

CAPITULO VII

Enfoques generales

Introducción

Parece innecesario decir que las técnicas de evaluación deben seleccionarse de acuerdo con su capacidad para señalar problemas concretos. Sin embargo, da la sensación de que la elección de los métodos es con frecuencia muy arbitraria y regida, no por su adaptación al problema, sino por la comodidad, o por lo que resulta compatible con las destrezas e intereses de los evaluadores. En el contexto de la evaluación del *currículum* en las escuelas, es comprensible, ya que es raro que las evaluaciones las efectúen evaluadores profesionales. No es extraño, por tanto, que si un científico se inclina por la utilización de un modelo experimental, un matemático opte por técnicas estadísticas, un fotógrafo aficionado haga un registro fotográfico, un profesor de inglés se muestre a favor del estudio de casos (aunque sea ficticio) o un director aplique la mayor parte del tiempo que dedica a la evaluación a crear estructuras de comisiones para la consulta al profesorado.

Respecto a la evaluación interna (Capítulo V), SIMONS (1981) indica que es preciso reconocer la relativa falta de dominio, de manera que las destrezas adecuadas para las actividades de evaluación se construyan sobre o que ya poseen los profesores. En general, la bibliografía sobre técnicas de evaluación no tiene en cuenta el grado de dominio previsible de los evaluadores ni las necesidades y objetivos prácticos que caracterizan los distintos contextos de evaluación. En otras palabras, gran parte de lo que se ha escrito resulta demasiado académico y alejado de las situaciones concretas, sobre todo cuando se trata de la bibliografía que se centra casi exclusivamente en los modelos de evaluación (p. ej.: Open University 1976, unidades 19 y 20; HAMILTON, 1976). Incluso DAVIS (1981), en un libro titulado *Teachers as Curriculum Evaluators*, recurre a gran cantidad de modelos para estructurar su exposición. Tratando de evitar esta línea de discurso, hemos organizado este capítulo en torno a una serie de cuestiones que surgen al considerar la elección de una metodología. Como no es frecuente que los modelos ayuden a los evaluadores a decidir qué hacer en la práctica, sólo se introducen cuando parecen ofrecer una solución a un problema concreto¹. (En realidad, muchos modelos aparecen como respuestas a problemas específicos.) Sin embargo, aunque nuestro tratamiento de los modelos no es exhaustivo, en este capítulo, hemos optado por centrarnos sobre todo en cuestiones teóricas, dejando para el Capítulo IX las de tipo más práctico.

Comenzamos reconsiderando qué objeto tiene la evaluación y la cuestión del centro de atención (es decir, qué evaluar), antes de pasar a estudiar el "cómo" de la misma. Expondremos los puntos teóricos y metodológicos fuertes y débiles de una serie de enfoques, examinando las distinciones entre investigación y evaluación y entre descripción y juicio.

Qué objeto tiene la evaluación

En la Primera parte, examinamos un conjunto de contextos en los que la evaluación parecía desempeñar un papel. Indicamos que diversas formas de evaluación facilitan una modalidad de rendición de cuentas (Capítulo Primero), promueven el desarrollo profesional y el perfeccionamiento institucional (Capítulo II) y facilitan la revisión curricular (Capítulo III). En la Segunda Parte, examinamos cómo las estrategias de evaluación tratan de cumplir aquellos objetivos. En distintos momentos señalamos que una estrategia que satisface un objetivo (p. ej., el de rendición de cuentas) no tiene por qué cumplimentar otros (p. ej., el desarrollo profesional). Tenemos que considerar, en consecuencia, si es posible definir en general para qué sirve la evaluación.

La mayor parte de las definiciones de evaluación se derivan de la evaluación de programas y se centran en la obtención de información para los responsables de las decisiones. Así:

...podemos definir la "evaluación" en general como la recogida y utilización de información con el fin de tomar decisiones respecto a un programa educativo.

(CRONBACH, 1963. pág. 672.)

¹ D. TAYLOR (1976) presenta una revisión muy breve y legible de los modelos de evaluación. Puede consultarse otra en LAWTON (1980).

La evaluación educativa es el proceso de delimitar, obtener y facilitar información útil para juzgar alternativas de decisión.

(STUFFLEBEAM y cols., 1971, pág. 43.)

La evaluación es el proceso de concebir, obtener y comunicar información para orientar las decisiones educativas en relación con un programa especificado.

(MACDONALD, 1973, págs. 1-2; citado en STENHOUSE, 1975, pág. 112.)

La evaluación del curriculum es el proceso de delimitar, obtener y facilitar información útil para tomar decisiones y hacer juicios sobre los currícula.

(DAVIS, 1981, pág. 49.)

Es evidente que las definiciones de la evaluación han sufrido pocos cambios durante los últimos veinte años. Lo que ha cambiado es la concepción de quiénes son los responsables de las decisiones. Aunque CRONBACH (1963) señalaba tres tipos de decisiones a cuyo fin se utilizaba la evaluación (es decir, perfeccionamiento de cursos, decisiones sobre individuos y reglamentación administrativa), la preocupación existente, tanto en los Estados Unidos como en el Reino Unido, en relación con la evaluación de programas durante los años sesenta y primeros setenta, ponía de manifiesto la identificación de los responsables de las decisiones con los patrocinadores de programas o equipos directivos de proyectos. Con la evaluación de proyectos curriculares, como el *Humanities Curriculum Project* (HCP), el centro de atención comenzó a cambiar, empezando a considerarse a los *consumidores* de programas educativos, es decir, a profesores y escuelas, como los responsables clave de las decisiones. En último término, la puesta en práctica de las innovaciones dependía de su juicio y acción. Hoy día la situación es aún -más compleja y el papel del responsable de las decisiones puede recaer en miembros de grupos muy diferentes, según los contextos concretos de evaluación. De este modo, en contextos de rendición de cuentas, pueden considerarse responsables de decisiones los políticos, los administradores, los inspectores, el electorado, los padres y los *governors*, pues todos ellos tienen la capacidad de limitar o aumentar los recursos y el apoyo que reciban las escuelas. No obstante, en términos de revisión del *curriculum*, los principales responsables de las decisiones son el director y el profesorado, mientras que, en el contexto del desarrollo profesional, es probable que las decisiones sobre la práctica correspondan a profesores individuales o a grupos de profesores.

La importancia concedida a la decisión en cuanto objetivo principal de la evaluación presupone que hay algo que exige el juicio y la acción. Esto nos hace volver un paso atrás en nuestra exposición: hay que decidir cuál es el problema. Como señala REID (1978), cuando se piensa en el *curriculum* y en la evaluación, con frecuencia se deja de lado esta cuestión. En algunos contextos, el problema no será muy específico, pudiendo percibirse como la necesidad de revisar la práctica con el fin de garantizar que todo se ajusta a como debe ser. Si los responsables de las decisiones quedan satisfechos, su única decisión consistirá en animar a seguir en la misma línea'. Algunas actuaciones relacionadas con la rendición de cuentas y la revisión curricular pueden concebirse en estos términos, aunque es difícil que la práctica satisfaga las aspiraciones en todos los aspectos. En la mayoría de los casos, "es posible mejorar".

Otras evaluaciones se centran en cuestiones específicas; un enfoque habitual en el contexto de las inspecciones formales (véase el Capítulo IV), en el del desarrollo curricular y en el de la formación permanente. Sin embargo, si las necesidades se convierten en el centro de atención, para comenzar la evaluación hará falta algún tipo de instrumento de análisis de necesidades, sobre todo si la actividad corre a cargo de los miembros de la institución (evaluación interna). Nos basamos para decir esto en el *Schools and In-Service Teacher Education* (SITE) *Evaluation Project*, que supervisó los considerables problemas que experimentaron las escuelas al desarrollar los programas de formación permanente basados en ellas mismas. Atendiendo al análisis de necesidades, afirma BAKER (1980):

Los métodos utilizados en la mayoría de los casos carecían de complicaciones: no se utilizó ningún modelo formal de técnica de análisis de necesidades (cf. HOUSTON y cols., 1978). A partir de las respuestas a los cuestionarios y a las entrevistas, los evaluadores (del SITE) concluyeron que en muchas escuelas el análisis de necesidades se realizó de manera superficial, dudándose de la validez de las mismas. Esto se puso de manifiesto en el tipo de peticiones remitidas a los coordinadores del proyecto, en las que a menudo sólo se indicaban áreas generales (p. ej., matemáticas o lenguaje), sin tratar de especificar detalles.

(BAKER, 1980, pág. 184.)

En esta cita, BAKER alude a la fuente de un modelo de análisis de necesidades (HOUSTON y cols., 1978). Otra puede ser una adaptación de la Nominal Group Technique (HEGARTY, 1977; O'NEIL, 1981). Esta técnica puede utilizarse para descubrir y clasificar problemas en orden de importancia, uno de los cuales puede estar constituido por las necesidades. También puede utilizarse para recoger datos facilitados por los profesores, así como para la deliberación sobre el curriculum y la revisión del mismo. (Los detalles de esta técnica aparecen en el Capítulo VIII.) No obstante, la característica más significativa de esta técnica es el procedimiento que permite

asegurar que las prioridades sean las del grupo total y no sólo las de los más vociferantes o poderosos. En otras palabras, pretende proteger la expresión de los puntos de vista individuales en el proceso de descubrimiento y formulación de una perspectiva conjunta. Entre los esquemas de las LEA del Reino Unido, el documento de Solihull (Solihull LEA, 1979) presenta un enfoque de la evaluación escolar que se orienta de alguna manera hacia este tipo de análisis. Se invita a los profesores a que clasifiquen los diversos aspectos de su trabajo escolar según una escala de prioridad de necesidades. Se les anima a continuación a que utilicen sus valoraciones personales como base para el diálogo y la planificación conjunta (véase también el Capítulo VI).

Qué evaluar

Una vez identificadas las necesidades o cuestiones normativas, el paso siguiente consiste en decidir qué aspectos del curriculum o de la organización son más adecuados para proporcionar la información más pertinente para adoptar decisiones. En el Capítulo V ya mencionamos la cuestión del centro de atención (o sea, qué evaluar), porque los enfoques de la evaluación interna se distinguen por su respuesta a la misma. A este respecto, los datos (JAMES, 1982) indican que las evaluaciones internas pueden agruparse grosso modo en dos tipos: las que se centran en los objetivos educativos, antecedentes (condiciones anteriores) y resultados, y las que se ocupan, ante todo, de los procesos educativos. Puede establecerse una distinción semejante respecto a otras estrategias de evaluación. Por ejemplo, los exámenes y los programas de pruebas de rendimiento se centran exclusivamente en los resultados de los alumnos, mientras que muchas listas de comprobación de las LEA parecen inclinarse hacia la evaluación de los procedimientos (o sea, condiciones antecedentes; véase Open University 1982, Bloque 2, 1º parte). Sin embargo, tradicionalmente, las inspecciones se ocupan tanto de los procesos como de los resultados y de los procedimientos internos de las escuelas.

HOLT (1981) observa variaciones semejantes en cuanto a los centros de atención. En las autoevaluaciones escolares, tanto las de iniciativa interna como externa, distingue tres tipos de evaluación: de resultados, de procedimientos y de procesos. El principal supuesto subyacente a las evaluaciones de resultados y de procedimientos consiste en que es posible especificar de antemano los objetivos, cuyo cumplimiento constituirá la medida de la calidad de enseñanza. Los objetivos pueden encuadrarse en relación con los resultados del aprendizaje de los alumnos o en términos de lo que “debe” ocurrir si la escuela funciona adecuadamente. Aunque el término “objetivo” suele utilizarse para aludir a la conducta que se pretende desarrolle el alumno, ambas formas de evaluación requieren la especificación antecedente de las intenciones.

Por supuesto, hay quienes creen, como SHIPMAN (1979), que la única forma de actuar consiste en formular objetivos de trabajo. Afirma que las tareas de enseñanza y aprendizaje y, en general, la oferta educativa, deben definirse de manera que adquieran un carácter operativo. Es más, esas definiciones han de incluir, por necesidad, una indicación de lo que se considera éxito y fracaso. Más adelante, estos indicadores se convierten en los criterios y normas que sirven para juzgar sistemáticamente los *curricula*.

La sencilla lógica del modelo de evaluación curricular por objetivos es engañosa porque es difícil negar la idea de que la enseñanza ha de juzgarse según el grado en que se acerca al cumplimiento de las intenciones especificadas. Nadie puede negar que el trabajo de profesores y escuelas debe tener metas y objetivos; en realidad, este supuesto está recogido en la *Education Act de 1980*. El problema está en si la mejor manera de evaluar el *curriculum* consiste en la evaluación de los resultados y procedimientos en relación con las intenciones u objetivos prefijados. Como señalamos en el Capítulo V, hay quienes dicen que esto no es posible. No pretendemos resumir aquí todos sus argumentos, sino señalar de nuevo sólo los que estimamos de especial importancia. En primer lugar, podemos considerar que los procesos curriculares tienen un valor intrínseco. Por eso, merecen que se les preste atención por sí mismos, o sea, sin referencia a objetivos o resultados. Si, como afirman JENCKS y cols. (1972), la enseñanza es ineficaz a la hora de mejorar las oportunidades de quienes se encuentran en desventaja, desde el punto de vista cultural, quizá sea importante evaluar si las experiencias escolares de los alumnos son, al menos, agradables y satisfactorias de por sí (cuestión que indica CLIFT: *Open University* 1982, Bloque 2, 1º parte, pág. 21). En segundo lugar, como señala EISNER (1979):

Los resultados de los programas educativos no son completamente previsibles, por lo que la evaluación exclusiva de los objetivos pretendidos puede llevar a desestimar otros de igual y, a veces, mayor importancia. La utilización de los objetivos como criterios de evaluación de los programas educativos cumple una función importante, pero, desde el punto de vista educativo, una concepción de la evaluación que se limite a lo planeado de antemano en términos de metas u objetivos es poco consistente.

(EISNER, 197, pág. 174.)

Por último y de mayor importancia, mientras que la utilización de los objetivos como criterios de evaluación permite emitir juicios en términos de éxito y fracaso, es incapaz de ayudar a diagnosticar por qué razones un curriculum aboca al éxito o al fracaso. En otras palabras, el modelo de evaluación por objetivos es incapaz de

proporcionar el tipo de pruebas a partir de las cuales puede desarrollarse el curriculum. El único tipo de acción que puede desencadenar consiste en mantener o no un programa o práctica existente. En consecuencia, de acuerdo con STENHOUSE (1975), “la crítica fundamental contra el modelo de objetivos es que valora sin explicar... Por eso, quien pretende desarrollar el curriculum no puede aprender nada de él” (pág. 120).

Como indicamos en el Capítulo V, estas críticas contra el modelo de objetivos han dado lugar a diversas alternativas que pueden denominarse en sentido amplio modelos de evaluación de “procesos” (p. ej., SIMONS, 1981). Éstos dirigen la atención al “curriculum en acción” (ELLIOTT, 1979; Open University, 1981 b). Sin embargo, este cambio del centro de atención no evita el problema de establecer criterios mediante los que juzgarlos “procesos” eficaces. A falta de indicadores de calidad de enseñanza empíricamente validados y aceptados de manera general (cuestión que mencionamos en el Capítulo VI), los evaluadores que opten por centrarse en los procesos vuelven a utilizar con frecuencia las intenciones prefijadas (si no objetivos más específicos) como referencias para examinar las experiencias concretas de aprendizaje². En consecuencia, no es raro que las “llamadas” evaluaciones de procesos se ocupen fundamentalmente de los “desequilibrios en la actuación” o de las diferencias entre las “teorías en acción” de los profesores y sus “teorías aceptadas” (véase ELLIOTT, 1976). En cuanto tales, no difieren tanto de versiones más complejas de los enfoques de la evaluación por objetivos, es decir, de los que prestan cierta atención a lo que sucede entre input y output o los que tienen en cuenta tanto los objetivos de proceso como los de resultados³.

En 1967, Robert STAKE propuso un marco de evaluación que quizá constituya aún la respuesta más completa a la cuestión de qué evaluar. La primera y acaso más importante diferencia de la que se ocupa STAKE es la existente entre descripción y juicio (las dos actividades distintas aunque complementarias de la evaluación). Dejando por el momento de lado los juicios, atenderemos brevemente a los tipos de datos que considera importante recabar para poder describir por completo un curriculum o programa, antes de juzgar su valor. Presenta su análisis en forma de “matriz de descripción” [véase la Figura 7.1. (STAKE, 1967, pág. 533; reproducido en STENHOUSE, 1975, pág. 108; también, Open University 1982, Bloque 2, pág. 65)]. Primero, distingue tres aspectos en el proceso educativo:

1. Antecedentes, es decir, “cualquier condición existente antes de la enseñanza y el aprendizaje que pueda relacionarse con los resultados” (STAKE, 1967, pág. 528). Como ejemplos, pueden citarse los factores ambientales, los procedimientos de la escuela, los intereses de los alumnos o los aprendizajes anteriores.

2. Transacciones, como las interacciones que se producen entre profesores y alumnos, entre los alumnos mismos, entre los alumnos y los materiales y tareas curriculares o entre los alumnos y el medio físico, social y educativo.

3. Resultados, que han de interpretarse en su más amplio sentido, de manera que abarquen los “inmediatos y a largo plazo, cognitivos y conativos, personales y comunitarios”

(STAKE, 1967, pág. 528.)

En segundo lugar, en relación con cada uno de estos aspectos, propone que se recaben datos relativos tanto a las intenciones (evita las connotaciones conductuales de términos como “objetivos” y “metas”) como a las observaciones. Esto se traduce en seis categorías de datos potenciales: antecedentes pretendidos, antecedentes observados, transacciones pretendidas, transacciones observadas, resultados pretendidos y resultados observados (que constituyen las seis casillas de la “matriz de descripción” de STAKE).

La tarea siguiente consiste en establecer la naturaleza de la relación entre, por ejemplo, los datos correspondientes a las transacciones pretendidas y los que se refieren a los resultados pretendidos o entre los datos relativos a los resultados pretendidos y los datos acerca de los resultados observados. Con el fin de ayudar a efectuar este tipo de análisis, STAKE introduce otros dos conceptos: contingencia y congruencia. Es más, distingue la contingencia lógica de la contingencia empírica. La contingencia lógica tiene que ver con la valoración de la medida en que las intenciones relacionadas con los antecedentes (p. ej., los supuestos acerca de los aprendizajes anteriores de los alumnos) son coherentes, desde el punto de vista lógico, con las referidas a las transacciones (p. ej., el tipo de tareas de aprendizaje planeadas) y con los resultados que se pretenden

² Sin embargo, al menos un autor (SCRIVEN, 1971) afirma que la evaluación no debe contar con los objetivos. Según él, es probable que el conocimiento de los objetivos que se prevé que consiga un programa determinado deforme la percepción del evaluador y le impida percatarse de efectos imprevistos que pueden constituir los resultados más significativos del programa. No obstante, su insistencia en los efectos convierte a su enfoque en otra versión de la evaluación de productos. Es interesante observar que SCRIVEN afirma también que la emisión de juicios es responsabilidad del evaluador. Sin embargo, a falta de enunciados relativos a los objetivos del programa, tiene que buscar criterios alternativos mediante los cuales elaborar sus juicios. La solución de SCRIVEN a este problema consiste en proponer una ‘lista de comprobación letal’ de trece necesidades de un programa que cree independientes de los valores de los destinatarios. El incumplimiento de cualquiera de esos puntos hace dudar de la calidad del programa. En la evaluación escolar, la objeción contra este enfoque es evidente: los criterios de la escuela se sustituyen por los de la persona que realiza la evaluación, sin ninguna garantía de que tales criterios sean objetivos.

³ EISNER (1969) establece una distinción entre los “objetivos instructivos” y los “objetivos expresivos”. Dice que es perfectamente aceptable que los profesores planeen actividades que proporcionen a sus alumnos oportunidades para aumentar su experiencia y permitirles opciones personales. Sin embargo, como los resultados de esas actividades son imprevisibles por naturaleza, EISNER cree que hace falta un nuevo término. La noción de “objetivos expresivos” se adapta especialmente a las artes (EISNER es profesor de arte); una alternativa tan adecuada como ésta, con un campo de referencia algo más amplio, puede consistir en los “objetivos de proceso”.

(p. ej., los resultados previstos). A menudo, este tipo de valoración requiere alguna forma de análisis filosófico. Como indicamos en el Capítulo III, a veces, esta operación se denomina evaluación intrínseca, expresión utilizada por SCRIVEN (1967) para aludir a un juicio sobre la estructura de un programa educativo en sí mismo; por ejemplo, su diseño y supuestos previos. De igual modo, la valoración de la contingencia empírica puede denominarse como evaluación empírica. En este caso, el análisis de contingencia entre antecedentes y transacciones y entre transacciones y resultados se basa en los datos de las observaciones (es decir, en la

Datos descriptivos

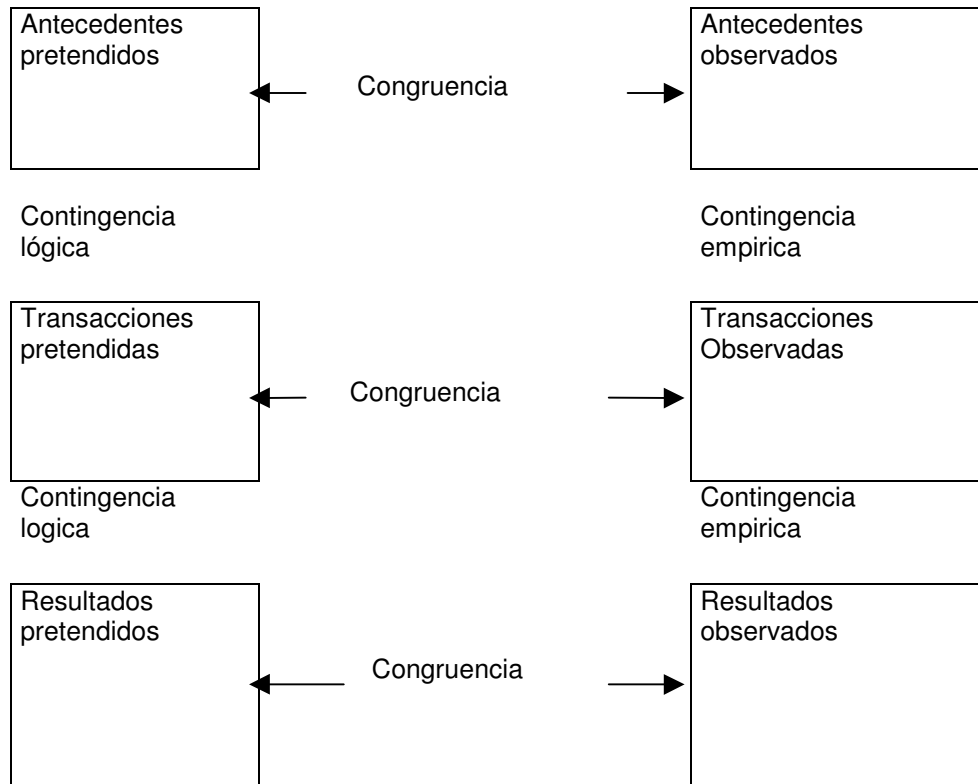


Figura 7.1.:Representación del procesamiento de datos descriptivos (STAKE, 1967, página 533)

evidencia empírica). Por ejemplo, pueden examinarse los procesos de clase en relación con: 1) la medida en que cumplen las previsiones derivadas de los datos de los aprendizajes anteriores de los alumnos, y 2) el grado en que los resultados pueden considerarse imprevistos.

