



7. Notas de evaluación de programas

En este documento repasamos algunos de los conceptos y herramientas básicos en la evaluación de programas. El esquema que seguimos no tiene carácter exhaustivo. Sólo pretende revisar y presentar algunos de los elementos más útiles en la valoración de intervenciones. A continuación, tratamos los siguientes temas:

- Tipos de evaluación de programas.
- Fases en la evaluación de programas.
- Técnicas más utilizadas en la evaluación de programas.

TIPOS DE EVALUACIÓN

Los primeros trabajos de valoración de programas se realizaron en el ámbito educativo, y consistieron en estudios pretest-postest para contrastar el grado de consecución de los objetivos pedagógicos operativos. En ese sentido, las aportaciones pioneras de Tyler en la década de los 40 giran en torno al criterio de eficacia de la formación, desarrollando uno de los aspectos centrales en la evaluación de resultados. Desde entonces se ha producido una notable evolución metodológica, así como una gran diversificación de los intereses de evaluación. Por un lado, los diseños experimentales y cuasi-experimentales han perfeccionado las estrategias de control de las amenazas a la validez interna, más allá de las comparaciones simples antes-después. Y, por otro lado, a la preocupación inicial por medir el logro de objetivos se han sumado otros elementos susceptibles de estudio, como el proceso de aplicación, la cobertura, la implantación, los costes y otras dimensiones que examinaremos más adelante.

En el curso de ese desarrollo, la evaluación de programas ha subrayado su carácter práctico, como herramienta al servicio de la toma de decisiones sobre la intervención. Con ello, a la vez que se introducían nuevos elementos de validez científica se ha ido dando más importancia a la utilidad social de la evaluación, puesto que no sólo sirve para conocer la efectividad de las acciones sino también para introducir mejoras en los programas, decidir sobre la continuidad de una línea de intervención, u organizar los recursos disponibles. Como ya hemos señalado, la evaluación de programas se mueve en una tensión constante entre su carácter de instrumento de intervención y las aspiraciones de fiabilidad y validez de toda investigación científica. Por eso, junto a las consideraciones de rigor metodológico, hay

que tener en cuenta la disponibilidad de tiempo y presupuesto, la credibilidad de los métodos utilizados o la participación de los implicados.

El resultado de esa evolución ha sido la emergencia de un enfoque más complejo de análisis, (1) que no suele ceñirse a un solo criterio de valoración, (2) que combina la utilización de técnicas cuantitativas y cualitativas de recogida de información, y (3) que tiene en cuenta múltiples dimensiones de un programa. Esa confluencia de perspectivas se ha utilizado como referencia para enunciar los diferentes “*tipos de evaluación*”:

1. Por ejemplo, en función de los criterios de valor, se distingue la evaluación de la eficacia, de la suficiencia, del progreso, etcétera.
2. Basándonos en la metodología, podemos diferenciar un enfoque cualitativo de uno cuantitativo.
3. Según la dimensión del programa en la que nos centremos, tendremos una evaluación de la implantación, de la cobertura, de los resultados, etcétera.
4. Dependiendo de quién lleve a cabo las actividades de valoración, se califica la evaluación de interna o externa.
5. Y de acuerdo con el momento y fines generales de evaluación, discriminamos una orientación formativa de una sumativa.

Ninguna de las categorías que se derivan de estos cinco esquemas de clasificación son incompatibles entre sí. Una evaluación de programas puede ser al mismo tiempo formativa y sumativa, interna y externa, de resultados y de la cobertura, y así sucesivamente. Además, existen tipos de evaluación que de hecho dependen de otros: por ejemplo –como mostraremos a continuación–, la valoración económica presupone la de resultados, mientras que ésta última tiene como requisito previo la evaluación de la evaluabilidad.

Hemos hecho referencia a los dos primeros esquemas de clasificación al definir el concepto de evaluación de programas. Seguidamente, exponemos los tres restantes, basados respectivamente en (1) qué se evalúa, (2) por parte de quién y (3) en qué momento del desarrollo de las actuaciones.

