

TEMA 3

METODOLOGÍA DE EVALUACIÓN

1

Asignatura: Evaluación de programas de intervención

Curso y Titulación: 2º de Psicopedagogía

Objetivo básico del tema:

Con el desarrollo del presente tema se pretende conseguir que el alumno pueda introducirse en el conocimiento de diferentes diseños de evaluación de programas de intervención psicopedagógica.

Otros objetivos del tema son:

- Comprobar la aplicación de los principios de evaluación de programas en diferentes diseños.
- Diferenciar los elementos esenciales que configuran los diseños de evaluación.
- Comprender el significado y utilidad de los estándares para la evaluación de programas.

Contenidos:

- Principios que regulan la evaluación de programas (pág. 3)
- Diseños de evaluación de programas (pág. 9)

3. Estándares para la evaluación de programas.

Desarrollo del tema:

Se apoya en la explicación de los conceptos ya indicados y en la presentación de un caso práctico que requiera la elaboración de selección de un diseño de evaluación de programas entre los presentados en el desarrollo del tema.

Materiales:

- Documento de desarrollo de los contenidos del tema.
- Ejemplos de aplicaciones de diseños de evaluación de programas.

Autoevaluación:

Una vez finalizado el tema, se tratará de determinar si el alumno domina el objetivo básico del módulo. Para ello, se utilizará una prueba escrita, con varias preguntas de respuesta breve. Además el alumno completará una práctica sobre la selección de un diseño de evaluación adecuado para evaluar un programa de intervención psicopedagógica.

2. DISEÑOS PARA LA EVALUACIÓN DE PROGRAMAS

Las decisiones que se adoptan sobre el diseño deberían tener en cuenta los principios de evaluación que se

enunciaron en el módulo 2 y, por supuesto, las implicaciones que tienen para evaluadores y audiencias. No debemos olvidar que un diseño es tan sólo una herramienta que se sustenta sobre unos principios de actuación; la configuración que finalmente tenga una propuesta evaluativa es fruto de una recreación concreta de esos principios en un contexto práctico, de modo que resulta más sencillo cambiar el diseño (reinterpretar el modo de abordar la realidad objeto de estudio) que renunciar a los fundamentos del mismo.

En este apartado vamos a agrupar los diseños de evaluación según un esquema que diferencia entre diseños de evaluación basados en las audiencias (tradicionalmente, cualitativos) y diseños establecidos por el evaluador (tradicionalmente cuantitativos). Son muchos los elementos que los unen, de modo que solo les separa la forma en que gestionan el proceso de evaluación: los diseños cualitativos se construyen con las audiencias, mientras que los diseños cuantitativos se elaboran a pesar de las audiencias.

2.1. DISEÑOS DE EVALUACIÓN BASADOS EN LAS AUDIENCIAS

La evaluación cualitativa o basada en las audiencias no siempre sigue un esquema racionalista en el que, como apunta Taylor (1961), se eligen y definen un conjunto de estándares, se especifica la clase de comparación y se deduce el grado con que el objeto satisface los estándares. Muchas evaluaciones cualitativas revelan estándares, comparaciones y juicios más implícitos que explícitos. Incluso algunas de estas evaluaciones –como las que se desarrollan bajo un modelo de evaluación iluminativa– se afanan más por lograr la comprensión de las audiencias que por emitir un juicio. No obstante, las evaluaciones cualitativas comparten cierta lógica común al enfrentarse al análisis, comprensión y, en su caso, valoración de los diferentes objetos sometidos a evaluación. Según Patton (1987) los diseños de evaluación cualitativa se caracterizan por ser naturalistas, inductivos y holísticos.

Los diseños de evaluación cualitativa son naturalistas: no intentan manipular el programa o a sus participantes para responder a los propósitos de la evaluación. Los evaluadores se ven implicados en una indagación naturalista que trata de comprender (y valorar) procesos y actividades que se desarrollan de una forma natural. Las actividades son "naturales" en el sentido de que no están planificadas ni manipuladas por el evaluador como ocurre en un experimento.

Este enfoque naturalista es apropiado cuando se parte de la idea de que los programas, servicios y actuaciones educativas son algo dinámico y en desarrollo, que se define continuamente en la práctica. En un programa las cosas se modifican con el tiempo, algo que también les ocurre a los participantes cuyas actitudes hacia el programa no se mantienen inalterables, como también varían la planificación y el desarrollo del propio programa. La naturaleza de estas variaciones no puede siempre predecirse o anticiparse. Una indagación naturalista puede ser, en este sentido, abierta y sensible a las derivaciones de los planes, a las variaciones no anticipadas, y a aspectos particulares que tienen su importancia en la puesta en práctica del programa.

Los evaluadores que utilizan la indagación naturalista se ocupan de captar los procesos de programa, en documentar las variaciones, y en explorar las diferencias individuales más importantes entre los resultados y las experiencias de los diversos participantes. La diferencia fundamental tiene lugar en términos de los grados relativos de manipulación calculada. Una estrategia de indagación naturalista se selecciona para describir impactos y procesos del programa que se expone de una forma natural. Los diseños experimentales se seleccionan para comprobar los efectos de los tratamientos controlados, reducir la variación en las variables extrañas, y para centrarse en un conjunto limitado de medidas predeterminadas.