En contraste con esta descripción de contingencia, la idea de congruencia de STAKE se refiere a la relación entre las intenciones y las observaciones. En concreto, supone el análisis del grado en que se cumple realmente lo que se pretendía. Así, las transacciones observadas pueden compararse con las planeadas; los resultados observados pueden evaluarse en relación con los previstos, y las condiciones antecedentes reales pueden compararse con las que se suponían.

En pocas palabras, el marco de STAKE nos ofrece un cuadro casi exhaustivo de los centros de atención adecuados para la evaluación. En cierto sentido, incorpora aspectos de los modelos de “objetivos” y de “proceso” porque, aunque STAKE concede gran importancia a los “procesos” (transacciones), no rechaza los datos de entrada ni del producto. (A diferencia de SCRIVEN, 1971, considera importantes las metas o intenciones de los profesores o realizadores de programas.) Sin embargo, por su misma estructura, su marco presume que las intenciones son explícitas, o pueden ponerse de manifiesto con facilidad. Parece, por tanto, pasar por alto las intenciones “ocultas” o las “teorías en acción” (ARGYRIS y SCHÓN, 1974) que sólo pueden ponerse de manifiesto mediante un cuidadoso estudio de la práctica. Es necesario centrarse en éstas como paso previo a la toma de conciencia que, a menudo, es el principal fundamento de la evaluación de procesos.

Cómo evaluar

Aunque STAKE distingue seis tipos de datos pertinentes para la evaluación educativa y diferentes formas de análisis, se muestra a favor de un enfoque holista más que atomista. En otras palabras, cree que lo verdaderamente importante son las relaciones dinámicas entre los diversos elementos de un programa educativo. Es más, con el fin de representar toda la complejidad de cualquier experiencia educativa, considera esencial contar con una “amplia base de datos”, idea que apoyan PARLETT y HAMILTON (1972) cuando

defienden su “evaluación iluminadora” y SIMONS (1981) en sus propuestas para la evaluación de los procesos en las escuelas (véase el Capítulo V). De acuerdo con esta perspectiva, es permisible, y deseable incluso, basarse en todo el conjunto de fuentes de datos, desde las pruebas de rendimiento normalizadas, hasta las entrevistas no estructuradas y la observación participativa, pasando por los esquemas de observación “sistemática”, los inventarios y los cuestionarios precodificados. EL criterio principal para su selección debe consistir en su capacidad para informar los juicios acerca de una actividad determinada. En general, este tipo de eclecticismo resulta más aceptable en el contexto de la evaluación que en el de la investigación “pura”. Se discute la distinción entre ambos, pero la evaluación suele diferenciarse de la investigación por su orientación a la solución de problemas prácticos, frente al interés por la elaboración teórica de la investigación pura. En consecuencia, los métodos específicos se escogen más por su potencial utilidad en situaciones en las que hay que tomar decisiones que por su concordancia con un paradigma concreto de investigación ⁴. Sin embargo, esto no quiere decir que los evaluadores puedan permitirse ser menos rigurosos en su enfoque. Para que la evaluación que se realiza en las escuelas tenga credibilidad pública y para que proporcione un fundamento para la acción, es preciso que sea fehacientemente válida, fiable y, en la medida de lo posible, carezca de inclinaciones tendenciosas.

El debate sobre la distinción entre evaluación e investigación resalta el hecho de que la bibliografía de investigación ha influido en los enfoques de evaluación; en realidad muchas técnicas específicas utilizadas en evaluación (véase el Capítulo VIII) se derivan de la investigación. La diferencia consiste en que la elección de una técnica de evaluación no tiene porqué estar guiada por la teoría, factor que permite el tipo de eclecticismo al que nos referimos antes. Dadas las relaciones metodológicas entre ambas perspectivas, es conveniente dedicar algún espacio para considerar las principales características de los diferentes enfoques de investigación.

Consideraciones teóricas

Las investigaciones recientes, tanto en las ciencias sociales como en las de la educación, se han caracterizado por una serie de estilos derivados de dos perspectivas fundamentalmente diferentes: la que podemos denominar enfoque positivista y la que llamamos enfoque interpretativo. El enfoque positivista supone que el mundo social, como el natural, está regido por leyes que pueden investigarse mediante la observación sistemática y el experimento. Utilizando métodos científicos muy similares a los usados en las ciencias naturales, se cree posible establecer determinados hechos sociales objetivos. Por supuesto, existen serias objeciones prácticas y éticas en contra de la manipulación de individuos que requeriría la investigación, por lo que gran parte de la investigación social en forma positivista se ha orientado a la obtención de pruebas correlacionales precisas que, sin el valor predictivo de las relaciones de causa a efecto que se descubren en los experimentos controlados, poseen, sin embargo, una generalizabilidad que se alcanza mediante la inferencia estadística. El procedimiento normalizado de investigación consiste, por tanto, en la elaboración de una hipótesis a priori, la definición de categorías operacionales de observación, el desarrollo de métodos “objetivos” de obtención de datos y la realización de los análisis estadísticos adecuados. En el campo de la investigación educativa y en relación con su aplicación a la evaluación, el paradigma clásico ha recibido diversas denominaciones:

“científico” o “hipotético-deductivo”, haciendo hincapié en la comprobación de las hipótesis a priori, “nomotético”, resaltando la búsqueda de leyes generales; “experimental”, “agrícola” o “psicométrico”, indicando las relaciones metodológicas con los enfoques de investigación de las ciencias naturales, la agricultura y la psicología. En términos de la teoría social y de la educación, la insistencia positivista en la cuantificación y generalización ha llevado a crear una imagen del individuo que se distingue por su sometimiento a fuerzas poderosas que limitan su carácter único y hacen que sus acciones sean previsibles. Esto se aplica especialmente a algunos campos de la investigación social en los que las explicaciones de la acción individual tienden a situarse en el macronivel, o sea, en el seno de las estructuras interrelacionadas de la sociedad. Estas explicaciones han dado lugar a una serie de teorías sociales “estructuralistas” o “funcionalistas”.

Sin embargo, esta imagen del individuo y de su relación con el mundo social no ha carecido de oposición. Por ejemplo, entre los teóricos sociales hay quienes afirman que los sistemas sociales son creaciones significativas de los individuos por lo que no pueden reificarse, es decir, tratarse como objetos” con vida independiente de sus miembros. (Esta es una objeción fundamental en contra de la concepción de la “escuela creativa” de HOYLE, 1975; véase el Capítulo II.) En consecuencia, habrá que considerar los fenómenos sociales como “acciones” que deben interpretarse y comprenderse en términos de intenciones, motivos, razones, etc., y no como “hechos” que puedan contarse, explicarse o preverse. Esta alternativa al paradigma positivista suele llamarse con frecuencia “interpretativa” (y a veces, “humanista”): término genérico que engloba un conjunto de enfoques cuya principal diferencia radica en sus diferentes puntos de vista sobre el acto de interpretación. Así, tanto la fenomenología, el interaccionismo simbólico, la etnometodología como la filosofía analítica (enfoques

⁴ Sin embargo, hay cursos, como el DE 304: “*Research Methods in Education and Social Sciences*” (Open University, 1979), que apoyan también la postura de que los métodos de investigación deben seleccionarse de acuerdo con su adaptación al objetivo que se pretende.

desarrollados en sociología, antropología y filosofía, algunos de los cuales están relacionados) hacen hincapié en la necesidad de tener en cuenta lo que es o hace una persona, en relación con lo que ella o sus iguales *piensan* que es o hace⁵. Por eso, la investigación tiende a ser *idiográfica*, tratando de descubrir las características exclusivas del caso concreto (de ahí el “estudio de casos”), en vez de establecer leyes generales en relación con la conducta (como en la investigación nomotética). Este cambio de centro de atención favorece también una alternativa respecto al modo hipotético-deductivo de elaboración teórica. En vez de ocuparse de la “comprobación” de hipótesis a priori, se hace mayor hincapié en la generación de hipótesis durante el proceso de investigación. Por supuesto, toda la investigación científica “inductiva” atraviesa este proceso en algún momento, pero, en los enfoques interpretativos, la generación de hipótesis y el “descubrimiento de una teoría fundamentada” (GLASAR y STRAUSS, 1967) son relativamente más importantes que la comprobación de hipótesis a efectos predictivos. En otras palabras, se considera más importante la comprensión que la previsión.

Aunque el paradigma interpretativo pueda resultar muy atractivo a quienes no simpatizan con el positivismo, también ha recibido críticas. La insatisfacción con el conservadurismo implícito de los enfoques interpretativos que, al centrarse en la forma de verse las personas a sí mismas y a las demás, tienden a dejar de lado la cuestión relativa a lo que *pueden llegar a ser*, ha estimulado a ciertos científicos y filósofos sociales (sobre todo a los de la escuela de Frankfurt de Teoría Crítica) a señalar la necesidad de interpretaciones que critiquen de forma activa y trasciendan las propias autocomprensiones de las personas. El enfoque sigue siendo interpretativo o humanista en la medida en que trata de mejorar la comprensión de los *significados* que las personas atribuyen a su existencia, pero la *explicación* no se queda en el *micronivel*, es decir, en el del individuo o el grupo. Pretende, en cambio, relacionar los niveles micro y macro de análisis. En otras palabras, el análisis de los constructos personales de los individuos se somete a un tipo de meta-análisis que suele llevarse a cabo con referencia a alguna macroteoría, como la teoría social marxista o la teoría psicoanalítica freudiana (algunos miembros de la escuela de Frankfurt eran psicoanalistas). De este modo, los teóricos críticos demuestran su preocupación activa por desafiar los supuestos de los que parten las personas, suscitando la consciencia de los mismos, con el fin de aumentar su grado de autonomía. Según Jürgen HASERMAAS (1972), que quizá sea el miembro más influyente de la escuela crítica, este enfoque, a diferencia de los demás, se orienta a la acción en la medida en que está informado por un “interés emancipador”.

En cierta medida, todos estos enfoques de la investigación social han influido en el desarrollo de planteamientos generales de la evaluación educativa. Gran parte de la primitiva evaluación de programas, sobre todo en los Estados Unidos, se realizó de acuerdo con el modelo clásico “científico” o “experimental”. La elegancia del diseño de investigación de FISHER (1953) resultaba eminentemente atractiva, de modo que la respuesta obvia a la pregunta de cómo evaluar parecía ser: administrar pretests a los alumnos, someterlos a distintas experiencias y medir después los resultados mediante postests. En la actualidad, están muy claros los problemas prácticos y éticos derivados del “auténtico” método experimental, pero aún hay quienes defienden un enfoque “científico” de amplia base

(BYNNER, 1980)⁶

En 1972, PARLETT y HAMILTON evidenciaron los defectos del paradigma experimental, acusando a los tradicionalistas de dar por sentado que los programas educativos diferían poco de los “tratamientos” agrícolas u hortícolas, pudiendo, en consecuencia, evaluarse de la misma manera. Defendiendo que la evaluación debía ocuparse más de describir e interpretar que de medir y prever, propusieron un enfoque socioantropológico. Denominaron a este tipo de evaluación “iluminadora”, porque pensaban que las decisiones sólo podían tomarse eficazmente sobre la base de una profunda comprensión de la forma de operar los *currícula* en situaciones específicas. Consideraban especialmente importantes dos tipos de comprensión: la del *sistema instructivo*, es decir, el modo de relacionarse los planes curriculares formalizados con las situaciones docentes concretas, y la comprensión del *medio de aprendizaje*, o sea, el complejo ambiente sociopsicológico y material en el que trabajan los alumnos. En consecuencia, el cometido de la evaluación iluminadora consistía en revelar las interacciones internas del sistema instructivo y del medio de aprendizaje y las que se producen entre ambos términos.

Como es característico en los medios académicos, la formulación de la “evaluación iluminadora” de PARLETT y HAMILTON (1972) también recibió muchas críticas. Por ejemplo, PARSONS (1976) juzga el enfoque en relación con diversos aspectos. *En primer lugar*, y quizá de forma bastante injusta, dice que los defensores de la evaluación iluminadora no reconocen las relaciones de la misma con la ciencia social interpretativa y, al hacer hincapié en su utilidad para la adopción de decisiones, tratan de evitar la crítica en cuanto investigación.

⁵ Véase en las unidades 7 y 8 Open University (1977) un informe sobre el enfoque etnográfico del análisis de la enseñanza, y en Open University (1979) informes más completos sobre las diferencias entre los enfoques de investigación experimental, de encuesta y etnográfico.

⁶ BYNNER (1960) se manifiesta a favor de un enfoque de investigación que abarque orientaciones diversas, combinando elementos cualitativos y cuantitativos sin necesidad de un control aleatorizado. (En CAMPBELL y STANLEY, 1966 puede verse una formulación clásica.) Hay pruebas de que este modelo se utiliza aún en la evaluación de programas educativos. Por ejemplo, los módulos curriculares elaborados por el *Dental Health Study* del *Health Education Council* se evaluaron de este modo (en 1982-1984). El director del *Study* afirmaba que esta forma de evaluación gozaba de la máxima credibilidad ante los patrocinadores del programa, muchos de los cuales sólo estaban familiarizados con los modelos “científicos” o “médicos”.

En segundo lugar, y de forma más contundente, dice que la concentración casi exclusiva método, sobre todo en la recogida de datos, ha contribuido a la omisión casi completa de consideraciones sobre el papel de la teoría formal. En la medida en que las teorías existentes pueden ayudar a “dar sentido” a los datos en determinadas circunstancias, cree que deben tener un lugar en el enfoque iluminador. En lugar, señala que la importancia otorgada al hecho de facilitar información responsables de las decisiones “aquí y ahora” es implícitamente conservado aceptar la perspectiva de los participantes, los evaluadores corroboran el a *que* al que quizá debieran oponerse. Por tanto, en este sentido, los enfoques iluminadores de la evaluación pueden estar vinculados al contexto y ser ahistóricos.

En la crítica de PARSONS hay elementos que recuerdan mucho la crítica los teóricos críticos preocupados por la *causalidad* hacen de algunos enfoques *exclusivamente* interpretativos. En consecuencia, es interesante señalar que la evaluación educativa, así como en la investigación social, ha surgido una tercera tradición, dentro de la cual un conjunto de especialistas en educación están empezando a promover el concepto de la investigación-acción educativa como forma de dotar de teoría a las formas interpretativas de evaluación, admitiendo por tanto, la posibilidad de emancipación (KEMMIS, 1981a; BROWN, HENRY, HEENRY y McTAGGART, 1982; ELLIOTT, 1982b)⁷. (Como el núcleo característico de la investigación-acción es la acción misma, más que las técnicas específicas recogida y análisis de datos, nos ocuparemos de ella con más detenimiento en el capítulo XI.)

En este mismo capítulo, hemos indicado que los enfoques de la evaluación en concreto, la elección de los métodos, podían ser eclécticos porque, a diferencia de la investigación social, la evaluación no depende de la teoría. A la luz de la exposición anterior, este supuesto se hace problemático. Sin duda, los científicos sociales dirían que, en la medida en que los enfoques de la evaluación se derivan de la ciencia social, llevan consigo cierto bagaje teórico *implícito*. Así, cualquier intento de combinar distintos enfoques puede equivaler a procurar la reconciliación de puntos de vista fundamentalmente opuestos sobre la relación entre individuo y la sociedad, por ejemplo. En respuesta a esa objeción, debemos que, aunque nuestra postura pueda crear dificultades en el nivel de interpreta de los hallazgos, no provoca problemas graves en la etapa de recogida de datos. Así, podemos defender el enfoque ecléctico en el sentido de que el eclecticismo es de tipo *metodológico* más que teórico y pocas metodologías mantienen relación *necesaria* con perspectivas teóricas concretas. El positivismo suele primar los métodos cuantitativos, mientras que los enfoques interpretativos se inclinan a favor de los métodos cualitativos, aunque hay excepciones. Por ejemplo, aunque gran parte de la investigaciones arqueológicas, paleontológicas e históricas son positivistas, a menudo se ven obligadas a utilizar pruebas de casos únicos porque éstos son los disponibles. Como los “problemas” que trata de resolver la evaluación educativa son los propios de la normativa o de la práctica, cualquier método que ilumine la cuestión puede ser útil y. por regla general, “cuanto más, mejor”. A este respecto, es significativo que PARLETT y HAMILTON (1972) sugiriesen recabar datos mediante diversos métodos, incluyendo los que suelen relacionarse con el enfoque experimenta⁸. Aun entre los científicos sociales preocupados sobre todo por el desarrollo de la teoría formal (o sea, realizando investigación “pura”), hay quienes defienden decididamente un enfoque multimetódico. Por ejemplo, WEBB y cols. (1966) dicen que es difícil que alguna medida pueda abarcar todos los parámetros. Afirman, en cambio, que “la evidencia más persuasiva proviene de la triangulación de los procesos de medida” (pág. 3). Como veremos, la triangulación constituye un procedimiento de validación útil porque permite que los datos obtenidos mediante un método se contrasten con los recogidos de otra forma. En el próximo apartado nos ocuparemos con mayor detenimiento de algunos problemas relacionados con la elección de metodología.

Consideraciones metodológicas

Como se desprende de la exposición anterior, gran parte del debate que rodea los enfoques teóricos de la investigación y la evaluación tiene que ver con el tipo de hallazgos que se consideren más importantes. El enfoque clásico “científico” o “experimental” aspira a ser nomotético, haciendo hincapié en la necesidad de establecer reglas o leyes generalizables. Por otra parte, los diversos enfoques “interpretativos” o “iluminadores” tienden a primar el modo idiográfico de explicación porque tratan de describir y explicar las circunstancias exclusivas del “caso” único. En consecuencia, las ideas de tipicidad o representatividad tienen menor importancia y el único tipo de generalización adecuado consiste en lo que ocurre cuando parece pertinente la aplicación de las hipótesis generadas en un estudio de casos” a otro (una forma de efecto de “bola de nieve”). Estas diferentes concepciones de lo que trata de conseguir la investigación o la evaluación dan lugar a distintos puntos de vista respecto al tipo de datos que conviene recabar. Así, aunque todos los métodos derivados de la investigación social y educativa comparten algunos aspectos, como la concepción del muestreo, del análisis y

⁷ En las ciencias sociales, la investigación-acción tiene una larga historia. La inició Kurt LEWIN en los Estados Unidos a mediados de los años cuarenta. No obstante, KEMMIS (1981a) señala que hay infinidad de pruebas que indican que LEWIN estaba al corriente de los trabajos de la teoría crítica se estaban realizando en Alemania en la misma época. En otras palabras, la relación entre ambos desarrollos puede ser tanto cronológica como ideológica.

⁸ PARLETT y HAMILTON (1972) instan a recabar datos sobre cuatro áreas: 1) observaciones; 2) entrevistas; 3) cuestionarios y pruebas, y 4) fuentes documentales y ambientales.

de los criterios de validez metodológica, el tratamiento de estos aspectos del proceso de investigación difiere considerablemente. A este respecto, la distinción entre métodos o técnicas cuantitativos y cualitativos es fundamental.

En la investigación cuantitativa, la representatividad, que es el criterio de generalización de los descubrimientos, se consigue de diversos modos: estudiando la totalidad de la población objetivo (p. ej.: todos los chicos y chicas de 16 años que estudian ciencias) o seleccionando una muestra cuyas características reflejen o representen las de la población objetivo. La selección de una muestra puede realizarse también de dos maneras: mediante una auténtica selección aleatoria (como en una tómbola) o asegurándose de que aparezcan en la muestra determinadas variables importantes, mediante la subdivisión de la población en grupos adecuados (p. ej.: niños de distintos grupos socioeconómicos o escuelas de diferentes regiones geográficas) antes de efectuar la selección aleatoria. En la investigación educativa, es corriente encontrar este último tipo de muestras, conocidas como *muestras estratificadas*. Es obvio que el muestreo no tiene por qué aplicarse sólo a personas o instituciones. En la observación estructurada de niños durante un tiempo determinado, se utilizan con frecuencia las variaciones de *muestreo temporal* (véase el apartado sobre “Observación directa” del Capítulo VIII); y en los procedimientos de evaluación, las preguntas de los exámenes constituyen a menudo una *muestra del contenido de la materia o tema*. En todas estas formas de muestreo, se aplica el principio de representatividad y éste, como veremos más adelante, constituye el fundamento de la “validez de contenido” de una técnica.