Qué se evalúa

Alvira propone una tipología de evaluación que se asienta en los diversos componentes de un programa. Cada clase de evaluación responde a una pregunta específica sobre la intervención, como queda patente de forma resumida en la siguiente tabla.

Tipos de evaluación de programas según Alvira

| TIPO | CONTENIDO |
|--|--|
| Evaluación de necesidades | Es una estrategia de valoración encaminada a identificar los problemas sobre los que intervenir, así como a definir el sistema comunitario en el que se va a aplicar el programa. |
| Evaluación del diseño y la conceptualización | Consiste en analizar la lógica teórica del programa, a partir de los contenidos explícitos en el documento escrito. |
| Evaluación de la evaluabilidad | Tiene como finalidad determinar si el programa es susceptible de evaluación, atendiendo tanto a la calidad de su formulación como a posibles obstáculos de carácter externo. |
| Evaluación de la implementación | Informa de las actividades llevadas a cabo, y se hace efectiva a través de la comparación del diseño teórico con el funcionamiento real del programa. |
| Evaluación de la cobertura | Pone de manifiesto el número de participantes en las actividades del programa, así como los problemas de acceso de aquellos que no hicieron uso del mismo (desconocer la existencia del programa, y barreras físicas y de motivación). |
| Monitorización y seguimiento del programa | Es un sistema longitudinal de indicadores sociales, paralelo al desarrollo del servicio o programa. |
| Evaluación de resultados | Es un dictamen de la eficacia y efectividad de la intervención. |
| Evaluación del impacto | Analiza los efectos indirectos del programa en aquellos segmentos sociales que no forman parte de la población diana. |
| Evaluación económica | Determina el coste del programa en comparación con los resultados que consigue. |

La evaluación de necesidades no es una valoración del programa propiamente dicha –ya que se trata de un juicio sobre la situación comunitaria-, y por eso en nuestro caso la hemos tratado en un apartado diferente. Sin embargo, es una de las estrategias valorativas del ciclo de intervención planificada, y también combina un componente de investigación con otro de valoración y toma de decisiones.

El resto de categorías se organizan –en nuestra opinión- en torno a la evaluación de resultados, de cuya preeminencia teórica hemos dado cuenta tanto a través de la revisión histórica del área como de los criterios de valor propuestos por la Organización Mundial de la Salud. En la evaluación de resultados se presta particular atención al control de las amenazas a la validez interna, puesto que tiene por objetivo atribuir al programa de intervención determinados cambios observados en la comunidad o en el problema social en cuestión. Esta preocupación ha generado una batería de diseños metodológicos de evaluación aplicables en diferentes circunstancias.

Entre los requisitos de la evaluación de resultados se cuentan los juicios sobre la conceptualización, la evaluabilidad y la implantación del programa. La evaluación de resultados contrasta la relación entre las actividades de intervención y los resultados previstos, y conlleva algunos pasos previos:

1. Primero, el análisis del modelo teórico del programa –o, en su defecto, la teoría implícita que subyace al mismo- es una guía para seleccionar las variables a observar, y muestra el grado de justificación de las actividades. Este examen de la *conceptualización* es la base para decidir qué expectativa de relación hay que contrastar más adelante.
2. En segundo lugar, procede anticipar las oportunidades y límites que ofrece el programa para ser evaluado. En la medida en que cuente con una adecuada formulación de los objetivos, y una clara correspondencia entre objetivos y actividades, mayor será su *evaluabilidad*. Desde ese punto de vista, la consistencia interna del programa guarda una relación directa con las posibilidades de evaluación: en otros términos, anticipar cómo se valorarán los resultados de una intervención ayuda a formular un programa de calidad. Pero también hay que prever en el análisis de la viabilidad otros obstáculos que no tienen que ver directamente con la formulación del programa, sino con la disponibilidad de tiempo y presupuesto, las dificultades de seguimiento, o el eventual rechazo de algunos grupos a proporcionar información.

