

La Evaluación Comprensiva de Programas Educativos: ¿Un Nuevo Paradigma Teórico?

Comprehensive Evaluation of Educational Programs: A New Theoretical Paradigm?

Rafael López-Meseguer ^{1*}

Manuel T. Valdés ²

¹ Universidad Autónoma de Madrid y Fundación Europea Sociedad y Educación, España

² Universidad Complutense de Madrid, España

En el mismo año en que se cumplen cien años del fallecimiento de Max Weber, este trabajo explora la posibilidad de trasladar los principios de la “sociología comprensiva” a la evaluación de programas educativos. Para ello, en primer lugar y tras una revisión de la literatura sobre evaluación de programas educativos, se establece la relación entre el enfoque comprensivo de evaluación y otros enfoques evaluadores. En la segunda parte, se desarrolla la aportación teórica de este trabajo en términos de construcción del objeto de investigación-evaluación. En particular, se hace referencia al entendimiento de los programas educativos como relaciones sociales y a la evaluación de programas a partir del análisis de su coherencia lógica y su correspondencia causal. En la tercera parte, se definen los límites en cuanto a la realización de juicios de valor en la práctica evaluativa. Sobre este particular, en consonancia con Weber, se señala que, si bien las evaluaciones no deben realizar juicios de valor, ello no significa que no se puedan evaluar los juicios de valor de los participantes de los programas. En ese sentido, los juicios de valor de los participantes constituyen un elemento central dentro enfoque evaluativo que se presenta. Por último, se propone la articulación metodológica como la estrategia de evaluación más adecuada para dar cumplimiento a las prerrogativas teóricas de la evaluación comprensiva.

Palabras clave: Evaluación; Evaluación comprensiva; Sociología comprensiva; Programas educativos; Articulación metodológica.

In the 100th anniversary of Max Weber's death, this paper explores the possibility of translating the principles of "Comprehensive Sociology" to the evaluation of educational programs. To this end, we firstly review the literature on the evaluation of educational programs and establish the relationship between the comprehensive evaluation approach and other approaches. In the second part, we develop the theoretical contribution of this work in terms of the construction of the evaluation object. In particular, special attention is dedicated to the understanding of educational programmes as social relations and the evaluation of programs based on the analysis of their logical coherence and its causal correspondence. In the third part, we define the limits to value judgments in the evaluation practice. In line with Weber, we argue that, while evaluators should not make value judgements, it does not mean that participants' value judgements cannot be evaluated. In fact, they are a central element in the evaluation approach that we present. Finally, the articulation of qualitative and quantitative techniques is proposed as the most appropriate methodological strategy given the theoretical prerogatives of comprehensive evaluation.

Keywords: Evaluation; Comprehensive evaluation; Comprehensive sociology; Educational programs; Methodological articulation.

*Contacto: rafaelmeseguer@gmail.com

1. Introducción

Llevar a cabo una evaluación implica decidir sobre el procedimiento a utilizar entre una amplia gama de posibilidades, y de ahí se deriva, naturalmente, la labor de definir adecuadamente qué constituye para cada cual el oficio de evaluar y la relación con aquello que se evalúa. La tendencia general de los evaluadores de programas, sin embargo, ha sido la de optar por un enfoque pragmático: en lugar de tratar de entender las relaciones entre las variables y su relación con la evaluación de programas en general (o del programa evaluado en particular), los teóricos de la evaluación se han centrado casi en exclusiva en facilitar la tarea de la evaluación a través de nuevos enfoques y diseños para mejorar la práctica evaluativa (Scriven, 1998).

Esta peculiaridad ha dado lugar a un inmenso cuerpo teórico sobre modelos evaluativos. Stufflebeam y Coryn (2014), en su célebre trabajo *Evaluation, Theory, Models & Applications*, son capaces de distinguir entre 23 modelos diferentes de evaluación respaldados con ejemplos particulares de evaluaciones llevadas a cabo. Lo común de estos casos es que todos ellos optan por unos u otros criterios epistemológicos y metodológicos particulares a la hora de llevar a cabo la práctica evaluativa. Conceptos como evaluación basada en criterios (Ligero, 2011), evaluación experimental y cuasi-experimental (Shadish et al., 2002), evaluación de estudio de caso (Stake, 1994, 1995), son habituales a la hora de distinguir maneras distintas de evaluar. Lo que se desprende a menudo de esta clase de distinciones es una escasa precisión terminológica y conceptual ya que, muchas veces, esas distinciones no diferencian correctamente lo que sería una técnica de evaluación, un modelo de evaluación y una teoría de la evaluación.

Una técnica de evaluación (la entrevista, la encuesta) sería la aplicación particular de una herramienta de ciencia social a la práctica de la evaluación de programas. Por su parte, entendemos que un modelo de evaluación es una concepción idealizada sobre cómo ha de llevarse a cabo la práctica evaluativa de la que se derivan unas prácticas más o menos estandarizadas. Una teoría de la evaluación, sin embargo, es algo más exigente; como señalan Stufflebeam y Coryn (2014, p. 50), es un conjunto coherente de principios conceptuales, analíticos y éticos que conforman un marco general para guiar el estudio y la práctica de la evaluación de programas. Se aprecia, por tanto, que la principal diferencia entre un modelo y una teoría radica en el grado de profundización sobre el objeto y la práctica de evaluar del que dispone el evaluador. Sin embargo, lo que hace a una teoría verdaderamente reconocible es que se sirve de unos conceptos particulares que informan sobre los fundamentos y el modo de conducir las evaluaciones en general, es decir, con los que comprender la totalidad del proceso evaluativo.

Por otro lado, hay que señalar que la teoría de la evaluación ha tenido un papel residual desde la emergencia de la evaluación como práctica profesional y académica reconocida. En un primer momento, las teorías iban dirigidas a contrarrestar algunas de las filosofías dominantes del gremio. Es así como se empezó a hablar de evaluación interpretativa (Stake, 1976) frente a la evaluación basada en criterios; de evaluación libre de objetivos (Scriven, 1973) frente a la evaluación estandarizada; de evaluación basada en la contingencia (Cronbach, 1982) o evaluación naturalista (Guba, 1978), frente a enfoques positivistas, etc. Sin embargo, los desarrollos teóricos de la evaluación de programas se fueron haciendo progresivamente más sofisticados y, en lugar de dedicarse a la crítica de la evaluación, comenzaron a desarrollarse teorías de la evaluación de carácter complejo

como puede ser el enfoque de la evaluación responsiva (Stake 2006), la evaluación CIPP (Stufflebeam, 2003), o la evaluación participativa (Cousins y Whitmore, 1998).

Llegados a este punto, es lícito preguntarse: ¿qué diferencia existiría, en la práctica, entre una teoría y un modelo? ¿Se derivan efectos específicos de la distinción entre una y otra? Si así fuera, ¿Qué habría de expresar una teoría para ser una teoría y no un modelo? En resumidas cuentas: ¿cómo podemos evaluar una teoría de la evaluación, por paradójico que suene?

A estas preguntas tratan de responder, en parte, William Shadish, Thomas Cook y Laura Levinton (1991) en su obra *Foundations of Program Evaluation: Theories of Practice*. En ella se señalan una serie de criterios con los que validar una teoría de la evaluación: (I) programación social, (II) construcción del conocimiento, (III) valoración, (IV) uso del conocimiento, y (V) práctica evaluadora. Siguiendo –en gran parte– lo establecido por los autores, consideramos que una teoría fundamentada de la evaluación debería responder, al menos, a: (I) la novedad de la teoría dentro de la tradición académica correspondiente, (II) cómo se construye el objeto de evaluación, (III) cuál ha de ser el tratamiento de los juicios de valor en las evaluaciones, y (IV) qué metodología(s) emplea en la práctica. Estos criterios deberían servir tanto para la evaluación de teorías de la evaluación como para la construcción de nuevas teorías. Así pues, en la medida que cada uno de estos interrogantes sean respondidos de una forma diferente a cómo se han respondido por parte de otras tradiciones teóricas, la pregunta que surge espontáneamente es la de si estaríamos frente a un nuevo paradigma teórico en el ámbito de la evaluación de programas en general, y de programas educativos en particular.