En la investigación cualitativa, el muestreo se realiza en el proceso de análisis y en la búsqueda de explicaciones. Cuando empiezan a surgir los conceptos, categorías e hipótesis en el estudio de un caso concreto (p. ej., la escuela, la clase o el niño), el investigador trata de encontrar datos correspondientes a otros hechos, situaciones y grupos que puedan ayudarle a aclarar, perfeccionar o contradecir sus percepciones. Como muestreo y análisis están tan entrelazados y la selección de los datos se rige por su capacidad de informar la teoría, el proceso se denomina *muestreo teórico*. Vinculada a esta idea está la de *convergencia progresiva*. Esta forma parte también del proceso de análisis de la investigación cualitativa y alude al perfeccionamiento progresivo de las categorías de observación cuando se aclara más la importancia de determinadas cuestiones. Por último, lo que los etnógrafos llaman *inducción analítica* participa también en la estructuración de la teoría a partir de la observación de diversos casos o ejemplos”. Se parece mucho al método hipotético-deductivo de la investigación “científica” clásica, pero utiliza casos únicos en vez de grupos para comprobar o “refutar” las hipótesis^{9,10}

Como en la investigación cualitativa, el análisis cuantitativo supone también la elaboración de categorías y dimensiones. De modo parecido, ambos enfoques requieren que éstas queden definidas con claridad, sean mutuamente exclusivas y, en la medida de lo posible, expliquen todos los datos de manera exhaustiva. Sin embargo, en la investigación cuantitativa, la definición de las categorías *precede* a la recogida de datos, aunque suele probarse y perfeccionarse en un estudio piloto”. Es más, el conjunto del proceso se conoce de modo más habitual como *clasificación*. Si se identifican, clasifican y, en caso necesario, calibran más de una variable correspondiente a las personas, situaciones o hechos observados, es posible llevar a cabo análisis correlacionales y calcular la “significación” estadística (la base de la inferencia estadística). Por ejemplo, en contextos de evaluación, es importante descubrir si existe alguna correlación significativa entre la categoría socioeconómica de los padres y el aprovechamiento escolar de sus hijos^{11,12}.

Como ya hemos sugerido, a la hora de decidir la conveniencia de utilizar técnicas de investigación cuantitativas o cualitativas en contextos de evaluación sobre la base de los méritos relativos de cada una, el juicio suele basarse en la importancia que se otorgue a las generalizaciones. Tanto los enfoques cualitativos como los cuantitativos permiten efectuar comparaciones entre personas y hechos, pero éste es el objetivo principal de la

⁹ En BLOOR (1978) puede verse el informe etnográfico de su intento de poner en práctica el método analítico de inducción y las dificultades que experimentó.

¹⁰ Parece que existen más informes sobre cómo recoger datos cualitativos o etnográficos que sobre a forma de analizarlos. La obra clásica de GLASER y STRAUSS (1967) sigue siendo de las más útiles, así como la de PATTON (1980). El curso DE 304: *Research Methods in Education and the Social Sciences*, de la *Open University* (1979), cuenta también con dos módulos (6 y 7) dedicados al análisis de los datos, incluyendo los datos etnográficos y la inferencia estadística. Véase también en HAMMERSLEY y ATKINSON (1983) un informe muy lúcido sobre la inducción analítica y otros métodos relacionados con la etnografía.

¹¹ Véase en CHOPPIN (1981) una exposición de problemas relacionados con el uso de procedimientos de medición educativa en contextos de evaluación.

¹² Refiriéndose a la observación participante, la principal técnica del etnógrafo, BECKER (1958) señala tres etapas en el análisis de datos que no difieren de las de la investigación cuantitativa. 1) la selección y definición de problemas, conceptos e indicios; 2) la comprobación de la frecuencia y distribución de los fenómenos, y 3) la construcción de modelos del sistema social. La inclusión de la segunda etapa estimula al investigador a hacer enunciados cuasi estadísticos como: todos los niños informaron que...’, ninguno hizo preguntas...’, ‘se observó que cierta proporción del grupo...’. Según BECKER es muy aceptable que un observador participante se base en este tipo de apoyo cuantitativo para llegar a alguna conclusión, sin embargo, también hace hincapié en que, en la investigación mediante observación participante, el análisis se realiza de forma concurrente, es decir, partes importantes del análisis se hacen al mismo tiempo que se obtienen los datos. En este sentido, su enfoque se adapta al modelo etnográfico (muestreo teórico, convergencia progresiva, inducción analítica) y difiere de la investigación mediante encuesta.

cuantificación, que aspira a formular generalizaciones que sean aplicables a la totalidad de la población objetivo. No obstante, inevitablemente, la reducción de un “caso” a un número limitado de dimensiones que puedan identificarse en otros casos tiene la consecuencia de sacrificar la investigación de lo que es exclusivo con el fin de descubrir lo que es común. En algunas circunstancias, esto puede ser deseable o necesario, sobre todo en el área de la valoración de los alumnos en la que es importante su “clasificación”. En otras situaciones ocurrirá lo contrario. Por ejemplo, en relación de nuevo con la valoración de los alumnos, el informe cualitativo de la experiencia y aprovechamiento educativos de un individuo (mediante la recopilación de un expediente de rendimiento) resultará, en ciertas circunstancias, más valioso que una relación de resultados en pruebas graduadas o en exámenes. Del mismo modo, para comprender las razones de la oposición del profesorado de una escuela a cualquier iniciativa sobre las relaciones entre escuela y comunidad, por ejemplo, quizá sea necesario estudiar la combinación específica de factores de la situación que afectan al trabajo de la escuela (p. ej., el apoyo de los padres, los recursos locales, las actitudes y conocimientos de los profesores. la organización de la escuela y el ingreso de nuevos alumnos). En otras palabras, en ciertos casos, un enfoque cualitativo será más adecuado para obtener el tipo de información que se requiere para tomar decisiones. Esto nos lleva de nuevo al punto inicial de este apartado: en el campo de la evaluación, es aceptable, y necesario incluso, que la elección de los métodos esté regida por valores utilitarios. Debe estar siempre presente el “problema” de la evaluación, de modo que las técnicas se escojan de acuerdo con su capacidad para iluminar los problemas prácticos involucrados en la acción presente y en la futura. Es obvio que esto no significa que esas técnicas puedan dejar de lado el tipo de rigor propio de la investigación. Como señalamos antes, para que la evaluación tenga credibilidad, es preciso demostrar lo que NUTTALL describe como el criterio de “aceptabilidad metodológica” (véase el Capítulo Primero y *Open University* 1982, Bloque 1, parte A).

Validez y fiabilidad

En todos los enfoques de investigación se hace hincapié en la *validez y fiabilidad* de los métodos. En otras palabras, se prevé que los investigadores demuestren que las observaciones que registran y analizan correspondan a lo que pretenden registrar y analizar. En esto consiste el concepto de validez. En segundo lugar, tienen que asegurarse de que evitan las fluctuaciones aleatorias en las observaciones que registran. En ello consiste el concepto de fiabilidad. Ambos conceptos encierran una serie de variantes o tipos de validez o fiabilidad. Como la fiabilidad es más fácil de describir, la analizaremos en primer lugar.

Esencialmente, la fiabilidad está relacionada con la *consistencia* de los resultados y se refiere al requisito de que, al menos en principio, cualquier otro investigador, o el mismo científico en otra ocasión, podrá *replicar* la investigación original, consiguiendo unos datos o resultados comparables. En consecuencia, la fiabilidad puede ser de dos tipos:

1. *Fiabilidad interjueces*, que alude al grado de acuerdo entre dos investigadores que registran sus observaciones de fenómenos iguales o similares.
2. *Fiabilidad intrajuez*, que se refiere a la consistencia de las observaciones del mismo investigador en distintas ocasiones.

Como la fiabilidad, la validez tiene que ver con los “errores” que pueden darse en el proceso de investigación. Sin embargo, en este caso, el centro de atención está constituido por los “errores sistemáticos” (o sea, sesgos), en vez de “errores aleatorios”. Aunque la validez de un método consiste básicamente en la medida en que las observaciones registradas son las que el investigador se proponía registrar, el concepto es bastante más complejo de lo que parece a tenor de este enunciado. Suelen distinguirse diversos tipos de validez:

1. *Validez aparente*. Se trata del nivel más sencillo (y débil) de validez, pues sólo requiere que una medida, etc., *parezca* que mide lo que se pretende.
2. *Validez de contenido*. Se alude a ella a veces como principio de “inclusividad” y se refiere al requisito de que los datos producidos abarquen todos los aspectos pertinentes de la cuestión de que se trate. Como señala DEALE (1975, pág. 30), constituye un aspecto muy importante de la validez para el profesor que pretenda construir pruebas de rendimiento o exámenes de clase, pues requiere que la prueba abarque de manera adecuada todo el programa.
3. *Validez referida a criterio*. Este concepto alude a la concordancia entre las puntuaciones de una prueba y algún otro criterio, como las estimaciones de capacidad de los profesores o los, resultados de un examen, por ejemplo. Este tipo de validez puede subdividirse en: a) validez predictiva, que requiere que la medida en cuestión prevea adecuadamente la conducta o los resultados, y b) validez concurrente, que requiere que la medida concuerde con otra medida de validez conocida o supuesta.

4. Validez de constructo. Se trata de un concepto muy importante en la teoría de las pruebas. En este contexto, “la validación de constructo comprueba el grado de adaptación de la prueba al conjunto de predicciones teóricas sobre una variable interviniente cuya existencia o utilidad puede no ser sino una hipótesis del creador de la prueba” (Open University 1979, bloque 6, pág. 6). En otras palabras, la medida debe reflejar el constructo (p. ej., inteligencia, ansiedad) que ocupa a los investigadores, sin que deba comprobar ninguna otra cosa. Del mismo modo, si la teoría prevé el constructo como un rasgo unitario, el análisis no debe escindir los resultados en varios factores independientes.

5. Validez interna. Esta expresión es de orden diferente a las anteriores, aunque, en algunos aspectos, las engloba. Se refiere fundamentalmente a la aceptabilidad de una explicación (o sea, silo que se interpreta como “causa” produce el “efecto”) y a la pertinencia del diseño de investigación y los instrumentos de medida para producir esta explicación.

6. Validez externa. En contraste con la “validez interna”, ésta se refiere a la capacidad de generalizar los hallazgos. También puede subdividirse. La validez poblacional refuerza la investigación mediante encuesta porque se refiere a las generalizaciones de poblaciones de personas, mientras que la validez ecológica refuerza la etnografía, dado que alude a la generalización desde un conjunto de condiciones (p. ej., el caso, el ambiente, la cultura) a otro distinto¹³.

La importancia de la validez y la fiabilidad en el contexto de investigación se funda en que aumenta la posibilidad de réplica del diseño de investigación y, por tanto, la verificabilidad de los descubrimientos de la investigación. En la investigación experimental de las ciencias naturales, el proceso de réplica desempeña un papel central; los avances en el conocimiento científico se logran tanto a través de la refutación de las teorías aceptadas hasta el momento como por el descubrimiento de nuevas teorías¹⁴.

En la educación y en las ciencias sociales, la réplica es más difícil. La conducta humana nunca es estática y, como el control de ciertas variables no sería ético ni natural, ningún estudio puede replicarse exactamente, con independencia de los métodos o diseños de investigación que se utilicen. Sin embargo, sigue en pie la necesidad de que los descubrimientos sean creíbles, lo que ha estimulado a los investigadores a buscar otros medios de conseguir que su trabajo sea válido y fiable. Aunque sea una simplificación exagerada, podemos decir que el establecimiento de la *validez* de los resultados constituye el principal problema de los enfoques cuantitativos, mientras que los problemas de *fiabilidad* predominan en los enfoques cualitativos.

Un procedimiento de validación de medidas cuantitativas, como los exámenes, consiste en pedir a ‘expertos’ que juzguen qué destrezas, objetivos o contenidos mide cada pregunta. Se trata de una validación de contenido y se consigue sobre todo mediante un análisis intrínseco. Por otra parte, las pruebas normalizadas se basan ante todo en la validez de constructo, que se consigue mediante la demostración de que la medida obtenida concuerda con las previsiones derivadas de la teoría en relación con las propiedades del constructo.

¹⁵ Como en el caso de la fiabilidad, esa demostración se realiza reuniendo muchas valoraciones o medidas de los fenómenos sometidos a observación, permitiendo un margen estadístico de error. Sin embargo, ni siquiera estas precauciones erradican todas las fuentes de error o de desequilibrio. Consideremos, por ejemplo, la administración de una prueba “de papel y lápiz”, que se supone validada, a un grupo de niños de enseñanza primaria cuyas clases se imparten de modo informal. El procedimiento de sentarlos solos y en silencio, dando sus respuestas por escrito puede resultarles tan extraño que desvirtúe gravemente tanto su ejecución como el tipo de *curriculum* que han experimentado en realidad. En otras palabras, la deformación provocada por el ambiente artificial puede invalidar los resultados obtenidos mediante un instrumento que, por lo demás, cumple todos los requisitos de la teoría clásica de las pruebas.

En los enfoques cualitativos, el hecho de que sólo se dé un “caso” del fenómeno que se investiga y que su aparición pueda ser fugaz, dificulta la aplicación de una batería de medidas. Sin embargo, la idea de que, a partir de un acontecimiento único, pueden recogerse distintos tipos de datos y que éstos pueden utilizarse posteriormente para contrastarlos con otros casos, constituye el fundamento del principal procedimiento de validación utilizado en la investigación cualitativa, que se conoce como *triangulación*.

De acuerdo con algunos investigadores que utilizan técnicas cualitativas y, sobre todo, etnográficas, la triangulación se ve afectada por problemas de validez (véase BLOOR, 1978). Por ejemplo, no existe garantía

¹³ En Open University (1979, bloque 8, “Evaluation of Research” aparece una descripción más detallada de estos diversos tipos de validez y sus puntos fuertes y débiles en las investigaciones experimental, de encuesta y etnográfica, véase también en DEALE (1975, págs. 29-38) una exposición respecto a la validez y la fiabilidad en la evaluación de los alumnos.

¹⁴ POPPER (1963, 1972) afirma que ninguna observación es “pura”, en el sentido de que este libre de carga teórica, se trata siempre de interpretaciones a la luz de teorías previamente aceptadas. En consecuencia, el esfuerzo científico ha de orientarse hacia la crítica de la teoría existente y a la evaluación del conocimiento (cf. El “cambio de paradigma” de KUHN, 1962).

¹⁵ Uno de los problemas que plantean las llamadas pruebas de CI consiste en que el constructo “inteligencia” sigue manteniéndose en el nivel de hipótesis. Por tanto, lo único que puede afirmarse con seguridad es que “¡los tests de CI prueban lo que prueban los tests de CI!”

absoluta de que un conjunto de fuentes de datos que se supone proporcionan evidencia relativa al mismo constructo lo hagan en realidad. En consecuencia, se han desarrollado procedimientos alternativos a la triangulación para establecer algún tipo de correspondencia entre los puntos de vista del investigador y los de quienes son investigados¹⁶. Ante la aparente naturaleza subjetiva de gran parte de las interpretaciones cualitativas, la validación se consigue cuando otros, en especial los sujetos sometidos a investigación reconocen su autenticidad. Una forma de hacerlo consiste en que el investigador ponga por escrito su análisis, en términos que puedan entender los sujetos de su investigación, se lo entregue y registre sus reacciones frente al mismo. Se conoce este procedimiento como validación de respuesta (BLOOR, 1978). En el campo de la investigación educativa, fue utilizado en el Cambridge Accountability Project (1981) y en nuestro análisis de actividades de auto-evaluación iniciadas en la escuela (JAMES, 1982).

La principal amenaza a la validez, tanto en los enfoques cuantitativos de investigación como en los cualitativos, radica en el concepto de reactividad. Se refiere este concepto a la tendencia de cualquier procedimiento de investigación a deformar la realidad que pretende investigar. Las condiciones en las que se administran las pruebas en clase (véase el ejemplo antes citado), en que se realizan los exámenes, en las que los inspectores llevan a cabo sus visitas o en las que un profesor efectúa una grabación magnetofónica de cuanto se dice en el aula, modifican el medio natural y provocan sesgos en los análisis de las observaciones. Este problema afecta especialmente a los enfoques positivistas y constituye el fundamento de la principal crítica que se les dirige desde el punto de vista ético. Por su parte, los enfoques interpretativos y, en especial, los etnográficos, aspiran a ser 'naturalistas' y procuran evitar a propósito la creación de condiciones artificiales de observación. Aun así, es raro el observador participante que puede entrar y salir del campo de investigación sin provocar trastornos (cuestión que trataremos con mayor detalle en el Capítulo IX).

La solución que suele dar el etnógrafo al problema de la reactividad consiste en adoptar una postura reflexiva. La reflexividad exige que los investigadores revisen constantemente, no sólo sus propias interacciones con los grupos que se investigan, sino también sus propios roles y las reacciones frente a lo que observan. En otras palabras, hacen un esfuerzo consciente para poner de manifiesto todo lo que pudiera inclinar de forma tendenciosa sus interpretaciones de los hechos. A veces, puede servirles de ayuda registrar primero sus observaciones utilizando descriptores de inferencia de bajo nivel (o sea, los que sirven para registrar la conducta sin inferir estados mentales ni hacer juicios) o mediante la grabación magnetofónica o magnetoscópica. Aunque esta última sigue siendo selectiva, su utilización permite comprobar (por triangulación) las impresiones iniciales de lo que ocurre al menos frente a algunos de los datos en bruto. No obstante, hay que sopesar esta ventaja en relación con el inconveniente que supone la intrusión adicional de los aparatos en el medio natural que, en realidad, puede incrementar la reactividad.

El problema de la reactividad y el proceso de reflexividad nos recuerda que las observaciones siempre son hasta cierto punto subjetivas. Con frecuencia se ha mencionado esto como objeción en contra de los enfoques interpretativos y cualitativos en la investigación y en la evaluación, aunque, en la actualidad, la mayoría reconoce que incluso los enfoques cuantitativos y positivistas no pueden considerarse "objetivos" en un sentido absoluto. Aun en relación con las ciencias naturales, hay una serie de filósofos que han arremetido contra la visión de la ciencia del empirismo lógico como una progresión objetiva hacia la 'Verdad' (KUHN, 1962 y 1970). De modo parecido, en la investigación educativa y social, se reconoce en la actualidad que las llamadas pruebas "objetivas" están limitados en cuanto a su objetividad. Como ha indicado NUTTALL (*Open University* 1982, bloque 1, parte C), lo único objetivo de esas pruebas es la *puntuación* de las mismas, en la medida en que pueden establecerse reglas que permitan efectuar esa tarea a no especialistas o encomendarla a los ordenadores. Todo lo que se aparta de esto lleva consigo algún elemento de subjetividad: la decisión de lo que ha de evaluarse, la elección de los ítems de la prueba o la de los criterios de valoración de la ejecución.

Dado el inevitable elemento de subjetividad de cualquier diseño de investigación, es probable que la idea de un informe reflexivo sea pertinente en relación con todos los enfoques. En consecuencia, con independencia de la tradición de investigación en la que se encuadre un estudio determinado, siempre es importante poner de manifiesto los supuestos sobre los que se basan las categorías de observación e interpretación, explorando las posibles fuentes de desviaciones tendenciosas. Sólo mediante la reflexividad puede abrirse el proceso de investigación a la clase de crítica pública que incrementa su credibilidad.

El hecho de que la mayor parte de este apartado se haya centrado en los enfoques de investigación se debe a que las fases de recogida y análisis de datos de la evaluación del *currículum* se ajustan a uno u otro paradigma de investigación, pudiendo validarse en consecuencia. Hay que señalar, sin embargo, que esta perspectiva sobre la relación entre evaluación e investigación social y educativa no goza de universal aceptación. Por ejemplo, EISNER (1980) dice que los métodos de evaluación derivados de la ciencia social (incluso de la

SHUTZ (1953) indica que deben cumplirse tres condiciones para que la investigación interpretativa satisfaga los requisitos de la teoría científica social. Los denomina postulados de: 1) consistencia lógica; 2) interpretación subjetiva, y 3) adecuación. El segundo postulado se refiere a que los constructos de segundo orden (o sea, los del investigador) deben incluir una referencia al significado subjetivo que la acción tiene para el actor (constructos de primer Orden). El tercer postulado indica que debe ser posible traducir la teoría al conocimiento de sentido común del que se deriva, de manera que permita a los investigados criticar las formulaciones del investigador y evaluar su plausibilidad.

ciencia social cualitativa) no han tenido la influencia que se pensaba. Propone, en consecuencia, que se considere la evaluación como una actividad artística con dos dimensiones: 1) conocimiento de la educación, y 2) crítica de la educación (EISNER, 1979). La primera se ocupa de la apreciación, es decir, la capacidad para discernir cualidades sutiles, mientras que la segunda actúa como mediadora de la percepción. En términos prácticos, infiere la necesidad de que los informes sobre la educación sean tanto evocadores como analíticos¹⁷. Aunque GIBSON (1981) no rechaza por completo la idea de que los métodos de crítica del arte sirvan para iluminar de algún modo los procesos que se desarrollan en la escuela y en el aula, se opone con vehemencia a una crítica del *currículum* de esa clase por: 1) suponer, erróneamente, que el *currículum* es una obra de arte y puede tratarse como tal; 2) aceptar supuestos equivocados respecto de la naturaleza de la crítica y dejar de lado su enorme variedad, y 3) suponer que, quienes aspiran a ser críticos, han de considerarse a sí mismos artistas para quienes el uso artístico del lenguaje constituye una necesidad. Según GIBSON, este último supuesto ha tenido efectos desastrosos, pues ha estimulado la elaboración de informes sobre currículos centrados en sí mismos, narcisistas, en apariencia omniscientes, de estilo grandilocuente y buscando conscientemente los golpes de efecto. (Conviene tener en cuenta esta última crítica cuando consideremos el estilo y la forma de los informes sobre estudios de casos —Capítulo X— y, en especial, los que utilizan enfoques ficticios; WALKER, 1981.) Si examinamos los datos disponibles sobre la práctica de la evaluación curricular en las escuelas (p. ej., la Segunda parte de este libro), su conexión con la crítica de arte o con la investigación social y educativa no es evidente en absoluto. Sólo las pruebas de valoración de los alumnos y los exámenes elaborados por agentes externos (véase el Capítulo IV) parecen haber adoptado una metodología estricta del tipo de investigación (de corte cuantitativo). Las demás formas de evaluación curricular no parecen guardar relación explícita con ningún procedimiento formalizado tomado de otros ambientes, dado que la principal actividad consiste en diálogos y reuniones en los que se comparten y juzgan experiencias o en el ejercicio del ‘juicio profesional’ individual (véanse los informes de actuaciones basadas en la escuela en el Capítulo V y las inspecciones en el Capítulo IV).

