos grupos según que estén expuestos a más de 8 horas/semana de trabajo extenuante o en menos de 8 horas/semana.



2. Este ejemplo se basa en:

KIM, B. J.; LEE, S.; RYU, W.; KIM, C. K.; CHUNG, J.; KIM, D.; YOON, B. 2013. «Excessive work and risk of haemorrhagic stroke: A nationwide case-control study». *International Journal of Stroke*, 8 (100 A), 56–61.

Puede encontrar otros ejemplos publicados de diseños de casos y controles en:

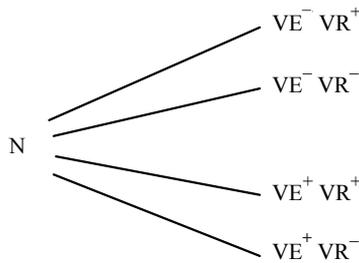
CREWS, M.; LALLY, J.; GARDNER-SOOD, P.; HOWES, O.; BONACCORSO, S.; SMITH, S.; MURRAY, R. M.; GAUGHRAN, F. 2013. «Vitamin D deficiency in first episode psychosis: A case-control study». *Schizophrenia Research*, 150 (2–3), 533–537.

MERLOTTI, E.; MUCCI, A.; VOLPE, U.; MONTEFUSCO, V.; MONTELEONE, P.; BUCCI, P.; GALDERISI, S. 2013. «Impulsiveness in patients with bulimia nervosa: Electrophysiological evidence of reduced inhibitory control». *Neuropsychobiology*, 68 (2), 116–23.

FICHA 4.5. DISEÑO TRANSVERSAL ANALÍTICO

En el contexto no manipulativo en que nos encontramos, los estudios de factores explicativos de un determinado fenómeno se basan, sobre todo, en diseños etiológicos de cohortes y de casos y controles. Sin embargo, no es infrecuente encontrar estudios basados en diseños transversales. Denominamos diseños transversal analíticos^{vii} a aquellos diseños transversales que tienen suficientes elementos de control para aportar alguna evidencia sobre factores explicativos. Hay que decir, sin embargo, que en la literatura inglesa generalmente desaparece el calificativo «analítico», por lo que se denominan genéricamente *cross-sectional study* o *cross-sectional design*.

De la misma manera que en los diseños de casos y controles, este diseño no puede garantizar el criterio de temporalidad, dado que tanto la exposición como el desenlace ya han ocurrido cuando se inicia el estudio. En el diseño transversal analítico, tampoco se inicia la investigación seleccionando muestras de participantes expuestos / no expuestos, o bien con / sin desenlace, sino que se miden simultáneamente todas las variables objeto de estudio y se clasifican los participantes según tengan una combinación u otra de valores en la VE y en la VR. Este diseño puede proporcionar información valiosa con respecto al estudio de factores de riesgo cuando se estudian factores que no varían, como el sexo o la carga genética, o bien exposiciones únicas que no cambian con el tiempo (por ejemplo, haber sido víctima de una violación). Al igual que el resto de diseños *ex post facto*, este diseño es propenso a la amenaza de selección.



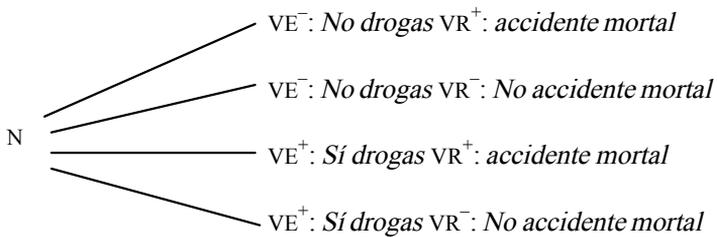
El procedimiento de aplicación de este diseño implica las siguientes fases:

1. Seleccionar una muestra a partir de la población de interés. Un aspecto especialmente crítico de este diseño es la representatividad de la muestra de estudio.
2. Medir las VE y VR objeto de estudio. Los símbolos utilizados en el esquema anterior mantienen la notación de las figuras anteriores e ilustran el caso más simple con una sola variable de cada tipo con dos categorías cada una.

Por ejemplo,³ en una investigación dirigida a contrastar la hipótesis «conducir bajo los efectos de drogas aumenta el riesgo de accidente mortal de tráfico», el desenlace o nivel crítico de la VR es la presencia de un accidente con víctimas mortales.

El procedimiento de aplicación de este diseño implicaría las siguientes fases:

1. La muestra se obtiene a partir de los controles de drogas realizados por los agentes de tráfico todos los sábados entre julio y diciembre de 2013, tanto a los conductores accidentados con resultado de muerte de algún ocupante, como a los no accidentados.
2. Se mide la presencia (VE^+) o ausencia de drogas (VE^-) a partir de pruebas como la espectrometría de masas y el análisis de sangre o de orina, y se registra si ha habido accidente mortal (VR^+) o no (VR^-).



3. Otros ejemplos publicados de aplicación del diseño transversal analítico se pueden encontrar en:

MURIS, P.; FOKKE, M.; KWIK, D. 2007. «The Ruminative Response Style in Adolescents: An Examination of Its Specific Link to Symptoms of Depression». *Cognitive Therapy and Research*, 33 (1), 21–32. Doi: 10.1007/s10608-007-9120-7.

WEIR, K. F.; JOSE, P. E. 2008. «A Comparison of the Response Styles Theory and the Hopelessness Theory of Depression in Preadolescents». *The Journal of Early Adolescence*, 28 (3), 356–74. Doi: 10.1177/0272431608314662.

ACTIVIDADES

Actividad de identificación de diseños

Identifique y represente esquemáticamente los diseños empleados en cada uno de los siguientes estudios.

1. Rasgo de personalidad pesimismo-optimismo y capacidad de ventas

Con la intención de contrastar una de las hipótesis que se desprenden de las teorías expuestas por el psicólogo Martin Seligman⁴ sobre la importancia del optimismo en el éxito personal, un equipo de psicólogos lleva a cabo una investigación sobre los vendedores de seguros de una empresa multinacional, en el que se quiere estudiar la relación entre el rasgo pesimismo-optimismo y las ventas conseguidas en un periodo de dos años. Con este objetivo, se escogen a 28 personas recién incorporadas a la compañía sin experiencia previa en la venta de seguros, 14 clasificadas como pesimistas y 14 como optimistas, de acuerdo con la puntuación de un test de personalidad. Al finalizar un período de dos años, se evalúa la productividad mostrada por cada vendedor. Se espera encontrar que los vendedores optimistas presenten un nivel más alto de productividad que los pesimistas.

2. Violencia filio-parental⁵

La violencia filio-parental (VFP) se ha definido como aquella donde el hijo/a actúa intencionalmente y consciente, con el deseo de causar daño, perjuicio y/o sufrimiento en sus progenitores, de manera reiterada, a lo largo del tiempo, y con la finalidad inmediata de obtener poder, control y dominio sobre sus víctimas para conseguir lo que desea, por medio de la violencia psicológica, económica y/o física. La VFP se ha relacionado con diferentes factores de riesgo individuales, familiares y comunitarios (por ejemplo, consumo de sustancias tóxicas, nivel socioeconómico y entorno familiar violento).

Un equipo de investigación interesado en los factores individuales se propone estudiar el efecto del trastorno por déficit de atención diagnosticado antes de los 12 años sobre las manifestaciones de VFP. Con este objetivo se selecciona una muestra de chicos adolescentes de 14 a 18 años que han sido denunciados por VFP y un grupo de chicos adolescentes de 14 a 18 años que no muestran actitudes ni comportamientos violentos hacia sus progenitores. Se utiliza la técnica del apareamiento para controlar el consumo de sustancias tóxicas, y todos los participantes proceden de familias con un nivel socioeconómico medio y que manifiestan no haber sufrido previamente violencia en

4. Martín Seligman, muy conocido por su teoría de la indefensión aprendida, por ser el principal impulsor de la psicología positiva y por haber sido presidente de la APA, realizó una experiencia sobre los vendedores de seguros de la Empresa MetLife. Seligman (1990). *Learned Optimism: How to Change Your Mind and Your Life*. Nueva York: Vintage Books.

5. Adaptado de: IBABE, I., JAUREGIZAR, J., DÍAZ, O. (2007). *Violencia filio-parental : conductas violentas de jóvenes hacia sus padres*. Vitoria-Gasteiz : Eusko Jaurlaritzaren Argitalpen Zerbitzu Nagusia = Servicio Central de Publicaciones del Gobierno Vasco.

la familia. Para cada participante se busca en su historia clínica la existencia o no de un diagnóstico de trastorno por déficit de atención antes de los 12 años.

3. Impulsividad y conductas de riesgo

En varios estudios se ha encontrado que la impulsividad es un factor explicativo de conductas de riesgo, como son el consumo de drogas y las relaciones sexuales sin preservativo. Sin embargo, se han encontrado algunos resultados que son contradictorios y que podrían ser atribuibles a los efectos del género de la persona. Los resultados de investigaciones previas han llevado a un grupo de investigadores a hipotetizar que los jóvenes del género masculino más impulsivos deberían ser los que más conductas de riesgo presentarían. En el caso de las chicas, no se esperaría ningún efecto diferencial entre las impulsivas y las no impulsivas.

