

## CAPÍTULO IV

### MODELOS REPRESENTATIVOS DE LA EVALUACIÓN EDUCATIVA

---

Existen una gran variedad de clasificaciones en las que se acoplan, más o menos holgadamente, los diferentes modelos o diseños de evaluación (Castillo y Gento, 1995; Cronbach, 1983; Doctors y Wokutch, 1980; Gardner, 1977; Guba y Lincoln, 1985; Guba, 1978; Guttentag, 1971; House, 1986, a; Lawton, 1980; Popham, 1975, a; Rafter, 1984; Rossi y Cronbach, 1976; Rossi y Freeman, 1986; Sandefur, Freeman y Rossi, 1986; Stufflebeam y Shinkfield, 1987; Stufflebeam y Webster, 1981; Timpane, 1982; Wortman, 1983)<sup>1</sup>. Sólo hay un rasgo común, previo a la presentación de estos “catálogos”, en las manifestaciones de sus autores: que ninguna de las categorías creadas es excluyente respecto de las demás y, en consecuencia, cualquier modelo podría encuadrarse simultáneamente en dos, o más, de ellas. Desde esta perspectiva, no nos ha parecido conveniente entrar en ejercicios de elu-

---

<sup>1</sup> Destacamos entre las más relevantes, sin ningún ánimo de ser exhaustivos, las siguientes: Castillo y Gento (1995): cuantitativos, cualitativos y mixtos; Doctors y Wokutch (1980): auditoría de procesos sociales, diseños experimentales y cuasi-experimentales, libre de objetivos, de sistemas, análisis costo-beneficio y responsabilidad; Gardner (1977): mediciones, juicio profesional, congruencia entre actuación y objetivos, toma de decisiones y libre de objetivos; Guttentag (1971): modelo experimental, jurídico y contextual; House (1986, a): análisis de sistemas, por objetivos conductuales, toma de decisiones, libre de objetivos, crítica artística, acreditación, por adversarios y transaccional; Lawton (1980): clásico, industrial, iluminativo, toma de decisiones, profesor investigador y estudio de casos; Popham (1975, a): consecución de metas, de enjuiciamiento que acentúan criterios intrínsecos, de enjuiciamiento que acentúan criterios extrínsecos y de facilitación de decisiones; Rafter (1984): científicos, interactivos e híbridos; Stufflebeam y Shinkfield (1987): falsas evaluaciones, cuasievaluaciones y verdaderas evaluaciones; Stufflebeam y Webster (1981): evaluaciones orientadas políticamente, orientadas a problemas específicos y orientadas a juzgar; Timpane (1982): experimentales, eclécticas, descriptivas y análisis costo-beneficio; Wortman (1983): experimentación social, metaanálisis, análisis costo-beneficio; etc.

cubración exagerados y hemos tratado de simplificar la cuestión al máximo. Para ello exponemos aquí dos grandes categorías en las que tan sólo hemos querido señalar una muestra de los existentes. El primero de estos grupos lo denominamos "modelos clásicos". Casi todos los modelos pertenecientes al mismo, en su mayoría, estarían encuadrados en la órbita empírico-racionalista y conductista. Manejan datos numéricos y los analizan, usualmente, a través de procesos estadísticos.

La segunda categoría, que hemos denominado "modelos alternativos" está formada por todos aquellos modelos que han aparecido, generalmente, como reacción a los de la categoría anterior. Casi ninguno maneja datos numéricos y/o responden a otro tipo de análisis y enfoque sobre la realidad evaluada.

Para el análisis que hemos efectuado, hemos seguido mayoritariamente las descripciones expuestas por Popham, en su obra publicada en español en 1980, *Problemas y técnicas de la evaluación educativa*, que hemos considerado de un gran interés debido a la imparcialidad y claridad de sus apreciaciones.

#### 1. MODELOS CLÁSICOS

- A) Modelos basados en la consecución de metas u objetivos.
- B) Modelos basados en la formulación de juicios.
- C) Modelos facilitadores para la adopción de decisiones.

#### 2. MODELOS ALTERNATIVOS:

- A) Evaluación como crítica artística.
- B) Evaluación libre de metas.
- C) Método del Modo Operandi.
- D) Evaluación comunicativa.
- E) Evaluación iluminativa.
- F) Evaluación respondiente.

## 1. MODELOS CLÁSICOS

### 1.1. Modelos basados en la consecución de metas u objetivos

Dentro de esta categoría podríamos establecer diferentes grupos, si bien nos ha parecido relevante no hacer nuevas subdivisiones que, sin duda, contribuirían a aumentar su complejidad y no arrojarían una mayor claridad en su comprensión. Señalamos unos cuantos modelos, eminentemente prácticos, que tratan de responder a las demandas de los profesores de acuerdo a sus necesidades.

#### 1.1.1. Modelo de Tyler

Comenzamos, siguiendo a Popham (op. cit., págs. 32 y 33), analizando el modelo de evaluación propuesto por Tyler. A este evaluador se deben muchos de

los enfoques de evaluación que durante años se han extendido, e influido, entre la mayoría de los educadores y evaluadores. Tal vez sea él uno de los primeros en establecer, y defender, la necesidad de una definición de los objetivos en términos de conducta -operativos- como el elemento central en el que se debe basar la evaluación. Tan grande ha sido su influencia que de él arranca toda una tradición (llamada “tyleriana”), que con el apoyo y desarrollo del conductismo, se ha instalado fuertemente tanto en los sistemas de enseñanza como en las mentes de muchos profesores. Su propio concepto de educación como proceso complejo e intrincado que va desde la formulación de objetivos, tan claramente como sea posible, hasta la asunción de decisiones sobre los cambios y mejoras necesarios en el curriculum, confiere a la evaluación un carácter dinamizador de los procesos de aprendizaje.

Las actividades básicas de un programa de evaluación serían: la identificación de los objetivos; la determinación de las oportunas actividades y experiencias de aprendizaje en función del conocimiento de los alumnos; y, por último, la evaluación del grado en que han sido alcanzados los objetivos. Para Tyler los objetivos han de formularse a partir del análisis de lo que considera “fuentes de las metas educativas” -el alumno, la sociedad y la naturaleza de los contenidos- y de la filosofía y psicología de la educación que las engloban. Véase la figura 9.

Los objetivos educativos representan los “criterios” en los que se ha de fundamentar la evaluación. La evaluación basada en “normas” es considerada como inadecuada e insuficiente, toda vez que la basada en estos criterios refleja un cono-

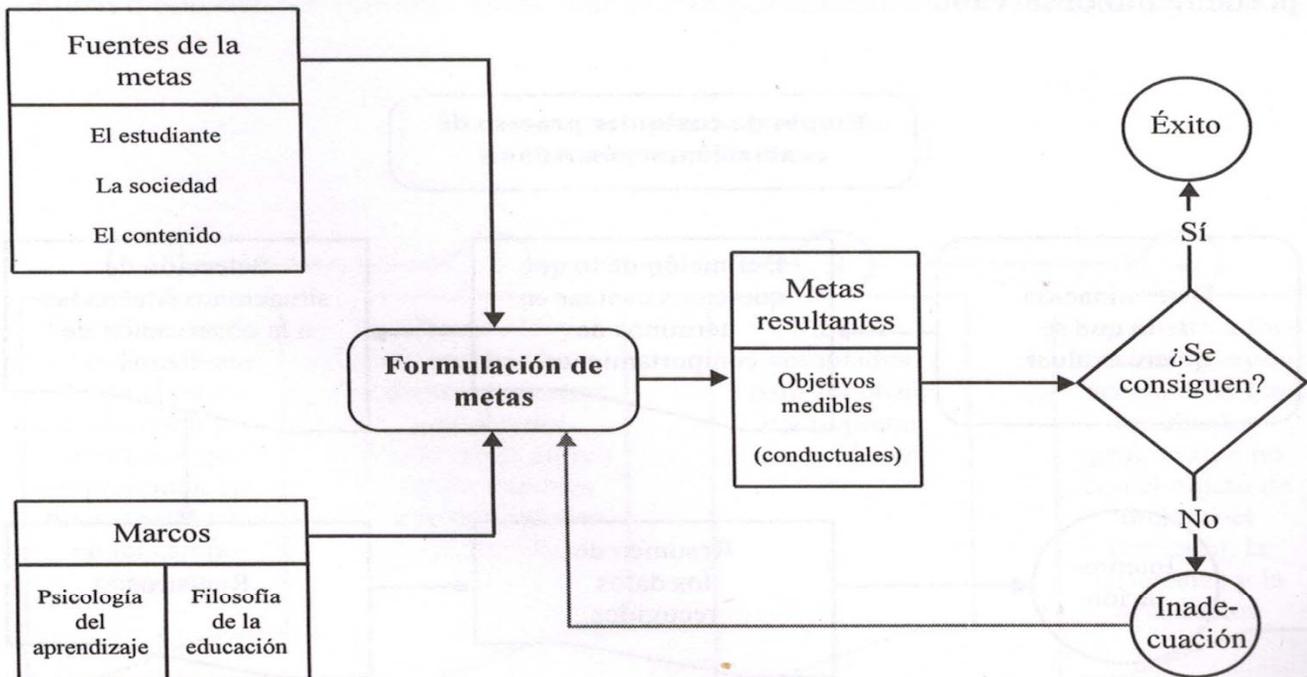


Figura 9: Enfoque de consecución de metas de Tyler.

cimiento mucho más preciso y objetivo acerca del rendimiento escolar, y posibilita la emisión de informes de carácter diagnóstico frente a las simples calificaciones, que a lo sumo comparan ordinalmente a los alumnos. Todos estos principios fueron ya reflejados en su famoso “estudio experimental de ocho años” (*Eight Year Study*).

### 1.1.2. Modelo de evaluación de Adams

Todos los modelos propuestos de evaluación, en cuanto procesos que son, se desarrollan siguiendo una secuencia de fases o estadios. Estas etapas poseen algunas diferencias relevantes de unos a otros autores, e incluso en ocasiones cuando se trata del mismo autor. A continuación vamos a exponer algunas de estas fases que debe encerrar cualquier proceso evaluativo. En primer lugar exponemos la secuencia expresada por Adams en su modelo de evaluación (1970, págs. 21 y 22). Véase la figura 10.

Para éste autor, aunque el proceso de evaluación está fundamentado básicamente en las mediciones, como ocurre con el uso de los tests, las etapas son las mismas cuando el proceso se centra en las observaciones del comportamiento.

En la primera etapa, habrá de tenerse en consideración el tipo de información a obtener, dado que si no fuera significativa afectaría a los juicios que pudiéramos emitir sobre el objeto que se desea valorar.

Para la segunda se propone la operativización del objeto de evaluación, en términos tales que permitan su medición. Normalmente se hará en forma de comportamiento observable.

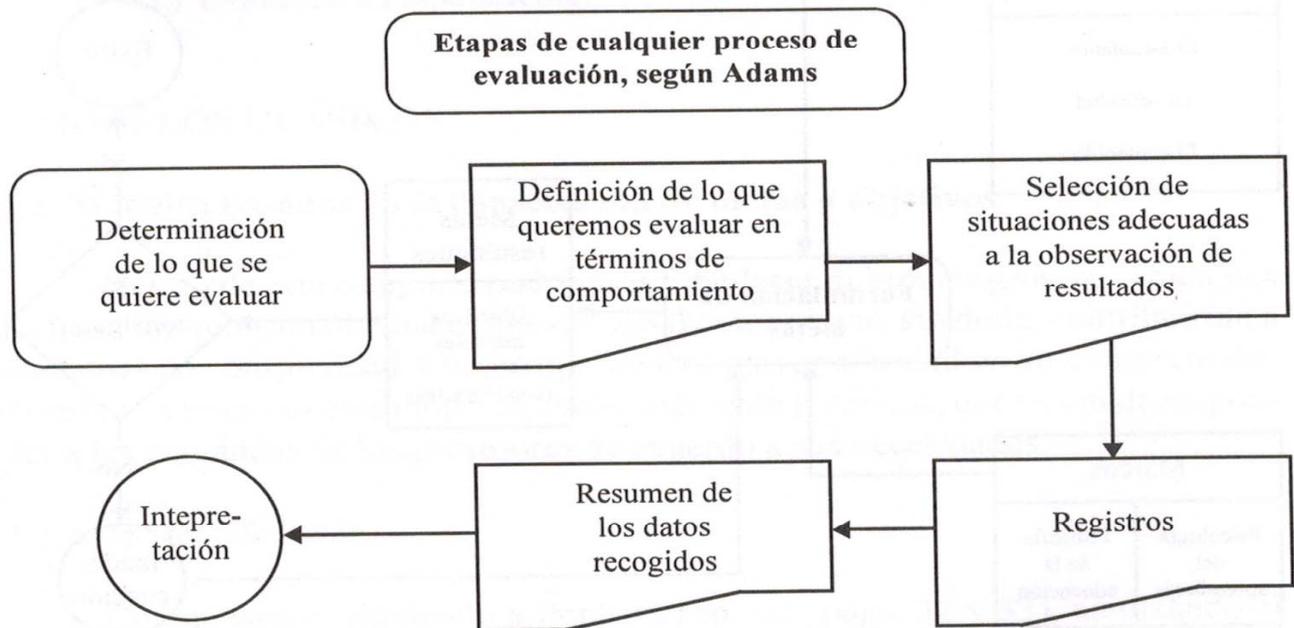


Figura 10.

La tercera etapa se centraría en seleccionar qué situaciones permiten la adecuada observación de los comportamientos anteriormente definidos. Estas situaciones bien pudieran ser las naturales del aula, o recreos, o la emisión de respuestas a determinadas preguntas, lo suficientemente “amplias y representativas como para que las inferencias basadas en los resultados del alumno sobre la muestra suministren un índice medianamente preciso de su nivel habitual” (Pág. 21)

La cuarta etapa implica la recogida de los datos, para lo que habría de determinarse qué aspectos son susceptibles de puntuación, cuáles las unidades de medida a utilizar, etc. Además se deben utilizar aquellos instrumentos de recogida de información más adecuados a la naturaleza de los propios datos: tests, grabaciones en vídeo, grabaciones en audio, observaciones sistematizadas de acuerdo a determinadas plantillas en unidades de tiempo, como las sugeridas por Flanders (1970), etc.

A esta etapa, continúa la síntesis de todos los datos obtenidos para permitir una adecuada interpretación de los mismos. Con esto finalizaría el proceso.

### 1.1.3. Modelo de evaluación de Taba

Nos vamos a fijar a continuación en otra de las secuencias propuestas para los procesos de evaluación. En este caso las sugeridas por Hilda Taba en su modelo (1983, págs. 410 y ss.). Véanse las figuras 11 y 12.