No obstante, con carácter previo al desarrollo de cómo el enfoque teórico de evaluación comprensiva que proponemos da respuesta a cada uno de los elementos anteriormente mencionados, consideramos necesario hacer una precisión acerca de la terminología empleada. En el contexto iberoamericano, es muy posible que el concepto de evaluación comprensiva pueda ser conocido en el ámbito académico. Ello se debe, sin embargo, a un criterio más de carácter lingüístico que evaluativo: y es que quién tradujo la obra de Robert Stake (2006), *Standards-based and responsive evaluation*, lo hizo como “evaluación comprensiva y evaluación basada en estándares”. Lo cierto es que el concepto “responsivo” o “respondiente” no existe en nuestro idioma, y en el momento de la traducción de la obra su utilización no era muy habitual por los científicos sociales ni por los evaluadores de programas. De ahí que, en los principales manuales de evaluación de programas en castellano, Stake haya quedado reflejado como el principal exponente de la evaluación comprensiva, cuando realmente debiera ser conocido como el exponente de la evaluación responsiva.

En ese sentido, conviene empezar señalando que nuestra utilización del concepto de comprensivo se basa en el intento de responder a los interrogantes teóricos mencionados desde las posibilidades que ofrece la “sociología comprensiva” de Max Weber (2014), en la que claramente se inspira, sumado a algunas aportaciones de otros importantes teóricos sociales (entre los que destacan Bourdieu, Chamboredon y Passeron) y de la evaluación (entre los que se incluye a Robert Stake, pero también muchos otros). Además, en el mismo año en que se conmemora el centenario del fallecimiento de Weber, este trabajo aspira a ser un ejemplo más de la actualidad del autor, con razón uno de los científicos sociales más afamados.

2. Teoría de la evaluación de programas

Existen numerosas formas de establecer distinciones teóricas en el ámbito de la evaluación de programas, aunque los criterios de esas distinciones varían de un autor a otro (Chen, 1990; Stake, 2006; Stufflebeam y Coryn, 2014). Como señala Stake (2006, p. 11), la primera labor que ha de realizar quien se adentra en el terreno de la teoría de la evaluación es exponer su propia clasificación de enfoques evaluadores. Este autor, por ejemplo, distingue entre evaluaciones “basadas en criterios y estándares” y evaluaciones “experienciales” o “interpretativas”. Otros autores, más interesados en disputas metodológicas, distinguen entre evaluación cuantitativa y cualitativa. Estos dos casos responderían a diferenciaciones epistemológicas y metodológicas.

Ramón Pérez Juste (2017), por su parte, distingue entre evaluaciones de procesos, evaluaciones de resultados y un último tipo de evaluaciones que califica como complejas. La evaluación comprensiva de programas educativos, como se verá más adelante, se circunscribe dentro de esta última categoría, por lo que el criterio clasificatorio que emplea este autor nos resulta más conveniente a efectos expositivos. En este caso, la diferenciación se establece a partir de un criterio teórico de base (el propio objeto de evaluación) y la distinción, por tanto, radica en el énfasis que se hace sobre uno u otro elemento del programa a la hora de valorarlo. Conviene señalar, no obstante, que, como todo criterio clasificatorio, se trataría de un artificio teórico con el que ordenar la realidad, sin que esta se corresponda del todo con dicho criterio.

Siguiendo con el argumento anterior, podríamos decir que los modelos de evaluación de resultados se dirigirían principalmente a corroborar si los resultados esperados ocurrieron y en qué medida son imputables al programa. Bajo esta rúbrica podemos ubicar, en primer lugar y por orden de importancia, los estudios causales experimentales y cuasi-experimentales y, segundo, las propuestas evaluativas que siguen estándares o criterios preordenados. Los modelos de evaluación de procesos, por su parte, tratarían de comprender los mecanismos en juego a la hora de vincular los procesos y resultados de una determinada intervención. Aquí se dan cita diferentes conceptos como los de “teoría del programa” (Weiss, 1972), “evaluación basada en la teoría” (Chen, 1990; Fitz-Gibbon y Morris, 1975) y, más recientemente, “teoría del cambio” (Rogers, 2014), que se diferencian entre sí por enfatizar o priorizar algunos elementos del proceso sobre otros.

2.1. *Enfoques teóricos de evaluación de resultados*

Dentro de los enfoques de resultados, los diseños experimentales son los que gozan de mayor popularidad. Su aplicación está dirigida a la atribución causal de los resultados observados al programa, esto es, tienen por objetivo analizar la relación de causalidad entre el cambio observado en una variable resultado y la aplicación de un programa de intervención (Stufflebeam y Coryn, 2014). Para ello, se construyen dos grupos de individuos, uno que pasará por el programa evaluado y que habitualmente se denomina “grupo de tratamiento”, y otro que no pasará por el programa, ya sea por pasar por un programa alternativo o por no pasar por ningún programa, que constituye el “grupo de control”. A fin de poder llevar a cabo esa atribución causal, la asignación de casos a los grupos de control y tratamiento debe ser aleatoria (Shadish, Cook y Campbell, 2002), garantizando así una distribución equilibrada en ambos grupos de posibles variables perturbadoras; es decir, permitiendo al evaluador descartar todas las explicaciones alternativas de los resultados observados excepto la propia aplicación del programa. Dicha aproximación es a menudo conocida como evaluación contrafactual, debido a que el grupo

de control funcionaría como un escenario donde observamos qué hubiese ocurrido con los casos tratados de no haber pasado por el programa (Imbens y Rubin, 2015).

No obstante, su implementación no es siempre posible, en tanto que requiere considerables recursos, un elevado compromiso de los implementadores del programa y un alto grado de colaboración por parte de los sujetos participantes en la evaluación (Stufflebeam y Coryn, 2014). Es por eso que es habitual recurrir a evaluaciones cuasi-experimentales, donde se relaja el supuesto de aleatorización en la asignación de los casos a los grupos de control y tratamiento. Sin embargo, al no haber distribuido los casos al azar, es posible que ciertos individuos de ciertas características se concentren en el grupo de tratamiento y, a la vez, que sean esos mismos individuos quienes manifiesten una mejoría en los resultados analizados. Dicha situación podría confundir al evaluador y hacerle pensar que es el programa quien causa los resultados cuando, en realidad, son esas ciertas características de los individuos del grupo de tratamiento las responsables del cambio observado (Imbens y Rubin, 2015). En tales circunstancias es necesario tomar medidas para poder atribuir causalmente los resultados al programa o descartar que éste tenga algún impacto, siendo habitual recurrir a técnicas de emparejamiento de casos tratados y de control, a diseños de regresión discontinua o a diseños de series temporales interrumpidas, entre otros procedimientos (Stufflebeam y Coryn, 2014).

Por último, es interesante destacar los modelos de evaluación por criterios, evaluación preordenada o evaluación basada en estándares, donde la valoración de un programa se hace en base a ciertos criterios y estándares preestablecidos (Ligero, 2011, p. 5), incrementando así la comparabilidad de los resultados alcanzados entre distintos equipos evaluadores, contextos geográficos, poblaciones sometidas al programa, etc. Este tipo de enfoques emergieron con fuerza durante los años sesenta en el ámbito de la cooperación al desarrollo y, con el paso del tiempo, fueron trasladándose hacia otros ámbitos. En el campo educativo, la creciente disponibilidad de bases de datos internacionales estandarizadas ha provocado una atención casi exclusiva hacia estos enfoques evaluativos, ya que permiten tener referencias nacionales, internacionales y regionales desde una perspectiva sistémica. Son destacables en ese sentido los esfuerzos realizados por la OCDE impulsando las pruebas PISA (*Program for International Student Assessment*) y PIACC (*Program for International Assessment of Adult Competences*), y por la IEA (*International Association for the Evaluation of Educational Achievement*) a través de las pruebas PIRLS (*Progress in International Reading Literacy Study*) y TIMMS (*Trends in International Mathematics and Science Study*).

Pese a que tales esfuerzos han supuesto una oportunidad inigualable para disponer de información que facilite la toma de decisiones en materia de política educativa, no puede dejar de reconocerse que no son evaluaciones que informen, por ejemplo, sobre los resultados que estén teniendo distintas formas de innovación docente o formas alternativas de relación entre la escuela y la comunidad. De ahí que, junto con las evaluaciones externas del sistema educativo (nacionales e internacionales) y los modelos de evaluación interna (a nivel de centro, del profesorado y del alumnado), deba establecerse un espacio para la evaluación de programas, de modo que sea posible evaluar cómo responde la lógica de un programa de intervención a su aplicación en distintos contextos educativos.

2.2. Enfoques teóricos de evaluación de procesos

Los modelos de evaluación de procesos surgen, en cierta medida, como reacción frente a la primacía de los modelos de evaluación de resultados, ya que tendrían serias dificultades para responder a interrogantes de suma importancia a la hora de valorar un programa como, por ejemplo: ¿a qué se debe la eficacia o ineficacia del programa? ¿Qué elementos del programa son los que hacen que funcione y cuáles no? ¿Qué lugar ocupa y qué importancia tiene la experiencia vivida por los participantes? De ahí que surjan una serie de propuestas evaluativas que ponen el foco en el programa en sí y en su proceso de implementación.