Estos investigadores seleccionaron al azar diez institutos de la provincia de Barcelona, entre los cuales escogieron, también aleatoriamente, doscientos hombres y mujeres de edades comprendidas entre los 15 y los 17 años que no hubieran tomado nunca drogas ni hubieran mantenido relaciones sexuales con penetración. A todos los jóvenes se les administró un cuestionario de impulsividad validado en población joven. De las puntuaciones obtenidas, se calcularon dos medianas, una para los chicos y la otra para las chicas, y se formaron así dos grupos de participantes diferentes para cada género: «impulsivos», que puntuaban por encima de la mediana correspondiente, y «no impulsivos», que lo hacían por debajo.

Un año después de esta primera evaluación, los investigadores se pusieron nuevamente en contacto con los mismos jóvenes y, siguiendo un entrevista estructurada, recogieron información relativa a la frecuencia con que en el último año habían consumido drogas, así como la frecuencia de relaciones sexuales sin preservativo.

Los resultados señalaban que, en el caso de los chicos, los que eran impulsivos presentaban una frecuencia de consumo de drogas superior a la del grupo no impulsivo, y se observaba el mismo efecto en el caso de las relaciones sexuales sin preservativo. Ambas relaciones fueron estadísticamente significativas. En el caso de las chicas, solo fue estadísticamente significativa la relación entre el consumo de drogas y la impulsividad.

4. El funcionamiento familiar y las rehospitalizaciones de los esquizofrénicos⁶

El objetivo de este estudio es analizar la influencia que puede tener la familia en el curso de la enfermedad de las personas que padecen esquizofrenia. Concretamente, se pretende investigar aquellos aspectos familiares que puedan estar relacionados con el número de reingresos hospitalarios de los pacientes esquizofrénicos.

A partir de los archivos de un hospital psiquiátrico, se selecciona una muestra de pacientes diagnosticados de esquizofrenia que han sido rehospitalizados, y de cada uno de los pacientes se obtiene información sobre el número de reingresos y el funcionamiento familiar. El instrumento empleado para la evaluación del funcionamiento

6. Adaptación de VARELA, C. B. 1999. «La influencia de la familia en el paciente esquizofrénico». <http://kunkaak.psicom.uson.mx/maestria/informacion/claudia1.html>. Consulta: 2 de febrero de 2013.

familiar es un cuestionario elaborado por Espejel (1997), el cual permite operativizar la variable funcionamiento familiar con cuatro categorías (MI: muy inadecuado; I: inadecuado; A: adecuado; MA: muy adecuado). El 80 % de estos pacientes son hombres y la media de edad es de 33 años. La media de tiempo con la patología diagnosticada es de 12,21 años, con una media de 4,09 hospitalizaciones. En referencia a la edad de los familiares, la media es de 52 años, con una escolaridad que corresponde a primaria incompleta y a ocupaciones del hogar.

5. Optimismo e ira, posibles causas del malestar físico

En un estudio se pretende analizar la relación entre el optimismo y la experiencia de ira con la frecuencia e intensidad del malestar físico percibido. Se selecciona a un grupo de jóvenes que presentan una alta frecuencia de malestar físico y a otro grupo que no tiene este malestar. Una vez seleccionados los dos grupos de participantes (con malestar físico / sin malestar físico), se mide, a través de las pruebas LOT-R y STAXI-2, su grado de optimismo disposicional y la experiencia de ira, respectivamente. El objetivo final es poner a prueba si estas dos últimas variables (optimismo e ira) son posibles causas del malestar físico.

6. Relación entre acontecimiento vital estresante y trastorno por estrés postraumático

Según indican los resultados de una investigación, el 20 % de las personas que sufren un acontecimiento vital estresante (AVE), como por ejemplo víctimas de violaciones, torturas, terrorismo o guerras, perpetúan el estrés resultante de este AVE, de forma que presentan lo que se denomina trastorno por estrés postraumático (TEPT) el resto de su vida.

En un estudio se escogieron treinta personas que habían sufrido un AVE y a treinta que no lo habían sufrido y que, además, nunca habían sufrido ningún trastorno psicológico importante. Al cabo de diez años se evaluó la presencia de TEPT. Los resultados indicaron que, con respecto a los participantes que habían sufrido un AVE, solo en el 80% de los casos habían desaparecido los síntomas de TEPT. Y por otro lado, tal y como era de esperar, ninguno de los participantes «sanos» no lo presentaba.

EJERCICIOS

Ejercicio 4.1: El nivel de información y la prevención del sida

Actualmente, muchas campañas de educación para la prevención del sida se dirigen a los adolescentes. En relación con este tema, un investigador desea estudiar el efecto del *nivel de información* (características de la enfermedad, estrategias de prevención, etc.) sobre la *adopción de precauciones* en las relaciones sexuales (en adolescentes sexualmente activos). Para llevarlo a cabo, el investigador diseña un cuestionario en el que incluye tres ítems para valorar el nivel de información (NI1-NI3). También incluye una pregunta para valorar el *optimismo irrealista*⁷ de los jóvenes (OI), dado que las investigaciones previas establecen que esta variable está relacionada con el interés por adquirir información y con el uso de protecciones. Por último, el cuestionario incluye una pregunta sobre el uso de preservativos en la última relación sexual (UP).

Para realizar el estudio el investigador parte de una muestra aleatoria de siete institutos de una comunidad autónoma. En cada uno de estos institutos se requiere la participación voluntaria de adolescentes sexualmente activos. A partir de la lista de voluntarios, se seleccionan al azar a siete alumnos de cada centro.

En la tabla 4.3 se incluyen los datos obtenidos en la investigación. Las variables NI1, NI2 y NI3 son indicadores del nivel de información que pueden tomar valores de 0 a 10. El *uso de preservativo* (UP) se ha definido como una variable binaria: se emplea el 0 para codificar la respuesta negativa y el 1 para codificar la respuesta afirmativa. Finalmente, la variable *optimismo irrealista* se ha definido también con dos categorías: presencia (código 1) y ausencia de optimismo irrealista (código 0).

CAS	NI1	NI2	NI3	NI	UP	OI	CAS	NI1	NI2	NI3	NI	UP	OI
1	0	4	5	0	1	1	29	1	1	1	0	1	1
2	1	3	0	0	1	1	35	6	4	9	1	0	1
3	5	9	6	1	1	1	36	6	6	6	1	0	1
4	1	1	1	0	0	1	37	3	3	3	0	1	0
5	1	2	1	0	0	1	39	8	8	8	1	1	0
6	2	2	2	0	0	1	40	3	9	9	1	1	0
9	9	9	3	1	0	1	41	5	5	7	1	1	0
10	3	4	1	0	1	0	42	1	9	5	1	1	0
11	3	4	5	0	1	0	43	0	9	9	1	1	0
12	8	7	5	1	1	0	44	5	5	5	1	1	0
13	6	6	5	1	1	0	45	5	6	5	1	1	0
14	7	5	4	1	1	0	49	2	1	8	0	0	1
15	4	8	9	1	1	0	51	0	9	0	0	0	1
16	5	6	6	1	1	0	59	7	5	1	0	0	1
17	7	7	9	1	1	0	60	5	4	5		0	1
18	7	6	5	1	1	0	69	5	5	2		0	1
19	8	8	4	1	1	0	70	3	4	2		0	1

7. Se denomina *optimismo irrealista* a la infravaloración de la propia vulnerabilidad en relación con la que se atribuye a los iguales.

CAS	NI1	NI2	NI3	NI	UP	OI	CAS	NI1	NI2	NI3	NI	UP	OI
20	0	1	5	0	0	0	79	9	0	0		0	1
21	6	5	6	1	0	0	80	6	7	5		0	1
22	7	6	7	1	0	0	88	2	3	3		0	1
23	8	8	8	1	0	0	89	2	3	1		0	1
24	8	7	9	1	0	0	90	5	5	5		0	1
25	.	1	.	.	0	.	91	1	1	1		0	1
26	5	5	5	1	0	0	99	2	3	4		0	1
27	2	3	2	0	0	1							

Tabla 4.3. Matriz de datos

1. La variable nivel de información se define como una variable binaria, creada a partir de los tres indicadores de información registrados (NI1, NI2 y NI3). Concretamente, se establece que el nivel de información será alto (código 1) si la puntuación media en los tres indicadores es igual o superior a 5; en caso contrario, se considera que el nivel de información es bajo (código 0). Teniendo en cuenta esta definición operativa, para cada sujeto se crea la variable nivel de información (NI) con dos categorías. En la tabla 4.3 se ha creado esta variable para los 38 primeros participantes (columnas sombreadas). Se trata de que complete la creación de esta nueva variable para los diez últimos casos.
 2. Clasifique las variables que intervienen en la investigación.
 3. Valore la estrategia de muestreo empleada.
 4. Identifique el diseño empleado.
 5. Complete la tabla 4.4 para estudiar la relación entre el nivel de información y el uso del preservativo.
 1. ¿Cuál es el porcentaje de participantes que utilizan el preservativo dentro del grupo de los que presentan un nivel de información alto?
 2. ¿Cuál es el porcentaje de participantes que utilizan el preservativo dentro del grupo de los que presentan un nivel de información bajo?
 3. Calcule la diferencia entre las dos proporciones y discuta los resultados obtenidos.
 6. ¿La variable optimismo irrealista ha actuado como una variable de confusión en esta investigación?
 1. Complete la tabla 4.5 para evaluar la relación entre el optimismo irrealista y el nivel de información.
 2. Complete la tabla 4.6 para evaluar la relación entre el optimismo irrealista y el uso de preservativos.
 7. Se trata de estudiar la relación entre el nivel de información y el uso de preservativos controlando el efecto del optimismo irrealista mediante un análisis estratificado.
 1. Realice la misma operación que en el apartado 5 dentro del grupo de participantes que manifiestan optimismo irrealista (tabla 4.7).
 2. Haga la misma operación que en el apartado 5 dentro del grupo de participantes que no manifiestan optimismo irrealista (tabla 4.8).

