

MODELOS DE EVALUACIÓN DE PROGRAMAS DE FORMACIÓN

Marisa Vázquez Mazzini, 2017

La evaluación es inherente al ser humano. Permanentemente estamos emitiendo juicios del tipo “me gusta”, “no me gusta”, “a la gente le sirve”, “a la organización le es provechoso”, etc. Todos estos y otros juicios se basan en un conjunto de criterios que orientan la mirada sobre el ente a evaluar. Guían tanto la elección de aspectos que merecen ser tenidos en cuenta como la forma de apreciarlos.

En la evaluación de programas orientados al aprendizaje es posible distinguir dos modelos apreciativos, que parten de premisas diferentes y definen el término “evaluación” de maneras diferentes. Están sintetizados en la tabla que sigue.

Paradigma cuantitativo	Paradigma cualitativo
Evaluar es medir: se arriba a un resultado parcial o final, y se califica tal resultado de acuerdo con una escala.	Evaluar es develar e interpretar procesos o resultados previstos o no.
Los evaluadores no deben contaminar con su subjetividad la recolección de los datos y su medición.	La evaluación es intersubjetiva. Los evaluadores negocian marcos valorativos y puntos de vista para analizar el programa.
Los instrumentos de evaluación deben ser objetivos y precisos.	La mirada atenta para percibir cualidades y cambios y el intercambio de perspectivas son los principales instrumentos.
El evaluador debe ser neutral respecto del programa y del contexto.	Los hechos no dicen nada si se los escinde del contexto en el que se producen.
Los resultados deben ser cuantificados.	Debe utilizarse un lenguaje fiel a los datos. No siempre los números reflejan lo ocurrido ni dan cuenta de los cambios.
Si se siguen las prescripciones anteriores, el informe de evaluación refleja la realidad tal como es.	El informe refleja la realidad tal como la ven los evaluadores. No hay un único modo de valorar.

Evaluación cuantitativa

Consideraremos dos modelos de evaluación cuantitativa, los más difundidos en Capacitación Laboral. Uno es el Modelo de los Cuatro Niveles de Donald Kirkpatrick (1924 – 2014). El otro es el que propone identificar el retorno de la inversión (“*return of investment*” o ROI en el original).

En 1959, Donald Kirkpatrick (de la Universidad de Wisconsin, Madison, y ex-presidente de la Sociedad Estadounidense de Capacitación y Desarrollo) publicó en la revista *Training and Development* una serie de cuatro artículos titulados *Técnicas para evaluar programas de capacitación*. En ellos presentaba un sistema de evaluación en cuatro niveles, cada uno de los cuales proporcionaba datos que permitían efectuar ciertas predicciones acerca del siguiente. Su propósito al escribir los artículos era desentrañar el significado de la palabra “evaluación”, dada la variedad de significados asignados al término en el campo de la Capacitación Laboral en empresas. En 1994 publicó su libro *Evaluating Training Programs - The Four Levels* (Berrett-Koehler Publishers). Allí describe más profundamente cada uno de los cuatro niveles, presenta con claridad cuáles son las preocupaciones que el modelo intenta contestar, y da a conocer casos reales de aplicación exitosa del modelo.

La principal característica de este enfoque es la que da lugar a su nombre: los cuatro niveles. Con el propósito de brindar a los profesionales de la capacitación herramientas para demostrar su contribución al rendimiento de la empresa, diferencia cuatro instancias de medición de resultados.

- Evaluación de reacción (nivel 1), que permite determinar las apreciaciones positivas o negativas de quienes reciben el programa.
- Evaluación de aprendizaje (nivel 2), que permite medir el grado de adquisición de conocimientos, cambio de actitudes o desarrollo de habilidades.
- Evaluación de comportamiento (nivel 3), que permite detectar en qué grado se aplica lo que se aprende.
- Evaluación de resultados (nivel 4), que permite identificar en qué grado el cambio en el comportamiento incide en los resultados organizacionales.