Para esta autora la educación es concebida como el proceso cuyo principal objetivo consiste en intentar modificar el comportamiento de los alumnos y, conse-

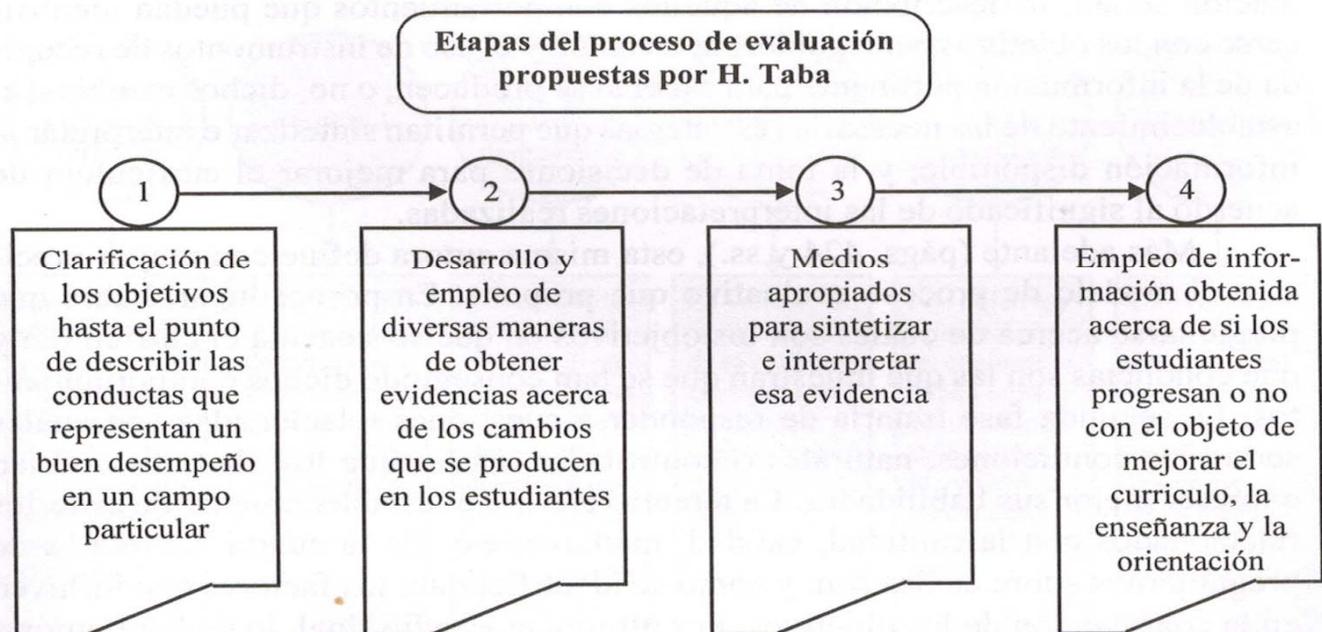
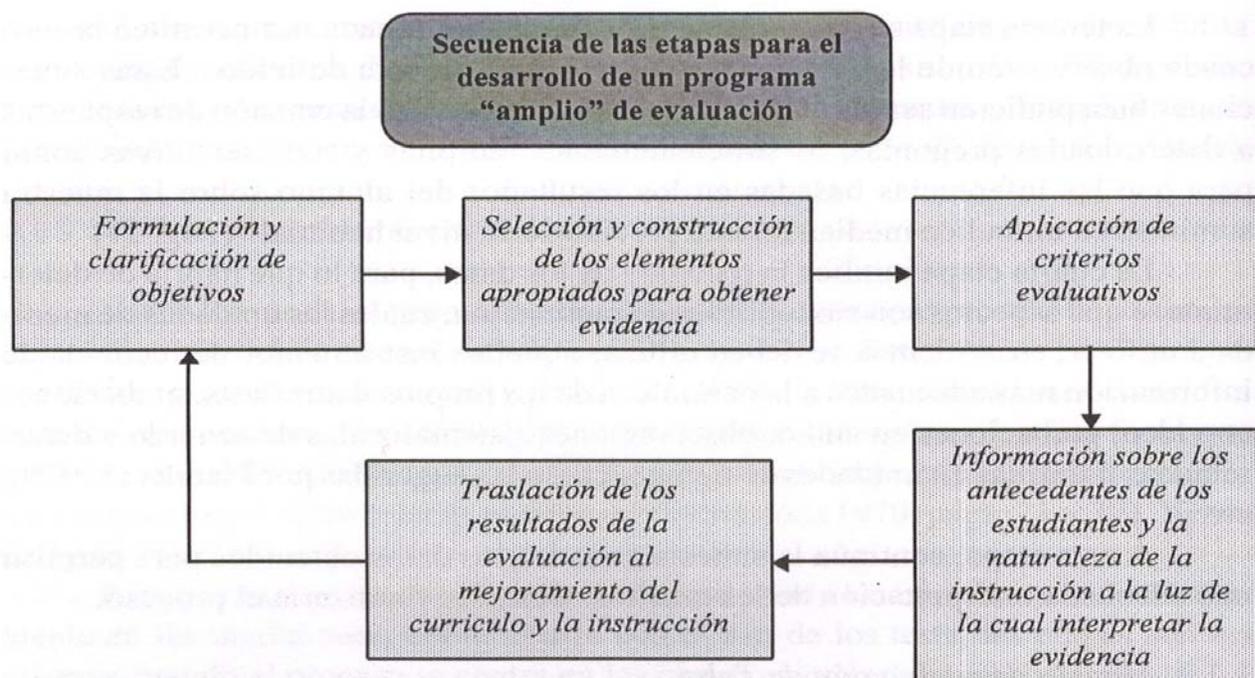


Figura 11.



**Figura 12.**

cuentemente, el proceso de evaluación consistirá en la determinación, y estimación, de los cambios producidos en relación a esos mismos objetivos preestablecidos, a través de determinados instrumentos: tests, observaciones, análisis de producciones, etc.

Por ello, los diferentes estadios que habrán de seguir los programas de evaluación serían: la descripción de aquellos comportamientos que puedan identificarse con los objetivos perseguidos; la creación y el uso de instrumentos de recogida de la información pertinente para saber si se producen, o no, dichos cambios; el establecimiento de las necesarias estrategias que permitan sintetizar e interpretar la información disponible; y la toma de decisiones para mejorar el curriculum de acuerdo al significado de las interpretaciones realizadas.

Más adelante (págs. 424 y ss.), esta misma autora define con mayor precisión el modelo de proceso evaluativo que propone. En primer lugar, habrá que preguntarse acerca de cuáles son los objetivos en que se sustenta el curriculum y qué conductas son las que muestran que se han conseguido dichos comportamientos. La segunda fase trataría de responder a cuestiones relacionadas con cuáles serían las condiciones, naturales o controladas, en las que los alumnos podrían expresar mejor sus habilidades. La tercera referiría al establecimiento de criterios relacionados con la cantidad, calidad, madurez, etc. En la cuarta habríamos de preguntarnos sobre cuáles son, y cómo se identificarían, los factores que influyen en la consecución de los objetivos. Por último, el estadio final, lo dedicaríamos a valorar las implicaciones que de los resultados pudiéramos extraer tanto respecto

de los objetivos planteados, como de la enseñanza en general. Apréciense cuanto llevamos manifestado en la figura 12.

Pasamos a continuación a enumerar las condiciones que según esta misma autora (págs. 414 y ss.) deben de reunir los programas de evaluación, véase la figura 13. El primer criterio, “compatibilidad de los objetivos” hace referencia a la congruencia que debe existir entre los planteamientos generales que inspiran el curriculum y aquellos sobre los que se fundamenta el programa de evaluación. Si existe una divergencia entre ambos, la evaluación perderá validez y fiabilidad, toda vez que estaremos valorando aspectos distintos de los pretendidos. Así ocurriría si en un sistema de enseñanza orientado, por ejemplo, hacia los “procesos”, llevamos a cabo una evaluación basada en los “logros” (como lo propuesto en nuestro “decreto de mínimos”). Igualmente suele ser muy frecuente este problema de “incompatibilidad”, cuando se traslada un modelo de evaluación que nos ha parecido muy interesante, a un sistema para el que no fue ideado. A menudo las divergencias que inspira a uno y a otro, harán fracasar la fiabilidad y validez de nuestras valoraciones. También puede ocurrir cuando los objetivos pretendidos no están lo suficientemente explicitados y claros como para que no originen ambigüedades con los considerados “implícitos” en nuestras prácticas pedagógicas.

La “amplitud” refiere a si los programas de evaluación cubren todo el dominio o espectro formulado en los objetivos educativos. Es fácil observar como para

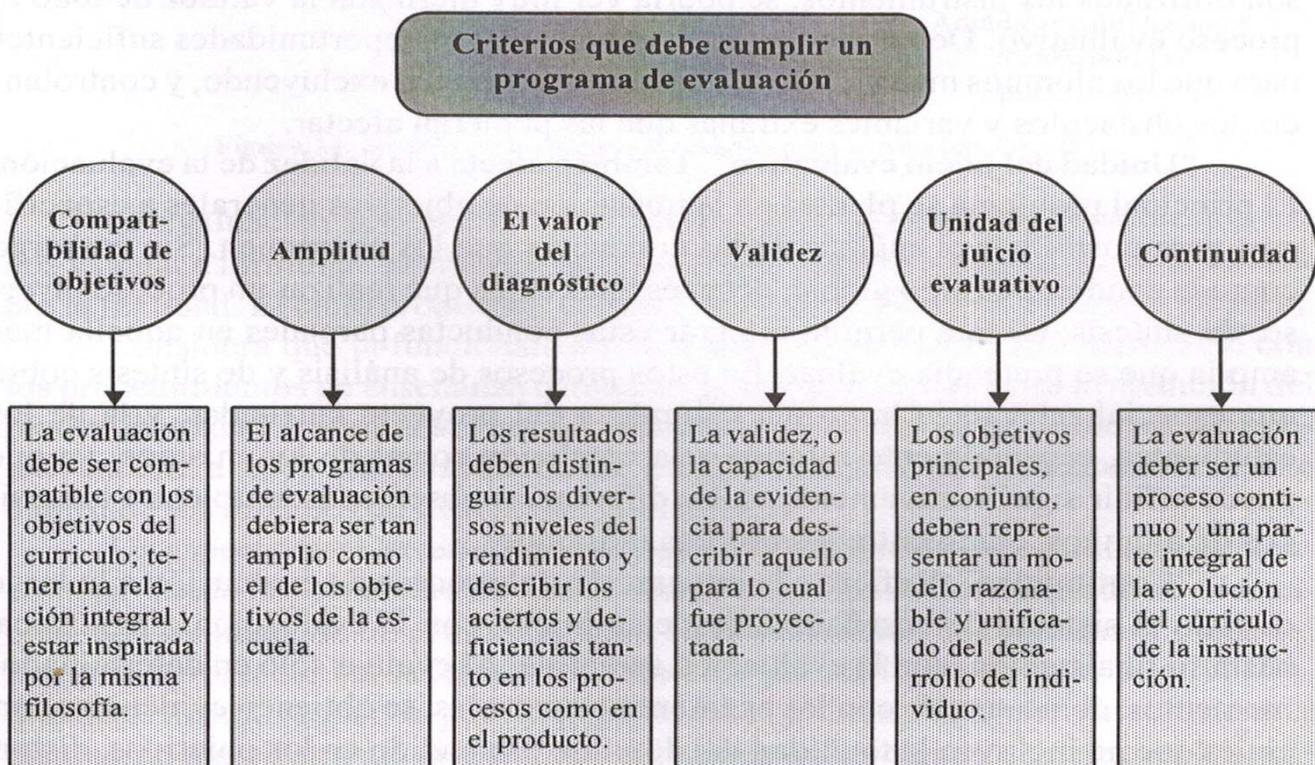


Figura 13.

aspectos concretos existen una gran cantidad de instrumentos que permiten su evaluación. Sin embargo no sucede con todos. Así, determinadas capacidades como las actitudes o la creatividad, por sólo poner dos ejemplos, resultan bastante difíciles de valorar, a lo que tendríamos que añadir la dificultad para poder encontrar las herramientas apropiadas. Eso hace, según Taba, que la enseñanza acabe orientándose preferentemente hacia lo que sí puede ser evaluado, en detrimento de aquellos otros aspectos que aun siendo valiosos, se descuidan en la evaluación.

“El valor del diagnóstico”. La evaluación debe facilitar la diferenciación no sólo de los niveles de rendimiento, sino también la distinción entre aquellos procesos que llevan tanto a la provocación de los errores como a la de los aciertos. De este modo la evaluación cobra una dimensión relevante al permitir valorar los niveles de comprensión sobre los que actúa el alumno y la capacidad y profundidad con la que realiza transferencias de sus aprendizajes.

La “validez”. Es directamente proporcional a la correspondencia entre los objetivos del curriculum y los instrumentos de evaluación. Razón por la que los objetivos deben estar claramente definidos. Su ambigüedad aumentaría la falta de coherencia y por ende, afectaría negativamente a la validez. Cuanto mayor sea la congruencia entre metas y utensilios, más validez tendrá el proceso de evaluación, y viceversa, la validez de la evaluación se verá seriamente afectada a medida que sea mayor la disparidad entre los fines y las herramientas utilizadas para su medición.

Además, si no se tienen presentes las situaciones y los momentos en los que son utilizados los instrumentos, se podría ver muy mermada la validez de todo el proceso evaluativo. De tal modo que se deben crear las oportunidades suficientes para que los alumnos manifiesten la conducta pretendida, excluyendo, y controlando, los obstáculos y variables extrañas que las pudieran afectar.

“Unidad del juicio evaluativo”. También afecta a la validez de la evaluación. El principal problema se plantea en la traslación de objetivos generales a específicos, para posibilitar el análisis de las conductas que los componen. Sin embargo como la conducta es algo global, después tendremos que realizar un proceso inverso, de síntesis, tal que permita integrar estas conductas parciales en aquella más amplia que se pretendía evaluar. En estos procesos de análisis y de síntesis cobra una especial relevancia la propia naturaleza del proyecto curricular, y la de los estudiantes, especialmente referida a la condición social de los mismos, ya que puede influir significativamente en las diferencias observadas tanto entre los grupos de alumnos, como entre los propios individuos.

“Continuidad”. Refiere a la integración del proceso de evaluación a lo largo de todo el sistema de enseñanza, desde su inicio, con la evaluación diagnóstica, hasta su finalización, con la evaluación sumativa. Al evaluar sólo en determinados momentos, como sucede con los exámenes o controles, se obtienen evidencias parciales que no incluyen la totalidad del dominio pretendido en los objetivos, distorsionando los resultados y disminuyendo la validez y eficacia de la evaluación.

#### 1.1.4. Modelo de evaluación de Carreño

Para este autor (1985, págs. 27 y ss.), los procesos de enseñanza-aprendizaje se dividen en tres fases muy estrechamente relacionadas entre sí, que son las que confieren coherencia y cohesión a su desarrollo: la planificación, la ejecución y la evaluación. Las “operaciones evaluativas” permiten que las otras fases puedan ser mejoradas, perfeccionándolas, ajustándolas a los cambios según se vayan éstos produciendo, corrigiendo los posibles errores e incidiendo en los aciertos.