El concepto de teoría del programa, del que emanan todos los enfoques centrados en los procesos, se popularizó con el trabajo de Edward Suchmann (1967), *Evaluative Research*. En él se presentan dos tipos de razonamientos que podrían explicar el fracaso de un programa: (I) que el programa falle en el intento de poner en marcha las actividades (*implementation failure*), y (II) que el fallo se deba a que el programa no genera los resultados previstos, ya que lo que prescribe no tenga consistencia (*theory failure*). Las evaluaciones de resultados, por sí solas, no proporcionan información alguna sobre ninguno de los aspectos mencionados. Es por eso que, aunque de modo minoritario, en la década de los setenta comienzan a aparecer una serie de trabajos que, desde diferentes perspectivas, enfatizan la importancia de analizar lógicamente las teorías que emergen de los programas, y tratan de desplegar mecanismos de evaluación que ayuden a averiguar la relación entre las actividades y los resultados del programa (Fitz-Gibon y Morris, 1975; Chen y Rossi, 1980; Weiss, 1997).

Así, por ejemplo, Chen (1990) sostiene que la teoría detrás del programa debe tener sentido en el marco de la teoría social. Para ello, considera necesario distinguir entre teoría normativa y teoría causal del programa. La primera serviría al propósito de establecer los objetivos y resultados que deberían ser perseguidos y posteriormente evaluados, mientras que la segunda se referiría al conjunto de proposiciones acerca de cómo funciona el programa, y que sería lo que se persigue validar científicamente.

Carol Weiss (1997), por su parte, distingue dos aspectos de la evaluación de programas: la teoría de la implementación (*implementation theory*) y la teoría programática (*programmatic theory*). La primera se centra en cómo se ha llevado a cabo el programa, es decir, testar si el programa ha sido conducido según lo planeado para llegar a los resultados esperados. La teoría programática, por su parte, indaga sobre los mecanismos que intervienen entre el desarrollo del programa y la generación de unos resultados esperados. Lo importante, para esta autora, no es tanto las actividades per se, sino la respuesta generada a tales actividades. Y es precisamente esa atención a la respuesta de los participantes de los programas la que ha abierto la puerta a las técnicas cualitativas de evaluación hasta convertirse en una práctica habitual de las evaluaciones de procesos. Por otro lado, cabe señalar que Weiss ha acabado por emplear la terminología de “evaluaciones de teoría del cambio” (*theories of change evaluation*) para designar el proceder evaluativo que combina teoría de la implementación y programática. Este concepto, además, se ha convertido en el más popular entre las evaluaciones de procesos.

Patricia Rogers es una de las autoras que más ha desarrollado esta perspectiva. Para la autora, “la teoría del cambio explica cómo se entiende que las actividades produzcan una serie de resultados que contribuyen a lograr los impactos finales previstos (...). En ocasiones, el término denomina de manera genérica a cualquier versión del proceso; por

ejemplo, a una cadena de resultados con una serie de cuadros de insumos vinculados a productos, resultados e impactos, o a un marco lógico que expone la misma información en una matriz” (Rogers, 2014, p. 1).

En definitiva, en lo que coinciden los modelos de evaluación de procesos es en la necesidad de plantear un marco lógico con el que ser capaces de entender cómo un determinado programa produce cambios en la realidad, y cómo relacionar esos cambios con los posibles efectos que dicho programa pueda generar. Y para que esto sea posible resulta necesario detenerse en cómo esos cambios son asumidos por los participantes del programa. Además, es importante destacar la importancia de estos enfoques en hacer posible la generalización de los procesos de implementación y posterior escalabilidad de los programas. En la medida que se dispone de información sobre el funcionamiento del programa surge la posibilidad de, por un lado, mejorar aquellos aspectos que no hayan funcionado correctamente y, por el otro, sistematizar los procesos de implementación a partir de la información recabada, de manera que pueda ser llevado más fácilmente a la práctica en otros contextos.

2.3. Enfoques teóricos de evaluación complejos

La forma en la que interactúan procesos y resultados, sin embargo, no agota la cuestión de los enfoques de evaluación: puede ocurrir que programas con una lógica bien establecida y bien diseñados técnicamente fracasen, y que la respuesta a dicho fracaso esté fuera de la lógica de los procesos y de los resultados. De ahí que, a menudo, para llevar a cabo la práctica de evaluaciones se requieran enfoques más amplios que llamamos, siguiendo a Pérez Juste (2017), modelos de evaluación complejos.

Dentro de este tipo de evaluaciones podríamos incluir al mencionado Stake (2006) y su modelo de evaluación responsiva, el modelo CIPP de Stufflebeam (2003), o la propuesta evaluativa del propio Pérez Juste (2017). Sin entrar al detalle de cada una de estas propuestas evaluativas, se podría decir que Stake pone un gran énfasis en la apreciación valorativa, tratando de deslocalizarla y situarla lejos de la apreciación del evaluador, dando lugar a diferentes “jueces” que se encargarían de analizar la “congruencia” y “contingencia” del programa, siguiendo su propia terminología. En ese sentido, como se verá más adelante, la importancia otorgada a los juicios valorativos de los participantes por este autor comparte una de las premisas fundamentales del enfoque de evaluación comprensiva que desarrollamos en este trabajo.

Stufflebeam, por su parte, opta por un enfoque más pragmático: su modelo CIPP (contexto, input, proceso y producto) hace un intento ciertamente fructuoso de síntesis entre enfoques de resultados y de procesos, a lo que suma una atención particularizada y atenta al contexto en el que se desarrolla el programa.

Por último, Pérez Juste, con una propuesta de estructuración muy próxima a la de Stufflebeam, periodiza la evaluación en cuatro momentos: la evaluación del programa en cuanto tal, la evaluación del proceso de implantación del programa, la evaluación de los resultados del programa, y la institucionalización de la evaluación del programa. Este último elemento es particularmente interesante, ya que considera que cada evaluación debe servir para la propia mejora de la práctica evaluadora, y que esa reflexión debe ser explícita e incorporarse al modelo.

En definitiva, podríamos señalar que lo que caracteriza a estos modelos y les dota de un carácter complejo es que parten de paradigmas teóricos que les permiten integrar los

análisis empíricos sobre los programas dentro de enfoques interpretativos más amplios. Para que ello sea posible, resulta necesario explicitar la relación entre el evaluador y el objeto de evaluación; ofrecer una tesis sobre el tratamiento de los juicios de valor, es decir, definir los elementos a partir de los cuales se puede decir que un programa es bueno o malo, eficaz o ineficaz; y, por último, señalar la metodología que se ha de seguir para llevar lo anterior a término. A continuación, se describe la forma en que la evaluación comprensiva concibe cada uno de esos elementos.

3. Construcción del objeto de evaluación: los programas educativos como relaciones sociales

En su célebre trabajo, *El oficio del sociólogo*, Bourdieu, Chamboredon y Passeron (2002, p. 50) afirman, en referencia a Saussure, que es el punto de vista el que crea el objeto. Esa definición es completada por Weber, quien sostenía que no son las relaciones reales entre cosas las que delimitan los saberes científicos, sino las relaciones conceptuales entre problemas.

Para adentrarse en el ámbito de la evaluación, el evaluador siempre debe operar con unos conceptos que le permitan romper con las formas habituales de entender el objeto de evaluación ya que, de lo contrario, estas formas de pensar espontáneas se introducirían en la perspectiva del evaluador de manera inconsciente. En este sentido, siguiendo a Durkheim, la ruptura consistiría en la elaboración de nociones científicas, es decir, la definición previa del objeto como construcción teórica “provisoria” destinada a “sustituir las nociones del sentido común por una primera noción científica” (Maus, *texto 5*; en Bourdieu, Chamboredon y Passeron, 2002).

Con los conceptos de ruptura (epistemológica) y construcción del objeto lo que se quiere poner de manifiesto es que todo procedimiento evaluativo necesita contar con una suerte de esquemas reflexivos previos para poder llevarse a cabo. Esos esquemas constituyen lo que Bourdieu, Chamboredon y Passeron denominan el “oficio del sociólogo”. Análogamente, nosotros hablamos de “el oficio del evaluador”. Pero ese oficio, sin embargo, no ha de interpretarse como si de una receta de cocina se tratase: se añaden los ingredientes, se llevan a cabo las elaboraciones en un sentido pautado y el producto es la evaluación. De lo que se trataría es de hacer explícita la visión del evaluador a la hora de aproximarse al objeto de evaluación. De ahí que, a nuestro juicio, una teoría de la evaluación deba ser entendida como un conjunto de esquemas reflexivos que atraviesan la relación entre el evaluador y el objeto de evaluación¹ y que, en el caso de la evaluación comprensiva, estarían basados en (I) el entendimiento de los programas educativos como relaciones sociales, y (II) la evaluación de los programas a partir del análisis de su coherencia lógica y su correspondencia causal.