3. En tercer lugar, conocer las circunstancias de aplicación del programa evita errores de atribución en la valoración de los resultados: si éstos no son los esperados no significa necesariamente que las actividades propuestas no fueran suficientes, puesto que ha podido ocurrir que no se hayan desarrollado en su integridad, o que no se ajustasen a lo especificado en el documento escrito. Para valorar la eficacia y la efectividad de la intervención es necesario constatar previamente que la *implantación* no se encontró –entre otros- con problemas de tiempo o presupuesto, de reactividad por parte de los usuarios o falta de competencia de los profesionales que lo aplicaron.

Con estos tres antecedentes se está en condiciones de hacer una evaluación de resultados de calidad. Complementariamente, se puede recoger información sobre el número de usuarios atendidos (cobertura), y/o los efectos indirectos en núcleos de población a los que no afectó directamente el programa (impacto). De igual manera, se pueden introducir consideraciones de carácter económico sobre la eficiencia de la intervención (evaluación económica).

La monitorización y seguimiento del programa es una categoría que parece discordar en la clasificación, ya que no hace referencia a un contenido diferente, sino a un procedimiento particular –basado en un sistema de indicadores- para gestionar y dirigir la intervención. De hecho, la información que se recoge puede ser sobre la cobertura, los roles profesionales, la calidad de la atención, etcétera. En España se aplica un sistema de estas características en los servicios sociales comunitarios: el Sistema de Información de Usuarios de los Servicios Sociales (SIUSS) es una aplicación informatizada en la que se recogen datos sobre cada persona que acude a solicitar servicios. En concreto, se describen las condiciones socio-familiares y de hábitat, la demanda planteada, los recursos aplicados y la valoración del profesional. Esta ficha social permite conocer los rasgos sociales y familiares de los usuarios de servicios sociales en general, y de cada programa en particular. También contabiliza el uso de los servicios y constata la evolución de la demanda. Todos esos datos sobre el funcionamiento del programa orientan el desarrollo de las actividades, y son útiles en la asignación de recursos y en la planificación de servicios.

Sin embargo, el SIUSS no informa directamente de la efectividad de la intervención, ya que no recoge datos sobre los resultados de cada programa. En general, en servicios sociales no es muy frecuente la valoración de resultados, sino que predominan los informes sobre otros aspectos, como la implementación o la cobertura: la práctica habitual consiste en una evaluación cualitativa interna, que resume en una memoria final las actividades llevadas a cabo, así como el número de usuarios atendidos. Se trata muchas veces de un informe parcial de la implantación, que se limita

a constatar el desarrollo de las actividades, sin especificar si se ajustaron o no a los planteamientos teóricos previos. Lo mismo ocurre con la cobertura, en la medida en que se cuantifiquen los usuarios sin informar de los problemas de accesibilidad de aquellos que no fueron atendidos.

Supongamos que llevamos a cabo un programa de Formación Profesional Ocupacional con madres solteras para facilitar su inserción laboral, y decidimos aplicar una metodología participativa en los cursos. En ese caso, la evaluación de resultados podría consistir en contrastar (a) la adquisición de aquellos conocimientos y habilidades que se querían potenciar, (b) la realización de comportamientos de búsqueda activa de empleo después de los cursos, o (c) el cambio de la situación laboral de las madres solteras. En lugar de este tipo de datos, a veces la evaluación se limita a informarnos de la opinión de las participantes en el curso, además de contabilizar el total de sesiones de formación y el número de participantes. Pese al interés que pueda tener esta información, no hace referencia directa a la repercusión de las actividades del programa en la situación de empleo de las madres solteras. Con frecuencia, las cuestiones de efectividad se resuelven con las impresiones cualitativas de los profesionales que aplicaron el programa (con los sesgos que ello puede conllevar, tanto por la metodología como por la fuente de información).