Tabla 4.4. UP * NI

	<i>Nivel de información (NI)</i>	
Uso de preservativos (UP)	Bajo (0)	Alto (1)
Sí (1)		
No (0)		
TOTAL	22	26

Tabla 4.5. OI * NI

	<i>Optimismo irrealista (OI)</i>	
Nivel de información (NI)	Ausencia (0)	Presencia (1)
Alto (1)		
Bajo (0)		
TOTAL		

Tabla 4.6. OI * UP

	<i>Optimismo irrealista (OI)</i>	
Uso de preservativos (UP)	Ausencia (0)	Presencia (1)
Sí (1)		
No (0)		
TOTAL		

Tabla 4.7. UP*NI. Presencia de optimismo irrealista

	<i>Nivel de información (NI)</i>	
Uso de preservativos (UP)	Bajo (0)	Alto (1)
Sí (1)		
No (0)		
TOTAL	18	6

Tabla 4.8. UP*NI. Ausencia de optimismo irrealista

Uso de preservativos (UP)	Nivel de información (NI)	
	Bajo (0)	Alto (1)
Sí (1)		
No (0)		
TOTAL	4	20

8. ¿Qué información, adicional a la que consta en el enunciado, necesitaría para valorar la validez de constructo de esta investigación?
9. ¿Qué información, adicional a la que consta en el enunciado, necesitaría para valorar la validez externa de esta investigación?

Ejercicio 4.2: Hijos inapetentes por influencia materna

Hace algunos años, un pediatra decía, con gran sentido del humor, que «si los niños comieran los pediatras no comerían», dada la frecuencia con que los padres consultan al especialista por una supuesta falta de apetito de sus hijos. Las investigaciones realizadas hasta el momento revelan que en la mayoría de los casos el niño no padece ninguna enfermedad y, recientemente, se ha propuesto que el problema puede estar relacionado con anomalías en la conducta alimentaria de la madre.

Para estudiar esta hipótesis, se entrevistó a todas las mujeres que acudían al servicio de ginecología de un determinado hospital y que cumplían las siguientes condiciones: 1) tener más de 15 años y 2) esperar su primer hijo. En función de los datos aportados durante la entrevista, se clasificó a las mujeres en dos grupos en función de si presentaban trastornos alimentarios o no. Los dos grupos se siguieron durante cinco años a partir del nacimiento del hijo evaluando diferentes aspectos de la conducta alimentaria del niño. Los resultados indican que el riesgo de presentar alteraciones en la conducta alimentaria de los niños es superior cuando las madres también los presentan.

1. ¿Cuál es la hipótesis que se pretende contrastar?
2. ¿Qué tipo de muestreo se ha utilizado? Comente las repercusiones que puede tener de cara a la generalización de los resultados.
3. ¿Qué diseño se ha utilizado? Indique sus ventajas e inconvenientes.
4. Indique una potencial variable de confusión que se ha controlado en esta investigación.
5. En el caso de encontrarse una relación significativa entre la conducta de la madre y del hijo ante la comida, ¿se puede concluir que la conducta de la madre es la causa de la inapetencia de los niños? Justifique la respuesta.
6. ¿Cómo se podría contrastar la hipótesis de esta investigación utilizando otro tipo de diseño *ex post facto*?
7. A partir de esta investigación, proponga un ejemplo de la amenaza a la validez interna denominada «mortalidad selectiva» o «pérdida no aleatoria de sujetos».

8. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez de constructo de esta investigación?
9. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez externa de esta investigación?

Ejercicio 4.3: Efectos del maltrato infantil

Lea el resumen del artículo de la revista *Child Abuse and Neglect* titulado «Childhood history of abuse and child abuse potential in adolescent mothers: a longitudinal study». Conteste a las siguientes preguntas:

1. Clasifique las variables que intervienen en la investigación.
2. Identifique la estrategia de muestreo empleada.
3. Comente las técnicas de control utilizadas por los autores.
4. Identifique el diseño empleado.
5. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez de constructo de esta investigación?
6. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez externa de esta investigación?

Resumen

Title Childhood history of abuse and child abuse potential in adolescent mothers: a longitudinal study

Author(s) de Paúl, Joaquín; Domenech, Leticia

Journal Child Abuse and Neglect

Collation vol. 24 n. 5 (2000) p. 701-713

Subject Child abuse potential; Adolescent mothers; Childhood history of abuse

Objetivos: el objetivo de esta investigación fue determinar si las madres adolescentes con hijos recién nacidos se encuentran en una situación de mayor riesgo para el maltrato infantil que las madres adultas, y analizar si las madres adolescentes con recuerdos de maltrato infantil tienen un mayor riesgo de maltrato infantil.

Método: se realizó un seguimiento de dos grupos (adultos y adolescentes) de madres embarazadas durante veinte meses, comenzando entre el quinto y séptimo mes de embarazo y finalizando cuando el/la niño/a tenía dieciocho meses. Los grupos de madres adolescentes (N = 24) y de madres adultas (N = 24) fueron emparejadas en todas las variables sociodemográficas. Durante el embarazo se evaluaron los recuerdos de maltrato infantil en todos los sujetos. Cuando el/la niño/a cumplió 1, 6, 12 y 18 meses, se evaluó la situación de riesgo para el maltrato infantil en todas las madres.

Resultados: las madres adolescentes y las madres adultas no presentaron diferencias en los recuerdos de haber sufrido maltrato físico o emocional en su infancia. Sin embargo, las madres adolescentes mostraron un mayor potencial de maltrato y puntuaciones más altas en sintomatología depresiva que las madres adultas. Las madres con recuerdos de haber sufrido castigos físicos severos mostraron un mayor potencial de maltrato infantil y las madres con recuerdos de maltrato físico

infantil mostraron mayor potencial de maltrato y puntuaciones más altas en depresión. Se observó un efecto de interacción entre las variables edad de la madre y maltrato físico en la puntuación en depresión.

Conclusiones: los resultados de este estudio indicaron que el potencial de maltrato infantil es mayor en las madres adolescentes que en las madres adultas y más elevado en las madres que han sido víctimas de maltrato físico que en las que no lo han sido. Asimismo, se aprecia que, entre las madres adolescentes, las que han sido víctimas de maltrato infantil constituyen el grupo de mayor riesgo para el maltrato infantil.

.....

Ejercicio 4.4: Un nuevo diseño para estudiar el comportamiento del consumidor

Revise el ejercicio 2.4 «Comportamiento del consumidor» y elabore una modificación del enunciado para convertirlo en un ejemplo representativo de un diseño *ex post facto* transversal analítico.

Ejercicio 4.5: Estilos de liderazgo

El modelo de Vroom-Yetton (1973, 1988) distingue entre cinco procesos de decisión empresariales que varían en función del grado de participación que la gerencia permite a los subordinados. En el caso menos participativo, la decisión la toma el gerente sin consultar a los subordinados, mientras que en el caso más participativo la decisión se toma en grupo. El modelo predice que, si un gerente adecua el grado de participación de los subordinados al tipo de situación de decisión empresarial, entonces se incrementa la probabilidad de éxito de la decisión.

Un equipo de psicólogos planifica una investigación para contrastar el modelo de Vroom-Yetton. Con este objetivo, seleccionan dos grupos de gerentes en función del porcentaje de éxitos logrados en las decisiones que han ejecutado durante los últimos seis meses. Un grupo está formado por gerentes que han alcanzado un porcentaje de éxitos inferior al 26 % y el otro por los que han alcanzado un porcentaje de éxitos superior al 74 %. A continuación, se realizan una serie de entrevistas para valorar si los procesos seguidos para tomar estas decisiones coinciden con los prescritos por el modelo de Vroom-Yetton. Los resultados apoyan la adecuación del modelo: los gerentes que logran un mayor porcentaje de éxitos en la mayoría de los casos han variado el grado de participación de sus subordinados de acuerdo con las prescripciones del modelo, mientras que el otro grupo se ha decantado por un estilo fijo de decisión (mayoritariamente autocrático o mayoritariamente consultivo).

1. ¿Cuál es la hipótesis que se pretende contrastar?
2. Identifique y clasifique las variables que intervienen en el estudio.
3. ¿Qué diseño se ha utilizado? Indique sus ventajas e inconvenientes.
4. Suponga que para contrastar el modelo de Vroom-Yetton se realiza una nueva investigación que os resumimos a continuación: «Los participantes son una muestra de estudiantes voluntarios. Esta muestra se organiza en grupos y en

cada grupo se escoge al azar al estudiante que actuará como líder (gerente). A continuación, a cada líder se le presenta un conjunto de situaciones de decisión ficticias y se le pide que las resuelva siguiendo los cinco procesos de decisión definidos por el modelo (desde el más autocrático hasta el más participativo). La presentación de los estímulos se realiza de acuerdo con un diseño experimental intrasujeto adecuadamente contrabalanceado. La medida de la variable dependiente es la calidad de las decisiones propuestas por los líderes». ¿Qué mejora supone en relación con el planteamiento inicial?