Sobre el primero de los aspectos, Max Weber (2014, p. 117) llama relación social al “comportamiento de varias personas en la medida en que el significado de la acción de cada una esté referido al de las otras y la acción se guíe por esa referencia”. Si tomamos como referencia esta definición, en el ámbito educativo podemos encontrar muchos tipos de

¹ William Shadish (1998) sintetizó esta idea de la mejor manera al señalar que “la teoría de la evaluación es lo que somos” (*evaluation theory is who we are*).

relaciones sociales susceptibles de ser evaluadas como programas: desde un proyecto que busque promover las competencias sociales y cívicas en el alumnado de secundaria a través del aprendizaje por proyectos, a un curso de formación que busque mejorar las habilidades de dirección de los equipos directivos; desde programas de implantación de títulos universitarios que busquen integrar unas competencias transversales comunes, hasta programas que busquen mejorar el clima entre el profesorado del centro. Lo fundamental, en ese sentido, no es tanto la casuística que cabe incluir bajo la etiqueta de programas, que puede ser mucha y variada, sino el “sentido” que guíe la acción (mejorar la competencia social y cívica de los adolescentes, las habilidades directivas, las competencias transversales de los universitarios, mejorar el clima entre el profesorado). De la consideración de los programas educativos como relaciones sociales se sigue, por tanto, un principio de reciprocidad con respecto a los fines del programa que debe ser identificable por los participantes.

En ese sentido, para que exista una relación social las partes que participan de ella deben estar sujetas a un cierto reconocimiento mutuo, es decir, que en mayor o menor medida reconocen que la acción que realizan está guiada por la expectativa de lo que aquella relación significa. Llevado al ámbito que nos ocupa, un programa prescribe un conjunto de actividades a realizar guiados por un fin que expresa aquello que se espera lograr. De ahí que, para que un programa tenga sentido, ese sentido debe ser reconocido por los participantes del mismo. No obstante, el grado y modo particular en el que ese reconocimiento se produce será, naturalmente, cambiante: cuando ese acuerdo es total, algo que ocurre solo en la teoría, nos encontraríamos frente a un tipo ideal² “puro” de programa (es decir, que el programa funcionaría a la perfección en relación a los fines que se propone); cuando ese acuerdo sobre los fines, actividades y logros del programa es suficiente, habría que valorar el grado de correspondencia causal de dicho programa, esto es, el grado de dicha suficiencia; cuando esa correspondencia no existe, significaría que el programa carece de consistencia y que, por tanto, no tiene sentido que pueda producir efectos.

Para situar esta idea en la práctica de programas educativos, consideremos el ejemplo anteriormente mencionado de un programa que busca mejorar las habilidades de dirección de una serie de equipos directivos de escuelas públicas. Imaginemos que ese programa tuviera entre sus fines el intento de mejorar las perspectivas de los equipos directivos acerca de la importancia de elevar las expectativas (personales, académicas, laborales) del alumnado susceptible de abandono educativo, y se les ofrece una formación específica sobre ello. Para que dicho programa fuera eficaz, en primer lugar, los participantes de dicho programa (los equipos directivos), deberían ser capaces de reconocer los fines de aquello que se les desea transmitir (la importancia de elevar las expectativas para evitar el abandono). En la medida que ello no ocurra, no se producirán efectos atribuibles al programa. En la medida que los equipos directivos reconozcan la validez de la idea, habrá que valorar en qué medida y cómo esa idea transforma sus significados acerca de la

² Recordemos que, Para Weber, los tipos ideales no son leyes generales de las cuales el fenómeno individual sea un ejemplo, sino un concepto abstracto, “relativamente vacío respecto a la realidad concreta” (Weber, 2006: 104). En ese sentido, el tipo ideal representa el grado de racionalidad perfecta que cabría esperar de un programa educativo.

importancia de las expectativas del alumnado susceptible de abandono, y en qué medida y cómo esta idea orienta sus acciones directivas después de recibir la formación.

Por otro lado, de la consideración de los programas educativos como relaciones sociales se derivan una serie de consecuencias epistemológicas. En uno de sus escritos fundamentales de teoría sociológica, *La objetividad del conocimiento en la ciencia social y en la política social*, Weber argumentaba que la sociología debía tratarse de una ciencia de la realidad (Weber, 2009, en Abellán, 2014):

Queremos comprender la realidad de la vida que nos rodea y en la que estamos inmersos en su peculiaridad, es decir, queremos comprender, por un lado, el contexto de sus fenómenos concretos en su forma actual y en su significación en la cultura, y, por otro lado, el motivo de que hayan sido así y no de otra manera. (p. 17)

Weber apuesta por una explicación causal de los fenómenos sociales, pero esa causalidad no ha de buscarse en la subsunción de fenómenos bajo el paraguas de leyes generales, al modo de las ciencias naturales. La comprensión (*verstehen*) de los fenómenos sociales y culturales requiere captar su individualidad (Abellán, 2014), y captar la individualidad de los fenómenos requeriría atender al significado subjetivo atribuido por los sujetos a las acciones. La tradición sociológica ha llamado a este modo de proceder individualismo metodológico. Aunque Max Weber no se ocupó de aplicar su teoría al campo educativo, su pensamiento influyó de forma significativa en algunos paradigmas teóricos cuya influencia en este campo ha sido notable, entre los que se encuentra el interaccionismo simbólico. Para los interaccionistas los seres humanos actuarían respecto a las cosas de acuerdo con los significados que éstas tienen para ellos³.

En ese sentido, la reconstrucción de los puntos de vista subjetivos de los participantes del programa sería uno de los objetos centrales de la evaluación comprensiva de programas educativos. Volviendo sobre el ejemplo anterior, diríamos que para considerar si el programa ha tenido efectos en cuanto a la elevación de las expectativas del alumnado susceptible de abandono por parte de los equipos directivos, un elemento fundamental de enjuiciamiento sería la propia consideración de los equipos directivos acerca del efecto que ha tenido el programa sobre las expectativas que depositan en su alumnado tras la formación específica. Ello encuentra justificación en que todo ser humano construye ideas, pensamientos y juicios sobre y a partir de sus propias experiencias, y eso le lleva a concebir el mundo de forma diferente al resto, lo que, a su vez, tiene consecuencias en su modo de actuar (Murillo, Hidalgo y Flores, 2016).

La segunda premisa es que la evaluación comprensiva basa su acción en el análisis de la coherencia lógica y la correspondencia causal de los programas. La conexión de los motivos de las relaciones sociales (o programas) con este tipo de evaluación la describe el propio Weber (2014):

Motivo es el conjunto de elementos que se le presenta al observador o al propio agente como fundamento que da el significado al comportamiento. Decimos que un comportamiento que se

³ Herbert Blumer, en su célebre trabajo, *Symbolic Interactionism. Perspective and Method*, resumió perfectamente el principal andamiaje teórico de una tradición por entonces emergente (1969, p. 2 en Flick, 2004, p. 32): “la primera premisa es que los seres humanos actúan con respecto a las cosas de acuerdo con los significados que éstas tienen para ellos. La segunda premisa es que el significado de estas cosas se deriva o surge de la interacción social que se tiene con los propios compañeros. La tercera premisa es que estas cosas se manejan en un acuerdo interpretativo utilizado por la persona al hacer frente a las cosas que encuentra, y este proceso las modifica”.

realiza coherentemente tiene coherencia lógica en la medida de que, de acuerdo con los usos promedio de nuestro pensamiento y de nuestros sentimientos, podamos afirmar que la relación existente entre los distintos elementos del comportamiento constituye un motivo típico (solemos decir que es un motivo lógico o correcto). Decimos, por el contrario, que una sucesión de fenómenos tiene correspondencia causal en la medida en que -de acuerdo con las reglas de la experiencia- existe la probabilidad de que esa sucesión se dé siempre en la realidad de la misma manera. (p. 90)

Los motivos que inspiran los comportamientos van de la mano de las acciones observables en la realidad social, y es precisamente esa conexión la que nos da una medida comprensiva del programa. De ahí que afirmamos que un programa es coherente lógicamente cuando (I) sea posible afirmar, en base al conocimiento científico existente y a las reglas de la experiencia común, que aquello que el programa prescribe tiene fundamento lógico y, (II) cuando los fines del programa hayan sido comúnmente reconocidos por los participantes en grado suficiente. Por otro lado, decimos que un programa tiene correspondencia causal cuando (I) se atestigua que los participantes del programa se comportan conforme a las prescripciones del mismo, es decir, que el programa produce cambios en los comportamientos o en la realidad educativa, y (II) cuando esos cambios se deben al programa objeto de evaluación.

