



International Journal of Clinical and Health Psychology

www.elsevier.es/ijchp



ARTÍCULO TEÓRICO

Reporting a program evaluation: Needs, program plan, intervention, and decisions

Salvador Chacón Moscoso^{a,*}, Susana Sanduvete Chaves^a, Mariona Portell Vidal^b,
M. Teresa Anguera Argilaga^c

^aUniversidad de Sevilla, España

^b Universidad Autónoma de Barcelona, España

^c Universidad de Barcelona, España

Recibido el 29 de septiembre 2012; aceptado el 17 de octubre de 2012

PALABRAS CLAVE

Informe;
Evaluación de
programas;
Necesidades;
Diseño;
Estudio teórico

Resumen Se encuentran diferentes posturas respecto a programas de intervención en función de la perspectiva metodológica adoptada, por lo que el profesional de la salud no dispone de unas directrices claras de actuación, dificultándose la acumulación del conocimiento. El objetivo propuesto es concretar los aspectos básicos/mínimos y comunes a explicitar en el informe de evaluación de cualquier programa, útil para el profesional, independientemente de la opción procedimental que se elija, fomentando de este modo la integración y la complementariedad entre metodologías, como respuesta a las circunstancias reales del contexto de intervención en cambio continuo. Estos aspectos se encuadran en las principales fases de evaluación: necesidades, objetivos y diseño (antes de la intervención), implementación (durante esta) y resultados (después de la intervención). En cada una de ellas, se explicita en qué elementos basar la toma de decisiones a partir de evidencias empíricas registradas mediante instrumentos en unas muestras siguiendo un procedimiento determinado.

© 2012 Asociación Española de Psicología Conductual. Publicado por Elsevier España, S.L. Todos los derechos reservados.

KEYWORDS

Report;
Program evaluation;
Needs;

Abstract The approach to intervention programs varies depending on the methodological perspective adopted. This means that health professionals lack clear guidelines regarding how best to proceed, and it hinders the accumulation of knowledge. The aim of this paper is to set out the essential and common aspects that should be included in any program evaluation report, thereby providing a useful guide for the professional regardless of the procedural approach used. Furthermore, the paper seeks to integrate the different methodologies and illustrate

* Dirigir correspondencia a: Facultad de Psicología, Universidad de Sevilla, Campus Ramón y Cajal, C/ Camilo José Cela, s/n, 41018 Sevilla, España.

E-mail: schacon@us.es (S. Chacón Moscoso).

Design;
Theoretical study

their complementarity, this being a key aspect in terms of real intervention contexts, which are constantly changing. The aspects to be included are presented in relation to the main stages of the evaluation process: needs, objectives and design (prior to the intervention), implementation (during the intervention), and outcomes (after the intervention). For each of these stages the paper describes the elements on which decisions should be based, highlighting the role of empirical evidence gathered through the application of instruments to defined samples and according to a given procedure.

© 2012 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.
All rights reserved.

Se encuentran posturas antagónicas respecto a programas de intervención justificados desde diferentes orientaciones metodológicas (Anguera, 2003; Wallraven, 2011), por lo que el profesional de la salud no dispone de unas directrices claras de actuación respecto al diseño, implementación y evaluación en contextos de intervención en cambio continuo, lo que dificulta la acumulación integrada de conocimiento. El objetivo propuesto es concretar los aspectos mínimos/básicos y comunes a explicitar en el informe de evaluación de cualquier programa, útil para el profesional, independientemente del planteamiento que tenga (por ejemplo, cualitativo o cuantitativo; o experimental, cuasi-experimental u observacional).

Este informe se basa en un planteamiento de diseño y evaluación en interacción continua y se explicita de manera operativa. Desde el momento inicial de la evaluación de necesidades hasta la evaluación final de resultados, se justifica un continuo de toma de decisiones que ha de estar basado en evidencias empíricas desde la metodología científica (Anguera y Chacón, 2008), independientemente de la opción metodológica o procedimental por la que se opte en cualquier fase del proceso. Este esquema de informe fundamentalmente pretende ser útil para el profesional de tal manera que, al explicitar los criterios de toma de decisión a lo largo de todo el proceso de intervención, se posibilitará la integración de conocimiento científico en el ámbito de la salud desde unos referentes comunes de actuación.

En la actualidad, la dispersión en las intervenciones y la forma de comunicarlas (e.g., Bornas et al., 2010; Gallego, Gerardus, Kooij y Mees, 2011; Griffin, Guerin, Sharry y Drumm, 2010) imposibilitan su evaluación sistemática y extrapolación de resultados (Chacón y Shadish, 2008). Otros autores definieron previamente los elementos a concretar en el informe de evaluación de un programa (Cornelius, Perrio, Shakir y Smith, 2009; Li, Moja, Romero, Sayre y Grimshaw, 2009; Moher, Liberati, Tetzlaff y Altman, 2009; Schulz, Altman y Moher, 2010). Se considera, sin embargo, que esta propuesta aporta las siguientes ventajas en comparación con lo encontrado en la literatura al respecto: (a) Es integradora, adaptable a todas las metodologías, en lugar de ser aplicable exclusivamente a los programas con diseño experimental y/o cuasi-experimental; (b) Se centra en aspectos metodológicos comunes a lo largo del proceso, lo que

permite extrapolar y generalizar su uso a diferentes temáticas concretas en el ámbito clínico y de la salud, en lugar de estar supeditada a un contexto específico; (c) El uso de un esquema común de tintes metodológicos aplicado a diferentes contextos facilita el estudio comparativo entre intervenciones y, por tanto, la acumulación sistemática del conocimiento; (d) No sólo permite el análisis del informe de intervenciones ya en marcha o finalizadas sino que, además, puede ser útil para la planificación, diseño, implementación y evaluación de futuros programas; (e) Resulta de fácil y rápida aplicación, tal como se puede constatar por el menor número de elementos a considerar en nuestra propuesta al compararla con otras; y (f) Presenta un lenguaje no excesivamente específico, lo que posibilita su aplicación por parte de profesionales sin formación especializada en metodología.