Los diseños de evaluación cualitativa están particularmente orientados hacia la exploración, el descubrimiento, y la lógica inductiva. Un enfoque de evaluación es inductivo hasta el punto de que el evaluador intenta explicar la situación sin imponer expectativas preexistentes en el marco del programa, la institución o la actividad educativas.

La evaluación puede ser inductiva de dos formas. Dentro de los programas, instituciones o servicios, un

enfoque inductivo comienza con cuestiones acerca de las experiencias individuales de los participantes. Cuando se evalúa un programa, el enfoque inductivo busca características institucionales únicas que contextualicen cada caso. Cuando se evalúa una institución, un enfoque inductivo busca las culturas que conviven en ella y analiza las diferentes "instituciones" creadas por cada una de esas culturas. En cualquier nivel, las extrapolaciones pueden emerger cuando se analizan los materiales del caso, pero el interés inicial está en alcanzar una comprensión de los casos individuales antes de que éstos se combinen o sumen entre sí. Esto significa que los resultados de la evaluación se generan en contextos específicos.

Los diseños de evaluación cualitativa son holísticos. Los evaluadores se esfuerzan por comprender las situaciones y los programas como un todo. El evaluador busca la totalidad, es decir, la naturaleza que constituye una unidad de los contextos particulares. También asume que una descripción y comprensión del contexto social y político de un programa, institución o actividad es esencial para una comprensión en su conjunto de tales objetos de evaluación.

Un enfoque de indagación naturalista está plenamente justificado si consideramos que en la evaluación de programas, instituciones o actividades educativas, no siempre es posible aislar un objeto de evaluación sin el riesgo de proporcionar una comprensión o un juicio simplista. La ventaja de los retratos cualitativos como marcos holísticos es la atención detallada que se presta al matiz, el marco, las interdependencias, las complejidades, las idiosincrasias o el contexto.

2.1.1. Puntos de decisión en un diseño de evaluación cualitativa

En cualquier diseño de evaluación, el propio evaluador o el evaluador y los participantes en el proceso adoptan una serie de decisiones –a veces incluso de un modo implícito– que les lleva a un desarrollo evaluativo determinado. Los modelos apuntados en el Módulo 1 son un buen ejemplo de las diferentes perspectivas que pueden adoptarse en relación con aspectos como los criterios de evaluación, los términos de comparación, el rol del evaluador, el valor de verdad de los juicios emitidos, etc. No obstante, existen aún una serie de decisiones más cercanas al desarrollo de cada evaluación que se concretan con independencia del modelo o marco general elegido. Las decisiones tienen que ver básicamente con la identificación de las audiencias–identificación de las necesidades, la recogida de información, la interpretación valorativa o el establecimiento de juicios. Revisemos su significado y examinemos algunas de las ideas que pueden ayudarnos a decidir entre las diversas alternativas.

1. Identificación de audiencias–identificación de necesidades

La identificación del binomio audiencia–necesidad (A–N) no persigue otro fin que el de asociar dentro de un mismo esquema interpretativo a las audiencias afectadas por, o comprometidas con, la evaluación y a los intereses y necesidades que esas audiencias tienen respecto de la evaluación. Una evaluación puede estar condenada al fracaso o desviarse en su finalidad fundamental cuando damos prioridad a los intereses, necesidades y valores de una audiencia particular, ignorando o minimizando las aportaciones de las restantes audiencias interesadas en la evaluación.

Son, por tanto, las audiencias y sus necesidades las que determinan el objeto, si es que puede hablarse de un objeto de la evaluación. Este último, a menudo, puede llegar a convertirse en un constructo sólo conocido por el evaluador y sin el más mínimo interés para las audiencias y participantes en la evaluación. El evaluador cualitativo invierte, entonces, su esfuerzo en la identificación del binomio (A–N) como una clave para acercarse al modo que una determinada realidad educativa está siendo explicada y/o valorada por determinadas audiencias. Quizás el evaluador cualitativo no tenga que valorar un programa, sino muchos programas diferentes; quizás no tenga que evaluar un desarrollo institucional, sino muchos desarrollos institucionales, etc.

Algunas ideas que puedan ayudarnos a identificar las audiencias y sus necesidades serían:

- a) Identificar a los líderes de la evaluación (personas que la han promovido, solicitado o financiado) para que éstos nos ayuden a identificar a todas las audiencias interesadas o afectadas por la evaluación (fuerzas sociales y políticas, colectivos profesio-nales, etc.).
- b) Establecer contacto con personas que representan o son representativas de las diversas audiencias previamente identi-fi-cadas, y determinar: su idea de la evaluación, el valor que le conceden a la misma, sus necesidades e intereses, el uso que pretenden hacer de los resultados de la evaluación, el modo en que esos posibles resultados les afecta, etc.
- c) Tener en cuenta la posible existencia de audiencias adiciona-les, no consideradas por los líderes de la evaluación.
- d) Negociar con los líderes de la evaluación la importancia dada a las diferentes audiencias y, a partir de esa negociación, ordenar las necesidades que tratan de cubrirse en el proceso de evalua-ción.
- e) Diseñar la evaluación de acuerdo con las principales audien-cias identificadas y con sus preocupaciones fundamentales. En este sentido, es conveniente que el diseño sea sometido a revisión por parte de esas audiencias.
- f) Determinar cuál es la utilidad de la evaluación para las diferentes audiencias.
- g) Determinar cuál es, según las diferentes audiencias, el programa, proyecto, servicio, actividad, etc. que se pretende evaluar (García Jiménez, 1993).