En el proceso de evaluación se deben conjugar tanto los datos obtenidos a través de las mediciones como los recabados de las observaciones, según puede apreciarse en la figura 14. Los primeros otorgarán una mayor objetividad, así como la facilidad que entraña el manejo de datos numéricos para su análisis y para la posterior emisión de calificaciones. Los derivados de las observaciones, eminentemente cualitativos, servirán para la evaluación; aunque representarán mayores dificultades para calificar.

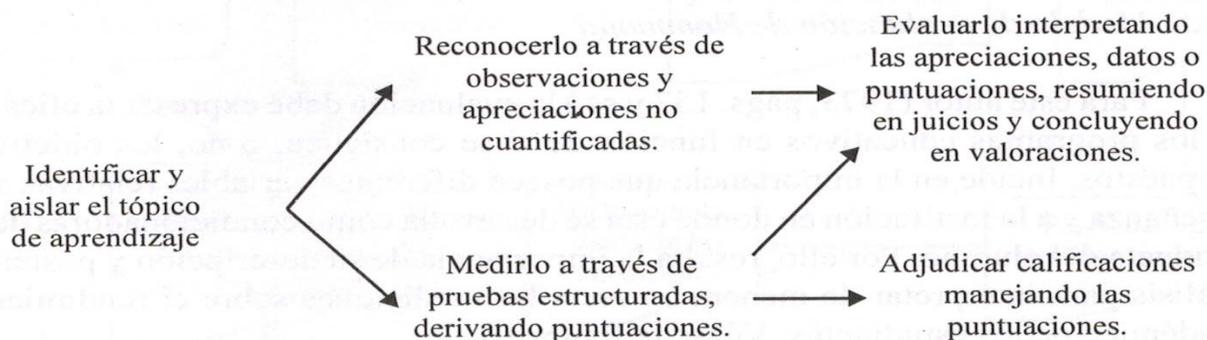
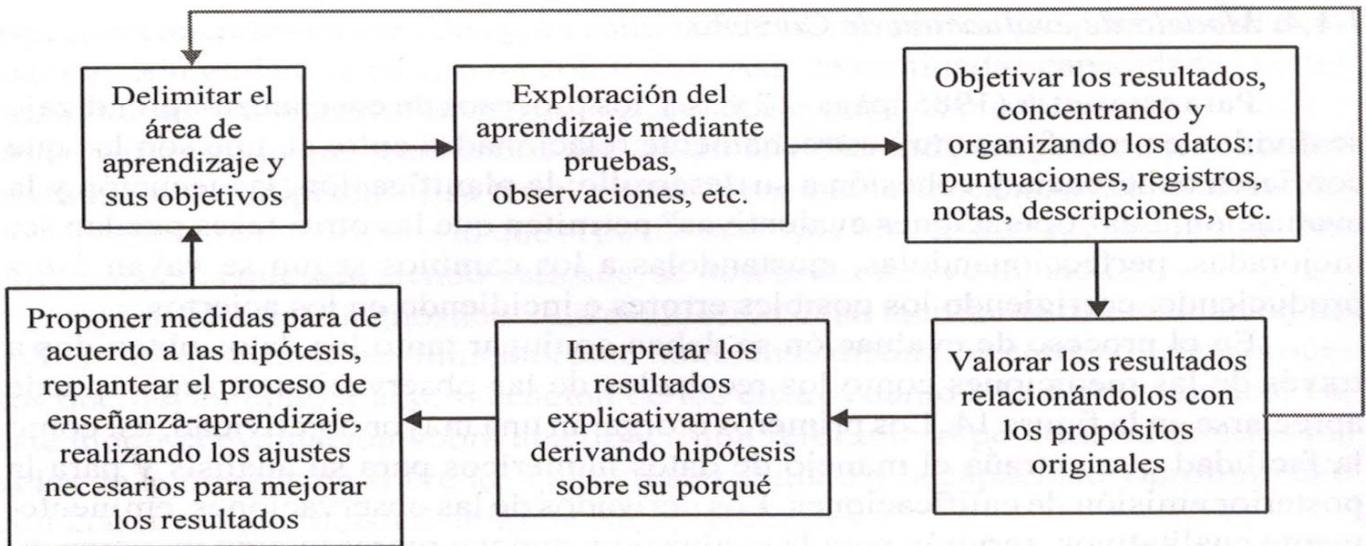


Figura 14: Modelo de evaluación expresado por Carreño (op. cit., 28).

La evaluación debe partir de las interpretaciones explicativas de los resultados y llegar a formular “proposiciones operativas” que permitan la intervención sobre la realidad. El ciclo propuesto, sería el reflejado en la figura 15 (Págs. 36 y 37).

Considera que la funcionalidad de la evaluación está relacionada tanto con los procedimientos de enseñanza como con su selección a través de la medición del rendimiento de los alumnos, pues comparando los resultados se puede preferir objetivamente a unos sobre otros. Desde esta perspectiva, podría encuadrarse éste dentro de los modelos de evaluación comparativa.

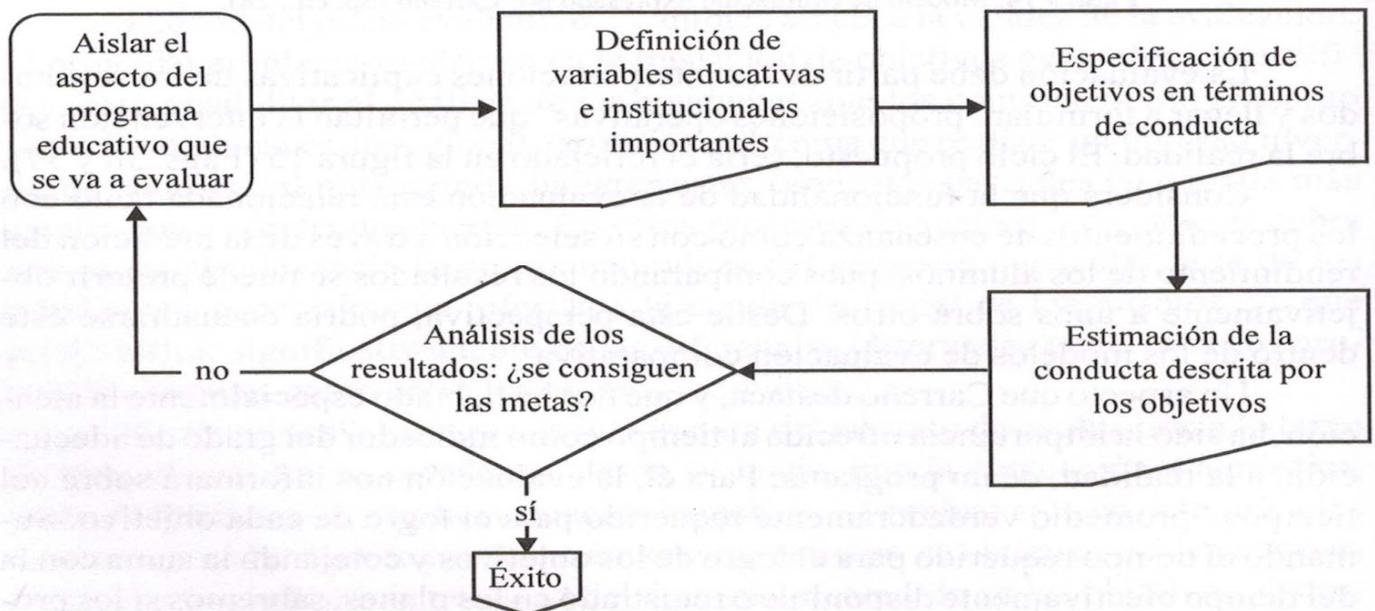
Un aspecto que Carreño destaca, y que nos ha llamado especialmente la atención, ha sido la importancia ofrecida al tiempo como indicador del grado de adecuación, a la realidad, de un programa. Para él, la evaluación nos informará sobre «el tiempo» “promedio verdaderamente requerido para el logro de cada objetivo. Sumando el tiempo requerido para el logro de los objetivos y cotejando la suma con la del tiempo efectivamente disponible o registrado en los planes, sabremos si los programas son realistas y factibles o ambiciosos e irrealizables” (pág. 32).



**Figura 15:** Ciclo propuesto por Carreño para poder realizar proposiciones de carácter operativo.

### 1.1.5. Modelo de evaluación de Hammond

Para este autor (1973, págs. 157 y ss.) la evaluación debe expresar la eficacia de los programas educativos en función de si se consiguen, o no, los objetivos propuestos. Incide en la importancia que poseen diferentes variables relativas a la enseñanza y a la institución en donde ésta se desarrolla como condicionadoras de la conducta del alumno. Por ello, resalta la importancia de su descripción y posterior análisis para interpretar de manera correcta las mediciones sobre el rendimiento académico de los estudiantes. Véase la figura 16.



**Figura 16:** Modelo de consecución de metas de Robert L. Hammond.

### 1.1.6. Modelo de evaluación de Metfessel y Michael

Tal vez lo más destacado de estos autores (1967, págs. 931 y ss.) sea la importancia que dan a los estadios primero -tratando de implicar en el desarrollo del programa a todos los participantes- cuarto y quinto de su diseño, con la utilización de múltiples criterios de medida para determinar hasta qué punto se han conseguido los objetivos. Véase la figura 17. Si bien, deberíamos de recapacitar acerca de la realización periódica de medidas en diferentes intervalos de tiempo, que afectaría directamente al criterio de “continuidad” especificado por Taba (véase la figura 13) y, por lo tanto, a la recogida de información de todo el dominio expresado con los objetivos.

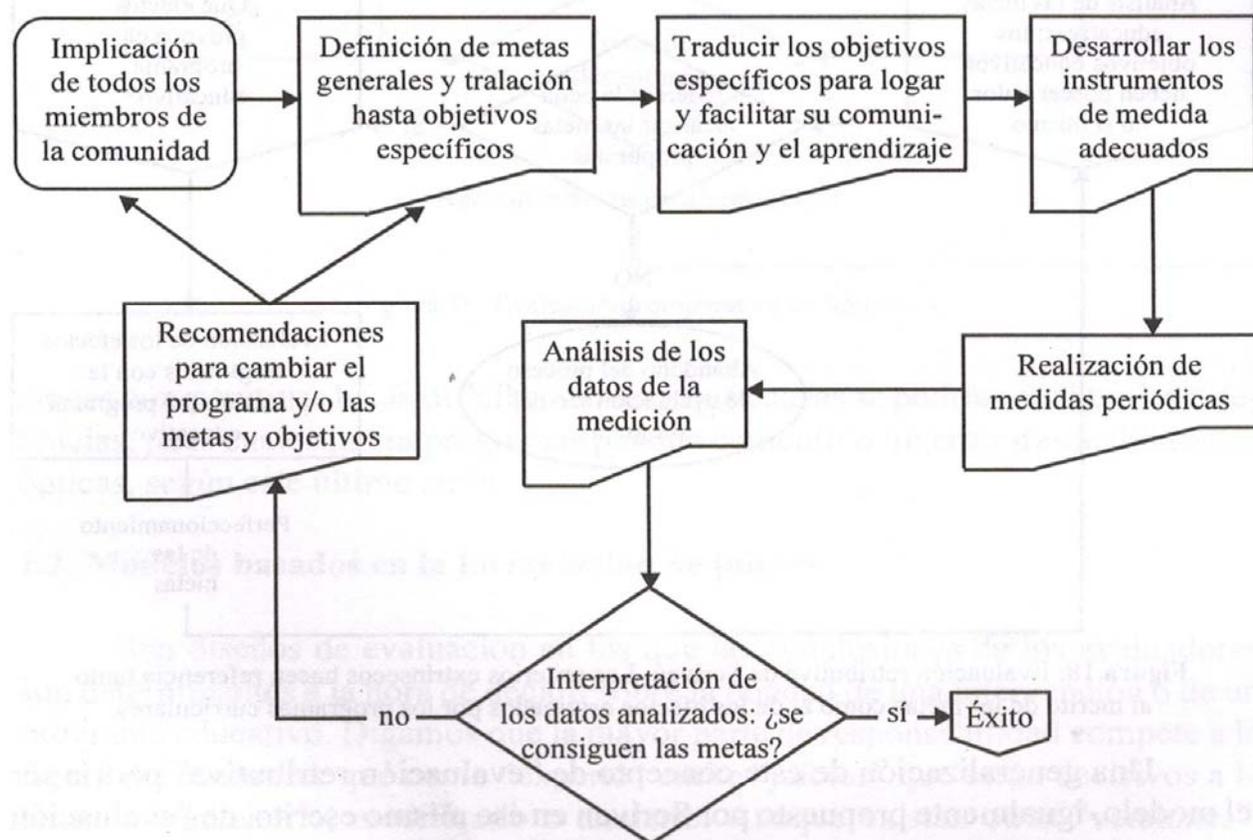
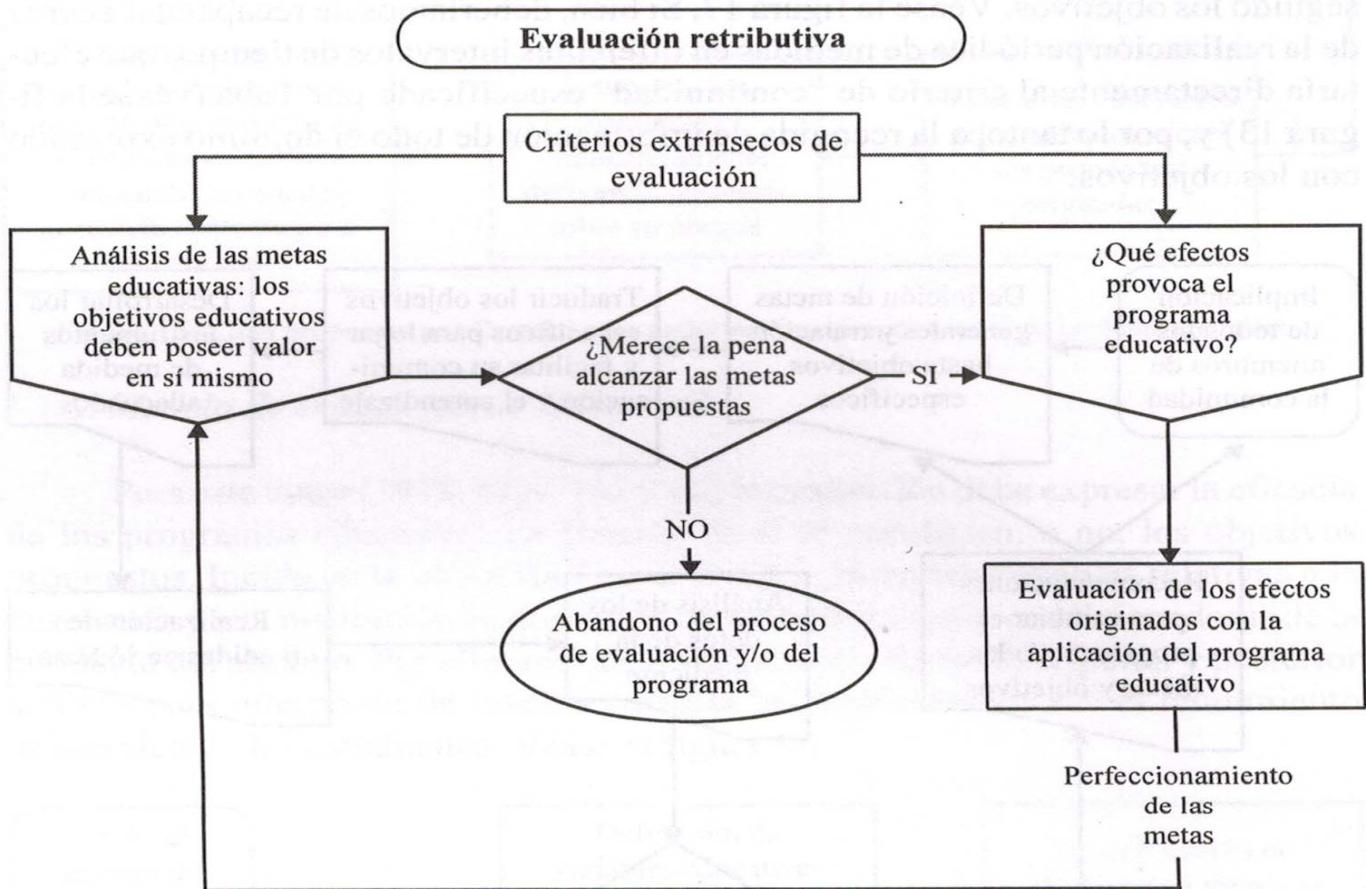


Figura 17: Modelo de consecución de metas de Metfessel y Michael.