En la misma línea, en el ejemplo anterior sólo se proporciona información superficial sobre la implantación y la cobertura. Sabemos el número de sesiones formativas que se impartieron, pero no sabemos si siguieron el carácter participativo especificado. Y conocemos el número concreto de mujeres que asistieron con asiduidad, pero desconocemos las dificultades que encontraron aquellas que no lo hicieron. Identificar el grado de aplicación del programa es esencial para valorar los resultados, mientras que detectar problemas de acceso genera orientaciones prácticas para mejorar el programa. En efecto, si descubrimos que muchas mujeres no asistieron por desconocer la existencia del curso, o por el coste -en tiempo y dinero- de desplazarse al lugar de formación, propondríamos estrategias de difusión y becas para aumentar el alcance del programa. Igualmente, si el hecho de ser madres condiciona la asistencia, complementar la intervención con una guardería para los hijos también mejoraría la cobertura.

Quién evalúa

La evaluación interna es aquella que realizan los mismos profesionales que aplicaron el programa, mientras que en la evaluación externa los encargados de emitir un dictamen sobre la intervención no participan directamente en las actividades del mismo. Cuando la evaluación se lleva a cabo “*desde dentro*” los evaluadores cuentan

con un conocimiento directo de las actividades, de los usuarios y del desarrollo del programa. Ese grado de familiaridad facilita tanto la recogida de información como la incorporación de cambios que se deriven del estudio, y hace menos costosa la evaluación.

Sin embargo, a menudo la valoración interna puede verse condicionada por la subjetividad de quienes la realizan, y en consonancia resulta menos creíble para los observadores externos. Para los profesionales puede ser difícil tomar distancia respecto a su propia actuación, y discriminar con neutralidad los pros y contras del programa. Además, con toda probabilidad, se sienten implicados en las consideraciones de éxito o fracaso en términos de resultados; y es menos común que utilicen estándares establecidos *a priori*.

La evaluación externa contrarresta en gran medida estas limitaciones, puesto que tiende a valerse de indicadores objetivos, diseños más rigurosos, y otras garantías de fiabilidad y validez. Esto dota a la evaluación “*desde fuera*” de mayor credibilidad social, extremo muy importante en este área, ya que favorece que sus datos sean tenidos en cuenta por quienes gestionan y por quienes implantan la intervención.

No obstante, como contrapartida, el papel de evaluador externo está más sujeto a problemas de definición: por ejemplo, puede ser percibido como un auditor, de cuya opinión depende la continuidad del presupuesto; o bien se asocia su informe a las futuras condiciones laborales de quienes aplicaron el programa. Estas interpretaciones pueden traducirse en el rechazo de participantes y profesionales a ser entrevistados y proporcionar información, o, en último extremo, pueden afectar a la calidad de los datos obtenidos.

Cuándo se evalúa

Se suele distinguir entre evaluación formativa y sumativa dependiendo del momento en que se lleva a cabo, aunque ambas también difieren en las funciones que cumplen en el proceso de intervención. La evaluación formativa o de proceso se realiza durante la aplicación del programa, mientras que la evaluación sumativa o final tiene lugar después de concluidas las actividades. La primera tiene por objetivo la mejora y perfeccionamiento del programa, de modo que la información que genera retroalimenta la intervención, y pueden cambiarse las actividades y características del programa sobre la marcha. Por su parte, la segunda tiene como principal función contabilizar los resultados, centrándose en los efectos del programa.

Lógicamente, ambas perspectivas pueden combinarse, poniendo en relación el proceso de intervención con sus resultados. Una evaluación formativa y sumativa al mismo tiempo es quizá la mejor forma de obtener el máximo aprendizaje de la

experiencia. Pero a veces se pueden usar por separado, adaptándose a las circunstancias de la intervención: así, la valoración del proceso es muy apropiada en aquellos casos en los que el programa se esté elaborando o construyendo, como estrategia exploratoria con la que identificar los componentes de mayor interés. Al contrario, si un programa ya está consolidado, y hemos contrastado sus efectos en experiencias anteriores, quizá sea suficiente con una evaluación final para contabilizar los resultados.