La evaluación cualitativa suele apoyarse en la metodología de estudio de casos para integrar los diferentes aspectos del diseño. A veces, se considera a todo un centro o programa como un caso; en otras ocasiones, para evaluar un programa o actividad se consideran diferentes casos identificables con las distintas audiencias y contextos en que se desarrollan dicho programa o actividad. En esta última perspectiva mencionada se trabaja desde metodologías de estudios de casos múltiples. Todavía, en una tercera perspectiva, suelen emplearse metodologías mixtas (cualitativas–cuantitativas) dentro de propuestas de evaluación cualitativas, cuando las audiencias o el propio evaluador desean ejercer cierto control sobre determinados aspectos bien definidos de un programa, proyecto o institución educativas. Por tanto, es la propia identificación de audiencias–necesidades la que termina configurando en buena medida el tipo de metodología a emplear.

¿Por qué utiliza la evaluación cualitativa una metodología como la del estudio de casos?

Los estudios de caso, los más utilizados dentro de un enfoque de evaluación cualitativa, se convierten en particularmente útiles allí donde se necesita comprender con gran profundidad algún problema educativo concreto o una situación, y allí donde es posible identificar casos ricos en información; ricos en el sentido de que puede aprenderse mucho preguntando sólo a algunas personas o colectivos. Los estudios de caso únicos o múltiples son particularmente importantes cuando la evaluación persigue captar las diferencias individuales o las variaciones únicas existentes en un programa o proyecto desde un contexto a otro, o desde las experiencias de un programa a las de otro programa. Cuanto más persiga un programa u organización resultados individualizados o adaptados a las necesidades de ciertos colectivos, mayor será la adecuación de los métodos de caso cualitativos. Cuanto más enfatiza en un programa o institución la existencia de resultados comunes para todos los participantes, mayor puede ser la adecuación de medidas cuantitativas estandarizadas del cambio y del desarrollo.

2. La recogida de información

La recogida de información, dentro del proceso de evaluación, puede entenderse como una búsqueda de elementos explicativos (de argumentos) que permitan ilustrar una realidad educativa y/o que sean capaces de

fundamentar un juicio sobre el valor, el mérito y el éxito de un proyecto o programa. Esa búsqueda trataría de responder a las necesidades e intereses identificados en las diferentes audiencias.

La recogida de información en la evaluación cualitativa se apoya en el uso de procedimientos y técnicas variadas, si bien la idea central que se persigue en todas ellas es de acercarse a las diferentes perspectivas mantenidas por las audiencias. En este sentido, suelen utilizarse preferentemente procedimientos obstrusivos de observación y de encuesta tales como la observación no estructurada o la entrevista etnográfica. No obstante, algunos diseños de evaluación también incorporan otros procedimientos menos obstrusivos del tipo de cuestionarios, listas de control, sistemas de categorías, etc. Finalmente, también suele ser habitual el empleo de documentos personales como los diarios o biografías, autobiografías, historias de vida, etc.

El uso de un tipo de procedimiento de recogida de información determinado suele depender del tamaño de la audiencia a encuestar y del tipo de información que se desee recoger en cada caso: el cuestionario o la entrevista estructura, para audiencias numerosas y cuando deseamos recoger una información de tipo general, son más útiles que la entrevista; en cambio la entrevista etnográfica resulta más apropiada cuando deseamos profundizar en determinados argumentos, dentro de una audiencia relativamente restringida y previamente seleccionada por su importancia.

La utilización de un diseño de evaluación cualitativa no está reñido con el uso de criterios e incluso indicadores que orientan o guían de un modo más preciso la recogida de información. La cuestión no debe situarse en el plano en el que se oponga al diseño cualitativo (emergente, naturalista, etc.) un procedimiento que determine qué información debe recogerse y cómo debe valorarse, y en el que se identifican con claridad las dimensiones, criterios, indicadores, índices o descriptores de la evaluación. La cuestión fundamental no es si deben utilizarse o no dichos elementos sino quién los selecciona, por qué los elige, con qué grado de consenso cuenta dicho elección entre los implicados en la evaluación. La naturaleza del problema debe situarse en un plano de espacios de poder: quien establece los criterios es quien gobierna la evaluación.

Sea cual sea el procedimiento(s) y la técnica(s) de recogida de información seleccionadas y se utilicen o no criterios, indicadores, etc. , es conveniente:

- a) Mantener conversaciones informales con los representantes de las principales audiencias para acercarse a sus puntos de vista, intereses, necesidades, valoraciones ... del proyecto o programa.
- b) Actualizar los puntos de vista mantenidos por las audiencias manteniendo un contacto periódico con ellas.
- c) Revisar la información recogida de cada audiencia, de modo que pueda ser interpretada (e incluso modificada y mejorada) por esas mismas audiencias, a partir de sus representantes o personas representativas o a través de las personas que aportan la información más fundamental.