Se parte de la idea de que el objetivo de la evaluación de un programa es la emisión de juicios de valor acerca de la valía de dicho programa o de alguno de sus elementos (Anguera y Chacón, 2008). Para ello, se presentan diferentes fases de evaluación divididas en tres períodos: *antes* de la intervención, *durante* ésta y *después* de la intervención (Anguera, Chacón y Sánchez, 2008). El primer período se subdivide en evaluación de necesidades, objetivos y diseño, el segundo hace referencia a la evaluación de la implementación y el tercero a la evaluación de los resultados. En la Tabla 1 se presenta, en orden cronológico, el esquema del informe de evaluación de un programa y, a modo ilustrativo, se muestran ejemplos característicos en cada apartado.

Estas fases conforman una evaluación formativa y sumativa en interacción continua. Desde este planteamiento el programa, gracias a la evaluación continua a la que es sometido, es capaz de autorregularse y de aportar información útil a todos los grupos de implicados a lo largo de todo el proceso de intervención (Chacón, López y Sanduete, 2004).

La toma de decisiones que se realiza a lo largo de todo el proceso de evaluación (desde necesidades hasta resultados) ha de basarse en evidencias empíricas registradas mediante instrumentos apropiados en unas muestras explícitas siguiendo un procedimiento determinado. Es por ello que, antes de pasar a la descripción pormenorizada de los aspectos básicos incluidos en cada una de las fases de evaluación se presenta, conformando el método, una conceptualización general de muestra, instrumentos y procedimiento, como elementos comunes en todas las fases.

Tabla 1 Esquema del informe de evaluación de un programa.*Evaluación de necesidades*

- Conceptualización: problema a solucionar. Tipo de necesidad: normativa, percibida, expresada, comparada
- Muestra: proveniente de la población diana
- Instrumentos: e.g., encuestas, entrevistas, grupos de discusión, historias de vida
- Procedimiento: e.g., elaboración de instrumentos, aplicación, registro de datos
- Valoración: pertinencia, criterios de priorización de necesidades, cobertura y criterios operativos de inclusión/exclusión

Evaluación de objetivos

- Conceptualización: objetivos general/es (metas). Objetivos específicos (identificables, medibles/operativos, expresados temporalmente): ligados a objetivo general y a necesidades específicas
- Muestra: modelos teóricos existentes, estudios previos
- Instrumentos: e.g., datos secundarios, sistemas de codificación
- Procedimiento: e.g., revisiones sistemáticas, simulaciones
- Valoración: usabilidad, suficiencia

Evaluación del diseño

- Conceptualización: plan de trabajo donde las actividades, los recursos necesarios, la temporalización y momentos de registro están ligados a los objetivos específicos de acuerdo a necesidades (tipos de diseños; componentes de diseños y validez)
- Muestra: programas homólogos existentes
- Instrumentos: e.g., datos secundarios, sistemas de codificación
- Procedimiento: e.g., revisiones sistemáticas, simulaciones
- Valoración: coherencia interna, factibilidad, aceptabilidad, viabilidad

Evaluación de la implementación

- Conceptualización: ejecución de lo planificado
- Muestra: usuarios/as de los que se registran variables relevantes
- Instrumentos: e.g., entrevistas, cuestionarios, registros de observación, grupos de discusión, datos estandarizados
- Procedimiento: e.g., elaboración de instrumentos, aplicación, análisis de datos
- Valoración: calidad o bondad, solución a discrepancias, evaluabilidad

Evaluación de resultados

- Conceptualización: grado de consecución de los objetivos
- Muestra: participantes de los que se recogen datos
- Instrumentos: homólogos a los usados en la fase de evaluación de necesidades; e.g., encuestas, entrevistas, grupos de discusión
- Procedimiento: e.g., elaboración de instrumentos, aplicación, análisis de datos
- Valoración: eficacia, efectividad, eficiencia, continuidad, progreso, adecuación, utilidad, probidad, equidad

Método**Muestra**

En líneas generales, la muestra son los usuarios/as de los que se recogen datos para llegar a conclusiones en la evaluación (Anguera, Chacón y Sánchez, 2008). Así, en evaluación de necesidades, implementación y resultados son, respectivamente, las personas estudiadas para determinar las carencias sobre las que incidir en el programa, si se está llevando a cabo la intervención tal y como se planificó, y los efectos del programa. La explicitación de las características de los usuarios/as (la muestra) es fundamental de cara a una posible generalización de los resultados (Chacón, Anguera y Sánchez-Meca, 2008). Sin embargo, la muestra no siempre hace referencia a personas, sino a unidades de estudio (*units*), de tal manera que la muestra en la fase de evaluación de objetivos y diseño son los modelos teóricos y programas similares estudiados de los que se extraen posibles objetivos y elementos del diseño aplicables al programa que se está planificando.

Instrumentos

Los instrumentos son recursos metodológicos que permiten recoger la información empírica necesaria para llevar a cabo la evaluación (Anguera, Chacón, Holgado y Pérez, 2008). El profesional dispone de una gama de técnicas de recogida de datos que comprenden desde los datos secundarios, que requieren una interacción mínima con los usuarios/as del programa (como la revisión de datos archivados) y la observación, preferentemente a partir de sesiones grabadas, una vez se dispone del consentimiento del usuario o del responsable), hasta las que requieren una interacción activa con éstos (como entrevistas y grupos de discusión), pasando por las que implican una moderada interacción personal con la situación (como escalas, tests, cuestionarios y otros sistemas de registro de datos estandarizados), y que consisten en una elicitación de la respuesta del usuario; a su vez, este criterio se combina con el nivel de estandarización que presente el instrumento (Anguera, Chacón, Holgado et al., 2008).

Los instrumentos más comúnmente utilizados en la fase de evaluación de necesidades para la recogida de información son las encuestas a los potenciales usuarios/as, entrevistas en profundidad a implicados y expertos, grupos de discusión o las historias de vida; en las fases de evaluación de objetivos y diseño, prima la consulta de datos secundarios y los sistemas de codificación; los instrumentos más utilizados para la recogida de información en la fase de implementación son las entrevistas y cuestionarios realizados a los implicados, los registros de observación, los grupos de discusión y el sistema de registro de datos estandarizados como, por ejemplo, los sistemas de indicadores, las fichas, las listas de verificación o los autoinformes; finalmente, en la fase de resultados suelen utilizarse prácticamente todos los instrumentos previamente enumerados, teniendo especial interés el uso de los mismos u homólogos instrumentos que en la fase de evaluación de necesidades para facilitar la comparación de datos antes y después de la intervención (e.g., Casares et al., 2011).