Dada esta realidad, el motivo para incluir un amplio apartado sobre procedimientos de investigación en este capítulo (y en los Capítulos VIII y IX) se basa en la percepción de que, aun en el caso de que los procesos siguieran sin organizarse, toda evaluación incluye los pasos de recogida y procesamiento de observaciones y juicios. A la luz de las objeciones de GIBSON contra la crítica del *currículum*, en su formulación presente, creemos que los métodos de investigación tienen todavía mucho que ofrecer, entre otras cosas porque parten de la premisa del rigor. Si la evaluación educativa aspira a lograr la aceptación de una amplia audiencia no puede considerarse exenta de las comprobaciones y balances normales que regulan la investigación.¹⁸

Sin embargo, una importante diferencia entre evaluación e investigación (aparte de que la primera se centre en problemas prácticos más que teóricos, como señalamos antes) tiene que ver con la situación de la principal fuente de datos. En el contexto de la mayor parte de las investigaciones educativas y sociales, el investigador descubre sus datos ‘afuera’, en el mundo social. Aunque pueda considerarse a sí mismo parte de ese mundo social, no suele tenerse por fuente primaria de datos. En la evaluación curricular, sobre todo en la autoevaluación o revisión del *currículum*, los profesores que realizan evaluaciones pueden considerarse así mismos como la fuente principal de información. Por tanto, uno de los cometidos de la evaluación consiste en poner de manifiesto la información que se deriva de las experiencias propias y de otros, y escrutar los supuestos y valores que llevan consigo. En este contexto, la puesta en común de informaciones en diálogo constituye una forma legítima de obtención y análisis de datos¹⁹. No obstante, lo más importante es que se produzca un intento deliberado de manifestar las posturas de valor con el fin de emitir juicios que, en la medida de lo posible, estén libres de amenazas e inclinaciones tendenciosas. En ciertos aspectos, esta concepción aspira a las características de lo que HABERMAS (1972) describe como situación ideal de conversación” (fundamento de la teoría crítica). Según él, ésta es una forma de diálogo que se ha liberado de manera tan clara de los intereses ideológicos y de poder que las consideraciones racionales por sí solas dictan el resultado de los intercambios entre individuos o grupos. Ya nos hemos referido a procesos semejantes de deliberación y diálogo en relación con la revisión del *currículum* (Capítulo III) y nos ocuparemos nuevamente de ellos en el Capítulo VIII (véase en especial el apartado sobre el “*Curriculum* oficial”).

El juicio

La importancia del diálogo y la deliberación no es, sin embargo, exclusiva de ciertos tipos de evaluación, como

¹⁷ Esta visión de la evaluación como crítica del *currículum* ha sido elaborada también por WILLIS (1978).

¹⁸ No quiere decir esto que los métodos de investigación tengan que ser sólo los referentes a la experimentación científica social. Por ejemplo, STENHOUSE (1979a) señala el papel que pueden desempeñar los métodos de los historiadores.

¹⁹ Enfoque adoptado por LEWY (1977). El enfoque de ‘puesta en común de los informes’ se explora más a fondo en un breve estudio de casos de la actividad de evaluación de una escuela primaria (Open University, 1982, Bloque 2, 2. parte).

la revisión curricular. Como indicamos antes, refiriéndonos al marco de STAKE (1967), el juicio constituye una parte esencial de cualquier evaluación curricular. En la medida en que ese juicio supone el desarrollo y la aplicación de criterios de valor (o calidad), es inevitable alguna forma de deliberación, debiendo dirigirse con el mismo grado de cuidado y reflexividad que en las fases de recogida y análisis de datos. Parece ahora extraño que las definiciones de evaluación dadas al comienzo del capítulo hicieran hincapié en las actividades de adquisición de información y de decisión, mencionando poco o nada el acto mismo del juicio. En realidad, la valoración constituye la misma esencia de todo el proceso.

Según STAKE (1967): “Para comprender completamente un programa educativo, es preciso describirlo y juzgarlo por completo” (pág. 525). Sin embargo, esto no significa que el evaluador tenga que ser el juez final de la calidad de un programa. Algunos evaluadores creen que deberían juzgarlo (p.ej., SCRIVEN, 1967), mientras que otros manifiestan que el papel del evaluador consiste en proporcionar los datos descriptivos que permitan juzgar a otros (los responsables de las decisiones). La insistencia de STAKE en la “representación” (STAKE, 1972) indica que se inclina a favor de esta última postura, aunque entiende la “representación” como descripción y, por tanto, como alternativa al análisis “en profundidad”, más que como alternativa al juicio. No obstante, es obvio que, desde su punto de vista, el evaluador tiene la responsabilidad de “procesar” los juicios (STAKE, 1967). En otras palabras, considera que los juicios subjetivos de diversos individuos y grupos de intereses son datos” pertinentes y forman parte de cualquier representación del mismo modo que los datos de tipo más “objetivo”.

MACDONALD (1977^a), con su concepto de “evaluación democrática”, ha desarrollado más esta idea. A diferencia de la evaluación burocrática, que corresponde a los intereses de los organismos gubernativos, patrocinadores de programas y políticos, o de la evaluación autocrática, al servicio de la comunidad universitaria de investigación, la evaluación democrática reconoce “la pluralidad de valores y trata de representar un conjunto de intereses en su formulación de valores. El valor básico consiste en una ciudadanía informada y el evaluador actúa como intermediario en los intercambios de información entre grupos diferentes” (MACDONALD, 1977a, pág. 226). La importancia concedida por SIMONS (1981) al descubrimiento de los valores y supuestos que subyacen a la práctica de las escuelas (véase el Capítulo V) también forma parte de esta tradición.

A diferencia de los programas de evaluación a gran escala, en la evaluación basada en la escuela, el evaluador y el responsable de las decisiones pueden ser la misma persona o grupo, en cuyo caso el debate respecto a si el evaluador debe o no juzgar se queda en una cuestión académica. No obstante, sigue teniendo sentido la postura de STAKE, que asigna igual importancia a las actividades complementarias (y quizá inseparables) de descripción y juicio, como también la idea de MACDONALD respecto a que la evaluación, para que sea de verdad democrática, ha de reconocer la pluralidad de valores presentes en la escuela y en la comunidad.

El debate sobre quién debe emitir los juicios ha distraído la atención de una cuestión mucho más fundamental sobre la misma naturaleza del juicio. Somos conscientes de que la cuestión del juicio puede resultar insoluble y no estamos totalmente convencidos de que la siguiente exposición la aclare mucho. Carecemos de pruebas de que las ideas que presentamos hayan sido puestas en práctica, aunque nos llaman la atención las posibilidades que ofrecen.

La falta de claridad en el uso de los términos es en parte responsable del problema. Las expresiones “valores”, “normas”, “criterios” y “objetivos” suelen utilizarse casi como equivalentes (y nos tememos que también nosotros tengamos parte de culpa) y contribuyen a la extendida confusión sobre lo que supone el acto de juzgar. En un artículo sobre los orígenes y funciones de los criterios de evaluación, SADLER (1985) intenta poner orden en este campo (¿de minas?). Distingue primero los criterios de las normas, diciendo que los primeros se refieren a dimensiones importantes de una evaluación, mientras que las normas aluden a determinados niveles utilizados como puntos de referencia para efectuar atribuciones positivas o negativas de valor. Por ejemplo, el juicio sobre un conjunto curricular puede quedarse en los criterios relativos a la congruencia con las metas de la escuela, la viabilidad del mismo respecto a determinado grupo de alumnos o su coste. Sin embargo, las normas incluyen ideas de congruencia suficiente, bastante viabilidad y costes aceptables, lo que supone la calificación de niveles que operan como puntos de ruptura en relación con su aceptación o rechazo.

La mayor parte de las evaluaciones formales hacen referencia a uno o dos fundamentos del juicio: son relativos o absolutos (STAKE, 1967). Según SADLER, los juicios relativos se refieren sólo a los criterios, porque el tema en cuestión se compara con la distribución de otros casos en relación con determinado criterio. Los casos se clasifican (X se considera mejor que Y), de modo que carece de interés la especificación de un punto de ruptura. En cambio, los juicios absolutos hacen referencia tanto a los criterios como a las normas, puesto que el caso se compara con lo que definen las autoridades como cualidad criterial. En este supuesto, entran en juego las ideas de suficiencia y aceptabilidad.

Es interesante señalar que el resto del artículo de SADLER se ocupa de identificar los criterios y no las normas. Quizá le resulta inabordable el problema de éstas. ¿Acaso sólo son, como sugiere MACDONALD (1978), destilados subjetivos e idealizados del pasado? Pensemos en esta afirmación: “Los niveles [normas] eran más altos cuando yo estudiaba”. ¿Son, en realidad, relativos, en la medida en que las mejores actuaciones del pasado se convierten en el nivel calificador para el juicio de todas las actuaciones futuras o, en cierto sentido, las normas son

—o deben ser—objetivas o absolutas y establecidas por acuerdo mediante el discurso racional? Aunque los

conocimientos necesarios para generar tales normas sean en gran medida los de los filósofos, esta última opción parece constituir la mejor alternativa.

En relación con los criterios, SADLER puede ser más útil. Afirma que están organizados en sentido jerárquico y que la mayor parte de los criterios para el juicio educativo se sitúa en los niveles medio y bajo de un sistema tripartito, puesto que incluyen la referencia a los contenidos. Sin embargo, también tienen relación con criterios de alto nivel o últimos que representan los valores subyacentes.

Nos parece que esta organización queda bien representada en el marco para la evaluación del producto elaborado en la *Secondary Science Curriculum Review* (SSCR). Los productos en cuestión consistían en propuestas curriculares (materiales, estrategias docentes, esquemas de evaluación, etc.) elaborados por los grupos de trabajo de profesores. Esas propuestas se evaluaron en niveles locales, regionales y nacionales, utilizando un conjunto de criterios elaborados por los miembros del equipo central (EBBUTT y WEST, 1984). Los criterios se dispusieron en sentido jerárquico, de modo que cada uno de los cuatro criterios generales engloba un conjunto de criterios más específicos en términos de contenidos (véase, como ejemplo, la Figura 7.2).

En este ejemplo, aunque las “mayores oportunidades educativas” se sitúan en un nivel general, no son independientes de los contenidos, puesto que no constituyen “criterios básicos” como tales. Subyacente a este criterio está el valor genérico de la “justicia como equidad” (RAWLS, 1971). Por esta razón, los criterios generales de la SSCR pueden describirse como de nivel medio y los conjuntos de criterios específicos que de ellos dependen son de bajo nivel.

SADLER representa en tres fases la especificación de los criterios. *La primera* consiste en el acto evaluador primario, que precede a cualquier criterio como tal. Este reconocimiento de valor queda mejor expresado en la siguiente observación: “Puedo reconocer la calidad cuando la veo, pero no puedo decir por qué. Este tipo de valoración es holístico y no procede de la razón, sino de la interacción de los valores subyacentes con el contexto de experiencia. En otras palabras, la valoración primaria tiene carácter existencial.

La segunda fase de la especificación de criterios comienza cuando se ha reconocido el valor fundamental, “descomponiéndose”, por así decir, el valor abstracto para dar lugar a criterios con contenido (nunca se trata de un puro ejercicio abstracto porque siempre hace referencia al contexto existencial en el que tuvo lugar la valoración primaria). Según SADLER: “los criterios son útiles para hacer afirmaciones de valor, y necesarios para dotarlas de fundamento, porque determinan existencialmente el contenido de los valores axiológicos [es decir, subyacentes] aclarándolos e interpretándolos” (pág. 296). Por tanto, los criterios se refieren a los contenidos y son explícitos, mientras que los valores subyacentes son abstractos y con frecuencia están implícitos, lo que explicaría por qué los

Criterio 2 El producto debe ser compatible con el punto de vista del equipo central respecto al incremento de las oportunidades educativas,	Deben recomendarse los procedimientos de valoración y comunicación que se incluyen y, con independencia de qué sean sumativos o formativos, deben ajustarse a los objetivos, contenidos y estrategias docentes. El producto debe identificar las necesidades y procurar fomentar especialmente la participación de las niñas. El producto debe ser variado y estimulante, tanto para los profesores como para los alumnos. El producto debe facilitar especialmente la expresión de la diversidad cultural. El producto debe ocuparse especialmente de los estudiantes con necesidades especiales. - El producto no debe facilitar (de forma implícita ni explícita) la creación de estereotipos indeseables. La forma de organización del producto debe facilitar el crecimiento cognitivo de los alumnos. El producto debe facilitar que los alumnos superen el nivel mínimo de capacitación. El producto (incluyendo los diagramas) debe ser adecuado, claro y libre de ambigüedades con respecto a las personas a las que se dirige. El lenguaje del producto debe ser accesible a las personas a las que se dirige.
---	---

--	--

Figura 7.2: Un ejemplo de los criterios utilizados para evaluar los productos SSCR. (EBBUTT y WEST, 1984, pág. 13).

valores subyacentes se dan a menudo por descontados y han de redescubrirse a partir de los criterios establecidos (como en el ejemplo de la SSCR antes citado).

La tercera fase, que sigue a la de valoración primaria y a la de especificación criterial, consiste en el acto de prescripción, en el que los criterios se convierten en objetivos: “los objetivos son criterios que han sido paralizados y formulados de manera que operen como fines a la vista” (pág. 293). De este modo se convierten en centros de atención para la acción. A partir de aquí, SADLER indica que la concepción de que la evaluación consiste sin más en el proceso de medir hasta que punto se han alcanzado los objetivos preespecificados es muy natural porque cierra la relación circular valoración-criterios-objetivos.

Sin embargo, indica también que los criterios especificados, explícitos, que, por naturaleza, son atomísticos, no suelen captar la naturaleza holística de las valoraciones a partir de las cuales se originan. En otras palabras, el todo es más que la suma de las partes; en realidad, puede haber conjuntos alternativos y contradictorios de criterios derivados del mismo valor genérico [McINTYRE, 1981, en un estudio originario de teoría moral, pone el ejemplo de principios de *adquisición* y *dominio justos* (ideología capitalista) frente a principios de *distribución justa* (ideología socialista), referidos ambos a la virtud supraordinada de la justicia]. Concluye SADLER:

aunque todo conjunto de criterios de evaluación educativa debe, en general, hacerse explícito, en principio es indeseable su especificación completa antecedente. La razón de esta afirmación se basa en cuatro fundamentos: a) La descomposición de los principios de valor siempre es parcial; b) los criterios surgen de la experiencia y no pueden deducirse de los principios de valor; c) no *pueden* hacerse explícitos todos los criterios, y d) hay criterios para utilizar los criterios (p. ej., principios que exigen la suspensión de un criterio en beneficio de otro) (pág. 297).

Por todas estas razones, SADLER está dispuesto a depositar cierta confianza en los conceptos de conocimiento experto (*connoisseurship**) y crítica, elaborados por EISNER (1979) y WILLIS (1978). Se basan éstos en la aceptación de quienes no actúan como jueces de las evaluaciones holísticas de otros, bien porque pueden interpretar los juicios del *connoisseur* a la luz de experiencias propias semejantes o porque acepten sin más los juicios autorizados.

Por desgracia, aunque quizá no fuera del todo imprevisible, SADLER no llega a indicar formas de realizar este análisis que sirvieran de orientación práctica para efectuar evaluaciones en las escuelas. Sin embargo, creemos posible extrapolar algunas indicaciones:

1. Es importante observar que la especificación antecedente de criterios puede constituir una contraseña y que está muy justificado recoger información en relación con el área de interés a efectos de evaluación de manera libre y abierta.
2. Es preciso reconocer también que los individuos elaboran juicios holísticos intuitivos sobre la base de la información que obtienen o reciben (el acto evaluador primario de SADLER). No obstante, debe estimulárseles para que organicen y aclaren sus juicios (sentimientos) por su cuenta o en el seno de un grupo de compañeros.
3. Estos juicios organizados pueden examinarse en relación con sus valores subyacentes, planteando la cuestión: ‘¿por qué creo que esto es bueno o malo?’ Esto debe llevar a la conclusión de que merece la pena conservar esos valores subyacentes o que, por el contrario, no son sino simples prejuicios. El proceso es similar al de tomar conciencia de algo y sólo resultará eficaz para el individuo si hay otros que participen en él, bien en calidad de “amigo crítico” (véase el Capítulo y), bien como grupo cooperativo (véase el Capítulo XI).
4. Una vez examinados los valores y los juicios, pueden formularse como criterios con contenido y utilizados como base del juicio formal. Como indica SADLER, esos criterios pueden convertirse en objetivos en la medida en que puedan utilizarse como orientaciones de las acciones futuras aunque, por las razones antes citadas, evite la consecuencia de que sean necesariamente los criterios de los juicios futuros.

Aunque este proceso pueda parecer complejo, es en realidad de tipo naturalista, puesto que sólo pide a los evaluadores que hagan explícito lo que suelen hacer de forma intuitiva. Al ponerlo de manifiesto, los criterios

* El concepto *connoisseurship* se deriva del Francés *connoisseur* (en la actualidad *connaissance*), se traduce como conocedor”, pero con el matiz de “disfrute”, “gusto”, de “saborear” aquello en toque se es persona experta, un “*connaissanceur*” es alguien que además de conocer en profundidad algo, es también una persona apasionada, que disfruta mucho con esa temática que domina, por ejemplo: la música barroca, la cultura maya, o las setas, etc. Se diferencia de la persona erudita en que ésta domina muy bien una determinada parcela del conocimiento, pero no tiene por qué ser una persona apasionada, que se implica emocionalmente en ese proceso de conocer. (N. del R.)

antes ocultos se hacen susceptibles de debate público sin necesidad de especificar de antemano determinados indicadores de calidad que deforman la experiencia y niegan la realidad del pluralismo de los valores²⁰.

En muchos sentidos, el proceso en conjunto difiere poco del proceso de deliberación descrito en el Capítulo VIII. Como las observaciones de WALKER (1975) de la reflexión en grupo sobre las propuestas para la acción, los juicios pueden hacerse de manera autoconsciente y ordenada aunque, como en el caso de la deliberación, quizá sea más eficaz dentro de la estructura de los grupos cooperativos. Sin embargo, hay una diferencia: mientras que la acción exige el acuerdo, el consenso no es necesario en relación con el juicio. Por ejemplo, el profesorado de una escuela puede estar dividido respecto a la conveniencia de una iniciativa, aun después de un riguroso examen de sus juicios; estaría justificado presentar valoraciones diferentes (cf. la concepción del evaluador como “intermediario sincero” de MacDONALD, 1977). Sólo cuando se toma la decisión a favor de una acción en detrimento de otras, es preciso rechazar unos juicios en beneficio de otros. Esta actividad pone en juego otro nivel de juicio. Con independencia de que los evaluadores juzguen o no por su cuenta, esta categoría de juicios es prerrogativa de quienes tienen la responsabilidad de la decisión (aunque a veces los evaluadores independientes hagan recomendaciones). La cuestión de quiénes sean los responsables de las decisiones y cómo realicen esta tarea es competencia de la gestión (que tratamos brevemente en el Capítulo XI). Lo que interesa a este respecto, así como a las cuestiones estudiadas por SADLER, es que el acto del juicio, en relación con la evaluación educativa, es complejo y tiene aspectos muy diversos.

Esencialmente, en esta exposición nos hemos visto obligados a volver al concepto de evaluación intrínseca que exige poner en acción las destrezas del filósofo. Si, como indica STAKE, toda evaluación requiere una descripción sistemática y un juicio también sistemático, la tarea que se plantea a los profesores y a otros que traten de evaluar la práctica educativa es considerable. Si sus juicios han de tenerse por creíbles, justos y exhaustivos, no cabe duda de que necesitan ampliar sus roles tanto en el sentido de la investigación curricular como en el de la deliberación y lo que HABERMAS llama “razonamiento práctico”. A este respecto, la deliberación se aproxima a los métodos habituales de trabajo de los profesores y la investigación arranca de ellos. En el Capítulo VIII estudiaremos específicamente lo que sabemos sobre el uso que los profesores hacen de estos métodos, volviendo después a las cuestiones generales que rodean la acción de los profesores como evaluadores en el Capítulo IX.

CAPITULO VIII

Revisión de técnicas específicas

Introducción

Con independencia de la táctica que se adopte para la evaluación en una escuela, requerirá la utilización de determinadas técnicas. Es más, aunque algunas estrategias favorezcan el USO de ciertas técnicas (p. ej., las pruebas de rendimiento), suelen darse opciones diversas. Hasta aquí, el espíritu del libro se ha caracterizado por animar a las escuelas y a quienes participan en ellas a adoptar una visión amplia de la estrategia (de ahí el planteamiento del Capítulo VI). Esa misma postura se aplica en gran medida a las técnicas. Los principales enfoques de investigación expuestos en el Capítulo VII han dado origen a un importante conjunto de técnicas para realizar la recogida y el análisis de datos; en consecuencia, tanto los profesores como otras personas interesadas tienen que estar preparados para utilizar las más adecuadas con los problemas de evaluación a los que se enfrenten. La matriz de descripción de STAKE (1967) sirve de recordatorio del ámbito de fuentes potenciales de datos, por lo que utilizaremos sus categorías como marco general de este capítulo. No obstante, como ya hemos señalado, la evaluación no es sinónimo de investigación en cuanto a la selección de problemas, propósitos finales ni a su preocupación por la recogida de datos, lo que quizá sea aún más importante.

Al presentar su matriz de descripción, STAKE nos recuerda que la evaluación no sólo tiene que ser una

²⁰ Al final del capítulo vi, hay una nota que indica que se han realizado diversos intentos de establecer criterios de proceso que sirvan de indicadores de calidad de enseñanza (véanse: JENCKS y cols, 1972; RUTTER y cols., 1979; OLSON, 1970; SIZER, 1979). Aunque existe un acuerdo evidente respecto a que una escuela tiene éxito en la medida en que es eficaz, desde el punto de vista educativo, siempre es discutible lo que pueda constituir la calidad de enseñanza. En el mejor de los casos, puede llegarse a un “consenso de trabajo” entre los profesores o entre ellos y otras partes interesadas. Parece, por tanto, que la búsqueda de criterios absolutos es una yana esperanza, si no una pérdida de tiempo.

actividad empíricamente fundamentada (como indicamos en el Capítulo III). Hemos optado por examinar técnicas que abarquen tanto la evaluación empírica como la intrínseca, agrupadas en cuatro epígrafes básicos:

Curriculum observado.