3. La interpretación valorativa

La evaluación conduce a un esquema interpretativo que o bien sólo permite ilustrar algo o bien implica además un juicio acerca del valor, mérito o éxito de algo. En la mayoría de los casos en los que evaluemos proyectos o programas educativos esa evaluación será, además, parte inseparable de una situación *pública*, es decir, que nos afecta como ciudadanos. No es sólo un asunto interpersonal sino colectivo.

Tanto para ilustrar o iluminar ciertos procesos como para emitir un juicio sobre un proyecto o programa educativo, el evaluador interpreta la información recogida hasta construir argumentos que hagan "creíble" sus valoraciones. Este proceso de interpretación se realiza en la evaluación cualitativa considerando la información recogida desde las diferentes audiencias y teniendo en cuenta los intereses, las necesidades e interpretaciones hechas desde esas audiencias, así como el contexto (limitaciones, condiciones ...) en que se desarrolla el proyecto o programa educativos. En este sentido, los resultados de la evaluación no generan

proposiciones necesarias y definitivas. Desde el momento en el que el evaluador tiene en cuenta el contexto social (educativo, económico, político ...) en que tienen lugar el desarrollo de un proyecto o programa y la misma evaluación, los resultados de la evaluación tienden a aparecer equívocos. De hecho, la validez de una interpretación valorativa debe juzgarse por el grado en que cada audiencia particular (cada interés particular) acepta los resultados evaluativos, es decir, los entiende creíbles, plausibles y probables.

En el proceso de interpretación, el evaluador puede construir su propio marco de referencia teórico y/o experiencial o compartir éste con los participantes y audiencias implicados en la evaluación. El análisis de la información recogida se convierte, entonces, en una tarea fundamental en el proceso de evaluación. La elaboración de los propios sistemas de codificación puede verse como competencia exclusiva del evaluador o cómo tarea en la que median las perspectivas de otras personas y colectivos.

Sea cual sea la opción que finalmente se seleccione en la interpretación de la información recogida es conveniente:

- a) Considerar los diferentes criterios desde los que es posible interpretar los resultados: técnicos y relativos al mismo proyecto o programa (objetivos, contenidos, metodología, materia-les,...), metas institucionales, ideales democráticos, necesidades sociales, expectativas de colectivos o grupos sociales y profesionales, etc.
- b) Determinar la posibilidad de que coexistan diversas interpretaciones valorativas: de los líderes de la evaluación, de los evaluadores, de las diversas audiencias, de grupos de expertos ...
- c) Utilizar procedimientos complementarios para determinar la significación de la información recogida: grupos de expertos externos al proceso de evaluación hasta ese momento, las propias audiencias, estudio Delphi, etc.

4. Informe de evaluación

Un típico informe de evaluación cualitativa proporciona, según Patton (1987:8-9) la siguiente información: (a) una descripción detallada de la puesta en práctica del programa; (b) un análisis de los procesos fundamentales del programa; (c) una descripción de los diferentes tipos de participantes y de los diferentes tipos de participación; (d) las descripciones del modo en que el programa ha afectado a los participantes; (e) los cambios observados (o la falta de ellos), resultados, e impactos; y, (f) el análisis de los puntos fuertes y débiles del programa según la opinión de las personas entrevistadas (e.d. participantes, el equipo de trabajo que ha realizado la evaluación, los que financian la evaluación, los informantes-clave en la comunidad).

No obstante las indicaciones de Patton pueden seguirse utilizando esquemas argumentativos bien diferentes. Entendemos, sin embargo que cualquier propuesta de evaluación cualitativa debe plantearse la redacción del informe de evaluación como un acto que vincule su elaboración y difusión con las participantes y audiencias interesadas en la evaluación (García Jiménez, 1993).

De una parte, un informe de evaluación debe aludir al programa o proyecto que ha sido objeto de evaluación y su contexto, así como a los procedimientos y resultados de la evaluación. Su contenido, por tanto, debe responder a la premisa de que las audiencias entiendan sin dificultad lo que se ha hecho, por qué se ha hecho, qué información se ha obtenido, cuáles son las valoraciones resultantes y que tipo de decisiones (en su caso) se recomienda adoptar. De otra parte, la evaluación debe difundirse a todas las audiencias interesadas inicialmente por el proceso de evaluación y a aquellas otras que directa o indirectamente se ven afectadas o han financiado esa evaluación. En definitiva, en la evaluación de prácticas sociales como la educación, toda la comunidad debe conocer lo que en último término ha financiado.

En relación con los dos elementos destacados, en la redacción de un informe de evaluación cualitativa deberían considerarse los siguientes elementos:

- a) Centrar el informe en los elementos más importantes del proceso de evaluación, construyendo el discurso argumentativo en torno a una idea o tesis central de la que puedan derivarse otras ideas. Utilizar referentes que ilustren esas ideas y su desarrollo.
- b) Construir la argumentación anterior considerando, cuando los haya, equívocos o ambigüedades.
- c) Invitar a los representantes o personas representativas de las audiencias a que revisen el informe. Es pertinente que las audiencias tengan acceso a los informes interinos que se elaboran durante el desarrollo de la evaluación.
- d) Preparar informes adecuados (formato, nivel de desarrollo técnico) a las necesidades manifestadas por las audiencias. No obstante, debe considerarse la posibilidad de que todas las audiencias tienen derecho a conocer el informe total.
- e) Mantener el contacto con las audiencias tras la redacción del informe. A ser posible, es recomendable utilizar para ello diferentes técnicas de comunicación que complementen al texto del informe (mesas redondas, comunicaciones, radio, prensa, etc...).
- f) Asegurar que los informes pueden llegar, al menos, a las audiencias interesadas o directamente afectadas.
- g) Asegurar que los contenidos del informe que afecten de un modo negativo (prácticas ilegales, consecuencias no deseables a nivel profesional ...) a ciertos colectivos o a toda la sociedad se proporcione a los interesados, aún a riesgo de romper los lazos contractuales con el organismo o colectivo que financia la evaluación.