En el ejemplo que estamos analizando, diríamos, por tanto, que el programa tendría fundamento lógico en la medida que aquello que prescribe se adecúa al conocimiento sobre la materia (elevar las expectativas de los alumnos es uno de los principios más mentados en la literatura sobre liderazgo educativo); que los fines sean reconocidos por los participantes en grado suficiente, es decir, que reconozcan la importancia de elevar las expectativas de los alumnos susceptibles de abandono, aunque cada uno lo haga de manera diferente en función de las concepciones previas sobre tales expectativas; y que, fruto de la formación (correspondencia causal), tomen medidas que se dirijan a elevar las expectativas de sus alumnos como, por ejemplo, invitar a antiguos alumnos que hayan tenido éxito en sus trayectorias académicas o profesionales, con el objetivo de que los alumnos puedan reconocerse en la experiencia de otros que, estando en situaciones similares, se abrieron en paso en la vida.

La perfecta adecuación entre los objetivos del programa, la comprensión de los mismos por los participantes, las acciones llevadas a cabo y los resultados alcanzados constituiría, por tanto, el tipo ideal perfecto de grado de funcionamiento del programa, siendo lo demás desviaciones del tipo ideal. La coherencia lógica, en ese sentido, estaría en sintonía con la lógica de los enfoques de la evaluación de procesos, en cuanto que buscaría averiguar la correcta adecuación de objetivos, prácticas y resultados desde la perspectiva del evaluador y de los participantes. La comprobación de la correspondencia causal, por su parte, estaría más próxima a los enfoques de evaluación de resultados, ya que buscaría analizar el grado en que la situación típica, es decir, lo que fija el programa, ocurre en la realidad, ya sea a través de la perspectiva de los participantes o de otro tipo de pruebas estandarizadas con las que medir la eficacia del programa. Ambos elementos -la comprobación de la coherencia lógica y el análisis de la correspondencia causal- deben concurrir a la hora de valorar un programa desde un enfoque comprensivo de evaluación. En palabras del propio Weber (2014):

Si no hubiera coherencia lógica estaríamos simplemente ante una probabilidad estadística no susceptible de comprensión (o solo comprensible de manera imperfecta), y toda vez que una afirmación solo puede ser causalmente correcta en la medida en que se aporte la prueba de que existe una determinada probabilidad (de alguna manera calculable) de que la acción suele

adoptar en la realidad –por término medio o en el caso ideal– un desarrollo lógicamente coherente con una frecuencia determinada o de modo aproximado. (p. 91)

4. Los juicios de valor en la práctica evaluativa

Evaluar, al fin y al cabo, no deja de ser una práctica dirigida a valorar algo. De lo que se trataría, por tanto, es de poder afirmar si un programa es bueno o malo; eficaz o ineficaz; si produce o no resultados. Por su parte, las ciencias sociales, desde su configuración como ámbito científico particular, han buscado establecer juicios valorativos que no dependan de valores particulares. Sin embargo, como señala Stake y todo evaluador reconoce, el juicio valorativo y la percepción se entremezclan a la hora de llevar a cabo cualquier evaluación (Stake, 2006, p. 15). Así pues, si queremos establecer un procedimiento de valoración que sea coherente con lo visto hasta ahora y que cumpla con los estándares de calidad exigidos a cualquier disciplina social, debemos, en primer lugar, realizar una primera aproximación al significado de los juicios de valor en la práctica de la evaluación de programas; y, en segundo lugar, proponer una estrategia metodológica que nos permita dar respuesta a las prerrogativas teóricas planteadas a partir de esa primera aproximación. A continuación, señalamos la manera en que la evaluación comprensiva responde a ambas cuestiones.

La defensa de Max Weber de una economía y una sociología libre de juicios de valor ha sido recibida como una especie de asunción dogmática por parte de la tradición sociológica, y ha dado lugar a multitud de interpretaciones erróneas acerca del pensamiento del autor. Es por eso que, de manera sintética, trataremos de exponer las principales ideas de Weber sobre la posibilidad de realización de juicios de valor en las ciencias sociales con el objetivo de situarla en el marco de la evaluación comprensiva de programas educativos.

Conviene señalar, con carácter previo, que el enunciado weberiano de abstenerse de hacer juicios de valor ha de concebirse, en cierta medida, como un modo de practicar la ciencia social diferente de la que practicaban los economistas e historiadores de su época y, también en cierta medida, como una asunción de los principios de reflexión de la escuela neokantiana, centrada en la libertad individual, con la que Weber mantenía una relación estrecha (Pérez Díaz, 1980, p. 66).

La pretensión de Weber es la de diferenciar, como ámbitos separados, la esfera de la explicación de los hechos y la esfera de los juicios de valor, para lo que insiste en la necesidad de distinguir entre “juicios de valor” y “relación con los valores” culturales (*Wertbeziehung*) y, consecuentemente, entre “juicio de valor” y “análisis de los juicios de valor” (Abellán, 2010). La intención última del autor, tal y como señala uno de sus más autorizados comentaristas, es, por un lado, mostrar su rechazo a la aplicación de los procedimientos científicos-naturales a las ciencias sociales y, por otro lado, renunciar a los conceptos “esencialistas” de la tradición romántica (Abellán, 2010, p. 22). Frente a la ciencia natural, que busca refrendar leyes objetivas de carácter general, las ciencias sociales han de poner el foco en la individualidad, porque de lo contrario se eliminaría todo aquello que constituye la peculiaridad de lo social.

Así pues, la relación con los valores se referiría a los compromisos que adquirimos con diferentes modos de vivir en comunidad, pero cuya validez depende o, mejor dicho, se circunscribe, a un contexto social e histórico determinado. Por tanto, se trataría de un concepto empírico de “valor”, por el que se supone que los sujetos se adscriben a valores culturales promovidos por diversas instituciones (familia, derecho, religión, estado,

costumbres, economía...). El estudio de la relación de los individuos con los valores sería la tarea de la ciencia social, mientras que, por el contrario, la valoración de dichos bienes culturales correspondería a la filosofía, ya sea esta concebida de un modo historicista o de manera contemporánea.

Weber, por tanto, establece así una distinción entre los juicios de valor y el análisis de los juicios de valor, y señala que entre uno y otro tipo de conocimiento hay una ruptura insalvable. Sin embargo, eso no significa que los juicios de valor no puedan ser analizados científicamente, y de ahí que estableciese un procedimiento que buscaba prescindir de valoraciones en el estudio de los juicios de valor subjetivos (Abellán, 2010: 39). La primera cautela a la que hacía referencia era la del peligro de asignar un propósito a estos análisis, es decir, que el estudio no nos debe llevar necesariamente a un acuerdo entre los juicios en discusión, sino todo lo contrario: puede ocurrir, y de hecho es normal que ocurra, que de la investigación (evaluación) se evidencie la imposibilidad de llegar a un consenso entre los juicios de valor en juego. Retomando el ejemplo anterior, se podría dar el caso en que uno de los directores de escuela que están participando en el programa considerase que elevar las expectativas de los alumnos es una forma de engaño, les sitúa fuera de la realidad, y ello podría tener efectos perniciosos. Es decir, que podría ser contrario a los fines establecidos por el programa, sin que pudiera haber un punto medio de encuentro y desplegar algún tipo de efecto.

La segunda cautela que establecía tenía que ver con la necesaria reflexión acerca de lo que pueden hacer y no hacer las ciencias sociales y, por ende, las evaluaciones con carácter científico, con respecto a la relación con los valores. En palabras del propio Weber (WL 508; en Abellán, 2010):

En el terreno de los juicios de valor políticos -y especialmente en los juicios de valor de carácter económico o en los de política social-, si se quisieran deducir directrices en relación con una acción valiosa, lo único que una disciplina empírica puede ofrecer con sus propios medios es lo siguiente: 1) los medios indispensables (para esa acción valiosa), 2) las consecuencias inevitables y 3) las consecuencias prácticas de la competencia así generada entre los múltiples juicios de valor posibles entre sí. (p. 40)

La ciencia empírica no puede ofrecer determinados tipos de explicaciones ni ha de emitir juicios de valor, ya que eso pertenece al saber de otras ciencias y, en último término, al ámbito privado del individuo. Sin embargo, que el científico social deba abstenerse de realizar juicios de valor no significa que no pueda analizar juicios de valor⁴. De hecho, para Weber, el análisis de los juicios de valor es un elemento fundamental, sino el más importante, de las ciencias sociales.