Procedimiento

El procedimiento es el modo de ejecutar las acciones necesarias para llevar a cabo el registro y análisis de datos; por ejemplo, cómo se elaboró el instrumento de medida, cómo se aplicó dicho instrumento, en qué momentos se tomaron medidas, o cómo se analizaron los datos en función de los objetivos, de la naturaleza de los datos y del tamaño de las muestras (tanto de usuarios/as como de ocasiones de registro). Se recomienda la clara definición de los pasos a seguir, para posibilitar la replicabilidad de estas acciones. El procedimiento variará en gran medida en función de la metodología seleccionada (Anguera, 2003; Wallraven, 2011); sin embargo, este protocolo presenta una posición de complementariedad metodológica más que de contraposición (Chacón, Anguera et al., 2008).

Fases de evaluación

Una vez definido el método, conformado por los aspectos comunes en todas las fases de evaluación (muestra, instrumentos y procedimiento), a continuación se pasa a describir pormenorizadamente los demás aspectos básicos específicos a considerar respecto a la conceptualización (objetivo pretendido) y valoración (criterios de toma de decisión para evaluar las evidencias registradas) en cada una de dichas fases.

Evaluación de necesidades

- **Conceptualización:** problema a solucionar. La evaluación de necesidades, primera fase de evaluación de un programa, busca analizar las características esenciales del problema de intervención a solucionar. Es un procedimiento sistemático de identificación de necesidades, priorización y toma de decisiones de cara a la planificación de un programa para la superación de dichas necesidades (Chacón, Lara y Pérez, 2002; Sanduvete et al., 2009). En general, una necesidad es considerada como una discrepancia entre el estado actual de un grupo de personas y su estado deseado (Altschuld y Kumar, 2010). La evaluación de necesidades consiste en el análisis de estas discrepancias y una posterior priorización de las acciones futuras, de tal manera que se solucionen antes las necesidades más preocupantes, básicas y urgentes.

- **Tipo de necesidad.** En función de la fuente de información, se distinguen distintos tipos de necesidad (Anguera, Chacón y Sánchez, 2008):

- a) **Normativa:** es aquella que el experto o profesional define como necesidad en una situación determinada, basándose en unos criterios estándar o una normativa. El especialista establece un nivel “deseable” y lo compara con el nivel real. Si un individuo o grupo no alcanza este nivel, se considera que presenta la necesidad. Una definición normativa de la necesidad no es absoluta; quizá usando otras dimensiones, no se tendría concordancia con la necesidad normativa; también depende del momento temporal (así, se encuentran discrepancias si cotejamos en las administraciones públicas la construcción de hospitales basándose en ratios de número de camas por millón de habitantes ahora y hace veinte años).
- b) **Experimentada, sentida o percibida:** en este caso, necesidad equivale a carencia subjetiva, dato obtenido a través de un estímulo provocador de respuesta (la pregunta de un cuestionario, por ejemplo). Al evaluar una determinada necesidad de un servicio, se pregunta a los individuos si creen que lo necesitan; es decir, si lo perciben como tal necesidad. La necesidad experimentada o sentida, por sí sola, es una medida inadecuada de la necesidad real, pues es fluctuante y se ve modulada por la situación actual del individuo, tanto por defecto como por exceso.
- c) **Expresada o demandada:** es la propia de las personas que demandan un servicio. Se diferencia de la necesidad percibida en que, en este caso, la persona que siente la necesidad la expresa por iniciativa propia, sin necesidad de que se le pregunte o se le brinde la oportunidad explícitamente.
- d) **Comparada:** se establece a partir de la diferencia entre los servicios que existen en un área y los servicios de otra, aunque estas diferencias no tienen por qué corresponder a una necesidad ya que, a lo mejor, en el área en cuestión no se necesita dicho servicio; por otro lado, no hay seguridad de que todas las necesidades en el área de referencia estén satisfechas. Se ha utilizado con frecuencia con el fin de disponer de listados de características de pacientes que requieren atención especial (Milsom, Jones, Kearney-Mitchell y Tickle, 2009).

- **Valoración.** En base a las evidencias tomadas de la muestra (proveniente de la población diana), con unos instrumentos y procedimiento delimitados, se valoran cuatro aspectos fundamentalmente: (a) La pertinencia del programa: si realmente es necesaria la implementación; si es cierto que existen necesidades que se han de cubrir (Anguera y Chacón, 2008; Fernández-Ballesteros, 1995; World Health Organization [WHO], 1981); (b) Los *criterios de* priorización de necesidades: supone atender en primer lugar a aquellas necesidades más urgentes en función de la conjunción de distintos criterios previamente explicitados como pueden ser la gravedad en sus consecuencias o el número de implicados; (c) La cobertura del programa: la concreción de los usuarios/as y potenciales usuarios/as (Anguera y Chacón, 2008); y (d) Los criterios operativos de inclusión/exclusión al programa: es necesario explicitar

las razones concretas que guiarán la decisión de por qué unos potenciales beneficiarios participarán en el programa y otros no lo harán.