Para el autor, cada uno de estos niveles opera como condición de posibilidad para el otro: si los destinatarios están insatisfechos o incómodos no se dispondrán a aprender; si no aprenden, no podrán transferir los contenidos a sus comportamientos cotidianos; si no cambian su comportamiento cotidiano, no podrían producir mejoras en la organización.

El modelo se propone, además, delimitar responsabilidades en caso que pese a la implementación de un programa formativo el desempeño organizacional no mejore. Al respecto, Kirkpatrick señala que los niveles 1 y 2 corresponden al área de incumbencias del profesional de la capacitación. En los niveles 3 y 4, en cambio, las responsabilidades son

compartidas: ante la falta de indicadores de mejora en el desempeño de la compañía, no sería justo “culpar” al Departamento de Capacitación dado que otras variables inciden en el cambio de conducta y en la “traducción” de la nueva conducta a metas cuantitativas.

El cometido central de la evaluación es, para Kirkpatrick, ayudar al área de Capacitación a demostrar el éxito en sus funciones. No obstante, al reflexionar sobre cómo se desagrega esta misión fundamental, el autor suma otros propósitos:

- justificar la existencia del Departamento de Capacitación, que suele desaparecer en casos de reducción de estructura;
- decidir la continuidad o discontinuidad de un programa de formación, en función de la relación entre costos y beneficios;
- obtener información sobre el desempeño de los agentes educativos

Unas pocas pautas generales rigen el “cómo” de la evaluación más allá de las diferencias específicas entre niveles.

a) Buscar, en tanto sea posible, datos cuantitativos. Su procesamiento es más simple que en el caso de los datos cualitativos. Por otra parte, ése es el tipo de datos que esperan quienes toman las decisiones sobre los programas en función de la ecuación costo - beneficio.

b) Procesar cuantitativamente los datos de modo de construir parámetros, facilitar la comparación y generar homogeneidad entre la información concerniente a la capacitación y otra información útil para el negocio.

c) Diseñar e implementar instrumentos que cumplan con su cometido y que, a la vez, resulten económicos desde el punto de vista del tiempo y dinero que requieren.

d) Utilizar la información, ya que uno de los propósitos de la evaluación es la optimización de las decisiones.

e) Dar a conocer la información, ya que otro propósito de la evaluación es fundamentar la contribución del Área de Capacitación al resultado de la organización.

Más allá de estos principios generales, cada uno de los niveles supone procesos e instrumentos de evaluación propios.

1. Evaluación de reacción: diseñar una encuesta de fácil administración, que incluya todos los ítems que podrían generar satisfacción o insatisfacción en el destinatario (ejercicios, contenidos, tiempos previstos, instalaciones en caso de programas presenciales, comportamiento de los agentes educativos, etc.).

2. Evaluación del aprendizaje: diseñar un esquema de testeo antes de la formación y después de la formación, para comparar el estado inicial del destinatario con el estado final.
3. Evaluación del aprendizaje: comparar el desempeño del grupo de destinatarios capacitado contra el desempeño de un grupo de control, no capacitado. Asimismo puede ayudar implementar encuestas tiempo después, para preguntar a quienes fueron formados y a sus supervisores directos qué cambios identifican.
4. Evaluación de resultados: intentar aislar a la capacitación de otras posibles variables que intervengan en alguna mejora del desempeño organizacional. Si no se cuenta con pruebas sobre el resultado, conformarse con evidencias que induzcan a reconocer alguna contribución del programa.

Este modelo es objeto de numerosas críticas. La principal radica en que conduce a escindir aprendizaje (objeto del nivel 2) de transferencia (objeto del nivel 3). En otros términos: da por sentado que puede haber aprendizaje sin transferencia. La razón de ser de la Capacitación Laboral es, precisamente, la promoción de aprendizajes para la tarea, lo cual supone un proceso continuo que se inicia con la disposición a aprender y se concreta en determinados cambios en las capacidades de individuos y grupos. ¿Qué sentido tiene implementar programas "exitosos" en términos de resultados en las pruebas de contenido? Para los gerentes o quienes toman las decisiones presupuestarias tales resultados no constituyen una razón suficiente para justificar la inversión en formación. Para los jefes de los destinatarios, lo que ocurra durante un programa fuera del lugar de trabajo cuenta sólo si tiene valor predictivo con relación a lo que ocurrirá luego en el puesto. En suma: aunque pudiera demostrarse válida y confiablemente que los destinatarios han comprendido aquello que les fue enseñado, los datos no servirían para informar acerca de las mejoras en el desempeño. El modelo —tal como lo reconoce el propio autor— presenta debilidades en lo relativo a los niveles 3 (transferencia) y 4 (mejoras organizacionales y retorno de la inversión).