Ejercicio 4.6: Estudio sobre la timidez

En un estudio sobre la timidez, se quiere contrastar la hipótesis de que el estilo dominante de los padres podría ser un factor explicativo de la timidez. Para llevar a cabo esta investigación se entrevistó a un grupo 50 parejas que esperaban su primer hijo en un centro de atención primaria de Barcelona. La media de edad de los padres y madres era de 31 años. En una de las entrevistas se administró un test de personalidad que permitía clasificar a las parejas como muy dominantes o poco dominantes. Al cabo de tres años, se hizo un seguimiento de la personalidad del hijo, que se clasificó como tímido o no tímido.

1. Identifique y clasifique las variables que intervienen en el estudio.
2. Al inicio de la investigación, ¿está presente el acontecimiento de interés?
3. Identifique el diseño de investigación utilizado. Justifique la respuesta.
4. Indique las técnicas de control empleadas y las variables que aquí se han aplicado.
5. ¿De qué manera podría estar presente la amenaza a la validez interna denominada «mortalidad selectiva» o «pérdida no aleatoria de sujetos»?
6. ¿La amenaza anterior es muy habitual encontrarla en un diseño de casos y controles?
7. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez de constructo de esta investigación?
8. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez externa de esta investigación?

Ejercicio 4.7: Satisfacción en la pareja

Hay estudios que sugieren que la satisfacción en una relación de pareja antes de un episodio depresivo incrementa la probabilidad de remisión (recuperación) de un trastorno depresivo futuro. Para contrastar esta hipótesis, se seleccionó una muestra de 80 parejas estables, en las que uno de los miembros había sufrido un episodio depresivo en los últimos cinco años. Todas las parejas eran heterosexuales. Se midió la satisfacción con la relación del miembro con historial depresivo, y se clasificó en dos categorías: satisfecho y no satisfecho. A medida que los cónyuges con historial depresivo presentaban un nuevo brote, entraban a formar parte del estudio. La muestra final estuvo compuesta por 55 participantes, 30 de los cuales estaban satisfechos con su relación y 25 no lo estaban. Se hizo un seguimiento de la evolución de los síntomas

depresivos en todos los participantes y, al final del estudio, se registró el número de pacientes que se habían recuperado de la depresión.

1. Identifique y clasifique las variables que intervienen en el estudio.
2. Al inicio de la investigación, ¿está presente el acontecimiento de interés?
3. Identifique el diseño de investigación utilizado. Justifique la respuesta.
4. Indique las técnicas de control empleadas y las variables que aquí se han aplicado.
5. Identifique la estrategia de muestreo empleada.
6. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez de constructo de esta investigación?
7. ¿Qué información, adicional a la que consta en el enunciado, necesitará para valorar la validez externa de esta investigación?

Ejercicio 4.8: Los trastornos de la comunicación y del lenguaje en la infancia y trastornos psicopatológicos en la adultez⁸

La investigación longitudinal sobre el desarrollo de la salud mental más allá de la adolescencia en poblaciones no clínicas es deficiente. Este trabajo estudia hasta qué punto los trastornos de la comunicación y del lenguaje (TCL) en la infancia son predictivos de trastornos psicopatológicos en la adultez. Se seleccionó una muestra de 180 niños de 5 años con TCL de los cuales se obtuvo el consentimiento (de los padres) de 142 para continuar en el estudio. A continuación, se seleccionó otra muestra (con consentimiento de los padres) de 142 niños sin TCL emparejada con la primera por edad, sexo y nivel educativo. El diagnóstico de TCL se hizo a través del Peabody Picture Vocabulary Test, del Test of Language Development (TOLD) y del Goldman–Fristoe–Woodcock Test of Auditory Memory.

Los trastornos psicopatológicos se midieron a los 19 y a los 31 años con la entrevista estructurada University of Michigan Composite International Diagnostic Interview (UM-CIDI).

A los 19 años de edad se observó que la cohorte con TCL presentaba más trastornos psicopatológicos que la cohorte sin TCL, aunque la diferencia no fue significativa. A los 31 años ambas cohortes no presentaban diferencias.

1. Indique cuál/es son las variables de exposición.
2. Indique cuál/es son las variables de respuesta.
3. Identifique las variables extrañas.
4. ¿Qué técnicas de control se han aplicado?
5. ¿Cuál es el diseño utilizado?

8. Adaptado de: BAO, L., BROWNIE, E. B., & BEITCHMAN, J. H. (2016). Mental health trajectories from adolescence to adulthood: Language disorder and other childhood and adolescent risk factors. *Development and Psychopathology*, 28(02), 489-504. <https://doi.org/10.1017/S0954579415001054>

Ejercicio 4.9: Alteraciones del lenguaje en niñas y niños prematuros⁹

El objetivo del estudio fue estudiar si el nacimiento prematuro predispone a sufrir posteriormente alteraciones del lenguaje. Para ello se seleccionaron 20 niños de 6 meses de edad que habían nacido pretérmino y 22 los controles de la misma cohorte de nacimiento emparejados con el grupo de niños prematuros por edad, sexo, educación de la madre, lugar de residencia, orden de nacimiento y tipo de familia. A los 8 años de edad se evaluaron las habilidades lingüísticas de los 42 niños

En la prueba psicolingüística, las habilidades auditivas no difirieron entre los niños del grupo pretérmino y los niños del grupo de control. En cambio, los niños prematuros obtuvieron puntuaciones significativamente más pobres que sus controles en los tests visuales. Al estudiar las habilidades del lenguaje de los niños prematuros, sugerimos que se utilicen métodos cualitativos para analizar las habilidades lingüísticas. Recomendamos un seguimiento versátil de los niños prematuros hasta la edad escolar.

1. Indique cuál/es son las variables de exposición.
2. Indique cuál/es son las variables de respuesta.
3. Identifique las variables extrañas.
4. ¿Qué técnicas de control se han aplicado?
5. ¿Cuál es el diseño utilizado?

Ejercicio 4.10: Tipo de personalidad y patología vocal¹⁰

Objetivos. En este estudio, hemos intentado averiguar si el tipo de personalidad (tipo A o B) influye en la incidencia de patología vocal. También evaluamos el volumen del habla y la velocidad del habla para ambos tipos de personalidad.

Diseño. Se seleccionaron 50 participantes con patologías vocales y otros 50 con pliegos vocales normales, y se evaluó el volumen del habla y la velocidad del habla. A continuación, se sometió a evaluación la personalidad.

Resultados. Se encontró que los participantes con personalidad tipo A tenían una incidencia de patología vocal mayor que los participantes con personalidad tipo B ($p = .04$). Los otros dos parámetros (es decir, el volumen del habla y la velocidad del habla) fueron ambos más elevados en sujetos con personalidad tipo A que aquellos con tipo B, pero no alcanzaron significación estadística.

Conclusiones. Este estudio muestra que existe una relación muy estrecha entre el tipo de personalidad y la calidad de la voz, y la incidencia del abuso vocal y las patologías vocales subsecuentes están fuertemente regidas por los rasgos de personalidad de la persona.

9. Adaptado de: YLIHERVA, A., OLSÉN, P., SUVANTO, A. y JÄRVELIN, M. R. (2000). Language abilities of 8-year-old preterm children among the northern Finland 1-year birth cohort for 1985-1986. *Logopedics, Phoniatrics, Vocology*, 25(3), 98-104.
10. Adaptado de: NERURKAR, N. K., KAPRE, G. M. y KOTHARI, N. N. (2016). Correlation between personality type and vocal pathology: A nonrandomized case control study. *The Laryngoscope*, 126(9), 2063-2066. <https://doi.org/10.1002/lary.25810>

1. Indique cuál/es son las variables de exposición.
2. Indique cuál/es son las variables de respuesta.
3. Identifique las variables extrañas.
4. ¿Qué técnicas de control se han aplicado?
5. ¿Cuál es el diseño utilizado?

Ejercicio 4.11: Hogar multilingüe y TEL¹¹

El trastorno específico del lenguaje (TEL) es un trastorno del desarrollo común en niños pequeños. Para investigar la asociación entre el entorno doméstico multilingüe y TEL, llevamos a cabo un estudio en niños chinos de Hong Kong durante un período de 4 años en el Hospital infantil Duchess of Kent. Se revisaron los registros médicos consecutivos de todos los nuevos referidos menores de 5 años y se compararon los niños diagnosticados con TEL con aquellos con lenguaje normal y desarrollo general. Se consideró que un niño presentaba TEL si tenía un cociente de lenguaje de más de una desviación estándar por debajo de la media y por debajo del cociente de desarrollo general en niños con cociente de desarrollo general normal, pero sin enfermedades neurológicas u otras orgánicas. Utilizamos regresión logística binaria y ordinal para evaluar cualquier asociación entre exposición multilingüe en el hogar y TEL, ajustando por edad y género de los sujetos, edad de los padres, nivel de educación y estado laboral, número de hermanos, historial familiar de retraso del lenguaje y cuidador principal en el hogar. Las edades medias de los casos y controles fueron 2.56 y 2.89 años respectivamente. Los varones predominaron en ambos grupos (casos, 75.2%; controles, 60.2%). Los niños estuvieron expuestos a entre uno y cuatro idiomas en el hogar, los más importantes fueron el chino cantonés seguido del inglés.