Trasladándonos de nuevo al enfoque de evaluación comprensiva, la consideración de los enunciados anteriores traería consigo una serie de consecuencias prácticas en su proceder metodológico: la primera operación evaluativa consistiría en formular los objetivos (primeros y últimos) del programa en términos de medios-fines racionales. En segundo lugar, se trataría de deducir las consecuencias derivadas de los fines que establece el programa, en coherencia con la evidencia existente. En tercer lugar, habría que comprobar el grado de adecuación entre los fines que establece el programa y los motivos que aducen los participantes. De estos tres elementos se encargaría precisamente la comprobación de

⁴ De manera demostrativa y, por tanto, no exhaustiva, Weber estableció un procedimiento metodológico con el que analizar juicios de valor sin realizar juicios de valor (Abellán, 2010, p. 41).

la coherencia lógica del programa. En cuarto lugar, se trataría de observar si (I) lo que prescribe el programa ocurre, ya sea a través de la perspectiva de los participantes o de otro tipo de pruebas, (II) si ocurre en grado de probabilidad aceptable conforme a los fines propuestos, (III) si se han de tener en cuenta otras consecuencias inesperadas de su aplicación, o (IV) concluir que no hay una adecuación entre lo que el programa prescribe y lo que realmente ocurre. Esta operación es lo que llamamos el análisis de la correspondencia causal del programa.

No obstante, a pesar de ese valor central de los juicios de valor de los participantes, conviene señalar que no es el único método posible para atestiguar la correspondencia causal de un programa: en el ámbito educativo, así como en otras disciplinas sociales, suele ser habitual servirse de pruebas estandarizadas de tipo competencial, diagnóstica, u otras realizadas *ad hoc*, con las que medir los progresos de los participantes, en función de las características del programa. En el ejemplo que venimos utilizando, junto con la consideración por parte de los equipos directivos acerca de la mejora en términos de las expectativas académicas y profesionales del alumnado que le ha proporcionado el programa, podríamos tratar de establecer una medida sobre las expectativas que tienen los equipos directivos sobre el alumnado con carácter previo a la formación, y volver a medir con posterioridad a la misma, de tal forma que obtendríamos una medida de la mejora a través de un mismo indicador considerado en dos momentos distintos. Para asentar con mayor precisión el argumento, consideremos otro ejemplo hipotético de los anteriormente mencionados, como podría ser un plan dirigido a promover la competencia social y cívica de adolescentes a través del aprendizaje por proyectos. Junto con un test de competencias cívicas y sociales realizado antes y después de la implementación del programa, una evaluación comprensiva consideraría fundamental los juicios de los participantes (en este caso profesores y alumnos) acerca de los efectos que produce el programa.

En resumen, alrededor de los juicios de valor de los participantes del programa pivotan una serie de estrategias metodológicas que contribuyen a tener una medida comprensiva del programa. No obstante, el enfoque de evaluación comprensiva otorga una posición central a los juicios de valor de los participantes, sin los cuales no podría hablarse de evaluación comprensiva. La dificultad, por tanto, estribaría en cómo combinar todas estas estrategias de valoración. En ese sentido, no cabe duda de que el procedimiento que enuncia Weber, si bien todavía nos resulta útil para la reflexión epistemológica, requiere de algunas actualizaciones desde el punto de vista metodológico, pues hay que tener en cuenta los numerosos avances técnicos que han experimentado las ciencias sociales en los últimos cien años. A continuación, proponemos una estrategia metodológica con la que dar respuesta a los diferentes planteamientos teóricos expuestos hasta ahora.

5. Articulación metodológica de técnicas de investigación-evaluación

En términos metodológicos, la evaluación ha ido sirviéndose progresivamente de los avances técnicos de la investigación social, aunque adaptados a sus propios fines, que no siempre han de ser coincidentes. De ahí que la terminología empleada en los debates metodológicos en uno y otro campo sea muy parecida e incluso intercambiable. En lo relativo a la evaluación comprensiva, la perspectiva adoptada es la de la articulación metodológica, donde las necesidades de información son satisfechas a través de estrategias

metodológicas que combinan técnicas cualitativas y cuantitativas. En el plano académico, tales estrategias son habitualmente referidas bajo múltiples denominaciones: métodos mixtos de investigación (Burke, Onwuegbuzie, y Turner, 2007), investigación multimétodo (Serrano et al., 2009), complementariedad metodológica (Blanco y Pirela, 2016), integración metodológica (Ruiz, 2008), etc. No obstante, la idea de articulación metodológica que sostenemos tendría un sentido más amplio: a la referida combinación de técnicas cualitativas y cuantitativas se le une que, en todo caso, las técnicas empleadas se han de orientar al análisis de la coherencia lógica y la correspondencia causal del programa objeto de evaluación.

Como es sabido, la relación entre la investigación cualitativa y cuantitativa ha sido históricamente controvertida. No faltan hoy detractores al intento de articulación de ambas aproximaciones, considerando que ambas constituyen paradigmas de investigación y evaluación inconmensurables por estar asentados en supuestos metodológicos, epistemológicos y ontológicos irreconciliables (Kuhn, 2005). Sin un acuerdo mínimo sobre cómo producir conocimiento, qué es conocimiento y qué es cognoscible, no hay posibilidad de articular una estrategia metodológica coherente. Y dado que los paradigmas cualitativo y cuantitativo en ocasiones no ofrecen las mismas respuestas para esas preguntas, las posibilidades de articulación se ven dificultadas. Cuando desde estas posiciones se habla de integración metodológica, a lo que se hace referencia es al recurso a técnicas distintas dentro de un mismo paradigma. Es lo que Ruiz (2008) denomina integración metodológica intraparadigmática.

Del otro lado, es habitual encontrar autores que evitan por completo ese debate y tienden a pensar en las técnicas cuantitativas y cualitativas como meros instrumentos de recogida de información. No es necesario, por tanto, dedicar mayor tiempo a reflexionar sobre las dificultades que pueda entrañar el recurso a distintas aproximaciones metodológicas o sobre la mejor manera de articular tales técnicas. El debate se centra, por tanto, en el plano puramente técnico, donde las posibilidades de integración son plenas e independientes de las circunstancias que caracterizan a una evaluación particular. Este sería un caso extremo de lo que Ruiz (2008) califica como integración metodológica interparadigmática.

Entre ambas posiciones existe un extenso espacio intermedio donde caben múltiples posibilidades de articulación. La propuesta que aquí se realiza, y que es compartida por otros muchos evaluadores, parte de los dos planteamientos siguientes: primero, que las aproximaciones cualitativa y cuantitativa presentan especificidades que requieren de una profunda reflexión acerca de las posibilidades reales de articulación en la evaluación particular que se tenga entre manos; y, segundo, que los objetivos y preguntas de evaluación específicos de cada caso no abocan irremediabilmente a un determinado paradigma, de manera que sea incoherente recurrir a metodologías propias del paradigma alternativo. Las distintas técnicas cualitativas y cuantitativas, por idiosincráticas que sean, podrán ser empleadas de forma conjunta si el evaluador reflexiona lo suficiente sobre cómo pueden ser mejor satisfechos sus objetivos de evaluación.

Al respecto de las posibilidades de integración o articulación metodológica, Blanco y Pirela (2016) distinguen tres estrategias: la combinación, la complementación y la triangulación. La combinación metodológica pretende una validación de resultados a través del recurso a técnicas de investigación cualitativas y cuantitativas que, implementadas de forma independiente, pueden llevar a resultados convergentes que refuercen las conclusiones alcanzadas. En ese sentido, el análisis de la correspondencia causal en el ejemplo sobre el liderazgo de los equipos directos podría valerse de caminos

separados que empleen, respectivamente, técnicas cualitativas (grupos de discusión con los equipos de dirección) y cuantitativas (mediciones previas y posteriores a la formación sobre la labor de dirección) con el objetivo de que converjan finalmente en unas conclusiones coherentes sobre las posibilidades de atribución causal de los resultados.