Más allá de Kirkpatrick, diferentes autores (Pater Bramley, "Evaluating Training Effectiveness", McGraw-Hill, 1993; Jack Phillips, "Return of Investment", Routledge, 2011; Robinson, Dana y Robinson, Jim, "Training for Impact", Jossey-Bass, 1989, entre otros) se han ocupado de la cuestión del retorno de la inversión. Intentan responder a la pregunta relativa a cuál es el beneficio económico de los programas orientados a promover aprendizaje. Hasta el momento no hay acuerdo en el campo profesional respecto de la manera de efectuar este tipo de evaluación de manera precisa, fundamentalmente debido a la imposibilidad de aislar el factor "aprendizaje" de otros factores que puedan intervenir. Las recomendaciones de los autores incluyen cuestiones tales como las siguientes.

- Sistematizar los datos sobre los costos del programa.

- Recolectar datos que den cuenta de posibles contribuciones del programa (usando encuestas a los destinatarios y a sus jefes, o herramientas de análisis del rendimiento de la población capacitada).
- Procurar recortar el efecto del programa, comparando el desempeño de un grupo capacitado con el de otro no capacitado.
- Traducir las diferencias entre ambos grupos en indicadores cuantitativos: reducción de tiempos de producción, incremento de las ventas, disminución del ausentismo, disminución de la rotación, etc.
- A partir de tales indicadores cuantitativos, calcular los beneficios económicos del programa (productividad, rentabilidad, etc.).
- Calcular el beneficio neto del programa, que consiste en la diferencia entre beneficios económicos y costos.
- Calcular el retorno de la inversión: se divide el beneficio neto por el costo del programa, y al monto obtenido se lo multiplica por cien.

Las prácticas organizacionales —sobre todo las empresariales, en las cuales los análisis económicos resultan insoslayables— tienden a requerir este tipo de cálculos. Sin embargo, existe un acuerdo tácito acerca de su relativa confiabilidad. Más que de mediciones se trata de estimaciones que brindan orientaciones sobre una entidad difícil de aislar.

Evaluación cualitativa

Consideraremos ahora dos modelos de evaluación cualitativa. Uno de ellos es el que propone el Harvard Project Zero, especialmente relevante para monitorear el proceso de aprendizaje. El otro es el Modelo de la Evaluación Respondiente (*“Responsive Evaluation”* en el original) de Robert Stake, de la Universidad de Illinois. Resulta especialmente apto para monitorear programas institucionales desde una perspectiva colaborativa.

Una buena parte de los trabajos del Harvard Project Zero está destinada a revisar la perspectiva —instalada en el “imaginario social”— según la cual evaluar equivale a tomar pruebas formales y traducir los resultados a escalas cuantitativas de puntajes. Para esto sus investigadores utilizan una analogía entre la enseñanza y el deporte. Un entrenador de básquet —sostienen— puede comenzar la práctica pidiendo al equipo que ejercite ciertas estrategias. Mientras los jugadores participan del partido, estudia sus movimientos comparándolos con los parámetros de cada uno o con el rendimiento de expertos. Analiza los problemas cuando el grupo falla, indica cómo mejorar el desempeño. El juego termina no sólo con un resultado en puntajes que indica quién ganó, sino con reuniones de vestuario en las que se discute qué trabajar en adelante. La integración entre ejecución y *feedback* es exactamente lo que un aprendiz necesita cuando se aboca a comprender un tema en particular.

Este tipo de evaluación –que supone la integración entre actuaciones de los aprendices y *feedback* sobre tales actuaciones– tiene que ser constante y reunir un conjunto de aptitudes.