1. Indique cuál/es son las variables de exposición.
2. Indique cuál/es son las variables de respuesta.
3. ¿La variable de respuesta está operativizada?
4. Identifique las variables extrañas.
5. ¿Qué técnicas de control se han aplicado?
6. ¿Cuál es el diseño utilizado?

Ejercicio 4.12: Efectos de fumar en la salud y en la calidad de la voz¹²

Treinta y dos adultos (20 fumadores y 12 no fumadores) fueron examinados para determinar los efectos del tabaquismo sobre la salud (enfermedades e histología de la laringe), sobre la frecuencia fundamental (regularidad y jitter) y el nivel de estrés. El examen consistió en el análisis nasovideostroboscópico, la evaluación de la electroaringografía para el rendimiento en diferentes tareas y la autoevaluación del estrés.

11. Adaptado de: CHEUK, D. K. L., WONG, V., & LEUNG, G. M. (2005). Multilingual home environment and specific language impairment: a case-control study in Chinese children. *Paediatric and Perinatal Epidemiology*, 19(4), 303-314. <https://doi.org/10.1111/j.1365-3016.2005.00668.x>
12. Adaptado de: GUIMARÃES, I. y ABBERTON, E. (2005). Health and voice quality in smokers: An exploratory investigation. *Logopedics Phoniatrics Vocology*, 30(3-4), 185-191. <https://doi.org/10.1080/14015430500294114>

Aunque no es estadísticamente diferente, los resultados indican que los fumadores en comparación con los no fumadores muestran: 1) más de problemas de salud; 2) cambios histológicos de la laringe; 3) un mayor nivel de estrés; 4) una media F0 más baja para todas las tareas de voz.

Nota: F0 o frecuencia fundamental es una medida de cuán alta o baja es la frecuencia de la voz. Su correlato perceptual es el tono || Jitter: estabilidad de la voz.

1. Indique cuál/es son las variables de exposición.
2. Indique cuál/es son las variables de respuesta.
3. Identifique las variables extrañas.
4. ¿Qué técnicas de control se han aplicado?
5. ¿Cuál es el diseño utilizado?

Ejercicio 4.13: Relación entre audición y cognición¹³

Las últimas investigaciones señalan hacia una relación entre la pérdida auditiva y el deterioro cognitivo derivado del envejecimiento. En este sentido se puede considerar la presbiacusia como una discapacidad auditiva con una elevada vinculación con el rendimiento cognitivo.

El objetivo de la presente investigación ha sido explorar la relación entre audición y cognición durante el envejecimiento, valorando la incorporación de la técnica de la escucha dicótica en el protocolo clínico habitual de las personas mayores.

Se seleccionaron 80 participantes (40 hombres y 40 mujeres) de edades comprendidas entre los 65 y los 85 años, con dominancia manual diestra. Se excluyeron las personas que presentaban un mayor grado de deterioro cognitivo, con una puntuación en la Escala de Deterioro Global –GDS- superior a 3. También fueron rechazados aquellos individuos con una pérdida auditiva global superior a 70 dB (para 500 Hz, 1 kHz, 2 kHz y 4 kHz) y que mostrasen una diferencia entre la pérdida auditiva de ambos oídos superior a 20 dB.

Los participantes fueron clasificados en 4 grupos en función de la presencia o ausencia de Déficit Auditivo –DA- y de Deterioro Cognitivo Leve –DCL- (NoDA-NoDCL; NoDA-DCL; DA-NoDCL; DA-DCL). A continuación, se comparó la presencia / ausencia de DCL según que hubiera DA o no.

1. Indique cuál/es son las variables de exposición.
2. Indique cuál/es son las variables de respuesta.
3. Identifique las variables extrañas.
4. ¿Qué técnicas de control se han aplicado?
5. ¿Cuál es el diseño utilizado?

13. Adaptado de: IVERN, I., VALERO, J., SIGNO, S., VILA, J. M., CATALÀ, M. y TALLEDA, N. (2017). Relación entre audición y cognición durante el envejecimiento: la escucha dicótica como instrumento de evaluación. *Revista de investigación logopédica*, 7(1), 26-46.

LECTURA CRÍTICA

Busque los artículos del siguiente cuadro. Para hacer la búsqueda le sugerimos que utilice el Google Scholar y/o la base de datos Pubmed.

A partir del resumen y de la información del apartado «Método», haga un análisis metodológico del artículo y complete la ficha de lectura crítica. Excluya de este análisis todas las cuestiones relativas al análisis de datos y a la validez de conclusión estadística, ya que escapan a los contenidos expuestos en este texto introductorio.

Artículo 1

Título: «Full Breastfeeding Duration and Associated Decrease in Respiratory Tract Infection in US Children»

Autores: Caroline J. Chantry, Cynthia R. Howard y Peggy Auinger

Revista: *Pediatrics*

Año de publicación: 2006

Artículo 2

Título: «Factores que influyen en la duración de los tratamientos psicológicos empíricamente apoyados»

Autores: Francisco Ballesteros, Paola Fernández y Francisco Javier Labrador

Revista: *Anales de Psicología*

Año de publicación: 2013

Artículo 3

Título: «Influencia de variables sociodemográficas sobre los estilos de afrontamiento, el estrés social y la búsqueda de sensaciones sexuales en adolescentes»

Autores: María de la Paz Bermúdez, Inmaculada Teva y Gualberto Buena-Casal

Revista: *Psicothema*

Año de publicación: 2009

CRITICAL READING CARD: ARTICLE
(FICHA DE LECTURA CRÍTICA: ARTÍCULO)



Remember to thoroughly read the article once found and extract the following information. If you obtain a summary which does not provide enough information you may mark «lack of information» in the corresponding box.

-1- Write the article's **bibliographic reference** according to the APA's referencing style.

-2- Make a **list of the methodological terms** that appear in the method section not related with the study's subject. The first column includes the term in English and the other two the translation in Catalan and Spanish. Include the new terms in your English/Catalan/Spanish Vocabulary.

English	Catalan	Spanish

-3- Indicate which are the **objectives** and **hypothesis** of the research.

-4- Does the study address a clearly focused issue? An issue can be 'focused' in terms of the population studied, the intervention given, the comparator given or the outcomes considered.

-5- Identify which are the **exposure and response variables (EV and RV)**.

EV	RV

-6- Is exposure to EV levels measured in a standard, valid and reliable way?

-7- Are RV measured in a standard, valid and reliable way?

-8- Is it possible to identify the **sampling procedure used**? If so, just copy the information presented about it.

-9- What percentage of individuals refused to participate?

-11- Classify the design used according to the criteria presented in this chapter

-12- What factors other than the EV could affect the RV? Include potential confounding factors like differences in baseline characteristics between groups.

-13- Are individuals within groups blind to the group they are classified into AND are those researchers gathering information blind to the group the individuals belong to?

-14- Were the groups measured equally?

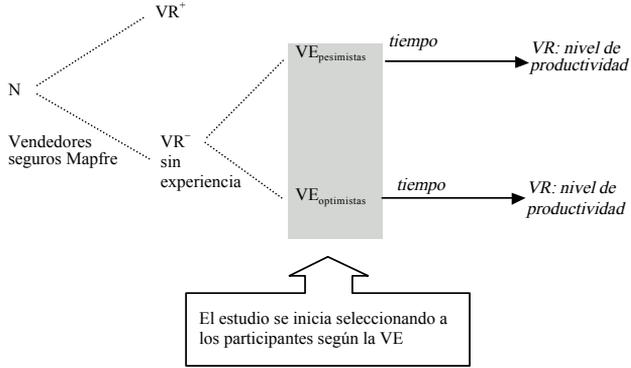
-16- Identify the **control techniques** used and which variables have been controlled with.

-17- Link the previous answers with the concepts of internal and external validity applied to the conclusions of this study.

Internal Validity	
External validity	

SOLUCIONES DE LA ACTIVIDAD DE IDENTIFICACIÓN DE DISEÑOS

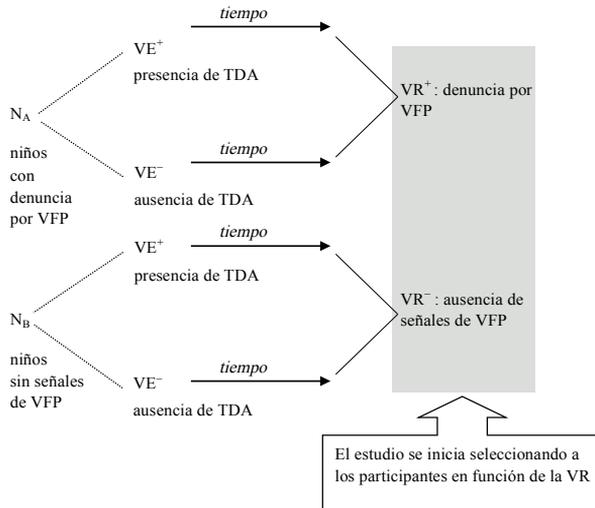
1. Rasgo de personalidad pesimismo-optimismo y capacidad de ventas
Diseño etiológico de cohortes



Fases:

1. Seleccionar una muestra de vendedores de seguros de la compañía Mapfre recién incorporados a la compañía, sin experiencia previa.
2. Medir la variable de exposición (VE, rasgo de personalidad pesimismo-optimismo) y agrupar a los participantes en función de las categorías o los niveles de la VE (VE, pesimismo-optimismo).
3. Efectuar un seguimiento de las cohortes.
4. Medir la VR (productividad).