Resultados del *curriculum*.

Curriculum oficial.

Contexto curricular.

En el caso del “*curriculum* observado” y en el de los “resultados del *curriculum*”, nos centramos en las técnicas de recogida y análisis de datos, pero el tratamiento del “*curriculum* oficial” y del “contexto curricular” es diferente. Pensamos que la evaluación del “*curriculum* oficial” requiere una evaluación intrínseca y, como las técnicas correspondientes a ella no están bien definidas, estudiaremos la naturaleza de esa evaluación, así como las técnicas que sean necesarias. Aunque los demás epígrafes hablan por sí mismos, la relación del de “contexto curricular” con la matriz de datos de STAKE no es evidente. En este caso, tenemos presente su idea de los antecedentes, aunque utilizamos el término “contexto” para hacer hincapié en la omnipresencia de las condiciones que constituyen el contexto de la enseñanza y el aprendizaje con independencia de que desempeñen o no un papel central en ellos. Por ejemplo, la arquitectura de una escuela constituye un aspecto evidente, aunque poco estudiado, del contexto curricular (WALLACE, 1980),

Al tratar las cuestiones relativas a la recogida y análisis de datos, encontramos una gran tradición de investigación en la que basarnos. Somos conscientes de los problemas metodológicos mencionados en el Capítulo VII, pero no pretendemos ser exhaustivos, pues ése sería el cometido de los libros sobre métodos de investigación. En cambio, respecto a técnicas de uso corriente, como las de entrevista, consideraremos los principales elementos del método y alguna experiencia de su empleo en evaluación a cargo de los profesores en las escuelas. Estas experiencias no son abundantes y, cuando carezcamos de información, expondremos los problemas que consideremos pertinentes para las escuelas. Como este capítulo consiste esencialmente en una *revisión*, trataremos de ofrecer una guía sobre estas cuestiones y sobre la bibliografía. Aunque existen algunos textos que abarcan ciertos aspectos de las cuatro áreas que hemos seleccionado, suelen ocuparse de manera específica de uno o dos (p. ej.: HOOK, 1981; GRONLUND, 1981), los tratan en general (p. ej.: HARRIS, BELL y CARTER, 1981; *Ford Teaching Project* 1975), o no se dirigen a los profesores (p. ej.: LEWY, 1977). El creciente interés por la figura del profesor como investigador ha promovido la aparición de algunos libros escritos específicamente para ayudar a los profesores en relación con los métodos de investigación, etc. (p. ej.: WALKER, 1985).

Tenemos presentes las constricciones y limitaciones de las situaciones en las que se encuentran los profesores, pero posponemos un estudio general de estos problemas prácticos hasta el Capítulo IX. Sin embargo, el hecho de que presentemos los detalles de un amplio conjunto de técnicas no supone una recomendación de ninguna de ellas (aunque nuestras preferencias serán patentes) ni pretendemos que los profesores se familiaricen con todas. La revisión trata de *ayudar* a elegir y a desarrollar técnicas apropiadas al problema que se investigue.

El curriculum observado

Introducción

Este apartado se refiere sobre todo a las técnicas de recogida de datos y se basa en las tradiciones de investigación nomotética e idiográfica. No las tratamos como competidoras, aunque la exposición de algunas cuestiones se oriente en ese sentido, sino *como fuentes* de distintos tipos de *información*. El apartado se subdivide de manera que refleje la preocupación por la *observación directa* de la conducta (p. ej., en el aula) y por la *observación indirecta*, a través, por ejemplo, de la reconstrucción de (a conducta a partir de los informes de los participantes (es decir, de los profesores y de los alumnos). Hasta cierto punto, nos hallamos ante unos límites artificiales porque, entre otras cosas, simplifican ciertas tradiciones teóricas en el campo de la investigación, pero los utilizamos por comodidad y para tratar de mantener presentes las sensibilidades de las tradiciones. El *curriculum* puede observarse en clase, en los pasillos de la escuela e, incluso, en las casas de los alumnos, aunque la mayor parte de las técnicas se experimentan en el aula. En consecuencia, la observación directa suele centrarse en clase y esto se refleja en nuestro tratamiento de la misma. La observación indirecta es adecuada dentro y fuera del aula. Resulta especialmente útil en aquellas situaciones, a menudo fuera del aula, no susceptibles de observación directa. En otras palabras, este apartado sobre la observación no trata de mantener el equilibrio de la exposición respecto a la evaluación en el aula y en el nivel institucional. En consecuencia, esperamos que el apartado sobre el “contexto curricular” corrija cualquier tendencia que se observe aquí a favor del nivel de clase.

Observación directa

Introducción

La preocupación creciente respecto a la observación que se percibe entre los evaluadores, sobre todo los relacionados con proyectos curriculares, procede de los partidarios del campo “iluminador” que se ocupan más de los procesos que de los productos del aprendizaje. Así, la experiencia con las técnicas de observación refleja los debates metodológicos de las dos tradiciones de investigación mencionadas en el Capítulo VII. En la investigación en el aula, esto se pone de manifiesto en dos enfoques: la observación estructurada (sistemática), y la observación no estructurada (informal).¹ Nos ocuparemos de ambos enfoques antes de considerar los problemas concretos que experimentan los profesores cuando observan, o sea, como participantes que se convierten en observadores.

Observación estructurada

Consiste en registrar la conducta que se desarrolla en el aula en relación con determinadas categorías especificadas de antemano y en reflejar la aparición de esas conductas en un cuadro de frecuencias o en sucesión. Proporciona así datos cuantitativos que pueden utilizarse a efectos de comparación de clases, práctica corriente en la investigación nomotética. Como puede suponerse, la observación estructurada tiene una historia de investigación más larga en los Estados Unidos que en la Gran Bretaña² y, en consecuencia, existen muchos sistemas relativos a este tipo de observación³. Muchos de esos sistemas se basan en las *Interaction Analysis Categories* de FLANDERS (1970), que codifican siete categorías de intervención oral del profesor (p. ej.: “alaba o estimula” y “hace preguntas”), dos de intervención del alumno (p.ej.: “responde al profesor y “el alumno habla por propia iniciativa”) y una de silencio o confusión. El sistema refleja un enfoque docente formal en el que domina la enseñanza dirigida a toda la clase. En este campo, los escritos generales dirigidos a profesores reflejan dos tradiciones: la relacionada con la evaluación del alumno, por ejemplo:

GRONLUND (1981) y TÉNBRINK (1974); y la relacionada con la observación de niños pequeños (también con fines de valoración), por ejemplo: BOEHM y WEINBERG (1977); COHEN y STERN (1978); GOOD y BROPHY (1978), y GOODWIN y DRISCOLL. (1980). En el contexto británico, GALTON (1978) ha documentado un conjunto de más de cuarenta sistemas que, en contraste con los elaborados en los Estados Unidos, tienen que ver con tipos más flexibles de organización de clase⁴; la escuela primaria constituye el principal Centro de atención de éstos⁵. Muchos sistemas de este tipo se centran en la actividad del alumno, en el contacto entre éste y el profesor y en la proporción del tiempo que el alumno utiliza “en su tarea”.

En realidad, hay diversas formas de describir los sistemas de observación estructurada, pero todos deben resolver los siguientes problemas:

1. El objetivo del estudio.
2. Cómo se elabora el plan.
3. Su centro de atención: el alumno, el profesor o ambos.
4. El número y tipo de categorías.
5. La frecuencia de registro.
8. El período de observación.
7. El número y la selección de personas observadas.

(Adaptado de BENNETT y McNAMARA, 1979, págs. 36-37).

EGGLESTON GALTON y JONES (1975) utilizan el “objetivo” como una dimensión de un mapa conceptual de la investigación mediante observación estructurada, aislando *tres* tipos distintos: proceso, para elaborar una descripción; comparativo, para comparar una clase o profesor con otros; proceso-producto, para relacionar la conducta con un cambio mensurable de los resultados. Todos ellos son pertinentes para la evaluación, aunque

¹ Son diversas las expresiones usadas en la bibliografía sobre observación para describirlas. La mas corriente es observación sistemática, pero no es correcto referirse a la observación no estructurada como “asistemática”. Del mismo modo, la división formal/informal es inadecuada; por eso preferimos utilizar las expresiones observación estructurada y no estructurada.

² HAMILTON y DELAMONT (1974) y DELAMONT (1976) estudian su desarrollo en los Estados Unidos en relación con la observación no estructurada y el estado de la investigación británica en el aula.

³ SIMON y BOYER (1974) y BORICH y MADDEN (1977) presentan detalles de gran número de sistemas de observación utilizados en los Estados Unidos, y ROSENSHINE y FURST (1973), BROPHY y GOOD (1974) y DUNKIN y BIDDLE (1974) revisan la investigación utilizando esos sistemas.

⁴ BOYDELL (1975) trata la cuestión de cómo realizar observaciones sistemáticas (estructuradas) en clases informales.

⁵ BENNETT y McNAMARA (1979) presentan una selección de sistemas de observación, utilizados en su mayoría en escuelas primarias. McINTYRE (1980) hace una selección no limitada a escuelas primarias

los estudios de procesos son los que están más al alcance de las escuelas⁶. El centro de atención del estudio y el número y tipo de categorías del sistema se relacionan con la conducta seleccionada para la observación. Esa muestra conductual puede clasificarse de acuerdo con el modo de desarrollo del plan: inductivo, comenzando con un conjunto amplio de conductas pertinentes y terminando con una teoría relativa a esas conductas: a partir de una teoría explícita de la conducta en clase que se operativiza mediante categorías o basada en objetivos curriculares que establecen cuáles son las conductas importantes. (Esto constituye la segunda dimensión del mapa conceptual elaborado por EGGLESTON, GALTON y JONES, 1975.) El muestreo de la conducta constituye también la base de tres tipos corrientes de sistemas:

1. Sistema de indicios: lista de conductas mutuamente exclusivas, aunque no exhaustivas, a partir de la cual se registra la aparición de una conducta (con independencia del número de veces que se produzca) o se contabiliza la frecuencia en que aparece durante un período específico de observación.
2. Sistema categorial: lista exhaustiva de conductas mutuamente exclusivas a partir de la cual puede calcularse la frecuencia de aparición y su sucesión.
3. Escala de clasificación: expresa un juicio, que supone un nivel inferencial elevado, como una puntuación en una escala⁷.

Las conductas que aparecen relacionadas en los sistemas de indicios y categorial deben ser directamente observables, con una inferencia mínima, garantía importante de fiabilidad.

Es obvio que la frecuencia de registro en cada uno de los sistemas indicados es diferente: los sistemas de indicios dependen de la aparición del hecho en cuestión, independientemente de *cundo* suceda; un sistema categorial codifica la conducta a intervalos temporales especificados; la escala de clasificación suele puntuarse una sola vez (cada ítem) en un mismo período, como el correspondiente a una clase. Por desgracia, este campo del muestreo, tanto temporal como de hechos, se caracteriza por el uso de términos muy diversos. Algunos son sinónimos, otros plantean diferencias sutiles, pero pocos autores tratan de unificarlos. (Sin duda, habrá quienes no estarán de acuerdo con nuestra tentativa.)

El *muestreo de hecho*⁸ consiste en hacer una señal cada vez que aparece una conducta, perdiéndose, por tanto, la sucesión conductual. Por ejemplo, si tomamos dos categorías de conducta, el registro podría ser el siguiente:

El profesor hace una pregunta

El profesor alaba o estimula

El *muestreo instantáneo*⁹ requiere la selección de una categoría a intervalos regulares que represente la conducta que se produce en el momento de la observación. Se trata de una forma discontinua de registro que mantiene la secuencia entre apariciones. Por ejemplo, utilizando las mismas conductas que antes, la anotación en cada minuto nos proporcionaría un registro como éste:

*registro de intervalo*¹⁰ significa que, a intervalos regulares, se anota la conducta ocurrida durante los mismos. Esto proporciona un registro continuo, aunque se pierda la sucesión de conductas dentro de cada uno. (Las consecuencias de ello dependerán de la longitud del intervalo¹¹) Para un adecuado muestreo conductual, es importante tener en cuenta el período de observación, su longitud y el momento en que aparece la conducta¹². El muestreo de alumnos puede realizarse de cuatro formas: mediante muestreo aleatorio; por observación de

⁶ Los estudios comparativos requieren procedimientos de investigación experimental con exigencias rigurosas imposibles de cumplir en las escuelas, por ejemplo, la necesidad de utilizar muestras. Las manipulaciones estadísticas para los estudios proceso-producto constituyen también un problema para las -escuelas que no disponen de programas informáticos adecuados (p. ej., para el análisis de regresión).

⁷ BORICH (1977) presenta un conjunto ligeramente distinto de definiciones de signo, categoría y clasificación; sin embargo, las que presentamos son las más corrientes. En todo caso, sus definiciones contienen los mismos elementos que las nuestras (o sea, exclusividad, exhaustividad y nivel de inferencia)

⁸ HOOK y cols. (1981) y BOEHM y WEINBERG (1977) lo utilizan, WRAGG y KERRY (1979) aluden a ello como una forma de muestreo natural", o sea, se elabora un registro cuando se produce algún cambio de categoría, tema o hablante (GALTON, 1978),

⁹ DUNKERTON y Cuy (1981) utilizan esta expresión. WRAGG y KERRY (1979) la denominan "muestra estática" y GALTON (1978) y COOPER (1974) a denominan "muestra temporal".

¹⁰ COOPER (1974) utiliza esta expresión. GALTON (1978) emplea "unidad de tiempo" y WRAGG y KERRY (1979) la de "muestreo de unidades".

¹¹ La longitud de los intervalos en el "muestreo instantáneo" y en el "registro de intervalos" depende de la naturaleza de las decisiones que ha de tomar el observador. Véase el debate entre DUNKERTON y GUY (1981) y EGGLESTON y GALTON (1981).

¹² GALTON (1979) señala que la representación excesiva de ciertas áreas curriculares puede darse cuando los alumnos escogen sus propias actividades en una clase de primaria.

todos ellos; por ciclos temporales, observando a cada uno durante diez minutos hasta incluirlos a todos en la observación; mediante observación continuada hasta la aparición, o sea, observando a un alumno o grupo hasta que se produce la conducta, pasando de inmediato a otro (BOEHM y WEINBERG, 1977).

Al desarrollar sistemas de observación estructurada, hay que tener en cuenta la importancia de su validez y fiabilidad. Uno de los puntos fuertes de estos sistemas consiste en su fiabilidad, que se garantiza mediante la definición estricta (o sea, sin ambigüedad) de la conducta que se codifica. El objetivo consiste en reducir los efectos derivados de cada observador concreto, de modo que dos observadores que codificaran la misma conducta produjeran, dentro de ciertos límites, registros idénticos de categorías. En consecuencia, la fiabilidad puede expresarse como correlación entre dos observadores (fiabilidad interjueces).¹³ Del mismo modo, también puede medirse la fiabilidad de un único observador que registre la misma conducta en distintos momentos (fiabilidad intrajueces). Para llevar a cabo una comprobación de este tipo, los profesores que actúan como observadores de su propio ejercicio docente necesitan alguna forma de registro independiente (p. ej., grabación magnetofónica). Como veremos, las críticas contra la observación estructurada se refieren sobre todo a su validez, que depende en parte del valor descriptivo y explicativo de los conceptos subyacentes al plan (McINTYRE y MACLEOD, 1978). Solucionado este problema, pueden considerarse cinco cuestiones relativas a la validez, planteadas por HURWITZ (1979)¹⁴:

1. ¿Los aspectos de la conducta en el aula que pretende medir el sistema difieren en realidad de clase a clase? (Validez de constructo.)
2. ¿Los aspectos de la conducta en el aula que pretende medir el sistema se mantienen relativamente estables en la misma clase? (Validez de constructo.)
3. ¿Cómo se relacionan las medidas de los aspectos de la conducta en el aula que pretende medir el sistema con otros aspectos de la conducta en clase que estén o no relacionados con ellos, desde los puntos de vista lógico o teórico? (Validez de constructo.)
4. ¿El sistema mide por completo los aspectos de la conducta de clase que pretende? (Validez de contenido.)
5. ¿Los aspectos de la conducta en el aula que el sistema pretende medir por completo existen en las clases reales? (Validez de contenido.)

(BENNETT y MCNAMARA, 1979, pág. 112.)

Naturalmente, todas las decisiones de muestreo antes mencionadas (de conducta, temporal, periódico, etc.) afectan a la validez del sistema de observación respecto al registro de la conducta representativa del aula. Como ya indicamos en el Capítulo VII, la validez y la fiabilidad se solapan; cualquier interpretación de la conducta que un observador tenga que hacer mientras codifica constituye una amenaza para la validez del sistema, al tiempo que afecta a su fiabilidad.¹⁵ Por tanto, en este aspecto, las escalas de clasificación presentan debilidades importantes. También se ve amenazada la validez en la fase en que se hacen inferencias, es decir, cuando se analiza el registro de conducta, aunque entonces puede utilizarse la triangulación (véanse los Capítulos VII y IX)¹⁶

Aunque la mayor parte de los sistemas de observación se ha elaborado a efectos de investigación, hay orientaciones respecto a cómo los profesores

pueden seleccionar de entre los existentes uno que se adapte a sus necesidades o elaborar uno propio. HOOK (1981) y BOEHM y WEINBERG (1977) dan consejos y ofrecen ejemplos específicos para profesores, basándose en la bibliografía tradicional de investigación cuando parece conveniente¹⁷. HARRIS, BELL y CARTER (1981) ofrecen orientaciones generales, aunque proporcionan pocos ejemplos. GRONLUND (1981) asesora sobre las escalas de clasificación, pero, como indicamos antes, se orienta hacia la valoración de los alumnos. Una idea importante, de especial relieve para los profesores que observan sus propias clases, es que los sistemas de observación no tienen por qué utilizarse para el registro “en vivo, es decir, de manera que un observador presente en el aula se encargue de ir codificando la conducta. En cambio, pueden utilizarse para analizar la conducta en el aula registrada por medios electrónicos, en cinta magnetofónica o magnetoscópica.

¹³ “McINTYRE(1980, pág. 15) dice que entre dos observadores puede desarrollarse una mutua comprensión implícita, adquiriendo mayor fiabilidad que el sistema mismo. En HÁRROP (1979) puede verse una exposición más técnica de los problemas con las medidas de fiabilidad.

¹⁴ Véase en BORICH (1977) un análisis de los motivos de invalidez en la observación estructurada. Presenta también una descripción general de los principios de la observación estructurada, pero se basa sólo en la bibliografía estadounidense.

¹⁵ Aspecto señalado por HOOK y cols. (1981, pág. 53) al referirse a la validez y la fiabilidad en el contexto de los profesores que investigan en sus propias clases.

¹⁶ HOOK y COLS. (1981, PAG 55) indican que la triangulación puede utilizarse para comprobar “observaciones”, aunque es improbable que resulte muy práctico; en general, la triangulaciones mas adecuada para comprobar las “inferencias” extraídas de las observaciones.

¹⁷ “El texto ‘clásico para construir una pauta de observación estructurada es el de MEDLEY y MITZEL (1963).

Observación no estructurada

En contraste con la observación estructurada, este enfoque contempla la conducta en sentido holístico y no como sucesos (o categorías) discontinuos. En este caso, la observación se centra progresivamente, procurando considerar la conducta desde el punto de vista de los actores (es decir, alumnos y profesores)¹⁸. Esta preocupación por cuestiones como las intenciones del alumno contrasta con la observación estructurada, que sólo considera la conducta observable. A los efectos de la evaluación, conviene distinguir entre la observación no estructurada, que utiliza métodos etnográficos, y la etnografía misma. Esta última se ocupa de elaborar teoría sobre los sistemas sociales (incluidas las escuelas y aulas), cuestión que no constituye una preocupación fundamental para la evaluación¹⁹. Sin embargo, hay características del enfoque de investigación utilizado para la elaboración de esa teoría, como la convergencia progresiva y la reflexividad, que sí son importantes para los evaluadores, sobre todo para los que trabajan dentro de su propio centro. (Volveremos sobre ello en este mismo apartado.)

En la observación participante (la principal forma de observación en etnografía), es fundamental tomar notas basadas en la observación directa. Es obvio que la bibliografía etnográfica, tanto relativa al ámbito educativo como a otros, contiene informes sobre esta técnica; aquí sólo nos ocuparemos de la bibliografía que pretende ayudar a solucionar los problemas de los profesores.

Las notas de campo constituyen la principal herramienta del etnógrafo para registrar observaciones directas. ELLIOTT (1978), Ford Teaching Project (1975) y HARRIS, BELL y CARTER (1981) presentan descripciones generales de notas de campo para profesores y ADELMAN y WALKER (1975) dan orientaciones en relación con la observación del aula en el contexto de la formación inicial. PATTON (1980) presenta un detallado informe sobre notas de campo realizadas por evaluadores profesionales (diferenciándolos de los investigadores), aunque se trata en general de agentes externos a la institución o programa investigado. Sólo HOOK (1981) se dirige expresamente y con cierto detalle a los profesores, incluyendo ejemplos de notas de campo.

La tradición de evaluación y el estudio de los niños pequeños en los Estados Unidos nos proporciona un enfoque análogo y útil: el uso de las anécdotas. Son descripciones concretas de incidentes y acontecimientos significativos, recogidas de forma sistemática y registradas en el plazo más inmediato posible al hecho. GRONLUND (1981) proporciona valiosas orientaciones para redactar las anécdotas²⁰. No obstante, la preocupación por la evaluación se plasma en el registro de muestras representativas de la conducta. En la evaluación, puede ser tan importante tener en cuenta hechos imprevistos o excepcionales como los que resulten significativos para el observador.