FASES DE EVALUACIÓN

Según Fernández Ballesteros, podemos resumir el proceso de evaluación en seis pasos sucesivos: (1) planteamiento de la evaluación, (2) selección de las operaciones a observar, (3), selección del diseño de evaluación, (4) recogida de información, (5) análisis de datos, y (6) elaboración del informe. El proceso se divide en dos grandes componentes: una primera parte de diseño del sistema de evaluación –compuesta por los tres primeros pasos-, y una segunda de aplicación del mismo –compuesta por los tres últimos. En la primera parte predomina la preocupación por controlar las amenazas a la validez interna y, al mismo tiempo, responder a las necesidades de los usuarios de la evaluación. En la segunda prevalecen consideraciones prácticas, encaminadas a implantar el diseño tal y como fue definido a priori, así como a superar eventuales barreras para entrar en contacto con la población objeto de estudio, y recabar la información necesaria.

El punto de partida de la evaluación de programas es una fase genérica de planteamiento, en la que se precisan cuáles son los intereses de quiénes encargan el estudio. También se analiza el contenido del programa, para familiarizarse con el mismo y determinar la viabilidad de la evaluación. Aunque se trata de pasos preliminares, este planteamiento previo condiciona las decisiones que se toman posteriormente, y puede ser determinante de la calidad de la evaluación.

1. El encargo contribuye a definir los objetivos de la evaluación, y se concreta en un diagnóstico de las necesidades que motivaron el estudio. Las características técnicas y valorativas de la evaluación tienen que adaptarse a tales necesidades, si quiere convertirse en una herramienta práctica en el ciclo de intervención. De ese modo, proporciona una guía para elegir los criterios de valoración del programa; sirve para seleccionar los procedimientos de investigación a los que se atribuye mayor credibilidad; y orienta otras decisiones sobre el enfoque de investigación, como optar entre la contabilidad de resultados o el examen del proceso de aplicación del programa.

Sin embargo, puede darse el caso de que los promotores de la evaluación no cuenten con una definición precisa de lo que esperan de la misma, y es frecuente que desconozcan las posibilidades que ofrece la evaluación de programas. En esas circunstancias, el evaluador asume el papel de exponer las alternativas metodológicas, junto a las dimensiones del programa susceptibles de análisis. En cualquier caso, sea con la formulación explícita del interesado o con las orientaciones y sugerencias del evaluador, es esencial determinar en este primer momento cuáles son las preguntas a las que debe responder la evaluación.

2. El segundo componente de esta primera fase consiste en familiarizarse con el programa: es decir, analizar en profundidad el documento en el que se resumen objetivos y actividades, entrar en contacto con el personal que lo aplica, y conocer el contexto organizacional en el que se va a llevar a cabo. Con ello se establecen los límites y oportunidades de la intervención desde el punto de vista evaluativo, y sirve para decidir los elementos a observar (fase 2) y elegir el diseño propiamente dicho (fase 3).
3. También es parte del planteamiento inicial evaluar la evaluabilidad. Es decir, partiendo de cómo esté formulado el programa, determinamos qué acciones de evaluación son posibles y cuáles no. Con objetivos mal definidos será difícil precisar la eficacia del programa; mientras que si falta correspondencia entre objetivos y actividades no sabremos en qué actividades centrarnos para valorar cada objetivo. También deben anticiparse en esta fase obstáculos de evaluación como la falta de presupuesto, la movilidad de la población o la reactividad de los participantes, entre otros.

Una vez que conocemos qué se espera de la evaluación, y qué oportunidades y limitaciones ofrece el programa para ser evaluado, estamos en condiciones de decidir las características metodológicas de la misma. En concreto, seleccionamos las operaciones a observar y el diseño de evaluación. Seleccionar las operaciones a observar consiste en elegir las variables dependientes que vamos a medir, así como las fuentes de información de las que obtendremos dichos datos. La elección del diseño es – por su parte- una propuesta metodológica para el control de las amenazas a la validez interna. Y finalizado el diseño, pasamos a recoger información, analizar los datos y elaborar el informe definitivo.