- FASES DEL PROCESO DE EVALUACIÓN

2.2. diseños establecidos por el evaluador

3. ESTÁNDARES PARA LA EVALUACION DE PROGRAMAS

Recogemos en este apartado algunas de las normas o estándares que se han elaborado para regular y orientar la evaluación de programas educativos. Con ello, queremos sentar un término de comparación esencial para analizar la aplicación de los principios de evaluación que se han presentado en el apartado anterior e, igualmente, ofrecer una guía para ordenar el diseño de un proceso de evaluación. Asimismo, las normas que aquí se presentan podrían servir como vía de reflexión sobre el modo en que planificamos, desarrollamos y difundimos las evaluaciones educativas.

3.1. Una breve referencia histórica

El establecimiento de normas para orientar la evaluación de programas, proyectos y materiales educativos, establecidas por comités de expertos y por organizaciones profesionales, responde a una tradición asentada hace cuatro décadas en el campo de la medición educativa. En 1954 un comité de la American Psychological Association (A.P.A.) elaboran las primeras *Technical Recommendations for Psychological Tests and Diagnostic Techniques*, con el objeto de regular el uso y las aplicaciones de los tests estandarizados y otras técnicas utilizadas en el diagnóstico. Tan sólo un año después, comités de expertos de la American Educational Research Association (A.E.R.A), del National Council on Measurement Used in Education (N.C.M.E.) preparan las *Technical Recommendations for Achievement Tests*, para regular el uso y aplicaciones de los tests de rendimiento. Con ambos documentos se elabora en 1966[] una edición mixta bajo el título de *Standards for Educational and Psychological Tests and Manuals*, auspiciada por la AERA, la APA y el NCME, que será revisada en la edición de 1974.

En la década de los 60, y en el propio contexto norteamericano en el que surgen las normas sobre los procedimientos de medición educativa, al hasta entonces prácticamente único objeto de evaluación –el alumno– se unen, respondiendo a las exigencias legislativas y a las demandas sociales, nuevos objetos de evaluación: profesores, programas, proyectos y materiales educativos. En la ya citada edición de 1974 de normas para los tests y manuales educativos y psicológicos, se intenta responder a la realidad evaluativa creada incluyendo una sección dedicada a la evaluación de programas y a los instrumentos que puedan utilizarse en dicha evaluación.

No obstante, el propio comité encargado de la edición de 1974 se declara incompetente para ampliar su ámbito de actuación a la evaluación de programas. Las tres asociaciones implicadas en la elaboración de normas invitan a nuevas organizaciones (National Association of Elementary School Principals, National School Boards Association, American Federation of Teachers, etc.) para constituir en 1975 el Joint Committee on Standards for Educational Evaluation.

Tras varios años de trabajo conjunto surge en 1981 una primera publicación bajo el título *Standards for Evaluations of Educational Programs, Projects, and Materials*[]. El documento contiene tres normas diferentes que se presentan agrupadas en cuatro grupos y corresponden a los cuatro problemas que, según sus autores, presiden cualquier proceso de evaluación: su utilidad, factibilidad, legitimidad y precisión. Cada norma se explicaba y analizaba mediante un comentario que incluía una aproximación general al problema suscitado, algunas guías para aplicar la norma, los errores detectados con mayor frecuencia, algunas advertencias en el uso de las normas y un ejemplo que ilustraba su aplicación.

En la publicación de 1981 no se incluyó la evaluación de las instituciones educativas y del personal a ellas asociado (profesores, administradores, directores, etc.). Esto hizo que algunos años más tarde se abordase por el Joint Committee la elaboración de nuevos estándares. Así, en 1988 se publica el documento *The Personnel Evaluation Standards*[] que contiene 21 normas para valorar y planificar sistemas de evaluación de profesores, administradores, orientadores y otro personal vinculado a la educación.

El proceso para desarrollar estándares seguido por el Joint Committee, en el que participan no sólo expertos sino también usuarios y personas preocupadas por la calidad de la evaluación educativa, recibe en 1989 la acreditación del American National Standards Institute (ANSI) que convierte a las normas desarrolladas por dicho comité en normas de aplicación en todos los estados de la federación.

Ese mismo año, el Joint Committee decide someter a revisión las normas de 1981 con el objeto de ampliar el campo de la evaluación educativa desde del ámbito escolar a otros como el universitario, el legislativo, el médico, el de la enfermería, el militar, el de los negocios, el de la administración pública y el de las agencias de servicios sociales. De este modo, en 1994 se elabora una segunda edición de las normas bajo el título *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs*. La nueva edición vuelve a considerar treinta normas fundamentales que también se agrupan en torno a los cuatro problemas de la evaluación ya considerados en la edición de 1981: utilidad, factibilidad, legitimidad y precisión.