Evaluación de objetivos

- **Conceptualización.** Los objetivos son aquellos aspectos que se quieren conseguir con la intervención en relación con la eliminación o disminución de la necesidad detectada; es decir, los resultados (*outcomes*) que se esperan obtener con el programa. Los objetivos generales o metas son constituidos por objetivos específicos, que han de ser definidos de manera concreta, identificable, medible/operativa y expresada temporalmente (Anguera, Chacón y Sánchez, 2008).
- **Valoración.** Dos aspectos básicos han de valorarse en esta fase, usualmente a partir de modelos teóricos existentes y programas similares: (a) la *usabilidad* del programa (Anguera y Chacón, 2008): medida en que se utiliza el programa preestablecido, modulado por un marco teórico, para cubrir las necesidades detectadas. Se pretende explicar por qué se toman unos objetivos concretos y no otros; se requiere para ello una justificación teórica, consistente en el estudio de la literatura (e.g., revisiones sistemáticas) con el que se detecten los objetivos de programas similares y se estudien los posibles problemas que puedan surgir, sus causas y posibles soluciones a éstos; se aconseja realizar simulaciones para estimar unos posibles resultados antes de la implementación del programa; y (b) en función del contexto de intervención, se han de tomar decisiones sobre la *suficiencia* del programa: si sus objetivos responden a todas las necesidades priorizadas sobre las que se ha decidido intervenir (Anguera y Chacón, 2008; Fernández-Ballesteros, 1995; WHO, 1981), además de si son potencialmente suficientes para superar o paliar dichas necesidades.

Evaluación del diseño

- **Conceptualización.** El diseño es el plan de trabajo que posteriormente se va a llevar a cabo. Ha de responder a los objetivos previstos y ha de ser tan detallado de forma que otro profesional que no sea quien lo ha planificado pueda ponerlo en práctica. La flexibilidad del diseño se considera una cualidad positiva pues, a la hora de implementarlo, podrían surgir algunos problemas a los que habría que responder reaccionando rápidamente y modificando lo planeado. La concreción del diseño se suele dar a partir de programas homólogos existentes obtenidos a través de búsquedas bibliográficas (e.g., revisiones sistemáticas); y previamente a la puesta en marcha del programa, se aconseja realizar simulaciones que posibiliten detectar posibles mejoras en dicho diseño. Se han de explicitar: (a) las actividades que se pretenden llevar a cabo para conseguir cada objetivo de acuerdo a las necesidades detectadas; (b) los recursos necesarios (humanos, económicos, materiales, etc.); (c) una temporalización concreta, que puede ser planteada en base a un PERT

(*Project Evaluation Review Technique*) o un CPM (*Critical Path Method*), consistentes en una calendarización en la que se indican las sincronías y diacronías que deben respetarse entre las diferentes actividades (Anguera, Chacón y Sánchez, 2008); (d) los momentos de registro de los datos para la evaluación; y (e) Cualquier otra particularidad de la intervención que pueda resultar de interés. Con todo ello, se obtienen distintos tipos de diseño, en los que se habrán de considerar unos componentes mínimos para potenciar la validez evaluativa.

- **Tipos de diseño.** En este trabajo se opta por una clasificación integradora de los diseños usados en evaluación de programas, basada en el nivel de intervención o control sobre el contexto de evaluación (Chacón y Shadish, 2001). Este criterio, planteado como un continuo, sitúa a los diseños experimentales en el extremo de máxima intervención y en el de mínima intervención a los diseños abiertos basados principalmente en metodología cualitativa. Sobre este continuo pueden posicionarse tres familias de diseños bien delimitadas estructuralmente: experimentales, cuasi-experimentales y observacionales*, ordenados de mayor a menor grado de manipulación y control. Los diseños experimentales (de intervención alta) son aquellos en los que una intervención es deliberadamente introducida para estudiar sus efectos, y en los que la muestra de usuarios/as es asignada aleatoriamente a las diferentes condiciones (Peterson y Kim, 2011; Shadish, Cook y Campbell, 2002). Los diseños cuasi-experimentales (de intervención media) básicamente difieren de los experimentales en que la muestra no es asignada aleatoriamente a las diferentes condiciones (Shadish et al., 2002). Finalmente, los diseños observacionales (de intervención baja) cuantifican la ocurrencia de la conducta sin introducir intervención alguna (Anguera, 2008; Anguera e Izquierdo, 2006). Desde un planteamiento exclusivamente metodológico y cuando el objeto de estudio es causal, se considera ventajoso el uso de diseños experimentales porque, en condiciones de aplicación óptimas, permitirían obtener una estimación insesgada del tamaño de efecto de la intervención (Chacón y López, 2008). En función de los momentos de medida (Chacón y López, 2008), y dentro de los diseños experimentales, se distingue entre diseños transversales (donde las variables son medidas en un momento puntual), longitudinales (donde se repiten las medidas de las variables de un grupo en más de un momento); y mixtos (donde se dan al menos una variable medida transversalmente y una de manera longitudinal). Dependiendo del número de variables independientes (VII) medidas (Chacón y López, 2008), los diseños se denominan unifactoriales (cuando presentan una única VI) o factoriales (cuando incluyen más de una VI); en los primeros se establece la diferenciación entre el uso de sólo dos valores (bicondicionales), donde sólo puede estudiarse la diferencia entre ellos; o más de dos valores (multicondicionales) donde es posible analizar más en profundidad la relación entre la VI y la/s variable/s dependiente/s (VD).

* En el término "diseños observacionales" se incluyen los planes generales de acción encaminados al registro sistemático y a la cuantificación del comportamiento tal como se produce en situaciones naturales o cuasi-naturales, siendo la metodología observacional la que da soporte a la recogida, optimización y análisis de los datos obtenidos a partir de estos diseños (Anguera, 1979, 1996, 2003).

De especial relevancia en estos diseños es el uso de técnicas de control de variables extrañas, con las que se trata de neutralizar las posibles influencias de factores ajenos al objeto de evaluación, para que no confundan la medida de su efecto (Chacón y López, 2008).

Sin embargo, no siempre es posible realizar una asignación aleatoria de los participantes a los diferentes grupos, sea porque éstos ya vengan conformados de manera natural, por falta de recursos o por cuestiones éticas, entre otras razones (Anguera, Chacón y Sanduete, 2008). Para estas situaciones se proponen los diseños cuasi-experimentales, cuyos principales tipos se presentan a continuación.