- Proporcionar evidencias sobre la comprensión de los contenidos.
- Proporcionar ocasiones para que, en el transcurso del programa, los destinatarios “usen” los contenidos en la resolución de problemas complejos.
- Proporcionar ocasiones para que anticipen la transferencia de los contenidos a otras instancias más allá del programa.
- Diluir la barrera entre “aprender” y “ser evaluado”, brindando tareas desafiantes en las que, a la vez, sea posible aprender y testear el propio nivel de comprensión.

En síntesis: la modalidad de evaluación propuesta por el Harvard Project Zero no difiere de aquella mediante la cual se aprende a pintar, a componer o a ejecutar una disciplina olímpica. El agente educativo experto guía al aprendiz en el hacer, favoreciendo grados crecientes de autonomía y dominio. Esta modalidad de evaluación no tiene por qué quedar restringida al *on-the-job training*. Puede perfectamente estar incluida en el sistema académico, en los programas presenciales de formación o en las modalidades virtuales, siempre y cuando el foco se desplace desde “hablar sobre un contenido” a “usar inteligentemente el contenido”.

Ocupémonos ahora del Modelo de la Evaluación Respondiente. Surgió en 1975, cuando Robert Stake llamó la atención sobre la naturaleza conflictiva de la evaluación. Las organizaciones no son comunidades homogéneas. Por el contrario, se caracterizan por la diversidad de voces. En ellas conviven —tal como sostiene Wenger, a quien nos dedicamos en la unidad 2— grupos de diferentes características, cada uno de los cuales adhiere a determinados intereses y valores.

Como ejemplo, consideremos una empresa cualquiera. Cuento o no con declaraciones institucionales de misión o visión compartida, “los de Marketing” darán prioridad a los aspectos comerciales; “los de Planta” a la producción, y “los de Finanzas” a la relación entre ganancias y pérdidas. “Los nuevos” intentarán cuestionar lo instituido mientras que posiblemente “los antiguos” lo defiendan. Todos estos focos de atención pueden coexistir armónicamente. Sin embargo, ante cualquier factor que pueda amenazar los valores e intereses centrales de cada grupo, la parte predomina sobre el todo. Marketing acusará a Planta de priorizar las especificaciones técnicas por sobre las necesidades del cliente. Planta acusará a Marketing de prometer al cliente lo que no se puede cumplir. Finanzas culpará a todos por malgastar recursos. “Los nuevos” acusarán a “los antiguos” de resistentes a la innovación, los operarios se quejarán de que los gerentes no conocen el “día a día” de la empresa...