3.2. Las normas y su utilidad

Las normas son afirmaciones que responden al esquema "tal aspecto de la evaluación *debe ser*". Por ejemplo, la norma sobre la claridad del informe de evaluación se presenta del modo siguiente: "El informe de evaluación debe describir el objeto por evaluar y su contexto, propósitos, procedimientos y resultados de la evaluación, de tal forma que las audiencias puedan entender con rapidez lo que se ha hecho, por qué se ha hecho, qué información se obtuvo, a qué conclusiones se llegaron y qué recomendaciones se efectuaron" (Joint Committee, 1988:58).

Cabe entenderlas como intentos para orientar el diseño, el desarrollo y el análisis de las evaluaciones de programas educativos. Si bien incluyen advertencias sobre el modo de valorar la adecuación de las actividades

evaluativas, no presentan criterios específicos para realizar tales valoraciones. Las normas se han creado para estimular y facilitar el entendimiento entre clientes, evaluadores y patrocinadores y, en su caso, entre los propios miembros del equipo de evaluación. En general, con las normas se trata de guiar a los evaluadores para que recogan información que (a) sea relevante para responder a las cuestiones planteadas por los clientes y demás audiencias y (b) sea suficiente para juzgar la utilidad, los costos, el nivel de respuesta a las demandas sociales, la factibilidad y el mérito del programa juzgado (Joint Committee, 1994:4).

Las normas propuestas por el Joint Committee (1981,1994) están organizadas en torno a cuatro atributos considerados necesarios y suficientes para cualquier evaluación bien realizada y justa: utilidad, factibilidad, legitimidad y precisión. A continuación reproducimos estas cuatro categorías de normas según se recogen en las dos ediciones citadas.

Las normas de **utilidad** guían las evaluaciones para que éstas sean informativas, oportunas y capaces de influir. Estas normas requieren que los evaluadores se familiaricen con sus audiencias, se percaten de las necesidades de información de éstas, elaboren diseños de evaluación que respondan a tales necesidades y remitan de manera clara la información adecuada cuando sea necesaria.

Las normas que se incluyen en esta categoría tienen que ver con la identificación de las audiencias, la credibilidad del evaluador, la selección y el alcance de la información, la identificación de los valores, la claridad del informe, la difusión y periodicidad de los informes y las consecuencias de la evaluación. En definitiva, las normas de utilidad vienen a definir si una evaluación satisface las necesidades prácticas de información de una audiencia.

Las normas de **factibilidad** reconocen que las evaluaciones se desarrollan en un marco natural, no en un laboratorio, y que requieren una financiación importante. Por tanto, los diseños de evaluación deben ser operativos en tales contextos naturales y las evaluaciones no deben consumir más recursos, materiales, personal o tiempo que el necesario para responder a las cuestiones de evaluación.

Las tres normas que recoge esta categoría aluden a los procedimientos prácticos, a la viabilidad política y a la relación costo-beneficio. En suma, las normas de factibilidad hacen una llamada para que las evaluaciones sean realistas, prudentes, diplomáticas y económicas.

Las normas de **legitimidad** reflejan el hecho de que las evaluaciones afectan a muchas personas de diferente manera. Estas normas responden a la idea de facilitar la protección de los derechos de las personas afectadas por una evaluación. Las normas tratan de despertar la sensibilidad y de avisar contra las actuaciones ilegales, sin escrúpulos, carentes de ética e ineptas de quienes llevan a cabo las evaluaciones.

Estas normas requieren que las personas que realizan evaluaciones conozcan y obedezcan las leyes relativas a la privacidad, la libertad de información y la protección de los derechos humanos. Se refieren a aquellos que llevan a cabo las evaluaciones para que respeten los derechos de los demás.

Las ocho normas que se recogen en esta categoría aluden a la orientación del servicio, los acuerdos formales, los derechos de las personas, las interacciones humanas, la revelación de los resultados, una apreciación completa y justa, los conflictos de intereses y la responsabilidad fiscal[.].

Las normas de **precisión** determinan si una evaluación ha generado la información adecuada. La evaluación de un programa debe ser comprensiva, esto es, los evaluadores deberían considerar como parte del programa el mayor número posible de las características que lo definen y deberían recoger información sobre aquellas características particulares que consideren importantes para valorar el mérito o el valor del programa. Además, la información obtenida debe ser técnicamente adecuada y los juicios rendidos deben tener una relación lógica con los datos.

Las doce[] normas de esta categoría tienen que ver con la documentación del programa, el análisis del contexto, los propósitos y procedimientos observados, las fuentes de información válidas, la información válida, la información fiable, la información sistemática, el análisis de la información cuantitativa, el análisis de la información cualitativa, la fundamentación de las conclusiones, la información imparcial y la metaevaluación. Estas normas se formularon con la intención de asegurar que una evaluación mostrase y proporcionase información precisa sobre el mérito y/o valor del programa. (Joint Committee, 1981:39; 1994:5–6).Revisemos, a continuación, el texto de estas normas según la edición de 1994:

Normas de utilidad: Las normas de utilidad tienen el propósito de asegurar que una evaluación va a satisfacer las necesidades de información manifestadas por los usuarios. Estas normas son las siguientes:

Identificación de la audiencia: Debería identificarse a las las personas implicadas o afectadas por la evaluación para que sus necesidades puedan satisfacerse.