Los diseños pre-experimentales (Chacón, Shadish y Cook, 2008) no son considerados propiamente cuasi-experimentales, pero son el origen de éstos. Con el diseño de un solo grupo con prueba posterior, se miden los resultados en un grupo de usuarios/as tras la intervención; no existen grupo control ni medidas tomadas antes de la intervención. Tomando como punto de partida este diseño, si se añade una medida anterior a la intervención, se obtiene el diseño de un solo grupo con prueba previa y posterior; si por el contrario se añade la medida de un grupo equivalente después de la intervención, se obtiene el diseño con grupo no equivalente y con sólo prueba posterior. Con estos diseños difícilmente se podrían obtener inferencias causa-efecto, ya que los resultados registrados podrían deberse a una gran cantidad de potenciales amenazas a la validez.

Añadiendo al diseño de un solo grupo con prueba posterior tanto un grupo control como una medida tomada en cada grupo antes de la intervención, se obtiene un diseño propiamente cuasi-experimental (Chacón, Shadish et al., 2008; Shadish et al., 2002): el diseño con grupo control no equivalente y prueba previa y posterior. Aunque resuelve numerosas amenazas a la validez, puede estar afectado por otras como la maduración. Los diseños que a continuación se presentan solventan este problema.

El diseño de cohortes básico consiste en la comparación entre medidas tomadas en grupos especiales denominados cohortes, que pueden definirse como grupos de usuarios/as (de un programa o de alguna organización formal o no formal) que se van sucediendo por los distintos niveles en los que se pueden estructurar dichas organizaciones. Este tipo de turnos regulares de entrada y/o salida de usuarios/as que se suceden en el tiempo suelen conformar grupos semejantes, aunque lógicamente disponibles en distintos momentos temporales; y en ocasiones, determinadas cohortes pueden recibir una intervención concreta frente a otras cohortes que la hayan precedido o sucedido, que no hayan podido recibir dicha intervención. La lógica que guía este diseño es la misma que presenta el diseño con grupo control no equivalente y prueba previa y posterior, con la ventaja adicional que supone la cuasicomparabilidad entre los grupos cohortes.

El diseño de discontinuidad en la regresión consiste en explicitar un criterio de inclusión a uno u otro grupo denominado punto de corte. Este valor concreto de la variable dependiente sirve de referencia, de tal manera que todas las personas que sobrepasan este valor comienzan a participar en el programa. Es el diseño considerado de mayor calidad de entre los diseños cuasi-experimentales, ya que comparte una característica con los experimentales y es que la regla de asignación a los grupos también es conocida; así, se ha demostrado que este diseño, en condiciones de aplicación óptimas, permitiría estimaciones

insesgadas de los efectos de un programa equivalentes a las que se pueden obtener con el experimento aleatorio (Trochim, 2002). Para que se pueda llevar a cabo este diseño, es necesario que se cumplan ciertas condiciones, por lo que su uso queda restringido: (a) han de existir relaciones lineales entre variables; y (b) se ha de asumir que la línea de regresión se prolongaría más allá del punto de corte si no se produjese la intervención. Además, presenta algunos problemas: uno de ellos es la dificultad de generalizar de unas situaciones a otras por lo específico que es cada diseño; el otro es la dificultad de disponer de un punto de corte adecuado.

El diseño de series temporales interrumpidas (Shadish et al., 2002) consiste en disponer de sucesivas medidas a lo largo del tiempo, bien de los mismos usuarios/as observados en distintos momentos, o bien de usuarios/as diferentes pero considerados similares; y conocer el punto específico en el que se implementa el programa para poder estudiar la influencia o impacto que éste ha tenido en dicha serie (Anguera, 1995). Los efectos del programa pueden apreciarse en las diferencias entre las series obtenidas antes y después de su implementación con respecto a: (a) un cambio de nivel: discontinuidad en el punto de implantación del programa, aunque la inclinación de ambas series sea la misma; o (b) un cambio de pendiente. A su vez, los efectos observados pueden mantenerse a lo largo de la serie o decaer; por otra parte, en ocasiones los efectos no se dan a la vez que se implanta el programa sino que se pueden observar con un cierto retardo. El hecho de efectuar múltiples medidas en el tiempo con anterioridad y con posterioridad a implementar el programa supone una ventaja en comparación con diseños con una sola medida antes y/o después de la intervención.

Finalmente, los diseños observacionales o de baja intervención (Anguera, 2008) surgen ante la extrema complejidad que presenta la realidad social: los individuos o colectivos participantes en programas de intervención no constituyen una realidad compacta, la dinámica de los procesos seguidos no es uniforme, existen serias dudas sobre presuntas relaciones de causalidad, y en ocasiones resulta verdaderamente difícil una recogida de datos que cumpla todos los requisitos de rigurosidad requeridos (Anguera, 2008).

Tradicionalmente, sólo se contemplaban las intervenciones en las que se ejercía un dominio sobre la situación a evaluar, y se daban consignas a los usuarios/as del programa con el fin de que su ejecución permitiera la implementación del programa según el plan previsto por los expertos. Sin embargo, cada vez es mayor el número de programas que se implementan sin la imposición de consignas claramente especificadas y en los contextos naturales y/o habituales para los usuarios/as del programa, aprovechando actividades espontáneas y/o cotidianas para ellos. Los diseños evaluativos de baja intervención se ajustan a las restricciones y a las oportunidades que impone la evaluación en estas condiciones. Con estos diseños, la prioridad se desplaza del control sobre potenciales fuentes de confusión (propio de la experimentación y de la cuasi-experimentación), hacia la minimización de los sesgos debidos a la reactividad y la representatividad contextual. En este caso, la inferencia causal puede apoyarse en el análisis de secuencias de conducta.

Los diseños observacionales se clasifican básicamente según tres criterios (Anguera, 2008): (a) En función del número de unidades consideradas en el programa, pueden ser ideográficos (con un solo usuario/a o un grupo considerado unidad de estudio, o centrado en un solo nivel de respuesta) o nomotéticos (con más de un usuario/a o

Tabla 2 Tipos básicos de diseños observacionales.