La confección de las notas de campo es semejante al registro de anécdotas, pero, además de las descripciones conductuales, pueden contener interpretaciones e impresiones del observador. Para mejorar su validez, deben redactarse lo más inmediatamente posible al hecho. Para un profesor, esto resulta difícil, pero, como indica ELLIOTT (1978), sí puede hacer sus anotaciones en un tiempo muerto de su ejercicio docente. Los profesores del Ford Teaching Project (1975) sugirieron que el profesor seleccionara el tema objeto de registro, de manera que revistiera un interés personal. No obstante, esto suscita ciertos problemas. En primer lugar, en la evaluación aparece siempre el problema de si el observador debe comenzar sin prejuicios o centrándose en una cuestión concreta. En segundo lugar, la selección puede estar basada no sólo en criterios claros, sino involucrar también las tendencias del profesor. Aparte de estos problemas especiales que afectan a los docentes que observan sus propias clases, cualquier observador tiene que decidir si toma notas durante el período de observación o después. La primera opción tiene la ventaja de su carácter inmediato y espontáneo, pero supone restar tiempo a la observación, por lo que algunos hechos podrían pasarse por alto. La redacción de notas después del período de observación permite realizar la tarea sin distracciones. Sin embargo, aunque pueda transmitirse una impresión general sobre el tiempo de clase, hay hechos que pueden caer en el olvido o recordarse de manera selectiva. Un posible compromiso sería anotar palabras sueltas para facilitar el recuerdo y establecer la sucesión de los hechos, para elaborar las notas más tarde. Los inconvenientes más generales de las notas de campo consisten en que: no pueden recoger las conversaciones con facilidad, aunque hay quienes afirman lo contrario (PATTON, 1980); resultan difíciles de analizar, y son subjetivas. Por otra parte,

¹⁸ El fundamento del interaccionismo simbólico consiste en la consideración de las intenciones e interpretaciones de la conducta de profesor y alumno, véase en DELAMONT (1976) un informe sobre este enfoque.

¹⁹ De esto no se deduce que la evaluación no tenga nada que ver con la teoría, cuestión sobre la que volveremos en el apartado sobre "El contexto curricular". En HAMMERSLEY (1980) puede verse una revisión de las teorías elaboradas por los etnógrafos de aula, véase una revisión sobre el uso de la técnica etnográfica en WILSON (1977).

²⁰ Basándose en los escritos anteriores de GRONLUND, CLIFT (Open University, 1981a, Bloque 4) presenta también un informe sobre los registros de anécdotas como instrumento para registrar la valoración del alumno.

tienen la ventaja de que son sencillas, no requieren la actuación de un extraño y son fáciles de revisar (operando como un recordatorio).

Las notas pueden tomarse de manera que abarquen un período largo, como un día, reuniéndolas a lo largo del trimestre o del año, por ejemplo, para constituir *un diario*. Es obvio que, *en esta forma, serán más selectivas, aunque pueden* estructurarse para dejar constancia de ciertas cuestiones. HOOK (1981) expone el uso de tarjetas de *muestreo de items* como forma de atender sistemáticamente a los problemas. El profesor guarda una tarjeta por cada problema de un *conjunto* que te parece importante. Al final de cada día, selecciona uno de ellos para escribir sobre él. Los problemas pueden escogerse aleatoriamente cada día barajando el “mazo” de tarjetas. Tanto ARMSTRONG como ENRIGHT (en NIXON, 1981) ponen ejemplos de diarios, aunque el de ARMSTRONG consistía en notas de campo redactadas después de las clases que *observaba* (como *observador* participante) en (a clase de otro profesor; ENRIGHT, por su parte, solía ser la única profesora presente en su clase.

Comparación entre la observación estructurada y la no estructurada

Como es lógico, se han producido debates metodológicos entre las dos tradiciones, en los que han participado diversos autores británicos a finales de los años setenta²¹ Los partidarios de *la observación* no estructurada han *mantenido* la ofensiva durante la mayor parte del tiempo, aunque McINTYRE y MACLEOD (1978) han contrarrestado las críticas y defendido las observaciones sistemáticas (estructuradas). Su defensa nos sirve como valioso resumen del debate²²

1. Los sistemas categoriales les muestran una concepción restringida del aula²³.

2.

McINTYRE y MACLEOD contrarrestan esta afirmación alegando que algunos de los defensores de *la observación* no estructurada parecen dar a entender que “saben” cuáles son los hechos significativos que dejan de lado Los sistemas categoriales. Sin embargo, McINTYRE y MACLEOD conceden que el uso de categorías especificadas de *antemano pasa por alto algunos datos* que no se registran y, en consecuencia, el crédito mayor se otorga a la capacidad descriptiva y explicativa de los conceptos que subyacen al sistema.

3. Los sistemas categoriales sólo proporcionan una descripción parcial de la conducta en el aula.

En respuesta a ello, McINTYRE y MACLEOD dicen que cualquier observación es parcial. Desde su punto de vista, la observación no estructurada no pone de manifiesto las razones que llevan a seleccionar un centro de atención respecto a determinados acontecimientos. Esto contrasta con las opciones explícitas que suponen las observaciones estructuradas²⁴.

3. Los sistemas categoriales, sobre todo los que reflejan las supuestas clases formales de los Estados Unidos, son inadecuados para una serie de contextos.

Por ejemplo, un sistema que resalta la importancia de las manifestaciones orales del profesor no es apto para categorizar la conducta grupal de los alumnos. Sin embargo, en la actualidad existe un amplio conjunto de sistemas adaptables a muy diversos contextos. En un nivel diferente, McINTYRE y MACLEOD reconocen que ningún sistema puede ocuparse de elementos de un contexto no compartidos públicamente en nuestra cultura, p. ej., las bromas entre un profesor y los alumnos basadas en un incidente anterior, pero, dicen, estas situaciones son raras y, por tanto, no constituyen un problema grave²⁵

4. La observación estructurada revela la falta de preocupación por los significados vinculados a la conducta de

²¹ HAMMERSLEY (1986 a y b), CHANAN y DELAMONT (1975) McALEESE y HAMILTON (1978) y BENNETT y McNAMARA (1979) presentan el trabajo de ambas tradiciones aunque el último se inclina a favor de la observación estructurada. STUBBS y DELAMONT (1976) representan el lado favorable a la observación no estructurada en el debate.

²² Hemos presentado las críticas que nos parecen más importantes; recomendamos al lector que consulte McINTYRE y MACLEOD (1978) si desea una exposición más completa de todas las críticas,

²³ McINTYRE y MACLEOD (1978) señalan que estas críticas de la observación estructurada están basadas a veces en las categorías de análisis de interacciones de FLANDERS como hemos indicado, se han elaborado sistemas mucho más complejos.

²⁴ Nos hemos referido a esta cuestión cuando hablamos de las notas de campo utilizadas en la observación no estructurada.

²⁵ Postura con la que no simpatizamos porque consideramos importantes la mutua comprensión y las *interacciones privadas de profesores y alumnos*.

alumnos y profesores.

Esta proposición recoge una preocupación fundamental de los observadores que utilizan el enfoque no estructurado, y McINTYRE y MACLEOD no niegan que constituya una limitación. Sin embargo, afirman que, a partir de la conducta observable, pueden hacerse algunas inferencias. En todo caso, hay otras técnicas, como la entrevista, que permiten acceder a esos significados²⁶.

5. La preocupación por la cuantificación es innecesaria, sobre todo para quienes están interesados en que los profesores investiguen sus propias clases.

McINTYRE y MACLEOD afirman que incluso los observadores no estructurados cuantifican cuando utilizan proposiciones del tipo: “la mayoría de los alumnos...”, pero tales expresiones son vagas. Por su parte, los observadores sistemáticos formulan enunciados precisos y utilizan la estadística para apoyarlos.

Aunque McINTYRE y MACLEOD neutralizan con acierto muchas de las críticas en contra de la observación estructurada, el debate refleja las distintas concepciones de la investigación que señalamos en el Capítulo VII. McINTYRE y MACLEOD no pretenden la hegemonía de la observación estructurada, pues dejan muy claro cuáles son sus limitaciones y esperan que se utilicen otros enfoques cuando se precisen otros tipos de información. Continúan exponiendo para qué puede servir. Haciendo hincapié en la conducta observable, reconocen que la observación estructurada no sirve para estudiar las percepciones de los alumnos y profesores. A este respecto, para ellos hay dos cuestiones importantes: ¿por qué aprenden los alumnos lo que aprenden? y ¿por qué enseñan los profesores como lo hacen? Respecto a la primera pregunta, aluden a tres tipos de pruebas: los informes de los alumnos sobre la conducta en clase; los procesos de pensamiento de los alumnos mientras están en clase, y los hechos observables en el aula. Sin embargo, afirman que los informes de los alumnos pueden no estar relacionados con las influencias sobre el aprendizaje, es decir, sobre su *forma* de aprender, y en la actualidad no contamos con procedimientos adecuados para descubrir que piensan los alumnos. Por supuesto, esto deja a los hechos observables como el mejor tipo de prueba. Considerando por qué enseñan los profesores como lo hacen, McINTYRE y MACLEOD dicen que sólo a través de la conducta manifiesta o controlando características observables pueden influir en el aprendizaje de sus alumnos. Ninguna de estas dos cuestiones requiere necesariamente el empleo de la observación estructurada, sino prestar atención a la conducta manifiesta. Es curioso que haya pocas críticas en contra de la observación no estructurada²⁷, aunque POWER (1977), examinando estudios de interacción en clase de ciencias, tenga que decir esto de la observación no estructurada: “el enfoque etnográfico muestra indicios de estar convirtiéndose en otra nueva religión. Sus profetas están a menudo más preocupados por denunciar la perversidad del mundo científico que por aclarar y resolver los problemas del paradigma interpretativo” (pág. 21)²⁸. Resaltando los significados atribuidos a la conducta, POWER afirma que la investigación etnográfica reduce la realidad a “lo que ocurre en el interior de la cabeza de las personas [y] lleva a perder de vista de qué manera las limitaciones externas configuran la interacción en clase... cuyas consecuencias están ahí, con independencia de que sean percibidas por profesores y alumnos. Es más, aunque las pautas de interacción pueden ser en parte el producto de las definiciones de la situación de los participantes, existe la posibilidad de que esos participantes tengan una falsa consciencia” (pág. 22). (Véase también la crítica de los enfoques exclusivamente interpretativos que presentamos en el Capítulo VII.)

El profesor como observador

La falsa consciencia se hace problemática por partida doble cuando uno más de los participantes, o sea, los profesores, se convierten en observadores. Estudiaremos ahora este caso especial de la observación. Como la mayor parte de la bibliografía refleja las preocupaciones de los investigadores, que vienen a la escuela desde el exterior, pocos autores han considerado los problemas especiales que afectan a los docentes cuando se convierten en observadores de sus propias clases. Por ejemplo, si un profesor deja de lado la enseñanza para observar, su abandono modificará la situación²⁹. Aunque STENHOUSE (1975) defendía que los profesores investigasen sus propias clases, no dio demasiadas orientaciones, salvo el rechazo de la observación estructurada. El principal texto detallado que se ocupa específicamente de las preocupaciones de los profesores que llevan a cabo observaciones procede de la Deakin University (HOOK, 1981).^{30 31} El material del

²⁶ Asimismo, McINTYRE (1980, pág. 14) dice que quienes construyen sistemas de observación estructurada sólo pueden validarlos relacionándolos con las pruebas acerca de cómo estructuran los participantes los acontecimientos pertinentes.

²⁷ Véase, no obstante, en DELAMONT (1981) una crítica interna, en especial del estrecho margen de lectura del trabajo que se realiza en otros países y en otras tradiciones teóricas de los etnógrafos de la enseñanza.

²⁸ McINTYRE (1980, pág. 3) se refiere también a esas denuncias del enfoque positivista, pero dice que no suelen basarse en argumentos sólidos, consistiendo sólo en ataques polémicos o en descalificaciones desdeñosas.

²⁹ SMETHERHAM (1978) lo considera como un problema de compromiso con el rol participante que produce cuando el profesor está demasiado ocupado con extraer datos en su calidad de investigador.

³⁰ Es estructurado, muy semejante a un texto de la Open University, con muchos ejemplos y

Ford Teaching Project, que hizo uso por primera vez de métodos de investigación adecuados para los profesores como investigadores, sólo contiene orientaciones generales y, hasta cierto punto, se inclina a favor de una forma concreta de cambio curricular, al centrarse como hace en la crítica de las estrategias. HARRIS, BELL y CARTER (1981) también son breves, aunque abarcan la mayor parte de las cuestiones. No obstante, prestan poca atención a las cuestiones metodológicas. BOEHM y WEINBERG (1977), aunque se centran en niños pequeños, proporcionan orientaciones detalladas sobre la observación. Como mencionamos antes, ADELMAN y WALKER (1975) dan orientaciones en el contexto de la formación inicial del profesorado, aunque esto no incluye consideración alguna de la observación estructurada.

Es obvio que, entre las rutinas de los profesores, está la de observar, aunque hace falta que sus observaciones sean más sistemáticas y continuadas (ADELMAN y WALKER, 1975; ANDERSON, 1979). Este reconocimiento hace hincapié en el elemento "participante" de la "observación participante" y supone que sólo el profesor puede hacer juicios informados en clase (Ford Teaching Project, 1975). Implica también que el docente tiene acceso a las razones para actuar³² y, en cuanto observador, es único, al permanecer en el aula durante un período de tiempo largo³³. Sin embargo, esta participación suscita dificultades: los profesores son, en efecto, participantes, pero con un papel específico que desempeñar que los excluye de parte de la acción asociada con la "cultura del alumno"³⁴. Ya hemos señalado el problema de abandonar la actuación docente para observar, modificando, en consecuencia, la situación (un conflicto entre dos roles, el de profesor y el de observador)³⁵. Puede que los alumnos no estén seguros de cómo interpretar la conducta del docente, provocando de este modo el 'observador' problemas de "reactividad". Como veremos, la grabación electrónica es una forma de reducir este efecto porque pospone la observación.

Asociado con el problema del rol está el de la inclinación tendenciosa. En su nivel más simple, la observación de un profesor puede adolecer del "efecto halo", es decir, el conocimiento antecedente de la situación puede utilizarse para ser aún más selectivo. Por ejemplo, los niños de familias pobres pueden hablar menos, parlo que se harán notar menos (BOEHM y WEINBERG, 1977), o los chicos pueden recibir un tratamiento diferente al de las chicas (DELAMONT, 1981). En otro nivel, se indica que los profesores pueden ser más proclives a hacer una interpretación tendenciosa de la conducta (HOOK, 1981, págs. 29 y 31). Más profundo aún es el problema de la "falsa consciencia" que señala POWER (1977). Es posible que un profesor malinterprete la conducta de un alumno, como indicará la acción que adopte como consecuencia de esa conducta (GOOD y BROPHY, 1978). Los profesores e investigadores del Ford Teaching Project (1975) hablaban de la intuición desorientada y de que era preciso contar con pruebas contundentes para fundamentar la acción. Es obvio que la reflexividad es el instrumento del etnógrafo para neutralizar los problemas de las tendencias infundadas y, por tanto, para contrarrestar las críticas respecto de la selección de los centros de atención efectuadas por McINTYRE y MACLEOD (1978) (véase la exposición precedente). Con respecto a la observación estructurada, el empleo de categorías especificadas de antemano reduce la posibilidad de que influyan tendencias infundadas (EGGLESTON, GALTON y JONES, 1975, pág. 58), y la manifestación explícita del fundamento del sistema de categorías constituye una forma de reflexividad. En un sentido más positivo, los esfuerzos para reducir las tendencias infundadas harán más conscientes a los profesores y más explícito su conocimiento tácito (ANDERSON, 1979).

Las técnicas de registro

Nos lleva esto a considerar las técnicas de registro que permiten evitar la selección de centros de atención y un acceso repetido a la conducta registrada, cuestiones importantes ambas para neutralizar las inclinaciones tendenciosas. Para los profesores que actúan como observadores, esos registros no constituyen la única

ejercicios: en realidad, como la Open University, la Deakin University enseña a distancia.

³¹ DAVIS (1980), también australiano (aunque sin relación con la Deakin University), se dirige específicamente a los profesores que desarrollan la evaluación en sus propias escuelas. Sin embargo, desde nuestro punto de vista, su enfoque esté excesivamente cargado de teorías y modelos de evaluación, relegando el examen de las técnicas a un apéndice.

³² SMITH y GEOFFREY (1968) lo considera uno de los puntos fuertes de la actuación del profesor de clase como observador participante. También se refiere a ello HOOK (1981, pág. 107). HAMILTON y DELAMONT (1974, pág. 9) hacen una observación similar.

³³ HAMILTON y DELAMONT (1974, pág. 7) hablan de la presencia prolongada de un observador en clase.

³⁴ La falta de conocimiento de las percepciones, etc., de los alumnos puede superarse entrevistándolos. Mejor aún, quienes trabajaron en el Ford Teaching Project (1975) descubrieron que la entrevista múltiple resultaba útil porque los alumnos se mostraban más abiertos y sinceros con un extraño que con su profesor. JEFFOCATE (1981) descubrió también que su papel como entrevistador externo era útil para evaluar el área delicada de la enseñanza multicultural.

³⁵ Conflicto señalado por HOOK (1981, pág. 59) y por HARRIS BELL y CARTER (1981, pág. 212, 1/6). Estos últimos autores y STENHOUSE (1975, pág. 155), citando un estudio doctoral de David HAMILTON, reconocen que este conflicto variará según el tipo de organización de clase. La cita de HAMILTON se refiere a que cualquier observador adulto puede ser considerado como profesor. suscitándose, por tanto, el mismo conflicto. STENHOUSE concluye que las clases "abiertas", con un rol de profesor más fluido, son más adecuadas para estimular el desarrollo del rol observador.

fuentes de datos, sino una ayuda para reconstruir la clase (ELLIOTT, 1978, pág. 79). ADELMAN y WALKER (1975, pág. 129) adoptan la perspectiva de que las grabaciones deben acompañar a la observación participante por lo que no eximen de tomar notas.

La grabación magnetofónica constituye el método básico para el profesor porque, interfiriendo menos, es más fácil de utilizar mientras se enseña (*Open University*, 1981a, Bloque 3, pág. 23). Es más, si el profesor está presente, es menos importante tener grabada toda la información y, se afirma, la enseñanza sigue siendo sobre todo verbal (ELLIOTT, 1978, pág. 80)³⁶ Los profesores que participaron en el Ford Teaching Project llegaron a decir que la grabación magnetofónica era la herramienta de investigación más importante, y BASSEY y HATCH (1979) abandonaron la grabación magnetoscópica en beneficio de la magnetofónica mientras elaboraban un sistema de observación estructurada para uso de los profesores. Cuando se utiliza un micrófono de radio o el profesor lleva un magnetófono de casete, todas las conversaciones que se produzcan en el radio de acción del micrófono quedarán recogidas. Sin embargo, también tiene inconvenientes: el coste del equipo; la reticencia de los alumnos, aunque ésta pueda reducirse teniendo el magnetófono en clase antes de que haga falta utilizarlo; el tiempo necesario para la transcripción; el carácter incómodo del equipo de grabación. La causa fundamental y universal de quejas al respecto es el tiempo necesario para las transcripciones!

En los siguientes textos pueden hallarse orientaciones respecto al uso de magnetófonos, incluidas las cuestiones relativas al equipo adecuado y a la mejor posición del micrófono:

ANDERSON (1979), que es una guía obligada para las grabaciones; *Open University* (1981a, Bloques, págs. 25-29), es también una guía, basada en parte en ejemplos extraídos de la experiencia de dos profesores. BASSEY y HATCH (1979), que dan breves indicaciones del equipo y disposición utilizados por ellos. HOOK (1981, págs. 237-239), que ofrece detalles generales y alude a los problemas de las grabaciones. BOYDELL (1975), que presenta una breve exposición de los problemas que plantean las grabaciones.

Otra técnica posible consiste en la grabación magnetoscópica, pero BASSEY y HATCH (1979) manifiestan que, en su caso, no mereció la pena. Sin embargo, si hay que tener en cuenta la conducta no verbal, el observador no conoce la situación (p. ej., no es profesor), o si hay que reproducir las grabaciones ante niños pequeños para calibrar sus percepciones, se hace necesaria la grabación magnetoscópica. Aunque ésta proporciona un registro más completo de la conducta en el aula (dependiendo, por supuesto, de las cualidades técnicas del equipo), tiene, sin embargo, una serie de inconvenientes:

1. La cámara proporciona una visión selectiva, sobre todo si se utiliza más de una, puesto que esto requiere hacer el montaje de tomas³⁷
2. El análisis de la grabación es complejo³⁸.
3. El equipo es costoso e incómodo³⁹.
4. La elaboración de informes con los datos así recogidos es difícil.

Los detalles de la grabación magnetoscópica para la observación de clase no suelen aparecer en textos sobre la cuestión. Por ejemplo, *Open University* (1981b, Bloque 3) sólo lo menciona en términos muy generales y HOOK (1981), aunque presenta mucha información útil sobre el uso del vídeo, proporciona pocos detalles técnicos en contraste con su tratamiento de la fotografía. BASSEY y HATCH (1979) dan las orientaciones más detalladas, basándose en su propia experiencia, y ADAM y BIDDLE (1970) describen, en líneas generales, cómo utilizaron un conjunto de dos cámaras para su investigación.

El uso de diapositivas, sobre todo si están sincronizadas con una grabación magnetofónica, supone un compromiso útil⁴⁰, pero requiere la presencia de alguien ajeno para registrar la conducta del profesor, salvo que

³⁶ Sin embargo, los profesores del Ford Teaching Project (1975) consideraron la falta de claves visuales como una desventaja.

³⁷ ADELMAN y WALKER (1975) indican que las convenciones de la realización de programas de televisión se desarrollan con frecuencia al modo del registro observacional. Es un problema experimentado en la *Open University* realizar programas sobre la vida en clase. Véanse, por ejemplo, los tres programas que acompañan al curso E364 (*Open University*, 1981a). Con un equipo de dos cámaras es posible efectuar una grabación utilizando la técnica de pantalla partida, con lo que se evita la selección debida al montaje (utilizada por ADAMS y BIDDLE, 1970).