Credibilidad del evaluador: Las personas que llevan a cabo la evaluación deberían ser creíbles y competentes para desarrollarla, de modo que los resultados de la evaluación resulten lo más creíbles y aceptables que sea posible.

Selección y alcance de la información: La información recogida debería ser seleccionada con la amplitud suficiente para que pueda enfocarse hacia las preguntas pertinentes sobre el programa y responder a las necesidades e intereses de los clientes y demás audiencias identificadas.

Identificación de valores: Habría que describir cuidadosamente las perspectivas, los procedimientos y la racionalidad utilizada para interpretar los resultados, de modo que estén claras las bases sobre las que se establecen los juicios de valor.

Claridad del informe: Los informes de evaluación debería describir con claridad el programa que está siendo evaluado, incluyendo su contexto, los propósitos, procedimientos y resultados de la evaluación, para que pueda disponerse de información esencial y fácilmente comprensible.

Difusión y periodicidad del informe: Los resultados parciales significativos y los informes de evaluación se deberían difundir a los usuarios interesados, de suerte que pueden utilizarlos cuando sea necesario.

Consecuencias de la evaluación: Las evaluaciones deberían planificarse, llevarse a cabo y difundirse siguiendo los continuos requerimientos de las audiencias, de esto modo aumentaría la probabilidad de que las evaluaciones resultasen útiles. (pp. 23–24).

Normas de factibilidad: Tienen el propósito de asegurar que una evaluación es realista, prudente, diplomática y frugal. Las normas son las siguientes:

Procedimientos prácticos: Los procedimientos de evaluación debería ser prácticos, de forma que mientras la información se recoge haya un mínimo de interferencia en el desarrollo del programa.

Viabilidad política: La evaluación debería planificarse y realizarse teniendo en cuenta las diferentes posiciones mantenidas por los distintos grupos de intereses, de esto modo se aseguraría su cooperación, y sería posible anticipar, contrarrestar o evitar los posibles intentos que se realicen para restringir, sesgar o aplicar equivocadamente los resultados.

Relación coste–beneficio: La evaluación debería ser eficiente y producir información valiosa para justificar los recursos empleados. (p. 63)

Normas de legitimidad: Las normas de legitimidad tienen el propósito de asegurar que una evaluación se

llevará a cabo legal y éticamente, y en beneficio de quienes están implicados en la evaluación, así como de aquéllos que están afectados por sus resultados.

Orientación del servicio: Las evaluaciones deberían diseñarse para ayudar a las organizaciones a dirigirse y prestar un servicio efectivo en función de las necesidades de los participantes a los que están destinadas.

Acuerdos formales: Las obligaciones de las partes formales respecto de la evaluación (lo que debe hacerse, cómo, por quién, cuando) deberían acordarse por escrito, para que esas partes estén obligadas a cumplir todas las condiciones del acuerdo o a renegociarlas formalmente.

Derechos de las personas: Las evaluaciones deberían diseñarse y llevarse a cabo respetando y protegiendo los derechos y el bienestar de las personas.

Interacciones humanas: Los evaluadores deberían respetar la valía y la dignidad humana en sus interacciones con otras personas vinculadas con la evaluación, de modo que los participantes no se vean amenazados o perjudicados.

Apreciación completa y justa: La evaluación debería ser completa y justa en su examen y registro de las debilidades y fortalezas del programa evaluado, de modo que pueda avanzarse a partir los puntos fuertes del programa e identificarse las áreas problemáticas.

Revelación de los resultados: Las partes formales de una evaluación deberían asegurar que todos los resultados de la evaluación, junto con las limitaciones pertinentes, resulten accesibles a las personas afectadas por la evaluación y cualesquiera otras con derechos legales reconocidos para recibir los resultados.

Conflicto de intereses: Los conflictos de intereses deberían enfrentarse con apertura y honestidad, de modo que no comprometan al proceso y los resultados de la evaluación.

Responsabilidad fiscal: La asignación del evaluador y el empleo de los recursos deberían reflejarse convenientemente a través de procedimientos de control de cuentas y, por otra parte, deberían ser prudentes y éticamente responsables, de modo que los gastos se justifiquen y resulten apropiados. (pp. 81–82).

Normas de precisión: Las normas de precisión se han promovido para asegurar que una evaluación presente y proporcione información técnicamente adecuada sobre los rasgos que determinan el valor o mérito del programa evaluado. Estas normas son las siguientes:

Documentación del programa: El programa que está siendo evaluado debería describirse y documentarse de forma clara y precisa, para que pueda ser identificado con claridad.

Análisis del contexto: Debería examinarse con todo detalle el contexto en que se desarrolla el programa, de modo que puedan identificarse las posibles influencias sobre el programa.

Propósitos y procedimientos observados: Los propósitos y los procedimientos de evaluación deberían describirse y controlarse con bastante detalle, de forma que puedan identificarse y valorarse.

Fuentes de información válidas: Las fuentes de información utilizadas en la evaluación de un programa deberían describirse con todo detalle, para que pueda apreciarse la validez de dicha información.

Información válida: Los procedimientos de recogida de la información deberían elegirse o elaborarse y, más tarde, aplicarse de modo que pudiesen asegurar que la interpretación consiguiente resultase válida para el uso pretendido.