		Momentos de registro	
		Puntual (1)	Seguimiento (>1)
Número de unidades	Ideográfico (1)	II.	I. Diseños diacrónicos
	Nomotético (>1)	III. Diseños sincrónicos	IV. Diseños mixtos/ <i>lag-log</i>

plural nivel de respuesta); (b) En función del momento de registro, pueden ser puntuales (cuando se analiza un momento concreto) o de seguimiento (con más de un momento); y (c) En función de las diferentes dimensiones o aspectos sobre los cuales incide la intervención débil que se aplica, pudiendo ser unidimensionales (una dimensión) o multidimensionales (varias dimensiones). La Tabla 2 presenta los tipos básicos de diseños que se generan en tres de las posibles combinaciones de los criterios (a) y (b), dado que el cuadrante II no da lugar a ningún diseño, por no disponer de suficiente potencial de datos. A cada uno de estos diseños se aplica posteriormente la complejidad añadida que supone si, de acuerdo con el criterio (c), son multidimensionales.

Los diseños diacrónicos (cuadrante I) se dividen, en función de qué tipo de datos se registran, en: (a) Extensivos (se registra la frecuencia de aparición del evento) y (b) Intensivos (además de la frecuencia, se registra el orden de aparición y, de manera optativa, la duración). (a) Dentro de los extensivos se encuentran, en función del número de momentos de registro, los diseños de panel (con 2 momentos), de tendencia (entre 3 y 49 momentos) y de series temporales (50 ó más momentos); y (b) Entre los intensivos se encuentran los secuenciales (prospectivos cuando se analiza qué conductas siguen al evento en estudio; y retrospectivos cuando se estudian las conductas precedentes) y los de coordenadas polares (con análisis a la vez prospectivo y retrospectivo).

Los diseños sincrónicos (cuadrante III) se dividen, en función del tipo de relación entre eventos que se estudie, en simétricos (cuando se estudia asociación sin direccionalidad en la relación) y asimétricos (que estudian causalidad).

Finalmente, los diseños mixtos o *lag-log* (cuadrante IV) pueden ser de 24 tipos al combinar 3 criterios: (a) el tipo de seguimiento (extensivo o intensivo); (b) la combinación del número de niveles de respuesta (1 ó más) y de usuarios/as (1 ó más) estudiados; y (c) el tipo de relación entre unidades (independiente, cuando no hay relación entre niveles de respuesta y/o entre usuarios/as; dependiente, cuando existe relación asimétrica; interdependiente, cuando la relación es bidireccional; e híbrido, cuando no se ajusta a ninguna de las modalidades anteriores).

• Componentes del diseño y validez de la evaluación. Para potenciar la validez de la evaluación y así posibilitar la generalización de los resultados obtenidos, se han de considerar al menos los siguientes componentes en el diseño: usuarios/as-unidades (U), intervención/tratamiento (T), resultados/"*outcomes*" (O), contextos/"*settings*" (S) y momentos/tiempo (Ti), que conforman el acrónimo UTOSTi. A partir de éstos, podrían definirse los distintos tipos de validez del siguiente modo (Chacón y Shadish, 2008; Shadish et al., 2002): (a) la validez de conclusión estadística se refiere

a la validez de las inferencias acerca de la correlación entre "t" (tratamiento) y "o" (resultado/"*outcome*") en la muestra; (b) la *validez interna* se refiere a la validez de las inferencias acerca de hasta qué punto la covariación observada entre "t" (el supuesto tratamiento) y "o" (el supuesto resultado) refleja una relación causal; (c) la validez de constructo se refiere a la validez de las inferencias sobre constructos de orden superior que representan muestras particulares (generalización de las muestras a la población de referencia); y (d) la validez externa se refiere a la validez de las inferencias acerca de si la relación causa-efecto se mantiene en variantes de personas, contextos, variables de tratamiento, momentos y variables medidas (generalización de las muestras a otras poblaciones diferentes a la de referencia, o entre distintas sub-muestras).

• Valoración. Se valoran los siguientes aspectos (Anguera y Chacón, 2008): (a) el grado de coherencia interna: nivel de ajuste de las necesidades detectadas con objetivos que las cubra, actividades para alcanzar dichos objetivos y recursos disponibles para ejecutar dichas actividades con una temporalización adecuada; (b) factibilidad: en qué medida la implementación del programa es posible en función de las acciones previstas (Fernández-Ballesteros, 1995); (c) aceptabilidad: en qué medida cada grupo de implicados desempeña el papel que tiene asignado, sin interferencia entre ellos; y (d) viabilidad: los requisitos mínimos que habrían de cumplirse para llevar a cabo la implementación del programa.

Evaluación de la implementación

• Conceptualización. La implementación consiste en la puesta en práctica del programa previamente diseñado (Anguera, Chacón y Sánchez, 2008).

• Valoración. Se valoran los siguientes aspectos: (a) la calidad o bondad: en qué medida la naturaleza y planteamiento de cada una de las acciones del programa se desarrollan de forma adecuada según los criterios técnicos y de expertos (Anguera y Chacón, 2008). Se analiza el grado de ajuste entre lo que se diseñó en la fase anterior y lo que realmente se ha implementado. Estas posibles discrepancias y sus causas son buscadas en todos los componentes del programa: actividades desempeñadas, temporalización, medios humanos y materiales utilizados, usuarios/as del programa y momentos de registro, entre otros; (b) la solución a posibles discrepancias: cuando existen desajustes, se pone a prueba la flexibilidad de la planificación realizada porque es deseable que pueda amoldarse a la nueva situación ante la necesidad forzosa de incluir alguna modificación a lo previsto; y antes de pasar a la evaluación de resultados, (c) El grado de evaluabilidad del programa, entendido como la medida en que éste

puede ser evaluado desde distintos criterios que interesen (Anguera y Chacón, 2008; Fernández-Ballesteros, 1995).

En este contexto es importante diferenciar entre monitorización/seguimiento y evaluación formativa donde, respectivamente, se refieren bien sólo al registro de datos durante el proceso o a la posibilidad de rediseño del programa en función de los datos registrados en el proceso.