En segundo lugar, la complementación se basa en la elección de una perspectiva metodológica dominante, empleando la aproximación alternativa como una herramienta de profundización, matización o generalización. Es lo que se ha denominado como complementariedad asimétrica (Serrano et al., 2009), defendida por autores como Ibáñez (1986), o complementariedad por deficiencia (Ortí, 1995), enfatizando el conocimiento imperfecto de la realidad a que conducen por separado las aproximaciones cualitativa y cuantitativa (Martínez, 2005). En efecto, las técnicas de investigación cuantitativa podrían ser enormemente eficaces para estudiar la atribución causal de resultados a través de complejos procedimientos estadísticos. En el ejemplo sobre la formación en competencia cívica, podíamos llevar a cabo un diseño experimental con que tratar de atribuir la hipotética mejora al programa implementado. Pero dichos resultados solo serían inteligibles en el sentido de la evaluación comprensiva a través de las apreciaciones que permite el trabajo cualitativo, por ejemplo, con entrevistas a alumnos y profesores con los que indagar en el significado del concepto de competencia cívica y sobre si esta es o no enseñable.

En cuanto a la triangulación metodológica, el concepto busca acentuar la necesidad de complementar los resultados alcanzados por una vía a través de formas alternativas de investigación, donde sean empleados distintos evaluadores, perspectivas teóricas, fuentes de información y metodologías de análisis (Denzin, 1970; Denzin y Lincoln, 1994; Mertens y Hesse-Biber, 2012). Nótese que, si la combinación pretendía reforzar los resultados alcanzados llegando a las mismas conclusiones a través de aproximaciones metodológicas distintas, y la complementación buscaba emplear una de ambas aproximaciones como una herramienta auxiliar de la otra, la triangulación aspira a un esfuerzo de integración total y simétrico que ayude a superar reduccionismos y privilegios metodológicos. Por ejemplo, el análisis de la coherencia lógica podría valerse de un trabajo cualitativo exhaustivo que haga emerger los significados de los distintos participantes sobre la lógica del programa. Ese trabajo podría continuar después realizando un cuestionario que permita definir una tipología de acuerdo con variables como el sexo de los participantes, su edad y su rol en el programa.

Finalmente, consideramos muy relevante no ser excluyente en el análisis de la atribución causal de resultados y poner en valor lo que las técnicas cuantitativas y cualitativas pueden ofrecer en ese sentido. Sobre este particular, Howe (2012) distingue elocuentemente entre el estudio de las relaciones de causalidad mecánica y de causalidad agencial. Las primeras serían relaciones causales del tipo de las que ocurren en el mundo natural, donde A causa B de la misma manera en que la temperatura causa la evaporación. Este tipo de relaciones se encuentran siempre presentes en los programas de intervención y son aquellas típicamente estudiadas a través de evaluaciones contrafactuales. Pero esas no son todas las relaciones causales que conforman la lógica de un programa. También hay relaciones de causalidad agencial, basadas en el comportamiento intencional de actores, no en relaciones mecánicas. Desentrañar esas relaciones comporta un tipo de proceder evaluador diferente, donde las técnicas cualitativas pueden jugar un papel más relevante que las cuantitativas.

La conclusión última de todo ello es que, pese a que determinadas formas de evaluación han tendido a reducir el estudio de la causalidad a ciertas relaciones mecánicas, no existen barreras efectivas para que la articulación metodológica aborde, en el marco de una misma evaluación, relaciones causales mecánicas y agenciales a través de técnicas cualitativas y cuantitativas. De la misma forma se posiciona la propia Rogers (2014) al describir la teoría del cambio, para quien ésta “se sirve de un conjunto de datos cualitativos y cuantitativos para respaldar la triangulación de los datos obtenidos a raíz de una evaluación de métodos mixtos” (p. 7). Y es que desentrañar y comprobar la lógica de un programa y la correspondencia causal entre sus distintos elementos, en ocasiones implícitas para sus propios desarrolladores, no puede reducirse a la aplicación de una técnica de investigación particular, por poderosa que ésta pueda ser desde el punto de vista estadístico. Es necesario el desarrollo de una estrategia coherente que permita entender qué se esperaba del programa, cómo se esperaba que éste funcionase, qué ha ocurrido en la práctica y cómo se relacionan los resultados con su implementación real. La articulación metodológica sirve a dicho propósito.

6. Conclusiones

En este trabajo hemos abordado las posibilidades que nos ofrece la sociología comprensiva de Max Weber en el ámbito de la evaluación de programas. Para ello, hemos empezado distinguiendo entre una teoría y un modelo de evaluación, hasta el punto de proporcionar nuestra propia definición de lo que es una teoría en relación a la evaluación de programas, considerando como tal un conjunto de esquemas y principios reflexivos que atraviesan la relación que se establece entre el evaluador y el objeto de evaluación.

Sobre los diferentes enfoques teóricos en el ámbito de programas, hemos distinguido entre enfoques de evaluación de resultados, de procesos y enfoques de evaluación complejos, donde hemos encuadrado el modo de evaluación comprensiva de programas educativos. Un enfoque de evaluación complejo sería aquel que fuese capaz de (I) proporcionar una teoría particular acerca de la relación entre el evaluador y el objeto de evaluación, (II) ofrecer una tesis sobre el tratamiento de los juicios de valor en los procesos de evaluación, es decir, definir los elementos a partir de los cuales se puede afirmar que un programa es bueno o malo, eficaz o ineficaz; y, por último, (III) señalar la metodología que se ha de seguir para llevar lo anterior a término. A lo largo del texto hemos ido describiendo el modo en que la evaluación comprensiva da respuesta a cada uno de estos interrogantes.

A propósito de la relación entre evaluador y objeto de evaluación, hemos señalado que la evaluación comprensiva se orientaría a partir de una serie de premisas o prerrogativas: (I) el entendimiento de los programas educativos como tipos ideales de relaciones sociales, y (II) la evaluación de los programas a partir del análisis de su coherencia lógica y su correspondencia causal. La primera de las premisas nos ha llevado a reconocer la importancia de la reconstrucción de los puntos de vista subjetivos de los participantes del programa como uno de los objetos centrales de la evaluación comprensiva, de tal forma que pudiera valorarse el grado de reciprocidad en cuanto a los fines que dicho programa pudiera prescribir. De acuerdo con la segunda premisa, hemos señalado que un programa es coherente lógicamente cuando (I) sea posible afirmar, en base al conocimiento científico existente y a las reglas de la experiencia común, que aquello que el programa prescribe tiene fundamento lógico y, (II) los fines del programa hayan sido comúnmente reconocidos por los participantes en grado suficiente. Por otro lado, decimos que un programa tiene

correspondencia causal cuando (I) se atestigua que los participantes del programa se comportan conforme a las prescripciones del mismo, es decir, que el programa produce cambios en los comportamientos o en la realidad educativa, y (II) cuando esos cambios se deben al programa objeto de evaluación.

En referencia al alcance del término valoración en el ámbito de la evaluación de programas, hemos conectado esta problemática con el clásico debate sobre los juicios de valor en las ciencias sociales para terminar optando por una solución de tipo weberiano: que la ciencia social deba estar libre de juicios de valor no quiere decir que no pueda analizar los juicios de valor. A partir de esta máxima, hemos afirmado que los juicios de valor de los participantes del programa son un aspecto central de la evaluación comprensiva, que ha de ser combinado, en la medida de lo posible, con otro tipo de pruebas estandarizadas, como suele ser habitual en el ámbito educativo.

Para llevar a la práctica lo descrito anteriormente, hemos optado por un procedimiento de articulación metodológica de investigación-evaluación, ya que la cuestión del método no puede reducirse a la aplicación de una técnica de investigación particular, por poderosa que ésta pueda ser desde el punto de vista estadístico. Por el contrario, toda evaluación comprensiva de programas educativos debe combinar las técnicas cualitativas y cuantitativas más eficaces orientadas al análisis de la coherencia lógica y la correspondencia causal del programa objeto de evaluación.

Por último, quedaría abierta la discusión con la que titulábamos este trabajo: ¿constituye la evaluación comprensiva un nuevo paradigma teórico en la evaluación de programas en general y en la evaluación de programas educativos en particular? Nuestra pretensión ha sido la de construir una serie de esquemas reflexivos, epistemológicos y metodológicos, que ayuden a avanzar en el terreno de la teoría de la evaluación y, por tanto, la decisión última acerca de si esta propuesta teórica constituye un nuevo paradigma le corresponde tomarla a quienes se reúnen dentro del oficio de la evaluación de programas.

Agradecimientos

Queremos agradecer a la Fundación Europea Sociedad y Educación su compromiso con la evaluación de programas educativos. En particular, las reflexiones conjuntas en el marco de la unidad de Observación y Evaluación (OyE), han inspirado muchas de las ideas que se presentan en este trabajo.