Para ilustrar la selección de operaciones a observar, pensemos en los indicadores utilizados en la valoración de un programa de inserción socio-laboral con inmigrantes. Un componente de la intervención consistió en cursos de Formación Profesional Ocupacional, complementados con la acción de Centros de Orientación Socio-laboral. Para valorar los primeros se tomaron medidas de la situación psicosocial y de empleo de los participantes en la formación, utilizando la información proporcionada por los interesados y por los monitores de los cursos. En los Centros se atendió fundamentalmente a la estructura y clima organizacional de los mismos, así como a otros rasgos de funcionamiento. Aunque la evaluación fue más amplia, el esquema muestra cómo las variables a observar se organizaron de acuerdo con las dos áreas componentes del programa.

La elección de dichos indicadores también estuvo condicionada por el encargo de evaluación, y el documento en el que se reflejaban objetivos y actividades. Veamos, en primer lugar, el caso de los Centros de Orientación. Por un lado, los administradores del programa se mostraron especialmente preocupados por la implantación de un conjunto de Centros que fuesen homogéneos en sus pautas de trabajo, dado que pertenecían a diferentes organizaciones sociales. Por otro lado, en el documento escrito no se formulaban objetivos operativos de logro en el ámbito de la inserción laboral. Los primeros contactos con los orientadores pusieron de manifiesto, además, que las actividades estaban más influidas por la demanda concreta de los usuarios -centrada en su mayoría en la gestión de los permisos de residencia y trabajo-, que por los contenidos definidos en el programa. Todas estas consideraciones son en sí mismas conclusiones de evaluación, pero enfocaron a su vez el examen del programa en la implantación de los Centros (para valorar el grado de convergencia en roles, clima, pautas de organización, etcétera). Como consecuencia, el informe de evaluación apenas pudo incorporar resultados en el área laboral, pero -dado su carácter formativo- contribuyó a definir los roles y competencias propios de los orientadores laborales, conformando una guía para futuros proyectos.

La valoración de los cursos de formación también se vio afectada por este tipo de planteamientos previos. Si bien se estableció una meta genérica de “*inserción laboral de los inmigrantes*”, los objetivos no especificaban si se hacía referencia a la estabilidad laboral, las condiciones de trabajo o la ocupación de los desempleados. Tampoco se partió de un modelo teórico que identificase variables relevantes en las que incidir para cambiar la situación de empleo. En la evaluación se optó por describir las circunstancias socio-laborales de los participantes a lo largo de los cursos de formación, y complementariamente medir algunas variables psicosociales de interés según la literatura sobre el tema. De esa forma, al tiempo que se valoraba la efectividad, se comprobaba el papel de variables y modelos relevantes para futuros programas.

Pero, como hemos comentado al revisar las fases de evaluación, el diseño no termina con la definición de las operaciones a observar. En el ejemplo que estamos analizando, no basta con tener los datos sobre la situación de empleo tras los cursos, sino que hay que conocer cuál era el punto de partida. Probablemente también interese comparar a los participantes en la formación con un grupo de inmigrantes no implicado en el programa, para demostrar la existencia de diferencias significativas. Teniendo en cuenta que la evolución en términos de empleo podría asociarse a una disminución del paro en la región, podemos incorporar controles estadísticos adicionales. Todo este tipo de decisiones configura el diseño de evaluación: una propuesta que delimita la validez externa de las conclusiones, y el grado en que se han controlado las amenazas a la validez interna.

TÉCNICAS DE EVALUACIÓN DE PROGRAMAS Y SELECCIÓN DE OPERACIONES A OBSERVAR

Para valorar los programas sociales podemos utilizar las mismas técnicas de recogida y análisis de información que describimos en el capítulo de Evaluación de Necesidades. El evaluador de programas recurre en la práctica a muy diversos sistemas de colección de datos: a título ilustrativo, podemos mencionar los grupos de discusión con los usuarios de un servicio, las encuestas a la comunidad receptora de un conjunto de prestaciones, las entrevistas a profesionales o participantes en un programa, etcétera.