### 1.1.7. Evaluación retributiva y comparativa

Convendría aquí hacer mención de una manera especial a la figura de Michel Scriven (Popham, op. cit., pág. 35 y ss.). En su artículo de 1967: “The Methodology of Evaluation”, incorpora clara y explícitamente, por vez primera, la concepción ética como criterio de evaluación, a través de su idea de “evaluación retri-

butiva” o de evaluación centrada en los efectos producidos por los programas de la enseñanza. Véase la figura 18. Para este autor y filósofo, cuando las metas y objetivos educativos carecen de valor, no tiene sentido conocer hasta qué punto han sido conseguidos. El papel que debe de adoptar el evaluador debe de ser crítico y activo con respecto al valor de dichos objetivos, censurándolos e incluso rechazándolos abiertamente si atentan contra la dignidad de las personas.



**Figura 18:** Evaluación retributiva de Scriven. Los criterios extrínsecos hacen referencia tanto al mérito de las metas como al de los efectos originados por los programas curriculares.

Una generalización de este concepto de “evaluación retributiva” podría ser el modelo, igualmente propuesto por Scriven en ése mismo escrito, de “evaluación comparativa”, que perfectamente podríamos encuadrar dentro de aquellos otros modelos especialmente diseñados para ayudar a tomar decisiones. La idea clave de Scriven es que si en el ámbito de la educación hay que tomar decisiones sobre diferentes currícula, estas se facilitarán eligiendo desde distintas alternativas. No es necesario comprender las causas por las que se diferencia un programa de otro; sino que para decidir acerca de ellos es suficiente conocer sus desigualdades y el tipo de efectos que cada uno provoca. Véase la figura 19.

Esta opción implicó un enfrentamiento respecto de la postura mantenida desde unos años antes por Cronbach (1963), quien señaló que este tipo de evalua-

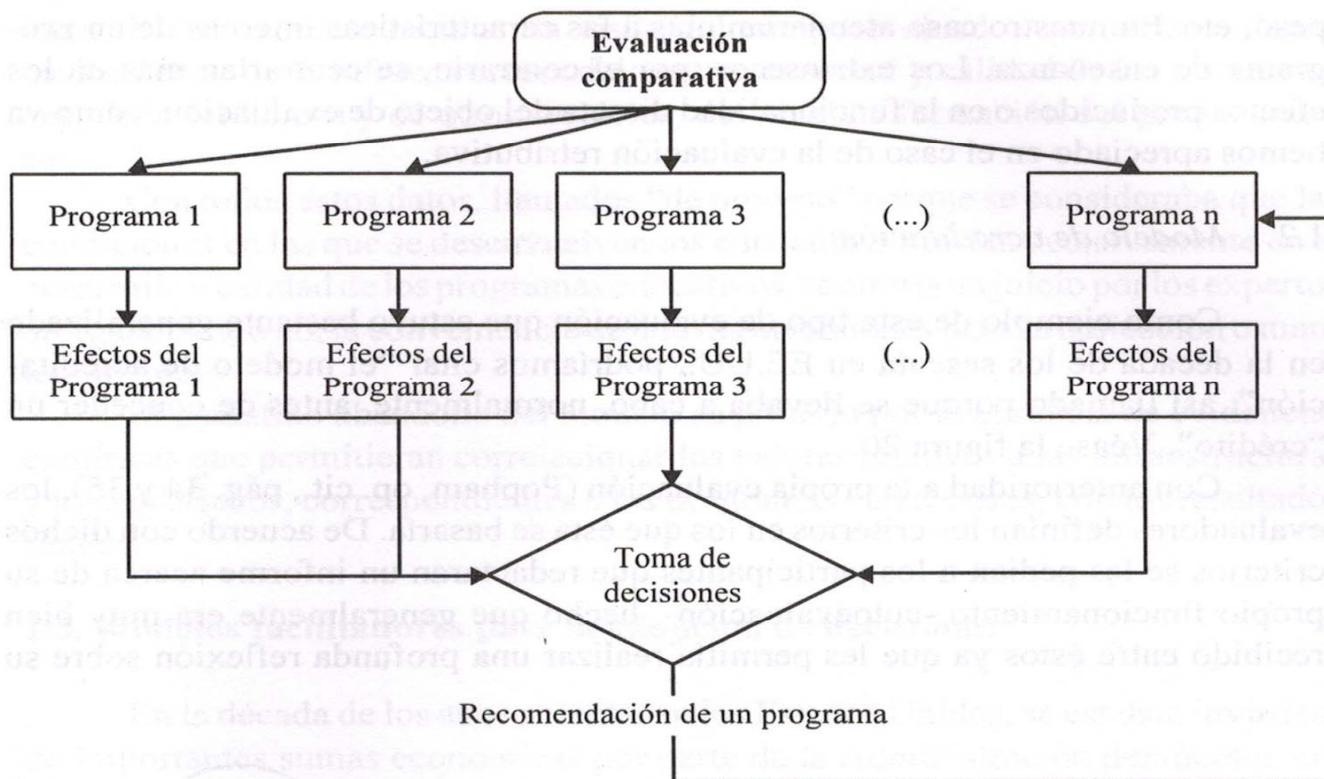


Figura 19: Evaluación comparativa de Scriven.

ciones comparativas hacía difícil entender a qué causas se podrían atribuir las diferencias, pues cualesquiera programas pueden coincidir o divergir desde diferentes ópticas, según este último autor.

## 1.2. Modelos basados en la formulación de juicios

Son diseños de evaluación en los que las conclusiones de los evaluadores son determinantes a la hora de decidir sobre la bondad de una intervención o de un programa educativo. Digamos que la mayor parte de responsabilidad compete a la figura del evaluador que como “experto” emite “juicios” que serán decisivos a la hora de mantenerlo, modificarlo o anularlo. Aunque existan otras “versiones”, como la ofrecida por Stufflebeam y Webster (1981, pág. 6), en las que se suaviza este poder del evaluador trasladando su responsabilidad a aquellos para los que evalúa: “... un estudio de evaluación educativa es aquél que es diseñado y llevado a cabo para ayudar a alguna audiencia a juzgar y mejorar el valor de algún objeto educativo”.

Para ello, unos diseños hacen mayor hincapié en unos atributos, que se toman como criterios de la evaluación que pueden ser inherentes al propio programa. Otros, sin embargo, agudizan el peso que ofrecen a los efectos que producen. Los criterios intrínsecos estarían asociados a cualidades del objeto evaluado: color,

peso, etc. En nuestro caso atenderían más a las características internas de un programa de enseñanza. Los extrínsecos, por el contrario, se centrarían más en los efectos producidos o en la funcionalidad directa del objeto de evaluación, como ya hemos apreciado en el caso de la evaluación retributiva.

### 1.2.1. Modelo de acreditación

Como ejemplo de este tipo de evaluación que estuvo bastante generalizado en la década de los sesenta en EE.UU., podríamos citar “el modelo de acreditación”, así llamado porque se llevaba a cabo, normalmente, antes de conceder un “crédito”. Véase la figura 20.

Con anterioridad a la propia evaluación (Popham, op. cit., pág. 34 y 35), los evaluadores definían los criterios en los que ésta se basaría. De acuerdo con dichos criterios se les pedían a los participantes que redactaran un informe acerca de su propio funcionamiento -autoevaluación-, hecho que generalmente era muy bien recibido entre éstos ya que les permitía realizar una profunda reflexión sobre su

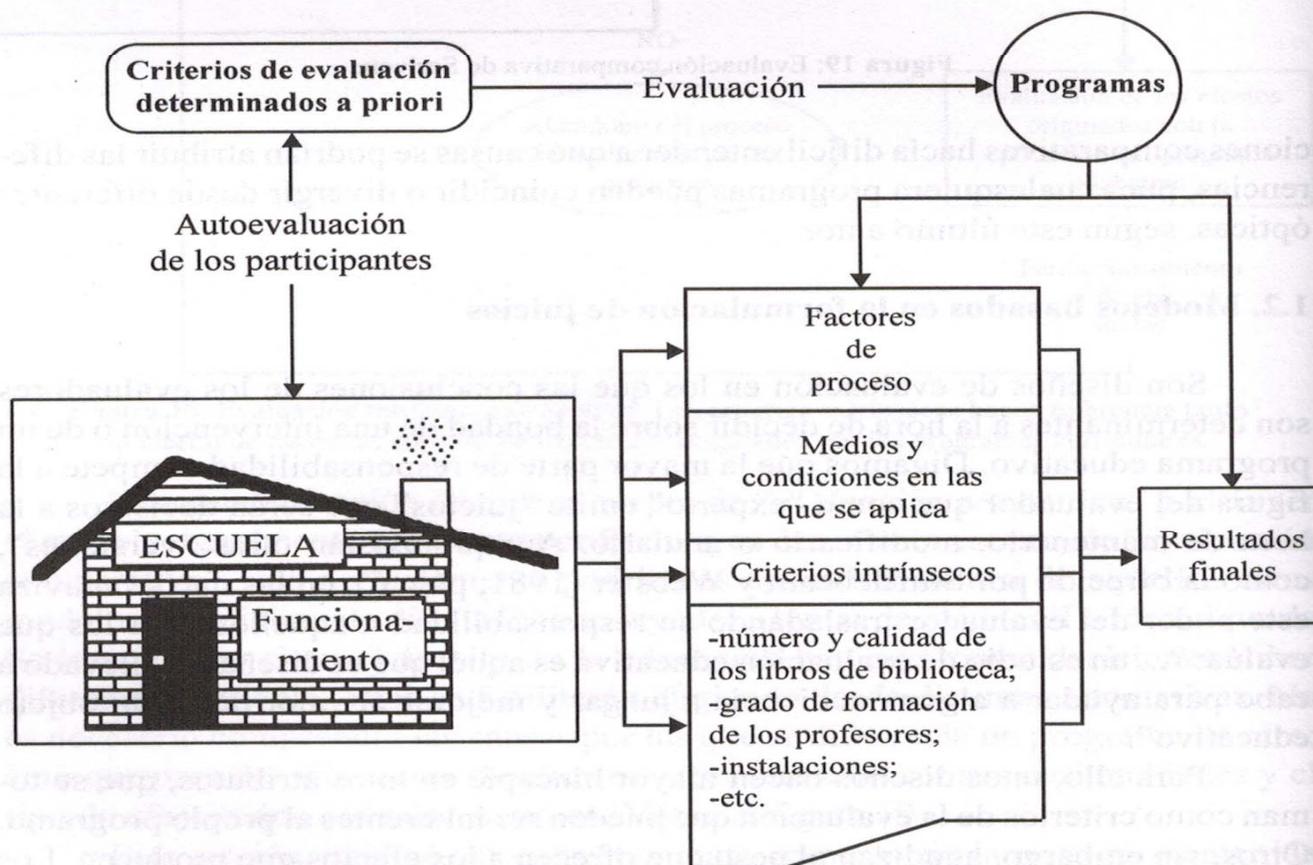


Figura 20: Modelo de acreditación.

labor. Posteriormente recibían la visita del equipo evaluador que obtenía datos, básicamente, de la infraestructura del centro: cantidad y calidad de las instalaciones, de la biblioteca y de otros servicios, del nivel de formación del profesorado, etc.

Con todos estos datos, llamados “de proceso” porque se consideraba que las condiciones en las que se desenvuelven los currícula influirían decisivamente en el desarrollo y calidad de los programas educativos, se emitía un juicio por los expertos otorgándoles, o no, la conveniencia de nuevas inversiones para su aplicación o mantenimiento.

El paulatino abandono del modelo se produjo por la ausencia de evidencias empíricas que permitieran correlacionar los valores relativos a las infraestructuras y los intrínsecos, correspondientes a los programas curriculares, con los resultados finales.

### **1.3. Modelos facilitadores para la adopción de decisiones**

En la década de los años sesenta, en los Estados Unidos, se estaban invirtiendo importantes sumas económicas por parte de la Administración demócrata, entonces en el poder, para el desarrollo de distintos programas educativos que incidieran en una mayor igualdad de oportunidades, o que trataran de compensar determinados déficits de algunas etnias, o de elevar el status cultural de aquellas clases sociales más desfavorecidas. Dentro de este marco de inversiones públicas en los servicios educativos se hacía patente, por un lado, la responsabilidad política de evaluar la calidad de dichos servicios y, por otro, la necesidad de obtener una adecuada información, tal que permitiera al Gobierno tomar decisiones sobre los mismos en función del binomio costes-beneficios.

Este marco social, político y económico es el caldo de cultivo que desde entonces hace fructificar toda una serie de modelos de evaluación cuya única finalidad radica en facilitar la toma de decisiones sobre nuevas inversiones (o el mantenimiento de las ya existentes) a políticos y a burócratas. Hoy en día, son los “indicadores educativos” los que, con su expansión y desarrollo, están ocupando este decisivo papel. Tal vez podamos considerar como los primeros antecedentes los modelos CSE de Alkin, el denominado PPBS (Sistemas de Planificación, Programación y Presupuesto) descrito por Rivlin (1971, pág. 3) como el proceso sistemático para ir obteniendo información substancial a medida que se toman las decisiones, y el CIPP, de Stufflebeam y Guba.