Referencias

- Abellán, J. (2010). Estudio preliminar. En M. Weber (Ed.), *Por qué no se deben hacer juicios de valor en la sociología y en la economía* (pp. 11-55). Alianza Editorial.
- Abellán, J. (2014). Estudio preliminar. En M. Weber (Ed.), *Conceptos sociológicos fundamentales* (pp. 11-71). Alianza Editorial.
- Blanco, N. y Pirela, J. (2016). La complementariedad metodológica: Estrategia de integración de enfoques en la investigación social. *Espacios Públicos*, 19(45), 97-111.
- Blumer, H. (1969). *Symbolic interactionism. Perspective and method*. University of California Press.
- Bourdieu, P. (2002). *El oficio de sociólogo: Presupuestos epistemológicos*. Siglo XXI Editores.

- Burke, R., Onwuegbuzie, A. J. y Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.
<https://doi.org/10.1177/1558689806298224>
- Chen, H.-T. (1990). *Theory-Driven Evaluation: A Comprehensive Perspective*. Sage.
- Chen, H.-T. y Rossi, P. (1980). The Multi-Goal, Theory-Driven Approach to Evaluation: A Model Linking Basic and Applied Social Science. *Social Forces*, 59(1), 106-122.
<https://doi.org/10.2307/2577835>
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. Jossey-Bass.
- Cousins, J. B. y Whitmore, E. (1998). Framing participatory evaluation. En E. Whitmore (Ed.), *Understanding and practicing participatory evaluation* (pp. 5-23). Jossey-Bass.
<https://doi.org/10.1002/ev.1114>
- Denzin, N. K. (1970). *Sociological methods: A sourcebook*. Watterworth.
- Denzin, N. K. y Lincoln, Y. S. (1994). *Handbook of qualitative research*. Sage.
- Fereday, J. y Muir-Cochrane, E. (2008). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5(1), 80-92. <https://doi.org/10.1177/160940690600500107>
- Fitz-Gibbon, C. T. y Morris, L. L. (1975). Theory-based evaluation. *Evaluation Comment*, 5(1), 1-4.
- Flick, U. (2004) *Introducción a la investigación cualitativa*. Morata.
- Guba, E. G. (1978). *Toward a methodology of naturalistic inquiry in educational evaluation*. University of California Press.
- Howe, K. R. (2012). Mixed methods, triangulation, and causal explanation. *Journal of Mixed Methods Research*, 6(2), 89-96. <https://doi.org/10.1177/1558689812437187>
- Ibáñez, J. (1986). Cuantitativo/cualitativo. En R. Reyes (Ed.), *Terminología científico-social. Aproximación crítica* (pp. 218-233). Anthropos.
- Imbens, G. W. y Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139025751>
- Kuhn, T. S. (2005). *La estructura de las revoluciones científicas*. Fondo de Cultura Económica de España.
- Ligero, J. A. (2011). *Dos métodos de evaluación: criterios y teoría del programa. Documento de Trabajo, Serie CECOD N°15*. CEU Ediciones.
- Löffler, E. (1996). *La modernización del Sector Público desde una perspectiva comparativa. Conceptos y métodos para evaluar y premiar. La calidad en el sector público en los países de la OCDE*. Documentos INAP.
- Martínez, M. (2005). *El paradigma emergente: Hacia una nueva teoría de la racionalidad científica*. Trillas.
- Mertens, D. M. y Hesse-Biber, S. (2012). Triangulation and mixed methods research: Provocative positions. *Journal of Mixed Methods Research*, 6(2), 75-79.
<https://doi.org/10.1177/1558689812437100>
- Murillo, F. J., Hidalgo, N. y Flores, S. (2016). Incidencia del contexto socio-económico en las concepciones docentes sobre evaluación. *Profesorado: Revista de Currículum y Formación del Profesorado*, 20(3), 251-281.
- Ortí, A. (1995). La confrontación de modelos y niveles epistemológicos en la génesis e historia de la investigación social. En J. M. Delgado y J. Gutiérrez (Eds.), *Métodos y técnicas cualitativas de investigación en ciencias sociales* (pp. 87-99). Síntesis.

- Pratt, D. D. (1992). Conceptions of teaching. *Adult Education Quarterly*, 42(4), 203-220. <https://doi.org/10.1177/074171369204200401>
- Pérez Juste, R. (2017). *Evaluación de programas educativos*. La Muralla.
- Pérez Díaz, V. (1980). *Introducción a la sociología: Concepto y método de la ciencia social en su historia*. Alianza Editorial.
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472-482. <https://doi.org/10.1016/j.tate.2010.09.017>
- Rogers, P. (2014). *La teoría del cambio*. Centro de Investigaciones Innocenti de Unicef.
- Ruiz, C. (2008). *El enfoque multimétodo en la investigación social y educativa: Una mirada desde el paradigma de la complejidad*. Upel.
- Scriven, M. (1998). Minimalist theory: The least theory that practice requires. *American Journal of Evaluation*, 19, 57-70. <https://doi.org/10.1177/109821409801900105>
- Serrano, A., Blanco, F., Ligeró, J. A., Alvira, F., Escobar, M. y Saén, A. (2009). *La investigación multimétodo*. Recuperado de https://eprints.ucm.es/30034/1/araceli%20serrano%20articulacion_metodologica._serrano_blanco_alvira.pdf
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19, 1-19.
- Shadish, W. R., Cook, T. D. y Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Sage. <https://doi.org/10.1177/109821409801900102>
- Shadish, W. R., Cook, T. D. y Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Shavelson, R. J. y Towne, L. (2002). *Scientific research in education*. National Research Council.
- Suchman, E. (1967). *Evaluative research*. Russell Sage Foundation.
- Stake, R. E. (1976). A theoretical statement of responsive evaluation. *Studies in Educational Evaluation*, 2, 19-22. [https://doi.org/10.1016/0191-491X\(76\)90004-3](https://doi.org/10.1016/0191-491X(76)90004-3)
- Stake, R. E. (1994). Case studies. En N. K. Denzin y Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 236-247). Sage.
- Stake, R. E. (1995). *The art of case study research*. Sage.
- Stake, R. E. (2006). *Evaluación comprensiva y evaluación basada en estándares*. Graó.
- Stufflebeam, D. L. (2003). The CIPP model for evaluation. En T. Kellaghan y D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 31-62). Kluwer. https://doi.org/10.1007/978-94-010-0309-4_4
- Stufflebeam, D. L. y Coryn, C. L. S. (2014). *Evaluation theory, models, and applications*. Jossey Bass.
- Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. En D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). Macmillan.
- Weber, M. (2009). *La "objetividad" del conocimiento en la ciencia social y en la política social*. Alianza Editorial.
- Weber, M. (2010). *Por qué no se deben hacer juicios de valor en la sociología y en la economía*. Alianza Editorial
- Weber, M. (2014). *Conceptos sociológicos fundamentales*. Alianza Editorial.

Weiss, C. H. (1972). *Evaluation research: Methods for assessing program effectiveness*. Prentice Hall.

Weiss, C.H. (1997). *Theory-Based Evaluation: Past, Present and Future*. Jossey-Bass Publishers.
<https://doi.org/10.1002/ev.1086>

Breve CV de los autores

Rafael López-Meseguer

Investigador doctorando en la Facultad de Ciencias Políticas de la Universidad Autónoma de Madrid, donde obtuvo su graduado en Ciencias Políticas y de la Administración, y cursó el Master en Democracia y Gobierno. Investigador asociado al Instituto de Estudios Sociales y Educativos de la Fundación Europea Sociedad y Educación, donde colabora en diversos proyectos de investigación. Sus temas de especialización son la Teoría Política y Social, la Teoría de la Evaluación, y la Investigación cualitativa en Educación. ORCID ID: <https://orcid.org/0000-0001-8488-2888>. Email: rafaelmeseguer@gmail.com

Manuel T. Valdés

Personal investigador en formación en el departamento de Sociología Aplicada de la Universidad Complutense de Madrid (ref: FPU16/02905). Egresado en Ciencias Políticas y Sociología por la Universidad Carlos III de Madrid y del Máster en Metodología de la Investigación en Ciencias Sociales de la Universidad Complutense de Madrid. Ha colaborado en investigaciones en el Centro Nacional de Medicina Tropical del Instituto de Salud Carlos III y en el Instituto Nacional de Estadística. Sus intereses de investigación se centran en la evaluación de políticas educativas, las desigualdades ante la educación y el papel de la educación en los procesos de movilidad social. Sus trabajos han sido publicados en revistas como la Revista Española de Investigaciones Sociológicas o la Revista Española de Sociología. ORCID ID: <https://orcid.org/0000-0001-8012-3956>. Email: manueltv@ucm.es.