2. Violencia filiofamiliar
Diseño de casos y controles

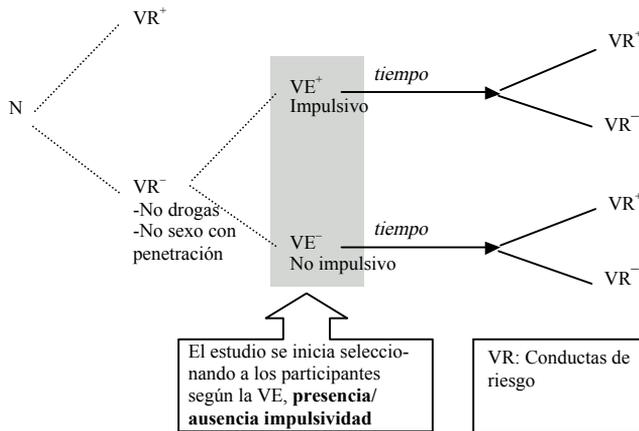


Fases:

1. Seleccionar una muestra de niños con denuncia de VFP.
2. Seleccionar una muestra niños sin señales de VFP.
3. Medir las VE (factores de riesgo relacionado con el TDA). En general, la medida de estas variables lleva a buscar información relativa al pasado de las participantes.

3. Impulsividad y conductas de riesgo

Diseño etiológico de cohortes

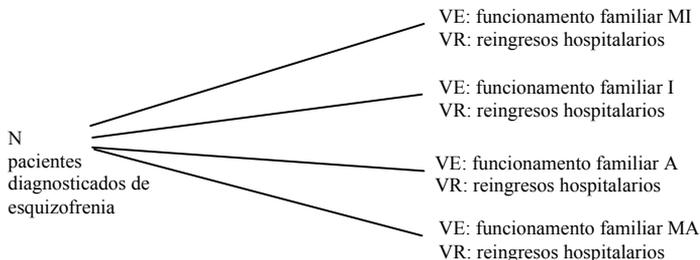


Fases:

1. Seleccionar una muestra de jóvenes que no hayan tomado nunca drogas ni hayan mantenido relaciones sexuales con penetración.
2. Medir la variable de exposición (VE, impulsividad) y agrupar a los participantes en función de las categorías o los niveles de la VE (VE, impulsivo/no-impulsivo).
3. Efectuar un seguimiento de las cohortes.
4. Medir la VR (consumo de drogas) y la VR (sexo sin preservativo).

4. El funcionamiento familiar y las rehospitalizaciones de los esquizofrénicos

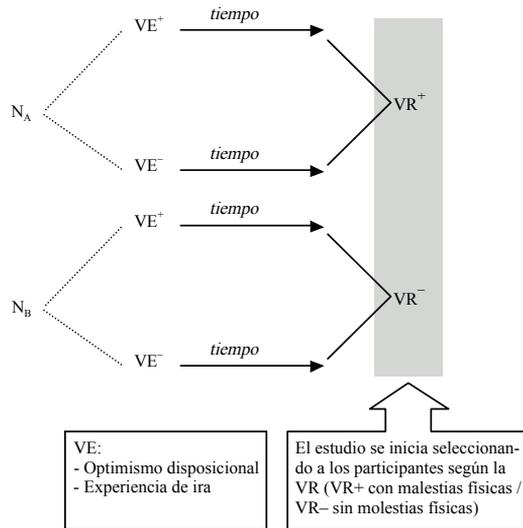
Diseño transversal analítico



Fases:

1. Seleccionar una muestra de pacientes diagnosticados de esquizofrenia de un hospital psiquiátrico.
2. Medir la VE (funcionamiento familiar) y la VR (reingresos hospitalarios) objeto de estudio.

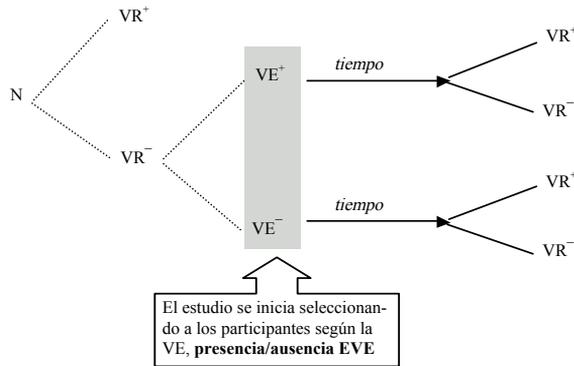
5. Optimismo e ira, posibles causas del malestar físico Diseño de casos y controles



Fases:

1. Seleccionar una muestra proveniente de la población NA de chicos con malestar físico.
2. Seleccionar una muestra de chicos con características similares a las anteriores pero sin malestar físico de la población NB.
3. Medir las VE (optimismo disposicional y experiencia de ira). En general, la medida de estas variables lleva a buscar información relativa al pasado de los participantes.

6. Relación entre acontecimiento vital estresante y trastorno por estrés postraumático
Diseño etiológico de cohortes



Fases:

1. Seleccionar una muestra de personas sin ningún trastorno psicológico importante, lo que implica no padecer TEPT.
2. Medir la variable de exposición (VE: AVE / no AVE) y agrupar a los participantes en función de las categorías o niveles de la VE.
3. Efectuar un seguimiento de las cohortes.
4. Medir la VR (TEPT).

Notas

- i Entre los autores que nos han precedido en el uso de la denominación diseños ex post facto podemos citar Coolican (2009), Hernández, Fernández y Bapista (2006), Kerlinger (1986), León y Montero (2015) o Sarrià (2001). Sin embargo, estos diseños, a menudo, junto con el resto de diseños no experimentales, se agrupan bajo el nombre de diseños observacionales. Puede encontrar ejemplos de este uso en la declaración STROBE (Vandenbroucke, 2007; Von Elm, 2007), a la que hacíamos referencia en la ficha 1.11. Otros ejemplos los hallará en algunos artículos que se citan como ejemplos a lo largo del capítulo.
- ii Es notoria la falta de consenso en la delimitación de los métodos cuantitativos alternativos a la experimentación. Cuando se opta por conceptualizarlos como un único bloque se corre el riesgo de ofrecer una imagen de falta de cohesión interna que amplifica sus puntos débiles respecto al diseño experimental y dificulta la captación de sus fortalezas. La delimitación del método selectivo tal como se hace en este documento se basa en aportaciones de diferentes autores (cfr. Arnau, 1978; Delgado y Prieto, 1997; Hernández et al., 2006; Kerlinger, 1973; Sarrià, 2001). De todos modos, debemos advertir al lector que se puede encontrar otras etiquetas para delimitar todo o parte de lo que aquí referimos como método selectivo (por ejemplo, método correlacional, método comparativo o método de encuestas). Sea cual sea la etiqueta que opte por utilizarse consideramos muy relevante delimitar grupos de diseños en función de sus fortalezas y debilidades respecto a un referente de validez justificado.
- iii Para un relato humorístico sobre las limitaciones que presentan los experimentos (y las revisiones sistemáticas y metanálisis de experimentos) y sobre la importancia de la investigación no manipulativa, recomendamos el artículo de Smith y Pell (2003).
- iv Para una exposición más detallada de estos diseños, puede consultar, entre otros, Margetts, Vorster y Venter (2002), León y Montero (2003), Mann (2003) o Song y Chung (2010). Otra manera de conocer las particularidades de estos diseños es revisando las herramientas de evaluación de la calidad de estudios ex post facto (véase, por ejemplo, Deeks et al., 2003; Jarde, Losilla, Vives y Rodrigo, 2013; Sanderson, Tatto y Higgins, 2007; West et al., 2002), así como las guías de publicación específicas de estos diseños, como la declaración Strobe (Vandenbroucke, 2007), ya que ambos tipos de recursos se centran en los elementos más relevantes de estos diseños.
- v Para una exposición detallada de los diseños etiológicos de cohortes, puede consultar, entre otros, el texto de Lazcano-Ponce et al. (2000).
- vi Para una exposición detallada de los diseños de casos y controles, puede consultar, entre otros, el texto de Armenian (2009).
- vii Para una exposición más detallada del diseño transversal analítico, puede consultar textos antes citados, como los de Mann (2003), Margetts, Vorster y Venter (2002) o Song y Chung (2010).