Evaluación de resultados

- Conceptualización. Los resultados suponen el análisis de los efectos del programa (Anguera, Chacón y Sánchez, 2008). Se determina empíricamente en qué grado se cumplieron los objetivos marcados, si se alcanzaron metas no previstas y/o con qué relación coste-beneficio.
- Valoración. Básicamente, se valoran los siguientes componentes (Anguera y Chacón, 2008): (a) eficacia: si se cumplieron los objetivos propuestos (WHO, 1981); (b) efectividad: si, además de los objetivos previstos, se obtuvieron algunos otros efectos inesperados, ya sean positivos, neutros o negativos (Fernández-Ballesteros, 1995; WHO, 1981); (c) eficiencia: es la relación coste-beneficio de los resultados obtenidos (WHO, 1981), además de la de coste-efectividad, y de la de coste-utilidad (Neumann y Weinstein, 2010), que debería siempre complementarla, y que se preocupa por conocer la satisfacción del usuario. Para determinar si realmente se pudieron conseguir los mismos objetivos con menos coste o lograr mayor grado de resultados positivos al mismo precio, se hace necesario comparar entre programas similares; (d) continuidad: grado en que se mantuvo la formulación inicial de los objetivos propuestos y las acciones programadas; (e) progreso: grado de implementación del programa evidenciado por la consecución de objetivos (Fernández-Ballesteros, 1995; WHO, 1981); (f) adecuación: en qué medida la evaluación se basó en información técnicamente apropiada (Fernández-Ballesteros, 1995); (g) utilidad: nivel de influencia directa y automática de los resultados obtenidos en la posterior toma de decisiones (Fernández-Ballesteros, 1995); (h) probidad: grado en que la evaluación se ha realizado de forma ética y legal, y con la debida atención al bienestar de los usuarios/as y demás implicados en la evaluación (Fernández-Ballesteros, 1995); e (i) equidad: medida en que la planificación y la implementación de un programa se efectúa con iguales estándares para todos los individuos que integran el colectivo de usuarios/as-potenciales usuarios/as y que, consecuentemente, gozan de igualdad de oportunidades, pudiéndose también aplicar al análisis económico de programas.

Discusión y conclusiones

En primer lugar, es destacable el hecho de que el esquema de informe propuesto se ha desarrollado enfatizando los aspectos mínimos básicos y comunes, al objeto de poder ser aplicado a la amplia casuística de la realidad interventiva. Por motivos de extensión, cada uno de los elementos no se ha presentado de una manera detallada.

Por otra parte, por razones meramente expositivas, se ha diferenciado entre los distintos tipos de diseños de evaluación. Sin embargo, se defiende un planteamiento de integración entre los diseños de alta, media y baja intervención por diversas razones (Chacón, Anguera et al., 2008): (a) porque la evaluación de programas debe percibirse como un proceso unitario e integrado; (b) porque los usuarios/as demandan unas acciones pertenecientes a un programa que se ha de evaluar con la metodología más adecuada, por encima de la diferenciación de modalidades procedimentales; y (c) porque en muchos programas convendrá complementar diferentes metodologías, o incluso introducir mutaciones en el diseño que asuman el tránsito de una a otra, en respuesta a una realidad cambiante en los usuarios/as, y en ocasiones en su contexto, a lo largo del proceso de implementación.

Se considera por tanto que, entre los intereses metodológicos y sustantivos que convergen en la evaluación de programas, son éstos últimos los que deben primar en tanto en cuanto deben cumplirse las normas éticas relativas a procurar evitar cualquier tipo de malestar físico o psíquico a cualquier individuo (Anguera, Chacón y Sanduete, 2008), quedando los intereses metodológicos en un plano secundario.

Dado el carácter limitado de los recursos disponibles, la evaluación de programas requiere, de manera imprescindible, un análisis económico que ponga en relación los costes generados por las acciones del programa, la eficacia obtenida a lo largo de todo el proceso y a su finalización, la eficiencia parcial y total de las acciones del programa, la satisfacción de los usuarios/as, así como diversos parámetros que permiten modular la interrelación entre todos los elementos en juego (Anguera y Blanco, 2008).

Por último, se considera necesario recalcar que no tiene sentido llevar a cabo el proceso de evaluación sin fomentar la participación de los implicados, puesto que es la mejor manera de: (a) obtener información útil para cada uno de los grupos; (b) aumentar la probabilidad de utilizar los resultados de la evaluación; (c) facilitar la implementación de lo diseñado; y (d) potenciar la validez de la evaluación (Chacón y Shadish, 2008).

Financiación

El presente trabajo forma parte de los resultados obtenidos en el proyecto de investigación PS12011-29587, subvencionado por el Ministerio de Ciencia e Innovación.

Los autores agradecen a la *Generalitat de Catalunya* el apoyo a nuestro grupo de investigación [*Grup de Recerca e Innovació en Disseny (GRID). Tecnologia i aplicació multimedia i digital als dissenys observacionals*], subvención 2009 SGR 829.

También agradecen al gobierno español el apoyo al proyecto "Observación de la interacción en deporte y actividad física: Avances técnicos y metodológicos en registros automatizados cualitativos-cuantitativos" (Secretaría de Estado de Investigación, Desarrollo e Innovación del Ministerio de Economía y Competitividad) durante el periodo 2012-2015 [subvención DEP2012-32124].