³⁸ Por supuesto, puede utilizarse sólo como estímulo para pensar, en vez de como fuente de datos que analizar. En el contexto de un estudio de investigación, ADAMS y BIDDLE (1970) estimaban necesarias entre 20 y 25 horas dedicadas a codificación por cada hora de grabación.

³⁹ Las dificultades experimentadas por los profesores para trasladar el equipo fueron la razón principal para que BASSEY y HATCH (1979) abandonaran el uso del vídeo en beneficio de la grabación magnetofónica.

⁴⁰ DENTON y SEXTON (1978) muestran una lista de las ventajas de la fotografía, en relación con su disponibilidad, sencillez de funcionamiento, fiabilidad técnica, versatilidad (p. ej., macrofotografía y uso sin flash), permanencia de los datos, atractivo en los informes y capacidad de proporcionar medios de comprobación de la validez.

un alumno maneje la cámara. Utilizando diapositivas semienmarcadas y una grabación magnetofónica, ADELMAN y WALKER (1975) son los exponentes mejor conocidos de esta técnica. Aunque es muy selectiva, dependiendo en gran medida de los juicios de quien maneje la cámara, la incertidumbre o ambigüedad inherente a las diapositivas parece constituir un instrumento muy útil para conseguir que profesores y alumnos hablen sobre la clase a *posteriori*. Sin embargo, es preciso que pase algún tiempo hasta que se reduzca la reactividad. Los detalles técnicos de este procedimiento pueden consultarse en ADELMAN y WALKER (1975)⁴¹. HOOK (1981) trata la fotografía en diapositivas con cierto detalle y en menor medida *Open Universiti* (1981a, Bloque 3).

El análisis

Los enfoques de observación estructurada y no estructurada tratan esta cuestión de forma diferente. Para poder estructurar un plan de observación es preciso haber dedicado algún tiempo a pensar sobre el análisis. Después de recoger los datos, sea en forma de marcas de frecuencia o de sucesión de categorías, hace falta analizarlos. En la observación no estructurada, el análisis comienza al mismo tiempo que la recogida de datos y continúa durante la convergencia progresiva. A intervalos y al final de la investigación, se revisan los datos y se realiza un análisis más reflexivo⁴². Si los planes de observación se utilizan para analizar la conducta registrada con anterioridad, p. ej., la conducta grabada en vídeo, en vez de registrada “en vivo”, el análisis puede llevarse a cabo desde diferentes ángulos, utilizando diversos planes (o modificaciones del original). Sin embargo, con independencia de cómo se utilice, en primera instancia, el análisis es una actividad mecánica que implica, por ejemplo, el cálculo de diversos índices. En el sistema de siete categorías de BASSEY y HATCH (1979), puede calcularse el porcentaje de frecuencias de “preguntas del profesoro” o de “orientaciones” o el número de interacciones largas y cortas contabilizadas. La determinación del significado de estos índices constituye un proceso creativo de especulación y juicio. No obstante, BASSEY y HATCH (1979) proporcionan ciertas orientaciones para evaluar la enseñanza por medio de una serie de cuestiones como éstas:

Considere las diferencias entre las interacciones largas y cortas. ¿Acaso en su clase las interacciones cortas sirven sobre todo para mantener ocupado a todo el mundo, mientras que las interacciones largas son más valiosas como acontecimientos docentes? ¿Sería deseable tener más interacciones largas?
(BASSEY y HATCH, 1979, pág. 137.)

McINTYRE (1980 pág. 18) dice que es difícil que el simple recuento de frecuencias sirva como indicio de una enseñanza eficaz. Según él, es más importante dilucidar cómo se relaciona el acontecimiento (que se contabiliza) con los que le preceden y con los que le siguen.

El tan difamado Interaction Analysis System de FLANDERS tiene un procedimiento muy desarrollado para el análisis que va mucho más allá del simple recuento de frecuencias, ofreciendo una matriz que permite representar cómo se desarrollan las interacciones verbales. Esto requiere la elaboración de un registro de la clase como sucesión de categorías. Por ejemplo, cualquier cambio desde la categoría 5 (el profesor imparte clase magistral) a la categoría 6 (el profesor da directrices) puede representarse como una anotación en una casilla de la matriz. FLANDERS ha elaborado también modelos de clases, como de “pregunta corta, respuesta corta” o de “investigación creativa”, que pueden identificarse a partir de la matriz completa.

Como señalamos antes, la observación estructurada supone empezar con algunas ideas sobre lo que se va a descubrir. No obstante, para los profesores que empiezan a realizar evaluaciones, este enfoque puede resultarles muy exigente. En consecuencia, cobran gran importancia los enfoques de evaluación cualitativa y como las grabaciones magnetofónicas se consideran una herramienta muy importante, nos centraremos en su análisis. El primer paso consiste en transcribir la grabación. Esta operación no sólo proporciona los datos sobre los que trabajar, sino que el mismo proceso hace surgir ideas para el análisis. No obstante, ha de establecerse un equilibrio entre la obtención de una transcripción completa, que lleva mucho tiempo (p. ej., unas diez horas para transcribir una grabación de una hora), y la tarea más viable de elaborar una transcripción parcial. ANDERSON (1979) aconseja al principiante que vaya haciendo una transcripción cada vez más detallada, terminando con una pequeña parte de la grabación “completamente” transcrita⁴³.

⁴¹ Tratan también de la foto fija sincronizada, auténtico compromiso entre la película y la grabación magnetofónica. No obstante, ésta es una técnica compleja de investigación que no está al alcance de la mayoría de las escuelas. Pueden verse los detalles técnicos de este enfoque en ADELMAN y WALKER (1971/1972).

⁴² El enfoque general de este análisis, característico de la etnografía, se expone en este mismo apartado. El aspecto fundamental consiste en la elaboración de hipótesis y a comprobación de éstas respecto a los datos, por ejemplo, buscando casos negativos.

⁴³ La transcripción completa es difícil porque algunos elementos del habla son difíciles de representar. En realidad, ANDERSON (1979) orienta acerca de las convenciones que deben adoptar para ayudar a perfeccionar esta representación.

De este modo, la transcripción comenzará intentando obtener una visión general (*Op en University, 1981a*, Bloque 3, pág. 30), antes de ocuparse de aspectos o sucesos concretos. Sean cuales sean los criterios utilizados para efectuar esa selección⁴⁴, es importante observar que constituyen un aspecto de la reflexividad. Esa percepción incluye el marco mental del analista, factor que influye en la transcripción (*Open University, 1981a*, Bloque 3, pág. 31). La transcripción puede utilizarse entonces para convertir lo conocido en extraño. Por eso es importante dejarla como está; el habla suele ser fragmentaria y “engañosa” (ATKINSON, 1981). En contraste con quienes trabajaron en el *Ford Teaching Project*, algunos autores hacen hincapié en la importancia de atender a la organización de la intervención oral, recomendando una minuciosa inspección de la transcripción⁴⁵. Ponen en guardia contra la tentación de esperar que surjan de los datos conceptos fundamentales, como “control social” (ATKINSON, 1981, página 101), y aconsejan estructurar el análisis a partir de los tecnicismos del habla en el nivel de cada enunciado, pasando a considerar después los episodios que se suceden dentro de la clase⁴⁶. Cuando los datos se toman de notas de campo, solas o con ayuda de una grabación magnetofónica, pueden utilizarse las directrices generales para el análisis cualitativo. PATTON (1980) proporciona orientaciones a quienes se ocupan de la evaluación más que de la investigación⁴⁷. Identifica las siguientes fases:

1. Descubrir un centro de atención para el análisis, ya sea partiendo de las preocupaciones de los responsables de las decisiones o de lo iniciado en la recogida de datos.
2. Organizar, comprobar y, por razones de seguridad, copiar incluso los datos.
3. Elaborar una descripción o narración cualitativa.
4. Elaborar, de forma inductiva, categorías, denominaciones y tipologías.
5. Desarrollar un análisis lógico de estas categorías para tratar de encontrar después clasificaciones cruzadas.
6. Expresar estas categorías, etc., cuando convenga, mediante metáforas⁴⁸.
7. Imaginar posibles causas, consecuencias y relaciones.

JAMES y EBBUTT (1981, págs. 90-91) describen los problemas que encuentra el profesor al analizar los datos y sugieren utilizar fichas de índice para las categorías que van apareciendo, permitiendo efectuar referencias cruzadas y buscar pautas.

La observación fuera de y en distintas aulas

La bibliografía sobre técnicas para este tipo de observación es mucho menos abundante, aunque, por supuesto, lo dicho hasta aquí sobre la observación no estructurada, sobre todo las notas de campo, tiene aplicación a nuestro tema⁴⁹. En consecuencia, describiremos brevemente dos estudios en los que se utilizó la observación estructurada. Uno de ellos versó sobre el “día del profesore” y el otro consistió en un estudio de seguimiento permanente de tres alumnos.

⁴⁴ “ Los miembros del *Ford Teaching Project* aconsejaban transcribir los datos importantes, basándose en una idea clara de las pruebas necesarias- En contraste con ANDERSON (1979), el equipo del proyecto no contempló la grabación magnetofónica como la principal fuente de datos, sino como base para reconstruir la clase.

⁴⁵ como es natural, estos autores (p. ej., STUBBS 1984, y ATKINSON, 1981) se ocupan del lenguaje de la clase. STUBBS (1981) habla de la selección no fundamentada de datos a partir de las transcripciones, que deja de lado la organización lingüística. Sin embargo, ofrece pocos detalles respecto a cómo tenerla en cuenta. HAMMERSLEY (1980, pág. 56), comparando los enfoques de los etnógrafos del aula con los de los sociolingüistas y etnometodólogos, no observa nada intrínsecamente erróneo en las selecciones a efectos de ilustración (lo que, en realidad, recomendaba el equipo del *Ford Teaching Project*), siempre que los criterios de elección fuesen claros.

⁴⁶ ANDERSON (1979) presenta una secuencia adecuada para este tipo de análisis, con diversos niveles que tienen cierta semejanza con *A System for Analysing instructional Discourse* (BROWN y ARMSTRONG, 1978). *Open University* (1981a, Estudio de caso 4) presenta dos estudios de casos de profesores que analizan sus clases mediante transcripciones. Ambos trabajan sobre todo en un nivel técnico detallado.

⁴⁷ A partir de la tradición de investigación etnográfica, BECKER y GREAR (1960) presentan un estudio clásico. *Open University* (1981a, Bloque 3) contiene una caracterización de la autoevaluación del profesor según las distintas fases de esa tradición.

⁴⁸ Es una técnica utilizada por David JENKINS en la evaluación de un proyecto curricular (SHIPMAN, 1974, capítulo VI). utiliza nueve “metáforas orientadoras” para describir cómo perciben los profesores el proyecto en las escuelas de prueba, por ejemplo: “intercambio de cualidades: el proyecto como obligación *recíproca*; “muestra libre: el proyecto como comercio” (pág. 99).

⁴⁹ como indicamos al principio del capítulo, la observación fuera de clase suele realizarse por métodos indirectos, por ejemplo, entrevistas, y es frecuente que se traduzcan en estudios de casos (véase “El contexto curricular).

-*Shadow study* Estudio de seguimiento “convirtiéndose en sombra”. La idea expresada es:

seguir o *mejor* observar a una persona, siguiéndola a todas partes sin *perderla de vista*, o sea ‘convertirse en su sombra’. La idea es parecida a “observación permanente”, pero la palabra *shadow* es más expresiva y, por otra parte, no se confunde con la expresión “observación permanente” propiamente dicha. (*N. del R.*)

El estudio del día del profesor (HILSUM y CANE, 1979) se ocupó de los compromisos de los docentes, más que de la enseñanza o de las relaciones con los alumnos. Se utilizó un registro continuo de la conducta de todos los profesores, de tiempo, sucesión y duración. Se codificó toda la actividad y se puso un número en una hoja en la que cada minuto del día estaba dividido en intervalos de cinco segundos, permitiendo colocar varios códigos en cada intervalo. Las anotaciones se hicieron de manera que pudiese medirse la duración de cada actividad. Asimismo, los profesores cumplimentaron un “diario” precodificado en el que señalaron el tiempo utilizado en determinadas actividades fuera de la escuela. El inventario permitía registrar también anotaciones cualitativas y comentarios.

Un *shadow study** es aquél en que los alumnos quedan sometidos a un seguimiento continuado casi absoluto durante cierto tiempo, generalmente entre un día y una semana. El HMI utiliza a menudo esta técnica⁵⁰. Las observaciones suelen ser no estructuradas, pero en el caso considerado por nosotros (el estudio paralelo realizado en la *Sir Frank Markham School*, 1981; véase el Capítulo V) se utilizaron inventarios. Tres alumnos de la misma clase, de capacidad elevada, media y baja, respectivamente, fueron sometidos a seguimiento continuado durante una semana. En cada clase se cumplimentó una serie de inventarios que abarcaban cuestiones como: formato de la clase (p. ej.: clase magistral, preguntas y respuestas, trabajo en grupo); materiales utilizados; actividades del alumno (p. ej.: lee en silencio, copia, escucha); preguntas del profesor. Algunos inventarios requerían efectuar una clasificación relacionada, por ejemplo, con lo aburrido que pareciera el alumno. Además de los inventarios de cada clase, se redactó una breve descripción cualitativa y se controlaron las tareas de los alumnos investigados, la atmósfera del aula, la confianza del alumno y las tareas para casa de los mismos. El observador, que ostentaba el cargo de subdirector de la escuela, era consciente de su condición de intruso y de los problemas de muestreo, en relación con la selección de los alumnos y el muestreo durante el período de observación⁵¹.

Una alternativa a este enfoque estructurado podría haber consistido en un estudio cualitativo que tratara de aprehender las sensaciones de cada alumno durante la semana e incluiría entrevistas. No obstante, el estudio estructurado reveló muchas ideas que podrían haberse investigado con detalle.

Problemas

Un problema fundamental consiste en decidir qué enfoque es más apropiado. Aunque muchos autores prefieren que los profesores utilicen enfoques no estructurados, hay ejemplos en los que los docentes han sacado provecho del uso de la observación estructurada⁵². Nuestra postura se parece a la de MCINTYRE y MACELEOD (1978) en que recomendamos que se utilice la técnica que parezca más adecuada para la evaluación concreta que se vaya a hacer. Aunque no es posible enumerar todo el conjunto de contingencias que pueden surgir, conviene considerar algunos factores que intervienen en la decisión sobre el enfoque adecuado. Es obvio que si la observación se realiza en una serie de clases de una escuela, con el fin de compararlas, la observación estructurada es viable. En este sentido, se trataría de un estudio de comprobación de hipótesis y no de generación de las mismas. Después de todo, los profesores tienen una amplia experiencia de clases individuales, por lo que puede resultar muy productivo compararlas⁵³. Pueden utilizar su experiencia para crear categorías preespecificadas, con mayor capacidad descriptiva y explicativa quizá que las creadas por los investigadores (cf. nota 26). Por otra parte, cuando su conocimiento de las clases sigue aún en un nivel implícito, pueden ser más útiles las observaciones no estructuradas y las descripciones narrativas⁵⁴.

Otro problema a tener en cuenta es el de las inclinaciones tendenciosas. Ya hemos comentado que los distintos enfoques tratan de forma diferente esta cuestión. Los problemas particulares experimentados por los profesores en cuanto observadores, a los que ya nos hemos referido, pueden imponer una pesada carga sobre su reflexividad, traduciéndose en un volumen considerable de notas de campo, de las que sólo una pequeña

⁵⁰ “*The Practical Curriculum*” (*Schools Council*, 1981, págs. 43-44) recomienda esta técnica a las escuelas.

⁵¹ El período en cuestión precedía inmediatamente a la fiesta de primavera de la escuela, o que se tradujo en actividades anormales en algunas clases.

⁵² Entre los defensores de los enfoques no estructurados se encuentran: STENHOUSE (1975), ELLIOTT (1978) y el *Ford Teaching Project* (1975). En *Open University* (1981 a, Estudio de caso 4) aparecen ejemplos de profesores que utilizan la observación estructurada. AINLEY y LAZONBY (1981) combinan el informe no estructurado del profesor sobre su clase con la observación estructurada a cargo de un investigador.

⁵³ POWER (1977) critica a los investigadores etnográficos que se imitan a un individualismo metodológico ya los estudios de casos de situaciones singulares[que, por tanto] están abocados a una falta de interés práctico” (pág. 22).

⁵⁴ ANDERSON (1979) defiende el análisis de las transcripciones de clases como método para poner de manifiesto el conocimiento que el profesor tiene del aula y advierte a los docentes para que eviten el análisis cuantitativo de las transcripciones en primera instancia. Si lo realizan corren el riesgo de perder de vista el contexto de la interacción, equivocar la sucesión de acontecimientos de la clase y, en todo caso, carecerán de transcripciones suficientes para generalizar.

proporción registren datos observables⁵⁵. La utilización de un observador externo (de lo que nos ocuparemos más a fondo en el Capítulo IX) y el uso de grabaciones magnetofónicas constituyen, por supuesto, alternativas al empleo de observaciones estructuradas para reducir las desviaciones tendenciosas. AINLEY y LAZONBY (1981) describen el uso de un inventario de observación estructurada a cargo de un observador externo, combinado con un informe del profesor sobre la clase. Este informe se utilizó después para validar el inventario⁵⁶. El problema del volumen de las notas de campo suscita quizá el problema más importante para un profesor que participa en una evaluación: el exceso de datos. Es un problema que afecta a todas las técnicas, sean de observación, de valoración o cualesquiera otras. Sin embargo, al tener que recoger datos observables de las clases que, por naturaleza, son rápidamente cambiantes y complejos, el problema se incrementa de manera notable. El consejo obvio en tales situaciones consiste en comenzar de manera sencilla, por ejemplo, con notas de campo redactadas después de clase, pasando más tarde a las grabaciones magnetofónicas, a la observación compartida o a ambas.

Observación indirecta

Introducción

Las técnicas a las que nos referimos en este apartado se precisan cuando no puede observarse directamente la conducta, como en el caso de que ésta ocurra con poca frecuencia o se desarrolle en privado. Incluso cuando la conducta es pública y no infrecuente, puede que el observador no esté presente. En consecuencia, los hechos han de reconstruirse a partir de informes sobre los mismos. Sin embargo, estas técnicas no se quedan en esto; proporcionan información sobre los motivos y actitudes respecto a determinados acontecimientos, así como acerca de las actitudes, valores y creencias en general. Los defensores de “modelos científicos” de investigación consideran estos métodos como sucedáneos pobres de la observación directa⁵⁷. Sin embargo, como indicamos cuando hablamos de la observación directa, hay investigadores como DELAMONT (1976), de la tradición interaccionista simbólica, que consideran fundamental para su metodología la revelación de las actitudes, valores y creencias⁵⁸. Como en el caso de la observación directa, la “estructura” constituye una clave para caracterizar las técnicas. Aunque el argumento metodológico respecto a la estructura tiene aquí su homólogo, en la bibliografía parece menos polarizado, permitiendo una elección más pragmática. Si, como investigadores, tenemos claro lo que nos interesa descubrir, podemos especificar y utilizar cierto grado de estructura⁵⁹.

Las técnicas de las que nos ocuparemos aquí son las entrevistas, los cuestionarios, los diarios de los alumnos y las medidas sin intrusión. Tratamos mas ampliamente las entrevistas y los cuestionarios porque su uso en evaluación es más obvio⁶⁰. Son también técnicas básicas de la investigación en ciencias sociales, por lo que la bibliografía que se ocupa de ellas es muy extensa. Aunque las escuelas plantean problemas a investigadores y evaluadores, no son menos difíciles que los que tienen que afrontar otros investigadores sociales que, por ejemplo, realizan entrevistas relacionadas con problemas como las violaciones o la brutalidad policiaca (SMITH, 1981). Sin embargo, en comparación con la extensa bibliografía sobre técnicas en general, es muy reducida la que trata específicamente de los profesores trabajando en sus propias escuelas⁶¹.

⁵⁵ Por ejemplo, un estudio de casos de la autoevaluación de un profesor, correspondiente sólo a un periodo de clase (*Open University*, 1981a, Estudio de caso 4), se tradujo a 18.000 palabras, como transcripción de quince minutos de grabación magnetofónica!

⁵⁶ Dicen, asimismo, que durante un período de tiempo largo, los niños transformaron el rol del agente externo. Este pasó de ser un observador no participante a observador participante (lo que planteó problemas para el registro de la conducta en la plantilla). Señalan que, utilizando a otro profesor de la escuela como observador, se acelera el proceso de implicación en las actividades de clase.

⁵⁷ Por ejemplo, SMITH (1981) señala esto mismo al referirse a las entrevistas y cuestionarios como elementos del enfoque de “encuesta”.

⁵⁸ A partir de este tipo de tradición, BECKER y GEER (1957) también se muestran a favor de la observación, pero de a observación participante, frente a la entrevista. No obstante, su postura no debe interpretarse como extrema, según se desprende de su diálogo con TROW (véase TROW 1957, y BECKER y GEER, 1958).

⁵⁹ Por ejemplo, se aprecia más la necesidad de utilizar métodos de triangulación véanse las referencias de la Nota 58.

⁶⁰ En principio, pensamos incluir aspectos de sociometría y constructos personales. Aunque suele recomendarse a los profesores la sociometría como herramienta de investigación (p. ej.: HOOK, 1981, y GRONLUND, 1981) y la manifestación de los constructos personales tiene aplicaciones evidentes en la evaluación (p. ej., para investigar las percepciones mutuas de alumno y profesor; véase NASH, 1976), no existen pruebas publicadas sobre su uso en la evaluación que orienten a los profesores. Para quienes estén interesados en estas técnicas, los trabajos siguientes les ofrecen explicaciones y orientaciones: sociometría: COHEN (1976); HOOK (1981), y GRONLUND (1981). Constructos personales: HALL (1980) hace una exposición básica, y BANNISTER y MAIR (1968) y FRANSELLA y BANNISTER (1977) son más avanzados. La teoría original del constructo aparece en KELLY (1955).

⁶¹ STENHOUSE (McCORMICK y cols., 1982, pág. 267) hace referencia a esta cuestión, indicando que las tesis de licenciatura constituyen una importante fuente de experiencia.