Como ya mostramos en la evaluación de necesidades, la elección de la técnica a utilizar está directamente asociada a la unidad y objeto que queremos evaluar. Por ejemplo, hay técnicas que son más apropiadas para el diagnóstico de necesidades percibidas, mientras que otras son exponentes de necesidades expresadas. Del mismo modo, en la evaluación de programas existe correlación entre las operaciones a observar y las técnicas aplicables en cada caso.

Una clasificación bastante útil del tipo de indicadores de que se hace uso en la evaluación de programas la proporciona la tipología de cuatro niveles de Kirpatrick. Se trata de una propuesta específica para la evaluación de las acciones formativas, pero en gran medida puede hacerse extensiva a otros tipos de intervención. Según este autor, para valorar los cursos de formación es posible recurrir a información sobre (a) la reacción de los participantes, (b) el aprendizaje de determinados contenidos, (c) los cambios de comportamiento después de la formación, y (d) el impacto en otros resultados de relevancia. Veamos con detenimiento cada uno de estos “*cuatro niveles de evaluación*”:

1. La *reacción* hace referencia en sentido estricto a la satisfacción de los participantes en los cursos de formación. Pero, por extensión, cabría definirla como la opinión que expresan los usuarios de un servicio sobre las prestaciones recibidas. Suele medirse a través de cuestionarios y escalas - inmediatamente después de finalizar la formación o el programa-, con los que se pretende conocer la percepción de la población diana.

La satisfacción de los alumnos de un curso no garantiza que se hayan obtenido los conocimientos deseados. Tampoco la percepción de los participantes en un programa tiene por qué coincidir con los resultados objetivos del mismo. Pero tanto la primera como la segunda constituyen un primer acercamiento a la valoración de un programa, que resulta relativamente fácil y económico en comparación con los demás niveles. Por otro lado, mostrarse satisfecho con la formación se relaciona por lo general con el aprovechamiento de la misma, en la medida en que la atención y motivación de los participantes es un requisito del aprendizaje.

2. El *aprendizaje* es el cambio en los conocimientos, habilidades o actitudes que se produce como consecuencia de la participación en un programa. Suele evaluarse con tests de conocimientos, escalas de actitudes, pruebas de desempeño y sistemas de observación. Ahora bien, mientras que la reacción se documenta suficientemente con una medición puntual, para evaluar el aprendizaje es recomendable contar con una medición previa al programa y, si es posible, con un grupo de comparación.

Se trata de un nivel de evaluación que resulta particularmente apropiado para las acciones de formación. Sin embargo, la adquisición de conocimientos o la modificación de actitudes no siempre se traducen en cambios de comportamiento en los contextos naturales. Pongamos por caso que un grupo de toxicómanos, gracias a varias sesiones de entrenamiento, llega a reconocer el intercambio de jeringuillas como una vía de contagio del SIDA y aprende algunas medidas de profilaxis al respecto. Esto no significa que vayan a llevar a la práctica las conductas de higiene en su vida cotidiana, donde se verán condicionados por cuestiones de tiempo y dinero, por la presión del grupo, etcétera. Este tipo de consideraciones justifica distinguir otro nivel de evaluación, que dé cuenta del grado de transferencia del comportamiento.

3. La *conducta* es la transferencia de los cambios en los conocimientos, habilidades y actitudes a los contextos naturales. En este caso es recomendable la observación, o incluso el informe de terceros, aunque también es válido el auto-informe de los participantes. Como decíamos en el nivel anterior, para constatar cambios conductuales debidos al programa, es oportuno tomar una medición previa al desarrollo de las actuaciones y utilizar un grupo de comparación. A ello hay que añadirle el interés de las mediciones de seguimiento, para contrastar que el cambio de conducta se ha instaurado de manera estable en el nuevo contexto. Por esta misma razón, no es conveniente tomar un indicador recién finalizadas las actividades, sino dejar cierto tiempo para que se produzca el cambio de conducta.