Información fiable: Los procedimientos de recogida de la información deberían elegirse o elaborarse y, más tarde, aplicarse de modo que pudiesen asegurar que la información obtenida resultase fiable para el uso pretendido.

Información sistemática: La información recogida, analizada y difundida en una evaluación debería ser revisada sistemáticamente y corregidos cualesquiera posibles errores detectados.

Análisis de la información cuantitativa: La información cuantitativa debería analizarse de forma sistemática y apropiada para que se pudiera responder de forma efectiva a las cuestiones de evaluación.

Análisis de la información cualitativa: La información cualitativa debería analizarse de forma sistemática y apropiada para que se pudiera responder de forma efectiva a las cuestiones de evaluación.

Fundamentación de las conclusiones: Las conclusiones alcanzadas en una evaluación deberían ser justificadas de forma explícita, para que las audiencias puedan valorarlas.

Información imparcial: Los procedimientos para informar deberían salvaguardar contra la distorsión originada por sesgos de cualquier parte y por los sentimientos personales, de modo que los informes de evaluación reflejen adecuadamente los resultados de la evaluación.

Metaevaluación: La evaluación en sí misma debería ser evaluada formativa y sumativamente con estas y otras normas que se consideren pertinentes, de modo que su realización pueda orientarse de un modo apropiado y, en cuanto se termine, las audiencias puedan examinar rigurosamente sus debilidades y fortalezas.

Un simple recuento de las normas revisadas nos muestra que el cuarenta por ciento de ellas, 12 sobre 30, son normas de precisión. Una primera interpretación de este dato nos lleva a considerar seriamente el peso histórico desempeñado por las anteriores ediciones de normas centradas en la medición educativa. Muchos de los miembros del Comité que elaboraron las normas para la evaluación de programas también forman parte de los comités que elaboran normas para tests y pruebas de rendimiento. Un análisis más detenido de las normas sobre evaluación de programas, proyectos y materiales revela que éstas, siendo un instrumento de consulta obligada y de orientación para cualquier evaluador educativo, han omitido algunos temas de interés para evaluadores y audiencias. Es el caso de las normas relativas a los referentes con los que se suelen comparar los programas durante el proceso de evaluación, a los criterios de evaluación que se utilizan, o a los indicadores que suelen emplearse para determinar el mérito o valor de un programa. Algunas de posibles respuestas a estos temas se encuentran precisadas en los modelos de evaluación que revisaremos más adelante.

BIBLIOGRAFÍA SOBRE MODELOS DE EVALUACIÓN CUALITATIVA

1. Crítica artística:

Eisner, E. W. (1985). The art of educational evaluation. A personal view. Londres: The Palmer Press.

Eisner, E. W. (1979). The educational imagination. On the design and evaluation of school programs. Nueva York: Macmillan.

2. Evaluación iluminativa:

Maxwell, G. S. (1984). A rating scale for assessing the quality of responsive/iluminative evaluations. Educational Evaluation and Policy Analysis, 6(2), 131–138.

Parlett, M., & Dearden, G. (Eds.) (1981). Introduction to illuminative evaluation: Studies in higher education, Guilford, Inglaterra: Society for Research into Higher Education. University of Surrey.

Parlett, M., & Hamilton, D. (1983). La evaluación como iluminación. En J. Gimeno Sacristán, & A. Pérez Gómez, La enseñanza: su teoría y su práctica, (pp. 450–466). Madrid: Akal.

3. Evaluación respondiente:

Stake, R. E. (1973). The countenance of educational evaluation. En B. R. Worthen, & J. R. Sanders (Eds.), Educational evaluation: Theory and practice, (pp. 106–128). Belmont, California: Wadsworth.

Stake, R. E. (1983). La evaluación de programas: En especial la evaluación de réplica. En W. B. Dockrell, & D. Hamilton, Nuevas reflexiones sobre la investigación educativa, (pp. 91– 108). Madrid: Narcea.

4. Evaluación democrática:

MacDonald, B. (1983). La evaluación y el control de la educación. En J. Gimeno Sacristán, & A. Pérez Gómez, La enseñanza: su teoría y su práctica, (pp. 467–478). Madrid: Akal.

MacDonald, B. y Stronach, I. (1989). The independent policy evaluation, Journal of Computer Assisted Learning, 5, 51.64.

5. Evaluación respondiente constructivista:

Guba, E.G. y Lincoln, Y.S. (1981). Effective evaluation. Improving the usefulness of evaluation results through responsive and naturalistic approaches. San Francisco: Jossey–Bass Pub.

Guba, E.G. y Lincoln, Y.S. (1989). Fourth generation evaluation. Londres: Sage.

American Psychological Association (1976). *Normas sobre tests y manuales educativos y psicológicos*. Madrid: TEA

La edición en español: Joint Committee on Standards for Educational Evaluation (1988). *Normas de evaluación para programas, proyectos y material educativo*. México: Trillas.

Joint Committee on Standards for Educational Evaluation (1988). *The personnel evaluation standards*. Newbury Park, CA: Sage.

Existen algunas variaciones entre las normas recogidas en la edición de 1981 y las de 1994.

Eran once en la edición de 1981 y no responden todas ellas a las formuladas en 1994.

4 Tema 2. Est