Referencias

- AMERICAN PSYCHOLOGICAL ASSOCIATION (2010). *Publication manual of the American Psychological Association (6a ed.)*. Washington: Autor.
- ANGUERA, M. T. (1989). Hacia una representación conceptual: Teorías y modelos. En J. Mayor y J. L. Pinillos (Eds.). *Tratado de psicología general. vol I: Teoría, historia y método* (pp. 543-580). Madrid: Alhambra.
- (2003). Observational methods (General). En R. Fernández-Ballesteros (Ed.), *Encyclopedia of behavioral assessment, vol. 2* (pp. 632-637). London: Sage.
- (2010). Posibilidades y relevancia de la observación sistemática por el profesional de la psicología. *Papeles del Psicólogo*, 31(1), 122-130.
- ANGUERA, M. T., BLANCO, A., & LOSADA, J. L. (2001). Diseños observacionales, cuestión clave en el proceso de la metodología observacional. *Metodología de las Ciencias Del Comportamiento*, 3(2), 135-160.
- ANGUERA, M.T., BLANCO-VILLASEÑOR, A., LOSADA, J.L., SÁNCHEZ-ALGARRA, P., & ONWUEGBUZIE, A.J. (2018). Revisiting the Difference Between Mixed Methods and Multimethods: Is It All in the Name? *Quality & Quantity*. doi: 10.1007/s11135-018-0700-2.
- ANGUERA, M.T., CAMERINO, O., CASTAÑER, M., SÁNCHEZ-ALGARRA, P., & ONWUEGBUZIE, A.J. (2017). The Specificity of Observational Studies in Physical Activity and Sports Sciences: Moving Forward in Mixed Methods Research and Proposals for Achieving Quantitative and Qualitative Symmetry. *Frontiers in Psychology*, 8:2196. doi: 10.3389/fpsyg.2017.02196.
- APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, 61, 271-285.
- APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63(9), 839-851. doi:10.1037/0003-066X.63.9.839.
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3-25.
- ARMENIAN, H. (Ed.). (2009). *The Case-Control Method: Design and Applications* (1a ed.). Nova York: Oxford University Press.
- ARNAU, J. (1978). La importancia de la observación en la investigación científica. En J. Arnau (Ed.). *Métodos de investigación en las ciencias humanas* (p. 1-22). Barcelona: Omega.

- (1989). Metodología de investigación y diseño. En J. Mayor y J. L. Pinillos (Eds.). *Tratado de psicología general. Historia, teoría y método* (p. 581-616). Madrid: Alhambra.
- (1995). *Diseños experimentales en esquemas*. Barcelona: Publicacions Universitat de Barcelona.
- ATO, M. (1991). *Investigación en ciencias del comportamiento*. Vol. 1. Fundamentos. Barcelona: PPU.
- (1995a). Conceptos básicos. En M.T. Anguera, J. Arnau, M. Ato, M.R. Martínez Arias, J. Pascual y G. Vallejo (Eds.). *Métodos de investigación en Psicología* (p. 45-70). Madrid: Síntesis.
- (1995b). Tipología de los diseños cuasi-experimentales. En M.T. Anguera, J. Arnau, M. Ato, M.R. Martínez Arias, J. Pascual y G. Vallejo (Ed.). *Métodos de investigación en Psicología* (p. 245-269). Madrid: Síntesis.
- ATO, M., LÓPEZ-GARCÍA, J. J., & BENAVENTE, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3). doi:10.6018/analesps.29.3.178511.
- ATO, M., & VALLEJO, G. (2007). *Diseños experimentales en psicología*. Madrid: Pirámide.
- BABBIE, E. R. (2011). *The basics of social research*. Belmont, CA: Wadsworth.
- BALLUERKA, N., & VERGARA, A.I. (2002). *Diseños de investigación experimental en psicología*. Madrid: Prentice Hall.
- BARON, R. M., & KENNY, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182. doi:10.1037/0022-3514.51.6.1173.
- BENNETT, C., KHANGURA, S., BREHAUT, J. C., GRAHAM, I. D., MOHER, D., POTTER, B. K., & M. GRIMSHAW, J. (2011). Reporting Guidelines for Survey Research: An Analysis of Published Guidance and Reporting Practices. *PLoS Medicine*, 8(8). doi:10.1371/journal.pmed.1001069.
- BERRA, S., ELORZA-RICART, J.M., ESTRADA, M.D., & SÁNCHEZ, E. (2008). A tool for the critical appraisal of epidemiological cross-sectional studies. *Gaceta Sanitaria*, 22, 492-497.
- BORDENS, K., & ABBOTT, B. B. (2011). *Research design and methods: a process approach* (8th ed.). Nova York: McGraw-Hill.
- BRYMAN, A. (2006a). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1), 97-113.
- (2006b). Paradigm peace and the implications for quality. *International Journal of Social Research Methodology: Theory & Practice*, 9(2), 111-126.
- (2012). *Social research methods* (4a ed.). Nova York: Oxford University Press.
- BUNGE, M. A. (1981). *La ciencia: Su método y su filosofía*. Buenos Aires: Siglo veinte.
- (1985). *La investigación científica*. Barcelona: Ariel.
- CAMPBELL, D.T., & STANLEY, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally Company.
- (1973). Diseños experimentales y cuasiexperimentales en la investigación social. Argentina: Amorrortu (traducción del original publicado en 1966).

- CASTAÑER, M., CAMERINO, O., & ANGUERA, M. T. (2013). Métodos mixtos en la investigación de las ciencias de la actividad física y el deporte. *Apunts. Educación física y deportes*, 2(112), 31-36.
- CHACÓN, S., & SHADISH, W. R. (2008). Validez en evaluación de programas. En M. T. Anguera, S. Chacón y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (p. 69-102). Madrid: Síntesis.
- CHACÓN, S., SANDUVETE, S., PORTELL, M., & ANGUERA, M. T. (2013). Reporting a program evaluation: Needs, program plan, intervention, and decisions. *International Journal of Clinical and Health Psychology*, 13, 58-66.
- COOLICAN, H. (2009). *Research methods and statistics in psychology*. Londres: Hodder Education.
- COOK, T. D., & CAMPBELL, D. T. (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Chicago: Rand McNally College Pub. Co.
- (1986). The causal assumptions of quasi-experimental practice. *Synthese*, 68, 141-160.
- COOPER, H., HEDGES, L. V., & VALENTINE, J. C. (Eds.). (2009). *The Handbook of Research Synthesis and Meta-Analysis* (2nd ed.). Nova York: Russell Sage Foundation.
- CRESWELL, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Method Approaches* (3a ed.). Thousand Oaks, CA: Sage Publications.
- CRESWELL, J. W. & PLANO-CLARK, V. L. (2007). *Designing and Conducting Mixed Methods Research*. Thousand Oaks: SAGE Publications.
- CHACÓN-MOSCO. S., SANDUVETE-CHAVES, S., & SÁNCHEZ-MARTÍN, M. (2016). The Development of a Checklist to Enhance Methodological Quality in Intervention Programs. *Frontiers in Psychology*, 7:1811. .3389/fpsyg.2016.01811
- DEEKS, J. J., DINNES, J., D'AMICO, R., SOWDEN, A. J., SAKAROVITCH, C., SONG, F., ... European Carotid Surgery Trial Collaborative Group. (2003). Evaluating non-randomised intervention studies. *Health technology assessment*, 7(27), 1-173.
- DELGADO, A. R., & PRIETO, G. (1997). *Introducción a los métodos de investigación de la psicología*. Madrid: Pirámide.
- DOMÉNECH, J.M., GRANERO, R., LOSILLA, J.M. & PORTELL, M. 1999. *Curs d'Anàlisi de Dades en Psicologia de la Salut*. Vol. 2: Contrast d'hipòtesi bivariable i introducció als models de regressió lineal múltiple en recerca no experimental. Terrassa: Cardellach.
- EVIDENCE-BASED MEDICINE WORKING GROUP (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*, 268(17), 2420-2425.
- GAMBARA, H. (2002). *Métodos de investigación en Psicología y Educación: Cuaderno de prácticas* (3a ed.). Madrid: McGraw-Hill.
- GENERALITAT DE CATALUNYA. DEPARTAMENT DE SALUT. (2009). *Guia de pràctica clínica sobre infeccions de transmissió sexual*. Barcelona: Departament de Salut.
- GLASS, G. V (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- GÓMEZ, J. (1990). Metodología de encuesta por muestreo. En J. Arnau, M. T. Anguera y J. Gómez (Eds.). *Metodología de la investigación en ciencias del comportamiento* (p. 237-310). Múrcia: Universidad de Murcia.