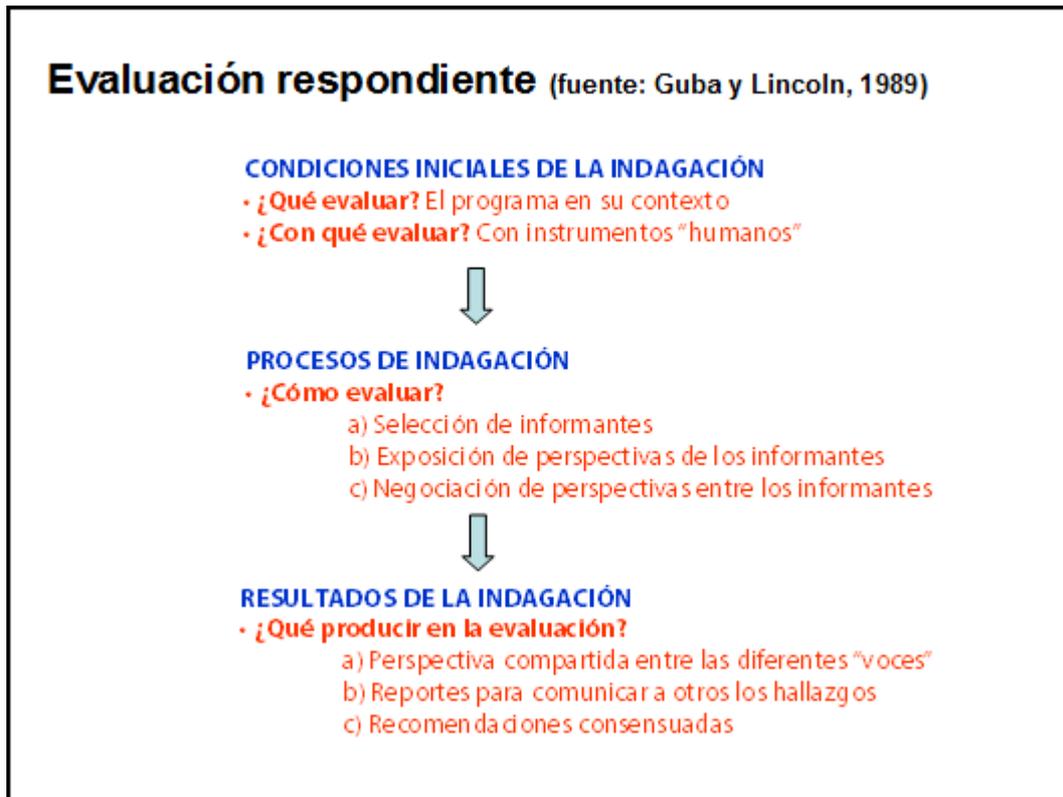
En un proceso de evaluación necesariamente se ven afectados los intereses y valores centrales de los grupos que de alguna manera están vinculados con el programa a evaluar. Por lo pronto, siempre hay un “alguien” que evalúa. Este “alguien” (un consultor, un profesional de Capacitación, RRHH, la gerencia general, un auditor, etc.) no puede prescindir de su propia filiación a alguno de los grupos presentes en la organización. Ya sea interno o externo, el

evaluador tiene *in mente* una serie de "instrucciones conceptuales" que le marcan lo que es bueno, malo, prioritario, importante o accesorio. Por consiguiente, los hallazgos de la evaluación pondrán en peligro los intereses de algunos grupos (los más ajenos al evaluador y a sus propios valores) y reforzarán los de otros (aquéllos a los que el evaluador adhiere). Para Stake, esta "toma de posición" no responde a la incapacidad o la falta de ética por parte de quien evalúa. Simplemente, es inevitable. Evaluamos dentro y con arreglo a un marco valorativo.

De ahí la necesidad de implementar una modalidad de evaluación "multivocal", en la que las distintas agrupaciones vinculadas de algún modo al problema a resolver y a la respuesta educativa elegida puedan aportar su mirada.

La Evaluación Respondiente concibe al proceso de evaluación como un proceso colectivo de construcción de "lentes". Qué datos buscar, qué significa "*resultado*", con qué criterios juzgar y cómo expresar los hallazgos no son cuestiones que puedan definirse *a priori*. Se van definiendo en las sucesivas negociaciones entre las partes involucradas. En la evaluación de un programa de capacitación, los actores "ponen a conversar" sus sistemas de valores y creencias. Como resultado del diálogo entre actores se llega a un nuevo sistema de creencias y valores compartido, inclusivo y más sofisticado que el que cada actor sostenía inicialmente. Con arreglo a este criterio: se interpreta la información obtenida (se otorga significado a los datos, se desentraña qué "dicen" los datos), se emiten reportes de evaluación, se acuerdan los cursos de acción para mejorar el programa.

El siguiente gráfico ilustra el proceso. Está elaborado sobre la base de Guba, Egon y Lincoln, Yvonna (1989), Sage Publications.



Aquello que se procura detectar es, en este modelo, cualquier cambio asociado al programa de formación: cambios en los destinatarios, cambios en el contexto, cambios en las creencias respecto de cómo encarar la capacitación, cambios en las actitudes hacia la capacitación. Resulta clave la perspectiva "multivocal" y colaborativa, porque diferentes actores organizacionales perciben diferentes aspectos. No aparecen aquí recomendaciones sobre mediciones o pruebas. Los autores proponen el empleo de instrumentos tales como las entrevistas no formalizadas de tipo etnográfico, el trabajo de campo, las conversaciones grupales. Todo indicio cuenta, en tanto el grupo de informantes seleccionados lo considere válido.