Nosotros, siguiendo a Popham (op. cit., págs. 43 y ss), sólo vamos a presentar dentro de esta categoría los dos modelos que consideramos más relevantes por la influencia que han podido ejercer en este ámbito: el modelo CIPP y el CSE.

### 1.3.1. El modelo CIPP

El modelo CIPP (evaluación del Contexto, de la Información, de los Procesos y de los Productos) fue ideado por Stufflebeam y Guba a comienzos de los años setenta (Stufflebeam et alius, 1971; Stufflebeam, 1974).

La filosofía a la que responde es que la evaluación debe ser una herramienta que debe utilizarse sistemáticamente de un modo continuo y cíclico. Para ello se describen tres estadios para la especificación, obtención y síntesis de la información que se ha de ofrecer a quienes toman las decisiones. Véase la figura 21.

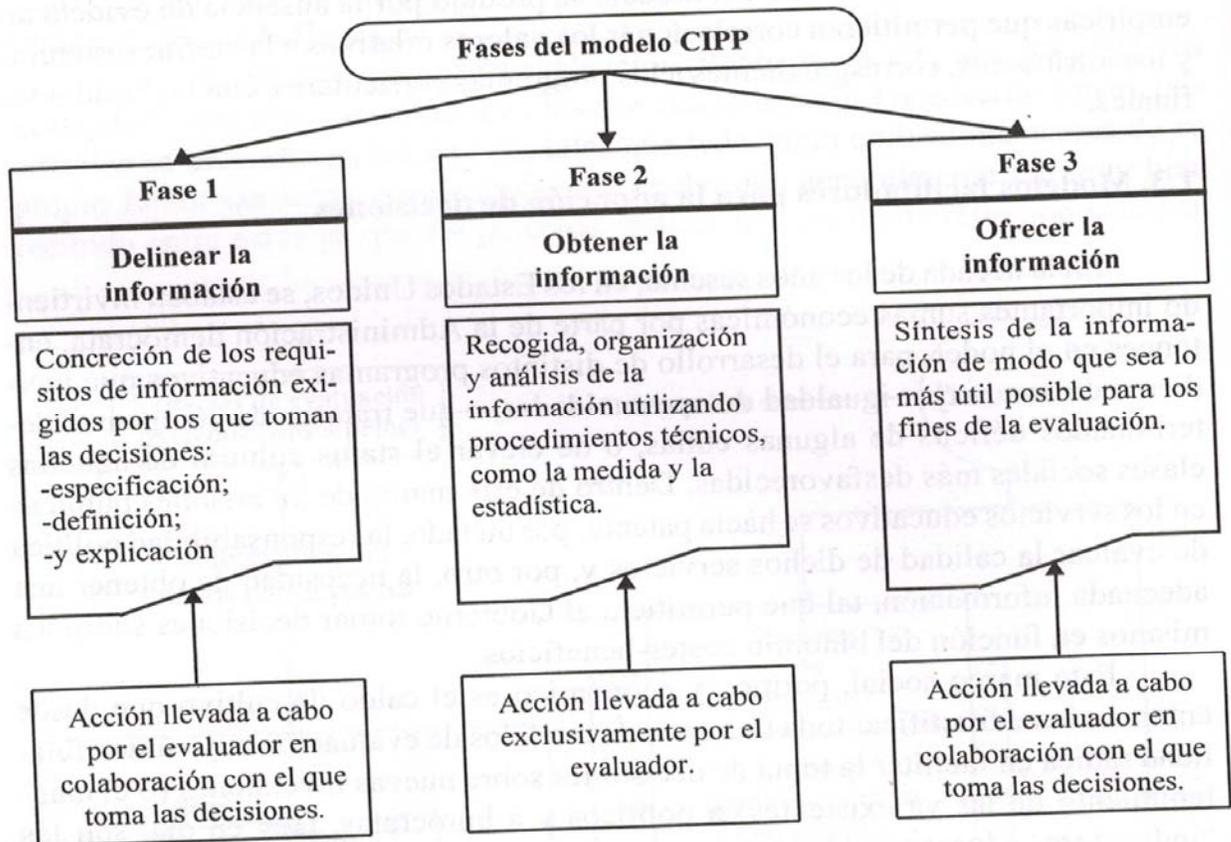


Figura 21.

De acuerdo a la variabilidad de los diferentes tipos de decisiones existentes, el programa de evaluación CIPP, distingue otras tantas situaciones en las que habrán de tomarse dichas decisiones. Véase la figura 22. Así, habrá unas exigencias de información distintas según se trate de mantener el equilibrio en el funcionamiento o mantenimiento del sistema (decisiones homeostáticas), de mejorarlo continuamente sin necesidad de grandes inversiones (incrementativas), de exigencia de un fuerte apoyo económico a la innovación (neomovilísticas) o, por último, aquellas otras que implican cambios totales dentro del sistema (metamórficas).

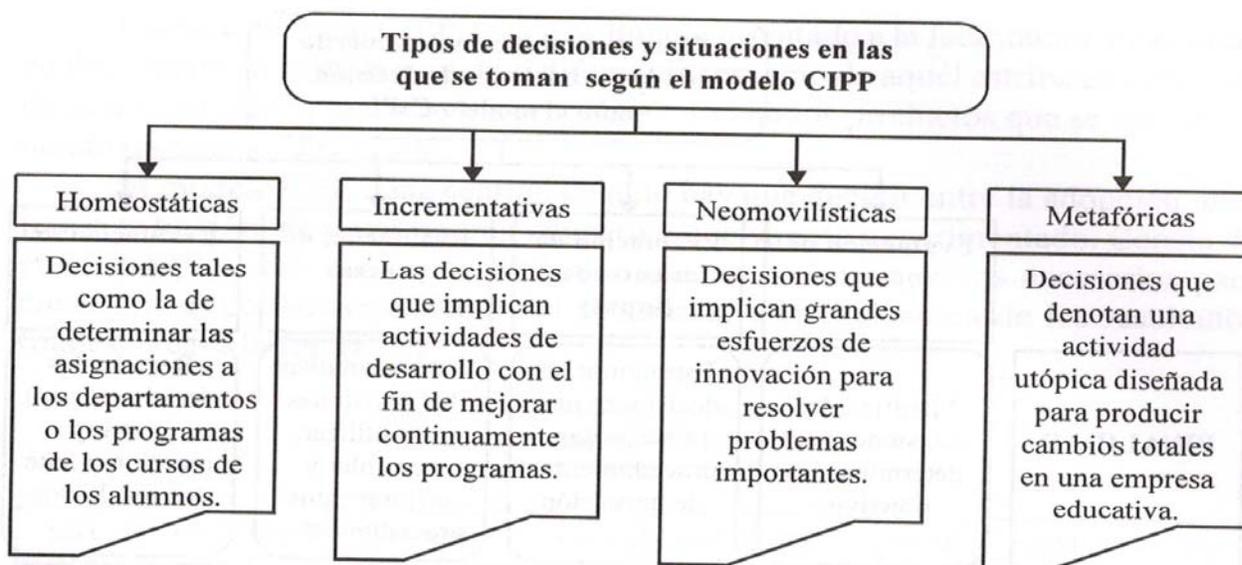


Figura 22.

Además de lo que ya hemos visto sobre el tratamiento de la información y de las situaciones en las que se habrán de tomar decisiones, en general, el modelo CIPP aplicado específicamente al marco educativo posee cuatro estadios que lo caracterizan, constituyendo en sí mismos el núcleo del programa. Es lo que podríamos denominar como las finalidades de los distintos tipos de evaluación. Véase la figura 23. En primer lugar la planificación de las decisiones orientadas a la determinación de los objetivos. Una segunda fase relativa a la organización de las decisiones de cara al diseño de la instrucción. El tercer paso consiste en la utilización, control y perfeccionamiento de los procedimientos utilizados en el diseño a través de las propias decisiones que interactivamente se vayan tomando. Por último, la reutilización de las decisiones de acuerdo a los resultados obtenidos.

Stufflebeam relaciona el modelo CIPP con la dicotomía formativa-sumativa propuesta por Scriven. Véase la figura 24. Distingue los que considera dos aspectos esenciales que debe cumplir cualquier programa de evaluación. La evaluación orientada a la facilitación de la información para los que han de tomar decisiones, que adquiere una naturaleza formativa y él denomina “proactiva”, y cuando su finalidad es la de rendir cuentas, en cuyo caso alcanza la característica de sumativa y él llama “retroactiva”.

### 1.3.2. El modelo CSE

El modelo CSE (Alkin, 1969) debe su nombre al hecho de haberse desarrollado en el Centro para el Estudio de la Evaluación de la Universidad de California, dirigido entonces por Marvin C. Alkin.

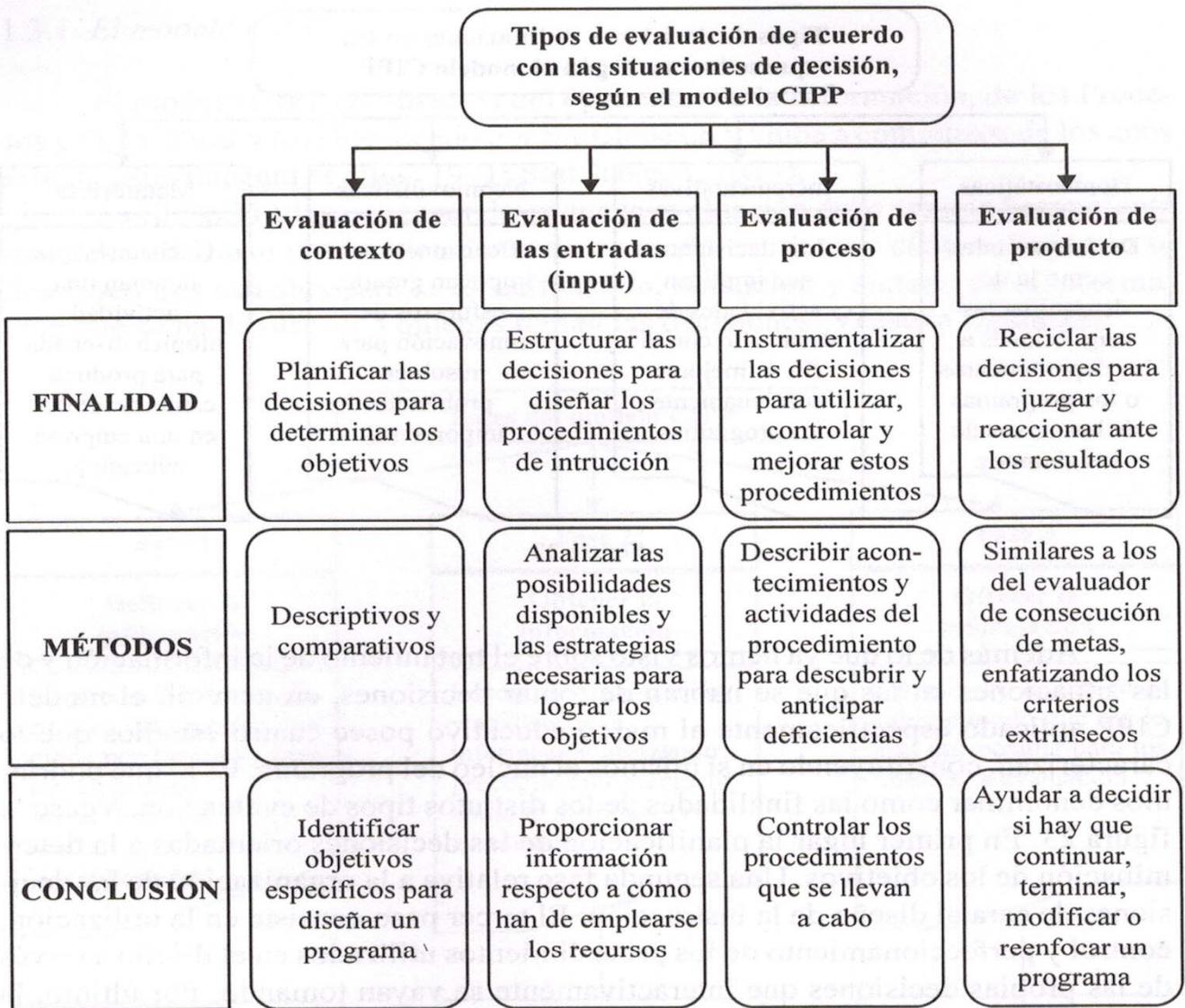


Figura 23.

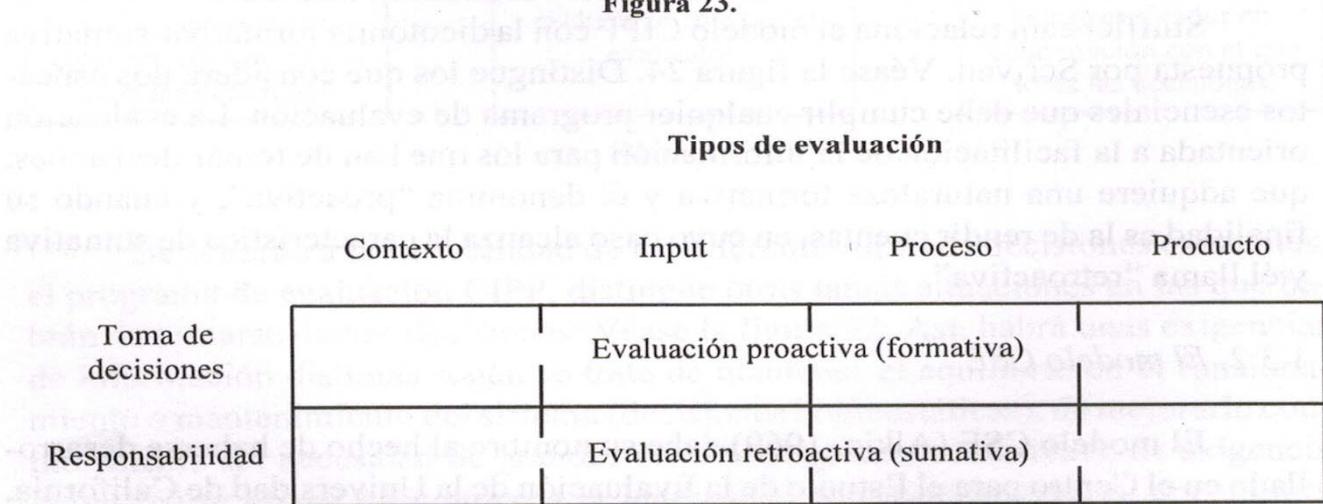


Figura 24: Extraído de Popham (op. cit., pág. 47).

Como el modelo CIPP, está igualmente orientado a la facilitación de la toma de decisiones, si bien, la principal diferencia respecto de aquél estriba en su consideración, por igual, tanto de los procesos como de los productos que se van originando durante su desarrollo.

El modelo CSE tiene sentido cuando hay que decidir entre la adopción, mejora o extinción, de un programa educativo que ya está experimentado. Consta de cinco estadios que van desde la determinación de los objetivos necesarios para conseguir el perfeccionamiento del programa, a la evaluación de los resultados finales. Véase la figura 25.