4. Los *resultados* son los cambios que se producen en última instancia en el problema que se quiere resolver o prevenir. En esta categoría entran indicadores como los siguientes: prevalencia del SIDA, número de accidentes laborales, proporción de toxicómanos en una zona geográfica, etcétera. Es el tipo de datos que permite extraer conclusiones más categóricas sobre la relevancia social del programa. Pero al mismo tiempo es el nivel que conlleva una evaluación más difícil y costosa. Desde el punto de vista metodológico, son válidas las mismas recomendaciones que en el nivel anterior.

No hay que confundir este nivel con la “evaluación de resultados” que hemos descrito más arriba: dependiendo del programa, la eficacia y la efectividad pueden llegar a evaluarse con indicadores de reacción, de aprendizaje, de conducta y de resultados indistintamente.

Como puede deducirse de la exposición anterior, los cuatro niveles están mutuamente relacionados. La satisfacción facilita el aprendizaje, que a su vez es un requisito necesario pero no suficiente para el cambio comportamental. Y son estos últimos cambios los que afectan al problema que nos interesa modificar.

Esto quiere decir que al observar que los indicadores de los cuatro niveles covarían entre sí, estaremos comprobando la suficiencia del programa (o la adecuación de su teoría implícita). Por tanto, puede resultar muy práctico combinar los cuatro niveles en la evaluación de un mismo programa.

Supongamos que queremos valorar un curso de prevención de accidentes impartido a los mandos intermedios de una planta de producción sidero-metalúrgica.

Siguiendo la distinción de Kirpatrick, podríamos utilizar uno o varios de los siguientes indicadores:

- (a) aplicación de cuestionarios a los participantes en el curso para conocer su opinión sobre los monitores y los contenidos (reacción),
- (b) realización de un examen sobre las prácticas de protección recogidas en el Manual de Seguridad de la empresa (conocimiento),
- (c) entrevista a los subordinados de cada mando intermedio, para determinar si ha cambiado en su comportamiento preventivo después de participar en la formación (conducta),
- (d) elaboración de un inventario de los accidentes acaecidos en la planta de producción antes y después del curso (resultados).

Si el objeto del curso de formación es en última instancia reducir el número de accidentes, ¿qué interés podría tener tomar las medidas a, b y c? En primer lugar, proporciona una evaluación más matizada de los resultados del programa. Pero además puede servir eventualmente para detectar por qué no funciona determinada intervención. En el ejemplo que estamos analizando, el fracaso en la prevención de incidentes podría atribuirse, entre otras, a alguna(s) de las siguientes causas:

- (a) que las actividades de enseñanza no hayan tenido éxito en la difusión de conocimiento,
- (b) que lo aprendido en el curso no se haya generalizado al contexto de trabajo, o
- (c) que los cambios en el comportamiento del supervisor no hayan sido suficientes para reducir la accidentabilidad.

El análisis de los cuatro niveles permite identificar donde “falla” el programa, y cómo mejorarlo en futuras aplicaciones.

Por otro lado, el esquema de cuatro niveles también puede utilizarse como guía para formular los objetivos de la intervención.



Isidro Maya Jariego

<http://www.personal.us.es/isidromj>

--

Para citar este artículo, utiliza por favor las siguientes referencias:

Maya Jariego, I. (2003). Diseño, evaluación e implantación de un proyecto de intervención social. En F. Loscertales y Núñez, T. (Coords.): *Comunicación e Intervención*. Módulo 10: Revisión del proyecto y redacción final de la Tesis de Maestría. CD Interactivo. Universidad de Sevilla.

Maya Jariego, I. (2001). La intervención social planificada. En F. Loscertales y Núñez, T. (Coords.): *Comunicación y habilidades sociales para la intervención en grupos*. Módulo 10: Evaluación y diseño del proyecto de intervención. CD Interactivo. Universidad de Sevilla.