Referencias

- Altschuld, J. W. y Kumar, D. D. (2010). *Needs assessment. An overview*. Thousand Oaks: Sage.
- Anguera, M. T. (1979). Observational typology. *Quality & Quantity. European-American Journal of Methodology*, 13, 449-484.
- Anguera, M. T. (1995). Diseños. In R. Fernández Ballesteros (Ed.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud* (pp. 149-172). Madrid: Síntesis.
- Anguera, M. T. (1996). Introduction. *European Journal of Psychological Assessment* (monograph on Observation in Assessment), 12, 87-88.
- Anguera, M. T. (2003). Observational methods (General). In R. Fernández-Ballesteros (Ed.), *Encyclopedia of behavioral assessment*, vol. 2 (pp. 632-637). London: Sage.
- Anguera, M. T. (2008). Diseños evaluativos de baja intervención. En M. T. Anguera, S. Chacón y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 153-184). Madrid: Síntesis.
- Anguera, M. T. y Blanco, A. (2008). Análisis económico en evaluación de programas. En M. T. Anguera, S. Chacón y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 259-290). Madrid: Síntesis.
- Anguera, M. T. y Chacón, S. (2008). Aproximación conceptual en evaluación de programas. En M. T. Anguera, S. Chacón y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 17-36). Madrid: Síntesis.
- Anguera, M. T., Chacón, S., Holgado, F. P. y Pérez, J. A. (2008). Instrumentos en evaluación de programas. En M. T. Anguera, S. Chacón y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 127-152). Madrid: Síntesis.
- Anguera, M. T., Chacón, S. y Sánchez, M. (2008). Bases metodológicas en evaluación de programas. En M. T. Anguera, S. Chacón y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 37-68). Madrid: Síntesis.
- Anguera, M. T., Chacón, S. y Sanduete, S. (2008). Cuestiones éticas en evaluación de programas. En M. T. Anguera, S. Chacón y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 291-318). Madrid: Síntesis.
- Anguera, M. T. e Izquierdo, C. (2006). Methodological approaches in human communication. From complexity of situation to data analysis. En G. Riva, M. T. Anguera, B. K. Wiederhold y F. Mantovani (Eds.), *From communication to presence. Cognition, emotions and culture towards the ultimate communicative experience* (pp. 203-222). Amsterdam: IOS Press.
- Bornas, X., Noguera, M., Tortella-Feliu, M., Llabrés, J., Montoya, P., Sitges, C. y Tur, I. (2010). Exposure induced changes in EEG phase synchrony and entropy: A snake phobia case report. *International Journal of Clinical and Health Psychology*, 10, 167-179.
- Casares, M. J., Díaz, E., García, P., Sáiz, P., Bobes, M. T., Fonseca, E., Carreño, E., Marina, P., Bascarán, M. T., Cacciola, J., Alterman, A. y Bobes, J. (2011). Sixth version of the Addiction Severity Index: Assessing sensitivity to therapeutic change and retention predictors. *International Journal of Clinical and Health Psychology*, 11, 495-508.
- Chacón, S., Anguera, M. T. y Sánchez-Meca, J. (2008). Generalización de resultados en evaluación de programas. En M. T. Anguera, S. Chacón y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 241-258). Madrid: Síntesis.
- Chacón, S., Lara, A. y Pérez, J. A. (2002). Needs assessment. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment*, vol. 2 (pp. 615-619). London: Sage.
- Chacón, S. y López, J. (2008). Diseños evaluativos de intervención alta. En M. T. Anguera, S. Chacón y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 219-240). Madrid: Síntesis.
- Chacón, S., López, J. M. y Sanduete, S. (2004). *Evaluación de acciones formativas en Diputación de Sevilla. Una guía práctica*. Sevilla: Diputación de Sevilla.
- Chacón, S. y Shadish, W. R. (2001). Observational studies and quasi-experimental designs: Similarities, differences, and generalizations. *Metodología de las Ciencias del Comportamiento*, 3, 283-290.
- Chacón, S. y 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* (pp. 69-102). Madrid: Síntesis.
- Chacón, S., Shadish, W. R. y Cook, T. D. (2008). Diseños evaluativos de intervención media. En M. T. Anguera, S. Chacón y A. Blanco (Eds.), *Evaluación de programas sociales y sanitarios: un abordaje metodológico* (pp. 185-218). Madrid: Síntesis.
- Cornelius, V. R., Perrio, M. J., Shakir, S. A. W. y Smith, L. A. (2009). Systematic reviews of adverse effects of drug interventions: A survey of their conduct and reporting quality. *Pharmacoepidemiology and Drug Safety*, 18, 1223-1231.
- Fernández-Ballesteros, R. (1995). El proceso de evaluación de programas. En R. Fernández-Ballesteros (Ed.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud* (pp. 75-113). Madrid: Síntesis.
- Gallego, M., Gerardus, P., Kooij, M. y Mees, H. (2011). The effects of a Dutch version of an Internet-based treatment program for fear of public speaking: A controlled study. *International Journal of Clinical and Health Psychology*, 11, 459-472.
- Griffin, C., Guerin, S., Sharry, J. y Drumm, M. (2010). A multicentre controlled study of an early intervention parenting programme for young children with behavioural and developmental difficulties. *International Journal of Clinical and Health Psychology*, 10, 279-294.
- Li, L. C., Moja, L., Romero, A., Sayre, E. C. y Grimshaw, J. M. (2009). Nonrandomized quality improvement intervention trials might overstate the strength of causal inference of their findings. *Journal of Clinical Epidemiology*, 62, 959-966.
- Milsom, K. M., Jones, M. C., Kearney-Mitchell, P. y Tickle, M. (2009). A comparative needs assessment of the dental health of adults attending dental access centres and general dental practices in Halton & St Helens and Warrington PCTs 2007. *British Dental Journal*, 206, 257-261.
- Moher, D., Liberati, A., Tetzlaff, J. y Altman, D. G. (2009). Preferred reporting items for systematic review and meta-analyses: The PRISMA statement. *BMJ*, 339, 332-336.
- Neumann, P. J. y Weinstein, M. C. (2010). Legislating against use of cost-effectiveness information. *The New England Journal of Medicine*, 363, 1495-1497.
- Peterson, C. y Kim, C. S. (2011). Psychological interventions and coronary heart disease. *International Journal of Clinical and Health Psychology*, 11, 563-575.
- Sanduete, S., Barbero, M. I., Chacón, S., Pérez, J. A., Holgado, F. P., Sánchez, M. y Lozano, J. A. (2009). Métodos de escalamiento aplicados a la priorización de necesidades de formación en organizaciones. *Psicothema*, 21, 509-514.
- Schulz, K. F., Altman, D. G. y Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *Annals of Internal Medicine*, 152, 726-732.
- Shadish, W. R., Cook, T. D. y Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.
- Trochim, W. M. K. (2002). *The research methods knowledge base*. New York: Atomic Dog Publ.
- Wallraven, C. (2011). *Experimental design. From user studies to psychophysics*. Boca Raton: CRC Press.
- World Health Organization, WHO (1981). *Evaluación de los programas de salud. Normas fundamentales. Serie Salud para todos*, 6. Geneva: WHO.