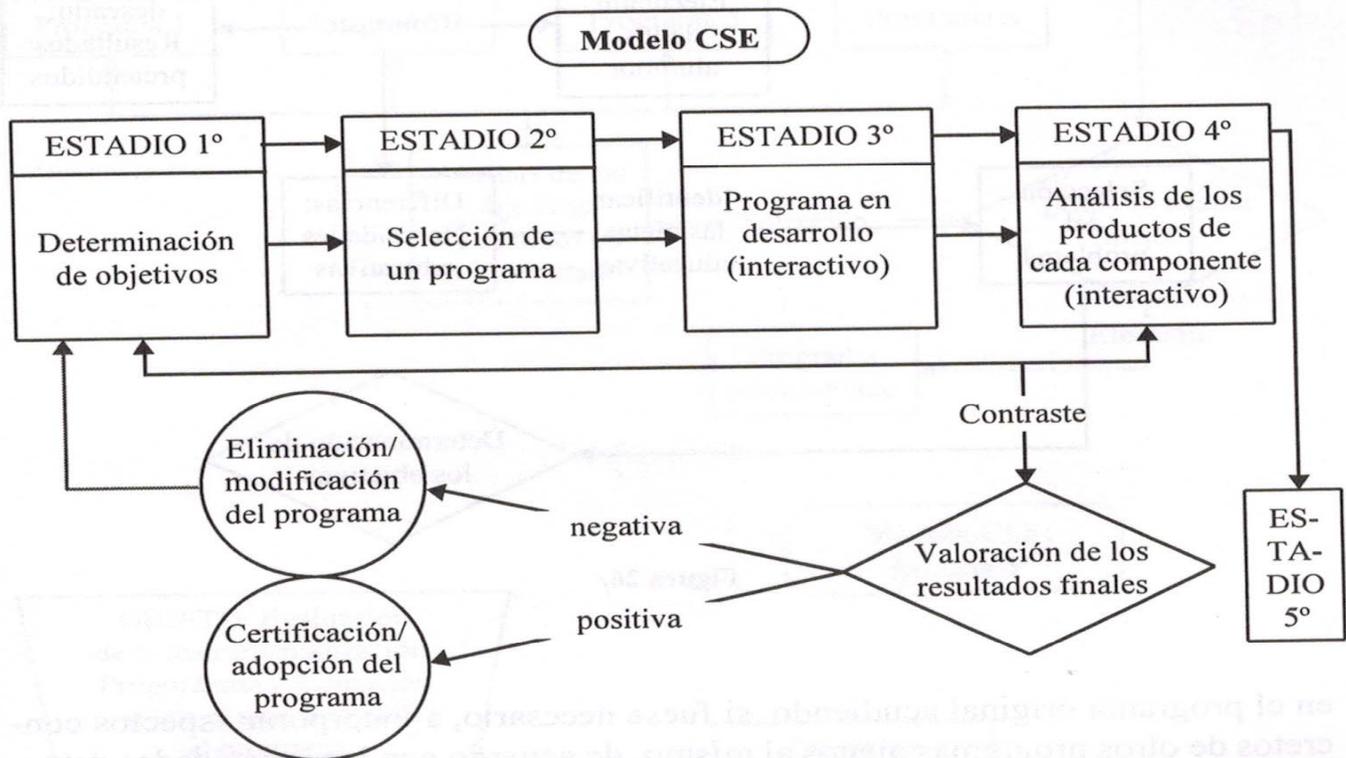


Figura 25.

Para lo cual, en primer lugar, se trata de establecer el nivel real de eficacia del programa y sus posibilidades de mejora. Así, el primer estadio está enfocado de cara a la determinación de los objetivos que nos propondremos a partir del contraste entre el nivel de ejecución real de los alumnos que siguen el programa educativo y el nivel de ejecución teórico (el deseado o perseguido tras la correcta aplicación de ese mismo programa). Los objetivos obtenidos como resultado de las diferencias entre ambos niveles, real y teórico, representan lo que el modelo llama “necesidades educativas”. Véase la figura 26.

En una segunda etapa, se ensayan diferentes “programas de segunda generación”, extraídos a partir de la introducción de diversas modificaciones de mejora

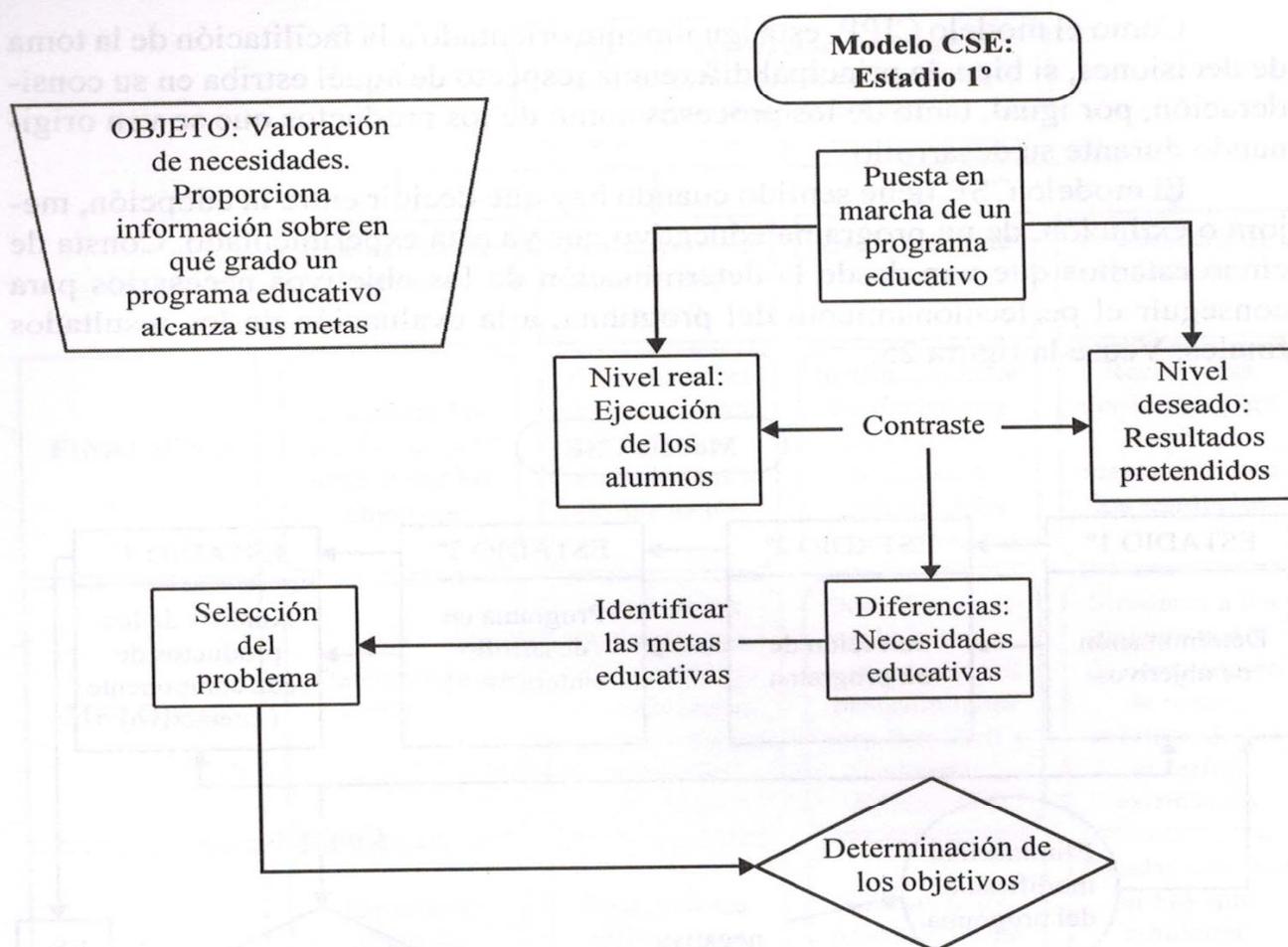


Figura 26.

en el programa original acudiendo, si fuese necesario, a incorporar aspectos concretos de otros programas ajenos al mismo, de acuerdo con las necesidades detectadas en la fase anterior. Se realiza una evaluación comparativa para determinar cuál de todos ellos se adapta mejor a los objetivos planteados inicialmente y, como resultado, se selecciona uno de estos programas. Véase la figura 27.

El tercer estadio trata de detectar, valorar y hacer las modificaciones oportunas para que durante la ejecución del programa seleccionado en la fase anterior se minimicen las posibles diferencias entre lo planificado y su desarrollo real (evaluación de los procesos). Como consecuencia de estos nuevos ajustes, fruto de una continua interacción, se obtendría lo que hemos denominado “plan ajustado”. Véase la figura 28.

El estadio 4.º evalúa hasta qué punto se están consiguiendo los objetivos propuestos, no sólo de una manera global, sino también de una manera más específica: en qué grado cada uno de los componentes curriculares está coadyuvando a

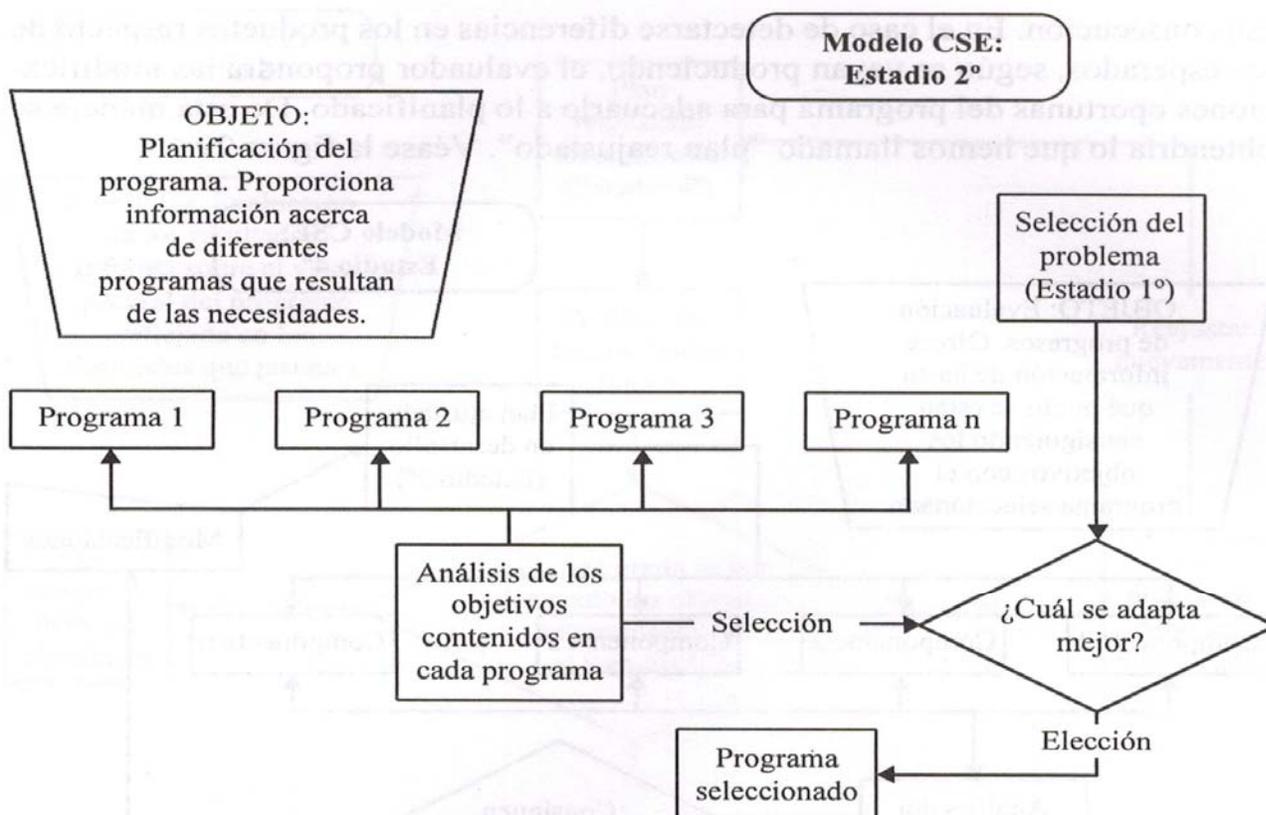


Figura 27.

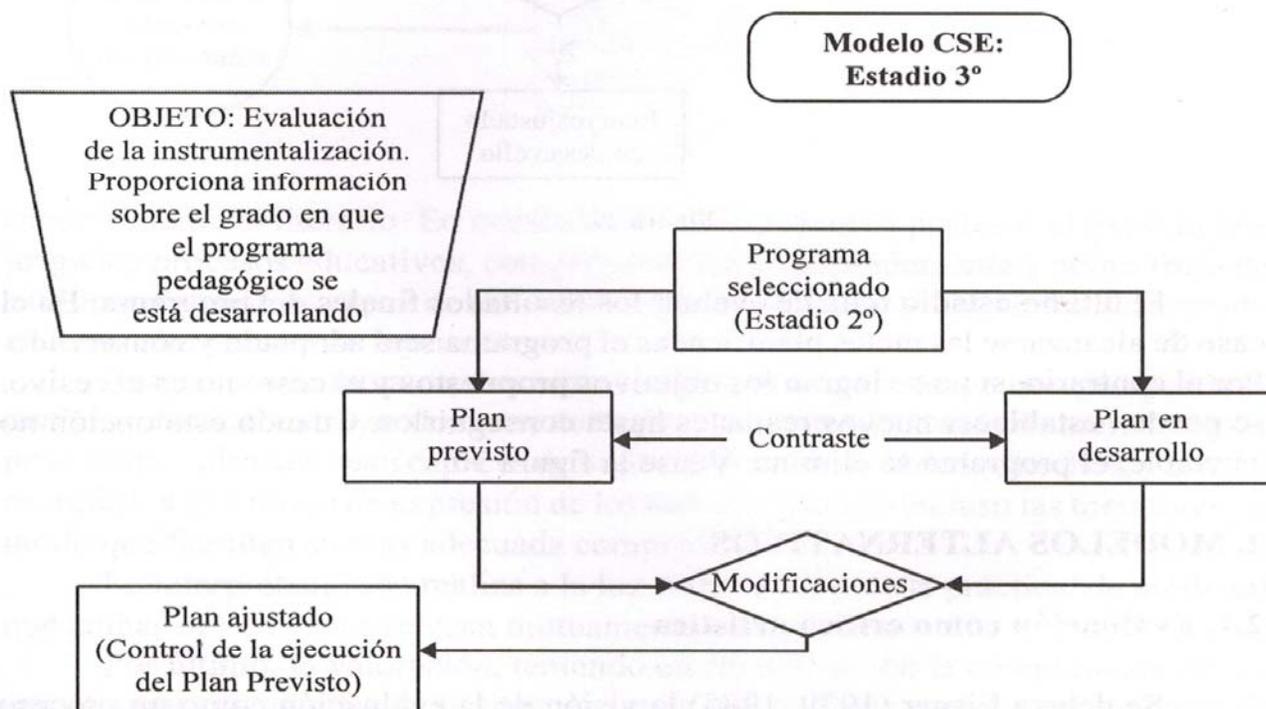


Figura 28.

esa consecución. En el caso de detectarse diferencias en los productos respecto de los esperados, según se vayan produciendo, el evaluador propondrá las modificaciones oportunas del programa para adecuarlo a lo planificado. De esta manera se obtendría lo que hemos llamado “plan reajustado”. Véase la figura 29.