- GROVES, R. M., FOWLER, F. J., COUPER, M. P., LEPKOWSKI, J. M., SINGER, E., & TOURANGEAU, R. (2009). *Survey Methodology* (2a ed.). Hoboken, NJ: John Wiley & Sons.
- HERNÁNDEZ, R., FERNÁNDEZ, C. & BAPTISTA, P. (2006). *Metodología de la investigación* (4 ed.). México: McGraw Hill.
- HIGGINS, J. P. T., GREEN, S., & Cochrane Collaboration. (2008). *Cochrane handbook for systematic reviews of interventions*. Chichester, England; Hoboken, NJ: Wiley-Blackwell.
- JARDE, A., LOSILLA, J. M., VIVES, J., & RODRIGO, M.F. (2013). Q-Coh: A tool to screen the methodological quality of cohort studies in systematic reviews and meta-analyses. *International journal of clinical and health psychology*, 13(2), 138-146
- JARLAIS, D. C. D., LYLES, C., CREPAZ, N., & the TREND Group. 2004. Improving the Reporting Quality of Nonrandomized Evaluations of Behavioral and Public Health Interventions: The TREND Statement. *American Journal of Public Health*, 94(3), 361.
- JOHNSON, R. B., & ONWUEGBUZIE, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33(7), 14-26. doi:10.3102/0013189X033007014.
- KERLINGER, F. N. (1973). *Foundations of behavioral research* (2 ed.). Nova York: Holt, Rinehart and Winston.
- KIRK, R.E. (2013). *Experimental Design: Procedures for Behavioral Sciences* (4a ed.). Los Angeles: SAGE.
- LABRADOR, F.J., VALLEJO, M.A., MATELLANES, M., ECHEBURÚA, E., BADOS, A., & FERNÁNDEZ-MONTALVO, J. (2003). *La eficacia de los tratamientos psicológicos*. Documento de la Sociedad Española para el avance de la Psicología Clínica y de la Salud. Siglo XXI. Noviembre del 2002. INFOCOP, 17, 25-30.
- LAZCANO-PONCE, E., FERNÁNDEZ, E., SALAZAR-MARTÍNEZ, E., & HERNÁNDEZ-AVILA, M. (2000). Estudios de cohorte. Metodología, sesgos y aplicación. *Salud Pública de México*, 42(3), 230-241.
- LIBERATI, A., ALTMAN, D. G., TETZLAFF, J., MULROW, C., GOTZSCHE, P. IOANNIDIS, J. P. A., MOHER, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*, 339, b2700-b2700. doi:10.1136/bmj.b2700.
- LEÓN, O., & MONTERO, I. (2001). Cómo explicar el concepto de interacción sin estadística: análisis gráfico de todos los casos posibles en un diseño 2x2. *Psicothema*, 13(1), 159-165.
- (2015). *Métodos de investigación en psicología y educación* (4a ed.). Madrid: McGraw-Hill.
- LEVITT, H., BAMBERG, M., CRESWELL, J.W., FROST, D.M., JOSSELSO, R., & SUAREZ-OROZCO, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73, (1), 26-46.
- LOHR, S. L. (2010). *Sampling: design and analysis*. Boston, MA: Brooks/Cole.
- LÓPEZ, J. S., BLANCO, F., SCANDROGLIO, B., & GUTMAN, I. R. (2010). Una aproximación a las prácticas cualitativas en psicología desde una perspectiva integradora. *Papeles del Psicólogo*, 31(1), 131-142.

- MANN, C. J. (2003). Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency Medicine Journal*, 20(1), 54-60. doi:10.1136/emj.20.1.54
- MARGETTS, B. M., VORSTER, H.H., & VENTER, C.S. (2002). Evidence-based nutrition – Review of nutritional epidemiological studies. *SAJCN*, 15(3), 68-73.
- MARTÍNEZ-ARIAS, R. (1995). Metodología de encuestas. En M. T. Anguera, J. Arnau, M. Ato, R. Martínez-Arias, J. Pascual y G. Vallejo (Eds.). *Métodos de investigación en psicología* (pp. 383-510). Madrid: Síntesis.
- MAYOR, J. (1989). El método científico en psicología. En J. Mayor y J. L. Pinillos (Eds.). *Tratado de psicología general. historia, teoría y método* (p. 3-70). Madrid: Alhambra.
- MONTERO, I., & LEÓN, O. G. (2007). A guide for naming research studies in psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847-862.
- MORENO, R. (1989). Explicación y causalidad. Enfoques y alternativas. En J. Mayor y J. L. Pinillos (Eds.). *Tratado de psicología general. Historia, teoría y método*. (p. 505-542). Madrid: Alhambra
- Moreno, R., Martínez, R. J. i Chacón, S. (2000). *Fundamentos metodológicos en psicología y ciencias afines*. Madrid: Pirámide.
- PARDO, A., & SAN MARTÍN, R. (2010). *Análisis de datos en ciencias sociales y de la salud II*. Madrid: Síntesis.
- PORTELL, M., & DOMÈNECH, J.M. (1997). Una propuesta para el análisis de respuestas binarias en diseños “cross-over”. *Psicothema*, 9(2), 407-415.
- RAO, K. (1995). *Estadística y verdad*. Barcelona: PPU.
- RIBA, C. (2007). *La metodología cualitativa en l'estudi del comportament*. Barcelona: Editorial UOC.
- SACKETT, D.L., ROSENBERG, W., MUIR, J.A., HAYNES, R.B., & RICHARDSON, W.S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 312, 71-72.
- SÁNCHEZ-ALGARRRA, P., & ANGUERA, M. T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: impact of recording and coding among predominating perspectives. *Quality & Quantity*, 47(2), 1237-1257.
- SÁNCHEZ-MECA, J., BORUCH, R.F., PETROSINO, A., & ROSA A.I. (2002). La colaboración Campbell y la práctica basada en la evidencia. *Papeles del Psicólogo*, 22, 83, 44-48.
- SÁNCHEZ-MECA, J., & BOTELLA, J. (2010). Revisiones sistemáticas y meta-análisis: herramientas para la práctica profesional. *Papeles del Psicólogo*, 31(1), 7-17.
- SANDERSON, S., TATT, I. D. & HIGGINS, J. P. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *International Journal of Epidemiology*, 36(3), 666-676.
- SARRIÁ, E. (2001). Métodos y diseños de investigación. En S. Fontes, C. García, A. J. Garriga, M. C. Pérez-Llantada y E. Sarriá (Eds.). *Diseños de investigación en psicología* (p. 95-122). Madrid: Universidad Nacional de Educación a Distancia.
- SCHULZ, K. F., ALTMAN, D. G., & MOHER, D. (2010). CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. *PLoS Medicine*, 7(3). doi:10.1371/journal.pmed.1000251

- SHADISH, W. R., COOK, T. D., & CAMPBELL, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- SHANEYFELT T. (2001). Building bridges to quality. *JAMA*, 286, 2600-2601.
- SHAUGHNESSY, J. J., ZECHMEISTER, E. B., & ZECHMEISTER, J. S. (2015). *Research methods in psychology* (10a ed.). Nova York, NY: McGraw-Hill.
- SILVA, L.C. (2000). *Diseño razonado de muestras y captación de datos para la investigación sanitaria*. Madrid: Diaz de Santos.
- SMITH, M.L., & GLASS, G.V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-60.
- SMITH, G. C. S., & PELL, J. P. (2003). Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ*, 327(7429), 1459-1461. doi:10.1136/bmj.327.7429.1459.
- SOLANAS, A., SALAFRANCA, L., FAUQUET, J., & NÚÑEZ, M.I. (2005). *Estadística descriptiva en Ciencias del Comportamiento*. Madrid: Paraninfo.
- TACQ, J. (2011). Causality in qualitative and quantitative research. *Quality & Quantity*, 45, 263–291.
- TASHAKKORI, A., & TEDDLIE, C. (2010). *Sage handbook of mixed methods in social & behavioral research*. Los Angeles, CA: SAGE Publications.
- TEDDLIE, C., & TASHAKKORI, A. (2008). *Foundations of mixed methods research : integrating quantitative and qualitative techniques in the social and behavioral sciences*. Londres: SAGE.
- TONG, A., FLEMMING, K., MCINNES, E., OLIVER, S., & CRAIG, J. (2012). Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology*, 12(1), 181. doi:10.1186/1471-2288-12-181.
- TONG, A., SAINSBURY, P., & CRAIG, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19(6), 349–357. doi:10.1093/intqhc/mzm042
- VALENTINE, J. C., & COOPER, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD). *Psychological Methods*, 13(2), 130–149. doi:10.1037/1082-989X.13.2.130
- VALLEJO, G. (1995). Diseños de series temporales interrumpidas. Dins M.T. Anguera, J. Arnau, M. Ato, M.R. Martínez Arias, J. Pascual i G. Vallejo (Eds.). *Métodos de investigación en Psicología* (p. 321-334). Madrid: Síntesis.
- VANDENBROUCKE, J. P., VON ELM, E., ALTMAN, D. G., GØTZSCHE, P. C., MULROW, C. D., POCKOCK, S. J., ... EGGER, M. (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE). *Epidemiology*, 18(6), 805–835. doi:10.1097/EDE.0b013e3181577511
- VIADER, M. (1996). *Diseño y análisis de experimentos en ciencias del comportamiento*. Barcelona: PPU.
- VIVES, PORTELL i BOIXADÓS (2009). *Mètodes d'investigació quantitativa. Episodi 0*. Barcelona: FUOC.
- VON ELM, E., ALTMAN, D. G., EGGER, M., POCKOCK, S. J., GØTZSCHE, P. C., & VANDENBROUCKE, J. P. (2007). The Strengthening the Reporting of Observational Stu-

- dies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Preventive Medicine*, 45(4), 247-251. doi:10.1016/j.ypmed.2007.08.012
- WEGENER, D., & FABRIGAR, L. (2000). Analysis and design for nonexperimental data addressing causal and noncausal hypothesis. Dins H. T. Reis i C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (p. 412-450). Nova York: Cambridge University Press
- WEIR, K. F., & JOSE, P. E. (2008). A Comparison of the Response Styles Theory and the Hopelessness Theory of Depression in Preadolescents. *The Journal of Early Adolescence*, 28(3), 356-374. doi:10.1177/0272431608314662
- WEST, S., KING, V., CAREY, T. S., LOHR, K. N., MCKOY, N., SUTTON, S. F., & LUX, L. (2002). *Systems to Rate the Strength of Scientific Evidence* (Evidence Report/Technology Assessment No. 47) (p. 64-88). Rockville; MD: Agency for Healthcare Research and Quality.
- WU, A. D. & ZUMBO, B. D. (2007). Understanding and Using Mediators and Moderators. *Social Indicators Research*, 87(3), 367-392. doi:10.1007/s11205-007-9143-1