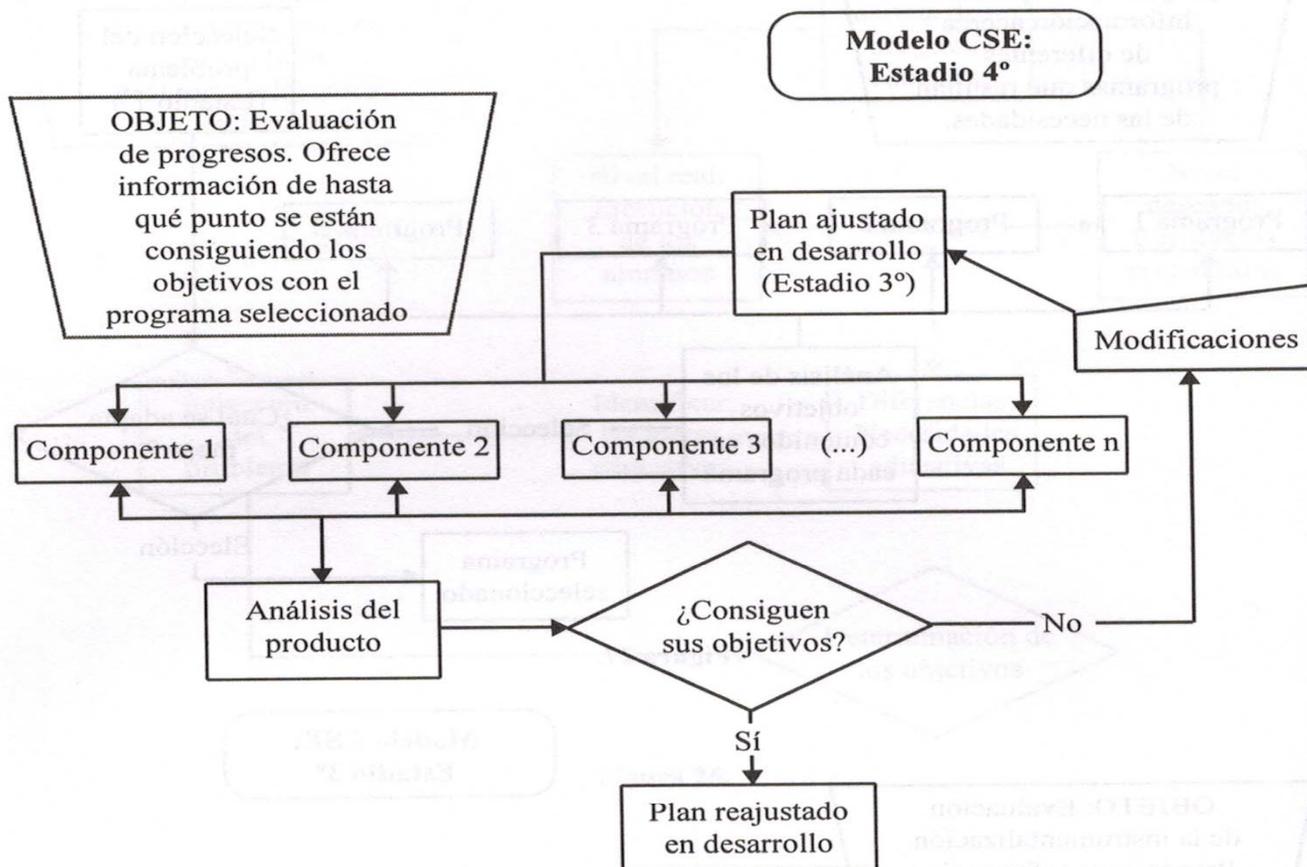


Figura 29.

El último estadio trata de evaluar los resultados finales del programa. En el caso de alcanzarse las metas planificadas el programa será adoptado y conservado. Por el contrario, si no se logran los objetivos propuestos y el costo no es excesivo, se pueden establecer nuevos reajustes hasta conseguirlos. Cuando esta opción no es viable, el programa se elimina. Véase la figura 30.

## 2. MODELOS ALTERNATIVOS

### 2.1. Evaluación como crítica artística

Se debe a Eisner (1979, 1985) la visión de la evaluación como un proceso similar al realizado por los expertos cuando enjuician el valor de una obra de arte,

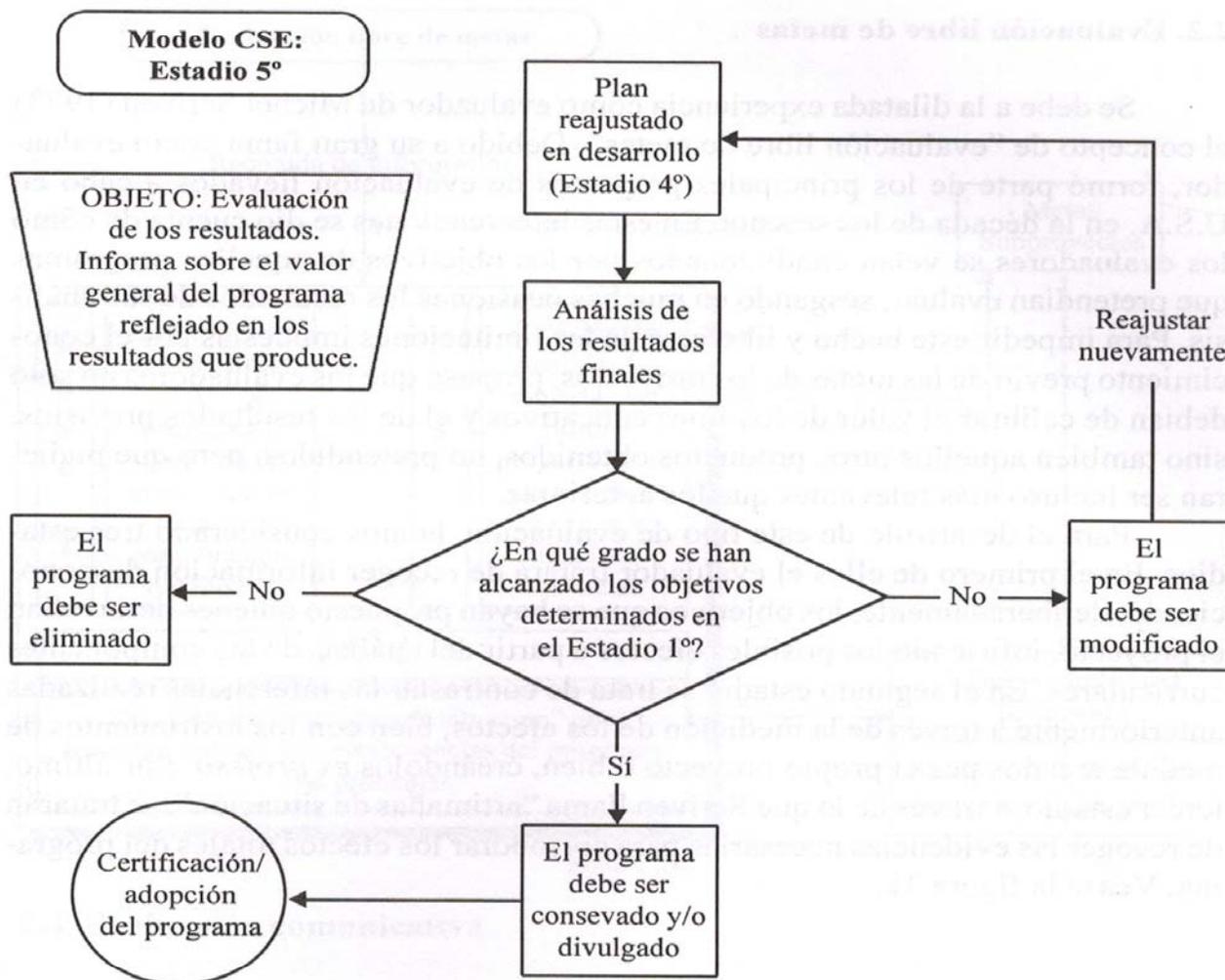


Figura 30.

especialmente el literario. En nuestro caso sería el propio profesor el experto que juzga los procesos educativos, comprendiéndolos profundamente y como fruto de esta comprensión es capaz de *describirlos, interpretarlos y valorarlos* respetando el desarrollo natural de la enseñanza.

Respecto a la tarea de describir se llevaría a cabo desde dos perspectivas distintas: fáctica, extrayendo datos de la realidad incluso a través de metodologías positivistas tales que sean capaces de explicar la naturaleza de los hechos, y artística, referida a la forma de expresión de los datos utilizando incluso las metáforas de modo que faciliten su más adecuada comprensión.

La interpretación se realiza a la luz de la teoría y de la práctica, de modo tal que ambas dimensiones se vean mutuamente enriquecidas.

Por último, la valoración, teniendo en consideración la comprensión de las características concretas de la situación en la que se desenvuelven los procesos de enseñanza.

## 2.2. Evaluación libre de metas

Se debe a la dilatada experiencia como evaluador de Michel Scriven (1973) el concepto de “evaluación libre de metas”. Debido a su gran fama como evaluador, formó parte de los principales proyectos de evaluación llevados a cabo en U.S.A. en la década de los sesenta. En estas intervenciones se dio cuenta de cómo los evaluadores se veían condicionados por los objetivos de aquellos programas que pretendían evaluar, sesgando en muchas ocasiones los resultados de sus análisis. Para impedir este hecho y liberarse de las limitaciones impuestas por el conocimiento previo de las metas de los proyectos, propuso que los evaluadores no sólo debían de calibrar el valor de los fines educativos y el de los resultados previstos, sino también aquellos otros productos obtenidos, no pretendidos, pero que pudieran ser incluso más relevantes que los anteriores.

Para el desarrollo de este tipo de evaluación, hemos considerado tres estadios. En el primero de ellos el evaluador tratará de recoger información desconociendo, deliberadamente, los objetivos que se hayan propuesto quienes desarrollan el proyecto, infiriendo los posibles efectos a partir del análisis de los componentes curriculares. En el segundo estadio se trata de contrastar las inferencias realizadas anteriormente a través de la medición de los efectos, bien con los instrumentos de medida creados por el propio proyecto o bien, creándolos *ex profeso*. Por último, tercer estadio, a través de lo que Scriven llama “artimañas de situación”, se tratarán de recoger las evidencias necesarias para corroborar los efectos totales del programa. Véase la figura 31.

## 2.3. Método del *modo operandi*

También se debe a Michel Scriven la creación de este modelo. Para ello se fijó en el modo de investigar la realidad en distintas ramas de la investigación social (antropólogos, historiadores, detectives, etc.). Observó como estos expertos siempre actuaban manteniendo unas pautas comunes. En primer lugar detectaban la existencia de determinados rasgos que hacían concordar, mediante hipótesis de trabajo, con determinados patrones de conducta. Tras recoger las evidencias necesarias corroboraban, o no, sus hipótesis iniciales. De este modo podían evidenciar determinadas relaciones de causalidad sin tener que acudir a métodos experimentales. En el caso de la evaluación educativa su aplicación serviría para detectar y evaluar las posibles causas que originarían efectos deseables, o perniciosos, en el desarrollo de un programa. Así el evaluador se convertiría en un investigador que trataría de encontrar determinadas cadenas de sucesos características a las que denominó como *modus operandi* (M.O.). La secuencia a seguir sería la indicada en la figura 32.

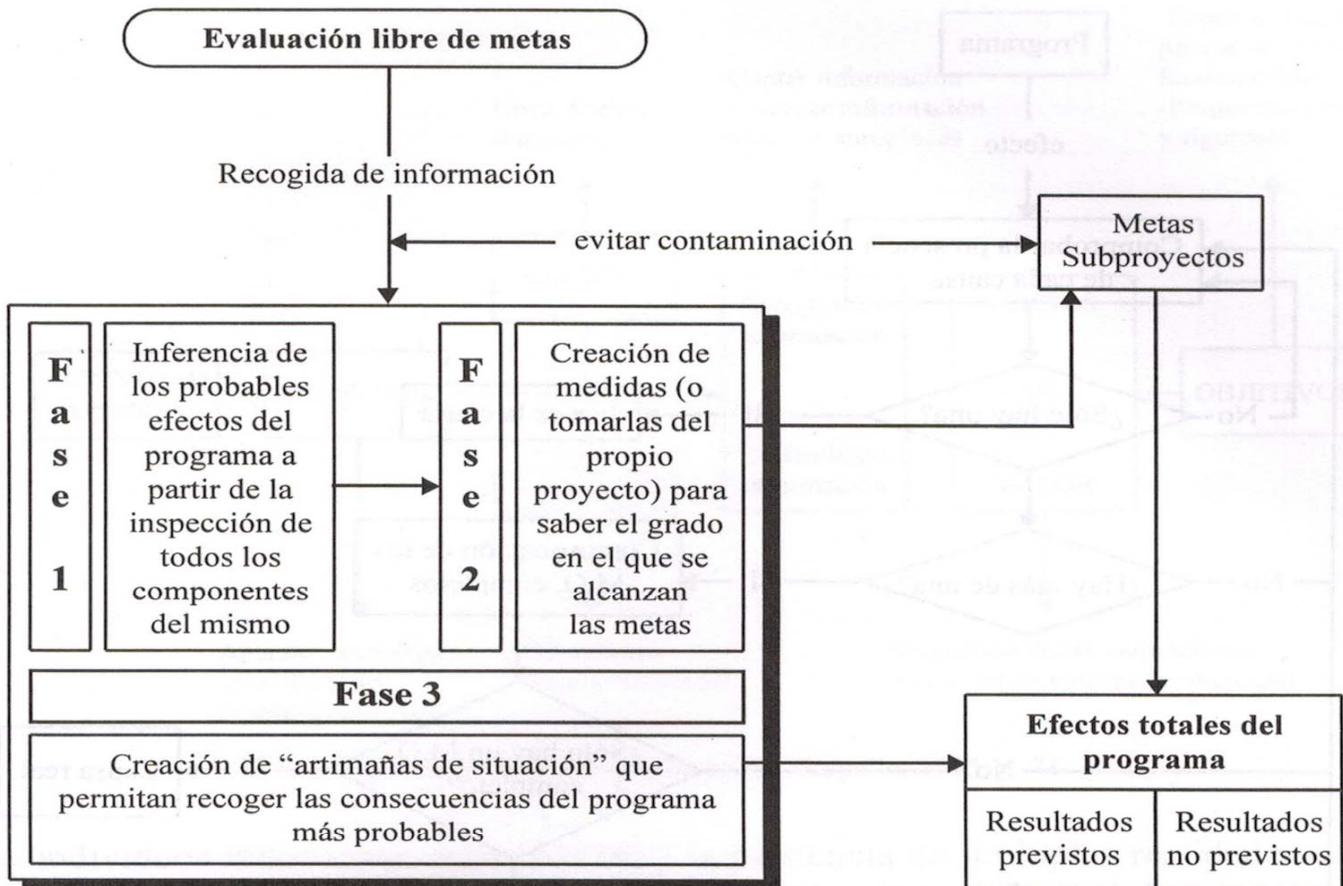


Figura 31.

## 2.4. Evaluación comunicativa

Nos hemos fijado en este modelo, propuesto por Jesús Rul Gargallo (1994, 1995), dada la frescura que introduce dentro del debate, siempre abierto, acerca de la evaluación. Para este autor, "hay evaluación si a través de la experiencia y de la comunicación se produce aprendizaje valorativo, con repercusiones individuales y colectivas, basado en datos e información objetivada" (Pág. 70). Considera a la valoración como una dimensión subjetiva sobre un objeto realizada *a priori* y la diferencia de la evaluación en que ésta es una actividad objetiva con una finalidad explicativa realizada *a posteriori*. Pero desde nuestra visión, es la importancia que le ofrece a la comunicación humana, con todo lo que ello conlleva, el aspecto más destacable como diferenciación característica respecto a otros modelos.

Dentro de su concepción de evaluación, sintetiza toda una serie de factores -véase la figura 33- que, sin duda alguna, influyen y la condicionan, como son: la dependencia respecto a la situación en la que se desarrolla, al contexto, la cultura, los valores (axiología), la subjetividad-objetividad, la sistematización en la recogida y tratamiento de la información, la facilitación para poder tomar decisiones y que éstas repercutan en los fines (teleología) de la evaluación, de tal manera que

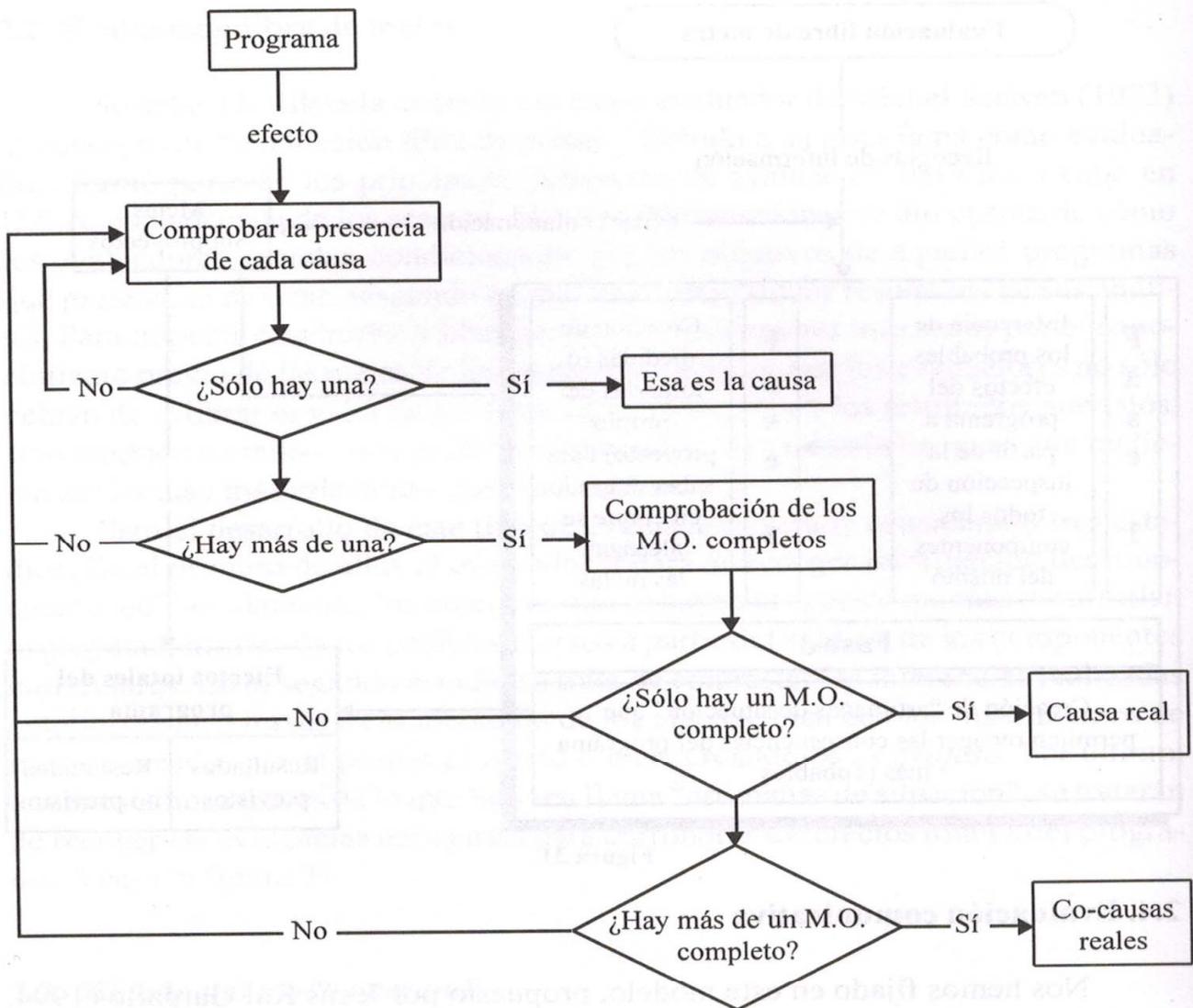


Figura 32.

conlleve al perfeccionamiento. Es por tanto, el suyo, un modelo interpretativo y sistemático que se centra en su dimensión objetiva: “evaluar es la atribución de valores intersubjetivos a un objeto como resultado de operaciones subjetivo-objetivas de tratamiento de datos y de información rigurosa en un contexto específico en relación con finalidades de conocimiento, toma de decisiones, optimización y desarrollo humano, individual y colectivo” (Pág. 73).

## 2.5. Evaluación iluminativa

Sus autores, Parlett y Hamilton (1972, 1976), rompen con la tradición empírico-racionalista para trasladar al campo de la evaluación educativa los métodos de trabajo propios de la investigación antropológica. El estudio holístico que proponen debe abarcar desde el análisis pormenorizado del contexto en el que los

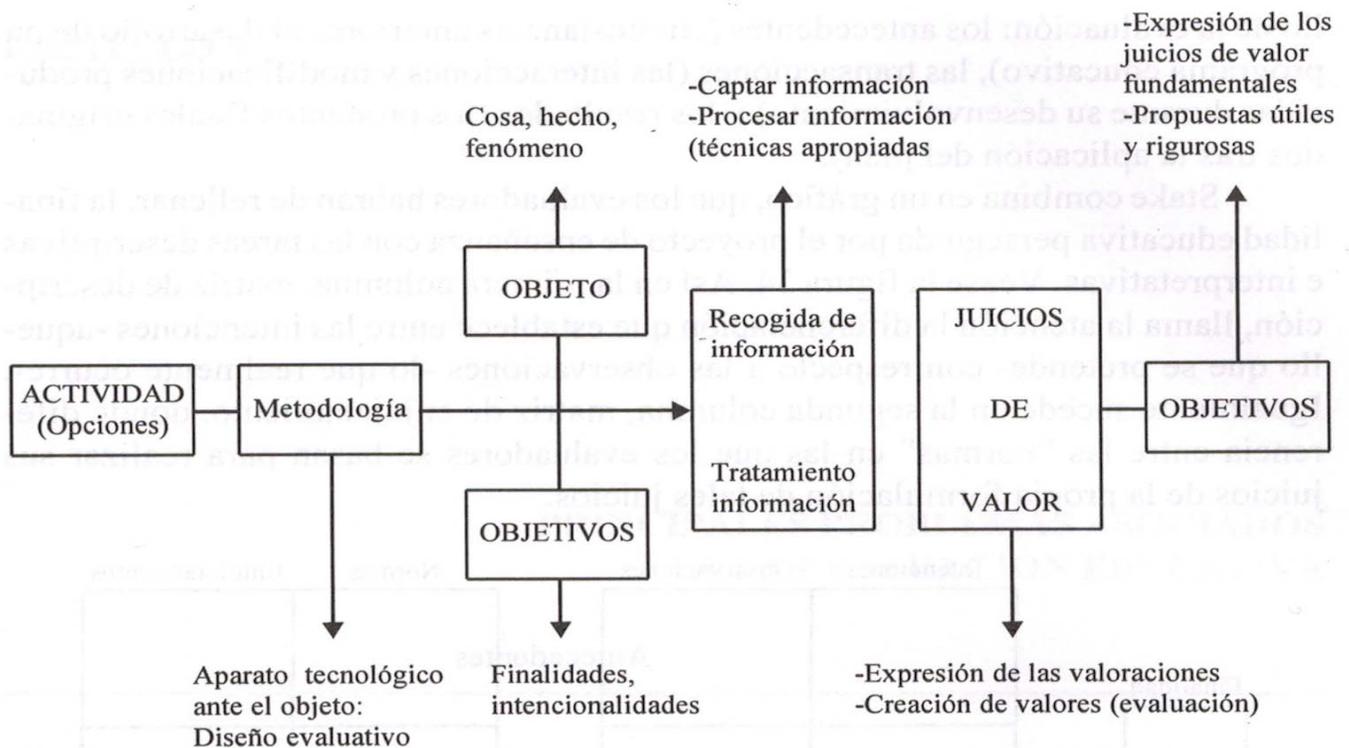


Figura 33: Cuadro extraído de Rul Gargallo, 1995, p. 74.

individuos intervienen, para descubrir la existencia de múltiples realidades que interactúan, asumidas o no por la colectividad, a los resultados, destacando de modo especial los procesos. Precisamente ésta es la causa de su desdén hacia las pruebas psicométricas y de basarse en una metodología cualitativa, observacional, flexible de acuerdo con el momento de la investigación y del contexto -con sus propios condicionantes psicosociales-, negociada con los implicados, centrándose y profundizando progresivamente en aquellos aspectos más significativos de los que en ese “escenario” tienen lugar.

## 2.6. Evaluación respondiente

Formulada por Stake en 1967<sup>2</sup> (1975, a; 1975, b; 1975, c; 1980) incide en la manera en la que los evaluadores realizan dos tareas básicas: las descriptivas y las relativas a cómo formulan sus juicios. Para ello, propone tres etapas en el desarro-

<sup>2</sup>Fue publicado originalmente en un artículo bajo el título de “El semblante de la evaluación educativa”. Por ese motivo también algunos autores lo denominan como “el modelo del semblante”. La referencia completa es: “The countenance of educational evaluation”. *Teacher's College Record*. N.º: 68. Págs.: 523 a 540. Igualmente se la conoce como el modelo de “réplica” o “respondiente” por cuanto trata de dar respuesta a los intereses y necesidades de todos los implicados en un proceso de evaluación.

llo de la evaluación: los antecedentes (circunstancias anteriores al desarrollo de un programa educativo), las transacciones (las interacciones y modificaciones producidas durante su desenvolvimiento) y los resultados (los productos finales originados tras la aplicación del plan).

Stake combina en un gráfico, que los evaluadores habrán de rellenar, la finalidad educativa perseguida por el proyecto de enseñanza con las tareas descriptivas e interpretativas. Véase la figura 34. Así en la primera columna, matriz de descripción, llama la atención la diferenciación que establece entre las intenciones -aquello que se pretende- con respecto a las observaciones -lo que realmente ocurre-. Igualmente sucede en la segunda columna, matriz de enjuiciamiento, donde diferencia entre las “normas” en las que los evaluadores se basan para realizar sus juicios de la propia formulación de tales juicios.

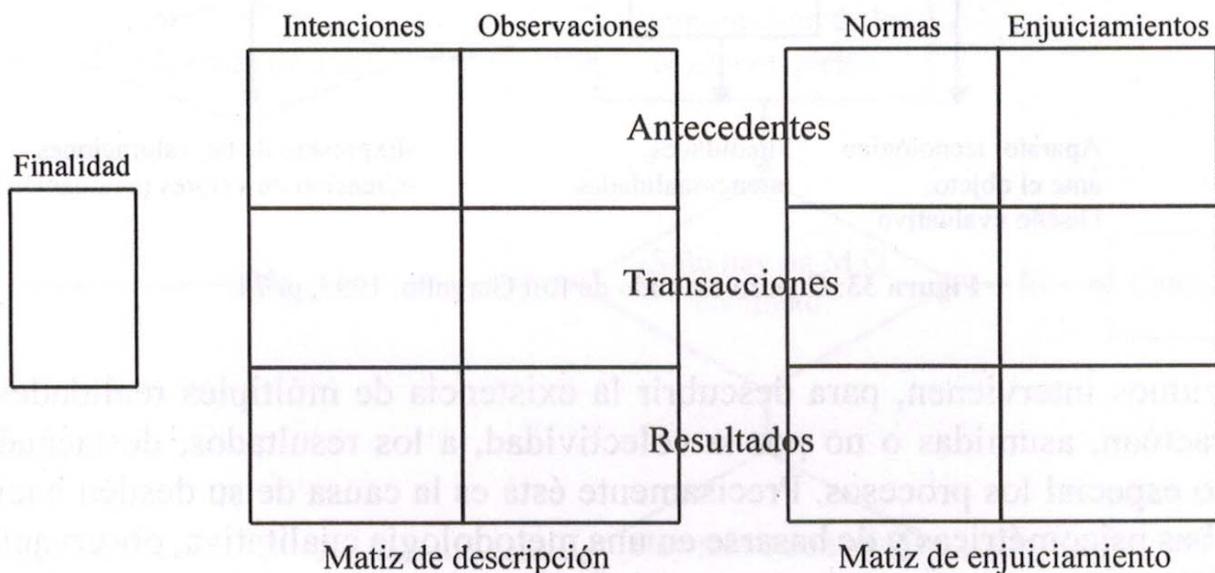


Figura 34. extraída de Popham (op. cit., pág. 41).

Pero tal vez lo más interesante de Stake no sea el cuidadoso aspecto técnico que le confiere a su modelo de evaluación, sino la filosofía que encierra. Stake considera que la evaluación, para que sea tal, ha de responder a los intereses y necesidades de quienes la patrocinan (la administración, generalmente) y de aquellos que están implicados en ella (alumnos, padres, profesores, etc.) de acuerdo con el nivel de conocimientos de cada uno. La manera de “responder” adecuadamente a las distintas, y a veces opuestas necesidades, está en recoger las diferentes visiones de cada sector. Esto le confiere al modelo un avance ético respecto de otros además de, por vez primera, un carácter ciertamente democrático.