Entrevistas o cuestionarios

Los investigadores pertenecientes a la tradición de investigación por encuesta (p. ej., SMITH, 1981) consideran los cuestionarios como entrevistas estructuradas autoadministradas, tratándolas del mismo modo que éstas. Naturalmente, quien responde a un cuestionario no puede pedir aclaraciones (por lo que las preguntas tienen que ser más sencillas y claras) ni el entrevistador sondear las respuestas; la motivación y el interés han de mantenerse por medio de artilugios de diseño en vez de a través de la formación de una relación personal. Por otra parte, los etnógrafos consideran que la entrevista no estructurada constituye un enfoque característico que forma parte de la observación participante. Aparte de estos conflictos metodológicos, quienes están en las escuelas han de tener en cuenta otros problemas. Por ejemplo, los niños pequeños o los que no escriben bien sólo pueden ser entrevistados (WRAGG, sin fecha) y es posible que los alumnos prefieran revelar información delicada sobre el trabajo de su profesor de un modo anónimo, a través de un cuestionario (Ford Teaching Project, 1975). Existe también un problema de muestreo, que supone decidir si todos los alumnos del centro participarán o no en la evaluación y de cuántos profesores se podrá disponer para entrevistar. En el nivel de aula, la población es pequeña y quizá no sea necesario pasar un cuestionario a todos los alumnos; además, puede que sea mejor entrevistar a una muestra bien escogida (ELLIOTT, 1981a). Con grupos que también constituyen un objetivo, como los padres, cuya dispersión (geográfica) es grande, los cuestionarios resultan atractivos. Con independencia de la técnica que se escoja, no debe iniciarse a la ligera. Las entrevistas con alumnos, aunque parezca algo obvio para un profesor, no sólo requieren determinadas destrezas personales, sino que plantean también problemas de rol (que consideraremos a continuación).

Entrevistas⁶²

Tipos

Hemos mencionado ya las entrevistas estructuradas y no estructuradas; son poios de un continuo de tipos. En un extremo del continuo encontramos las entrevistas estructuradas con cuestiones preparadas de antemano con precisión y respuestas especificadas con antelación (como las de un cuestionario); en el otro extremo aparecen las entrevistas no estructuradas, no directivas, en las que los entrevistados escogen los temas sobre los que hablar. Cada una de ellas representa el paradigma de dos tradiciones de investigación. Entre ellas se encuentran fórmulas de compromiso mayor o menor con uno de ambos extremos:

1. La entrevista centrada, cuyo tema es una situación o acontecimiento concreto, especificado de antemano y, de acuerdo con un conjunto de directrices generales, trata de explorar las percepciones que el entrevistado tiene en relación con la situación o evento (HOOK, 1981, páginas 138-139)⁶³
2. La entrevista semiestructurada, que supone un conjunto de temas cuidadosamente seleccionados y respuestas predeterminadas, combinadas con cuestiones abiertas que permiten a los entrevistados formular sus contestaciones, que el entrevistador debe anotar (WRAGG, sin fecha)⁶⁴
3. La entrevista informal, conversacional, que no da por supuesto nada de lo que pueda surgir. El entrevistador sigue el flujo de ideas del entrevistado y toma notas después de celebrada (PATTON, 1980)⁶⁵.

Como señalamos, en *ambas* tradiciones de investigación se hace un uso mucho más flexible de las entrevistas. Por ejemplo, SMITH (1981, págs. 175-179) habla de la “entrevista profunda” (equivalente a la “entrevista centrada”) con las víctimas de violaciones en el marco del enfoque de encuesta. Es evidente que este tipo requiere que el entrevistador siga los temas que señale el entrevistado. En general, cuando las

⁶² Somos conscientes de que, al tratar de abarcar todas las clases de entrevistas en la exposición que sigue, a veces, no reflejamos adecuadamente algunas de ellas. Hemos sacrificado hasta cierto punto la integridad de los enfoques en beneficio de la economía, tratando de añadir algunas puntualizaciones, cuando parecía necesario, procurando no pecar de pedantes. POWNEY y WATTS (1987) hacen una exposición detallada de la entrevista dirigida específicamente a los investigadores educativos.

⁶³ PATTON (1980) lo denomina “enfoque de entrevista de orientación”, aunque no se contra necesariamente en un hecho específico.

⁶⁴ PATTON (1980) presenta una versión algo diferente: la entrevista normalizada abierta. Las preguntas son normalizadas, pero las respuestas son abiertas.

⁶⁵ Existen ciertas dificultades para ordenar las denominaciones. Algunos etnógrafos se muestran partidarios de considerar este tipo de enfoque como integrante de la observación participante y no como entrevista (p. ej.: BECKER y GEAR, 1957; HAMMERSLEY, en Open University, 1979, Bloque 4). No obstante, siguen describiendo estas entrevistas como informales, no estructuradas y no directivas” (Open University, 1979, Bloque 4, pág. 149). SPRADLEY (1979) distingue las entrevistas etnográficas de los informadores (es decir, quienes ayudan al etnógrafo a comprenderla cultura) de la observación participante. En las primeras, informador y etnógrafo saben adónde van y el etnógrafo controla en diversos grados la situación. Al diferenciar la investigación mediante estudio de casos de la etnografía, STENHOUSE (1978) sólo habla del elemento de observación de la etnografía, dejando de lado la entrevista con los informadores.

hipótesis están claras, la entrevista puede ser estructurada (para comprobarlas); cuando hay que elaborar las hipótesis, deben ser no estructuradas⁶⁶. “No estructurada” no significa “improvisada” y sería inexcusable que un entrevistador asistiera a una entrevista desprovisto de cuestiones y temas de interés. Aun si la entrevista está explorando las características exclusivas de un individuo, el entrevistador debe tener criterios para juzgar qué tendrá que tomar y qué deberá dejar de lado.

Hay otro tipo de entrevista que no puede clasificarse con precisión en el continuo estructurada-no estructurada: las entrevistas en grupo. Pueden desarrollarse con diversos grados de estructuración, pero un elemento importante es la interacción entre los entrevistados. Con frecuencia se presentan como un medio apto para la evaluación en las escuelas, por lo que nos ocuparemos de ellas por separado.

Las preguntas en la entrevista

La naturaleza de la entrevista depende fundamentalmente del tipo y sucesión de las preguntas que se planteen. Para las entrevistas estructuradas, se elabora un inventario de preguntas y respuestas predeterminadas; para las entrevistas no estructuradas, se preparan directrices que contienen una colección amplia de temas que garanticen una cobertura suficiente y cierto grado de estructuración⁶⁷.

Con independencia de que se trabaje a partir de un inventario o siguiendo el flujo de ideas del entrevistado, el entrevistador ha de prestar una atención minuciosa a las cuestiones planteadas. Los detalles de todas las indicaciones que pueden darse sobre la entrevista salen del ámbito en el que se desarrolla este libro, pero resumiremos las cuestiones que aparecen en la bibliografía⁶⁸.

La característica más importante se refiere a si las preguntas deben ser abiertas o cerradas. Si se trata de clasificar una dimensión que se comprende con claridad, se utiliza una pregunta cerrada. Si se desconoce el conocimiento que el entrevistado tiene del tema o su concepción de la cuestión, se precisan preguntas abiertas (SMITH, 1981, págs. 156-157). Por supuesto, en una entrevista pueden utilizarse ambos tipos de preguntas, pero PATTON (1980) advierte que se tenga cuidado en no caer en la trampa de las cuestiones dicotómicas (p. ej., las que requieren respuestas de “sí” o “no”) en las entrevistas no estructuradas.

Una segunda característica de las preguntas de entrevista tiene que ver con su forma de expresión, tema del que se dispone de gran cantidad de orientaciones⁶⁹. PATTON (1980) señala los siguientes aspectos:

1. Singularidad, es decir, las preguntas sólo deben contener una idea⁷⁰.
2. Claridad; p.ej., el vocabulario debe situarse en el nivel correcto⁷¹.
3. Neutralidad; el entrevistador no debe expresar ninguna opinión sobre las respuestas a través de su forma de reaccionar. Tampoco deben formularse preguntas que sugieran la respuesta deseada, salvo cuando se utilicen para indicar a los entrevistados que puede discutirse la conducta aparentemente inaceptable.⁷²

Las preguntas de seguimiento y de sondeo requieren especial cuidado, dado que de ellas pueden obtenerse más detalles, conseguir que se elabore un tema o se aclare una respuesta. SMITH (1981, pág. 171) presenta una lista de sondeos no directivos para estimular nuevas respuestas sin conducir al entrevistado a un planteamiento concreto (por ejemplo: “¿qué quiere decir?”).

Esto nos lleva a la tercera característica: la sucesión de preguntas. Las de sondeo *pueden ir* a continuación de *una pregunta cerrada, en una sucesión simple*. En una entrevista estructurada o semiestructurada, hay que establecer una sucesión cuidadosa. Para las cuestiones complejas, una batería de preguntas puede abarcar los distintos aspectos⁷³. En una entrevista no estructurada, el entrevistador no debe “saltar” de un tema a otro

⁶⁶ Es posible comprobar hipótesis en una entrevista no estructurada, enfoque adoptado por PIAGET (1929) en su “método clínico”.

⁶⁷ SMITH (1981) y HOOK (1981) ponen ejemplos de guías de entrevistas.

⁶⁸ Sobre este tema existen diversos textos, pero limitamos nuestras referencias a los típicos de la bibliografía de las ciencias sociales. En BYNNER (Open University, 1979, Bloque 4) se hace una exposición introductoria de la entrevista de encuesta, y en SPRADLEY (1979), de la entrevista etnográfica. OPPENHEIM (1966), como SMITH (1981) —la referencia que utilizamos—, combina también la exposición sobre entrevistas y cuestionarios, dando algunas orientaciones sobre la bibliografía.

⁶⁹ PATTON (1980), partiendo de la tradición cualitativa, y SMITH (1981), a partir de la cuantitativa, ponen ejemplos detallados de lo que debe y no debe hacerse respecto a la redacción y las palabras que se utilicen.

⁷⁰ Es discutible fuera de las entrevistas estructuradas o semiestructuradas. Con frecuencia, el análisis de estos dos tipos de entrevistas carece de relación con el flujo de ideas que pueden hacer muy aceptable la falta de singularidad.

⁷¹ SMITH (1981, pág. 163) dice también que se mejora la claridad si, al preguntar sobre la conducta pasada, se menciona un momento específico. Este procedimiento es mejor que pedir generalizaciones. Así: “¿cuántas veces has faltado a la escuela en el mes pasado?” es preferible a: “¿cuántas veces al mes faltas a la escuela?”

⁷² Por ejemplo, si la pregunta de la Nota 71 fuese la primera sobre el tema, sería una pregunta de reclamo. Sin embargo, indica al alumno que el entrevistador no va a sorprenderse al descubrir que ha hecho novillos.

⁷³ “Una forma especial es el diseño en cinco dimensiones utilizado en cuestionarios y entrevistas semiestructuradas para suscitar, por ejemplo, respuestas sobre: 1) conciencia de los entrevistados respecto a una actitud: después, 2) sus actitudes generales, seguido de 3) su posesión de la actitud específica: 4) las razones para mantener esa actitud y, por último, 5) la intensidad de su sentimiento (SMITH,

para tratar de abarcarlos todos. ADELMAN y WALKER (1975) sugieren pasar de preguntas poco estructuradas a otras estructuradas, pero otros (p. ej., SMITH, 1981) proponen sólo un marco general que configure la sucesión de preguntas, comenzando, por ejemplo, con un “calentamiento” mediante preguntas descriptivas no problemáticas y terminando con preguntas del mismo tipo que faciliten el “enfriamiento” del entrevistado.⁷⁴ Es obvio que este tipo de consejos no siempre es apropiado para las entrevistas no estructuradas, en las que no pueden fijarse reglas generales.

Realización de la entrevista

No cabe duda de que la sucesión de preguntas constituye un aspecto de la práctica de la entrevista; en las entrevistas estructuradas, su curso está determinado de antemano. Con independencia del tipo de entrevista que se realice, es importante que el entrevistador establezca una relación con el entrevistado (Open University 1981a, Bloque 3, pág. 50). En la bibliografía aparecen una serie de términos para describir la actitud general que hay que adoptar: “simpatía”, “neutralidad”, “agrado”, “talante amistoso”. Debe transmitirse a los entrevistados la sensación de que se valoran sus respuestas, por ejemplo, permitiéndoles que hablen de lo que les interesa. Asimismo, el entrevistador debe merecerles confianza y crédito. La relación que se establezca tiene especial importancia para la entrevista no estructurada, aunque ambos tipos requieren lo que LAFEUR (DAVIS, 1981, página 138) llama “conducta atenta”, o sea, que la conducta verbal y no verbal (p. ej., contacto visual y expresión corporal) del entrevistador indique interés.⁷⁵

Como es natural, esta relación se establece al comienzo de la entrevista. Las manifestaciones iniciales deben explicar también el propósito de la entrevista: cómo y por qué ha sido seleccionado el entrevistado y el carácter confidencial de la información obtenida (HOOK, 1981, pág. 147).⁷⁶ Estas consideraciones son importantísimas para los profesores que entrevistan a sus propios alumnos, puesto que establecen un tipo de relación diferente, es decir, distinta de la de profesor y alumno.⁷⁷ Por supuesto, esta forma de actuar es contraria al modo normal de trabajar del observador participante.⁷⁸ Debe decirse a los alumnos, como a los entrevistados en cualquier estudio, quién dirige la operación, sea el conjunto del profesorado o un docente aislado. El establecimiento y el mantenimiento de la relación exigen destrezas interpersonales. Con respecto a esto, SMITH (1981, página 176) dice que quienes carecen de práctica de entrevistas deben reflexionar sobre tres aspectos: inclusión (hasta qué punto se sienten participantes en la entrevista o excluidos de ella); control (su percepción del grado de control que ejercen sobre la entrevista), y afecto (qué sienten hacia el entrevistado). Si los entrevistadores no creen tener derecho a hacer las preguntas que interesan (por ser de tipo personal, por ejemplo), se encontrarán incómodos, sensación que transmitirán a los entrevistados. Los entrevistadores pierden el control si los entrevistados toman las riendas y, por ejemplo, empiezan a hacer preguntas. El uso de un inventario o guión ayuda a mantener el control, pero sin esto, es decir, en una entrevista no estructurada, los entrevistadores tienen que ejercitar una destreza considerable.⁷⁹ El afecto permite que los entrevistados expresen sus sentimientos, normalmente reprimidos en las relaciones formales: un problema especial que afecta a los profesores que entrevistan a sus alumnos. Es curioso que los docentes pueden llegar a una entrevista con sus alumnos sin plantearse su derecho a hacer preguntas. Si asumen su rol controlador habitual, es difícil que establezcan el tipo de relación necesario para la entrevista.⁸⁰

1981, pág. 161). HOOK (1981, págs. 144-145) habla de “cuestiones embudo”, como fórmula para restringir progresivamente el centro de atención de la encuesta. Con respecto a la entrevista etnográfica, SPRADLEY (1979), sin recomendar una estructura tan rígida, sugiere una evolución en las preguntas.

⁷⁴ PATTON (1980) advierte contra la agrupación de todas las cuestiones concretas (p. ej., edad, ocupación, etc.) en una misma parte y aconseja distribuir las a lo largo de toda la entrevista.

⁷⁵ Es interesante señalar que hay pocas pruebas empíricas que indiquen las características que contribuyen a entrevistar bien o mal (SMITH, 1981, pág. 167).

⁷⁶ SMITH (1981) presenta una lista semejante en el contexto de las entrevistas estructuradas que forman parte de una encuesta.

⁷⁷ En este contexto, es interesante el consejo de STENHOUSE (1982, pág. 265) de que el entrevistador se siente al lado del entrevistado y no frente a él. Esta disposición puede hacer que la entrevista no se parezca tanto a un interrogatorio.

⁷⁸ RIECKEN (1956) expone el dilema con que se enfrentan los observadores participantes cuando estudian un grupo apocalíptico. Manifestar lo que son puede significar la imposibilidad de introducirse en el grupo, pero hacerse pasar por miembros de éste provocaría presiones que reforzarían las convicciones del mismo. Este refuerzo constituye una forma de sesgo del observador. Esta misma es la situación en la que no puede entrevistarse al informador, en términos de SPRADLEY (1979). Como diremos en el capítulo IX, las evaluaciones pueden encubrir investigaciones de difícil justificación desde el punto de vista ético.

⁷⁹ SIMONS (1977) recomienda desviar estas preguntas. SMITH (1981, págs. 177-178) presenta ejemplos de pérdida de control. No nos referimos aquí, por supuesto, a las conversaciones que se producen como elemento *natural de la observación participante*.

⁸⁰ JAMES y EBBUTT (1981, pág. 88) indican que es frecuente que los alumnos den por supuesto que el profesor tiene derecho a hacer preguntas; en consecuencia, les preocupa dar las respuestas “correctas”.

Además de los dispositivos verbales para sondear antes mencionados, el entrevistador tiene que utilizar la conducta no verbal para estimular las contestaciones (SIMONS, 1977), por ejemplo, permaneciendo en silencio tras una respuesta con la pretensión de estimular la elaboración de la misma. Por último, la entrevista debe acabar dando a los entrevistados la oportunidad de hacer preguntas⁸¹.

El registro de las entrevistas

Cuando se utiliza un inventario, éste mismo facilita el registro de la entrevista, junto con las anotaciones de las respuestas a las preguntas abiertas. En las entrevistas no estructuradas, la grabación magnetofónica es muy conveniente, pues proporciona un registro exacto, sin deformaciones, y permite hacer citas precisas en cualquier análisis (SIMONS, 1977, págs. 129-130). Además, evita las distracciones debidas a la necesidad de tomar notas durante la entrevista (HOOK, 1981, pág. 150), pudiendo dedicar más atención al entrevistado (PATTON, 1980). Sin embargo, la transcripción constituye un problema importante, aunque no tanto como en el caso de los registros de observación en clase. PATTON (1980) afirma que la proporción entre el tiempo de transcripción y el de grabación es de 4:1, pero es probable que esa razón se consiga si se cuenta con el apoyo de un secretario; WRAGG (sin fecha) menciona una proporción de 7:1. Es posible que el magnetófono inhiba a los entrevistados por su carácter intrusivo, sobre todo si les preocupa la confidencialidad. El equipo del *Ford Teaching Project* sugería la utilización de un magnetófono de casete con micrófono incorporado, situando el aparato o, en su caso, el micrófono fuera del campo de visión del entrevistado. A los niños pequeños puede permitírseles jugar con el magnetófono con el fin de que se habitúen a su presencia.

No obstante, el hecho de utilizar grabaciones magnetofónicas no significa abandonar las anotaciones, que pueden necesitarse para una entrevista posterior antes de que haya sido posible contar con la transcripción (SIMONS, 1977, página 130), o como registro de seguridad en caso de una mala grabación (HOOK, 1981, pág. 150). Asimismo, las anotaciones ayudan a desarrollar la entrevista, al permitir al entrevistador un seguimiento inmediato de lo que se va tratando en ella, actuando como recordatorio, por ejemplo, del hilo conductor de la misma cuando se agota un tema y evitando la interrupción del flujo verbal del entrevistado. Permiten también al entrevistador desviar la vista del entrevistado, facilitando a ambos cierto “espacio” para pensar (cf. la entrevista en posición lado a lado de STENHOUSE, 1978). Por último, las anotaciones resultan muy útiles para analizar la entrevista, sobre todo cuando se utiliza la escritura ordinaria (sin abreviaturas ni taquigrafía) en la transcripción.

Análisis de la entrevista

Los textos que se ocupan de la entrevista como técnica de evaluación tienen poco que decir al respecto. Algunos consideran el análisis como una cuestión de investigación general, utilizando el tipo de enfoque que expusimos en relación con la observación directa, basándonos en PATTON (1980).⁸² Como técnica analítica se aboga por el análisis de contenido (WRAGG, sin fecha) y de ella nos ocuparemos más adelante. Por el momento, nos referiremos a las precauciones que suelen recomendarse (SIMONS, 1977, págs. 130-131).

1. Los criterios de selección deben tener en cuenta las condiciones en las que se realiza la entrevista y el conocimiento que el entrevistado tenga de los propósitos de la investigación.
2. Cualquier inferencia debe ir acompañada de los detalles del contexto de la entrevista, con el fin de que otros puedan juzgarla.
3. Hay que tener mucho cuidado con el origen de las categorías que se establecen: si proceden del entrevistador o del entrevistado.

Con independencia del enfoque utilizado, es necesario algún tipo de comprobación, preferiblemente a cargo de un tercero⁸³. WRAGG (sin fecha, pág. 17) propone un modelo “en bocadillo”: una primera lectura rápida de

⁸¹ Aspecto que no suele mencionarse, pero que señala WRAGG(sin fecha).

⁸² BECKER (1958) hace un análisis de la investigación mediante la observación participante en tres fases. Otros dicen poco más aparte de que depende del objetivo (p. ej.: HOOK, 1981).

⁸³ Estando en el hospital, en calidad de paciente, ROTH (1974) envió resúmenes de sus notas de campo a diversos compañeros para que le ayudasen a desarrollar un estudio de la “vida” como paciente y a interpretar las observaciones. La posición del profesor es análoga, salvo en que ROTH tenía que pasar sus notas en secreto con el fin de que nadie del hospital las viese!

todas las transcripciones para decidir cómo desarrollar el análisis, seguida del análisis principal, haciendo después otra relectura para comprobar si se ha pasado algo por alto. Se trata del “método comparativo constante” de GLASER y STRAUSS (1967).

Como en el caso de la observación estructurada, la entrevista estructurada requiere que el análisis se planifique antes de la obtención de los datos, de manera que puedan comprobarse hipótesis específicas. Para efectuar esos análisis hace falta aplicar técnicas estadísticas de tipo descriptivo sencillo (p. ej., tabulaciones de respuestas) o de tipo correlacional. A este respecto, el análisis de las entrevistas estructuradas se parece al de los cuestionarios.

La entrevista en las escuelas

En el campo de la evaluación, se han publicado pocas experiencias sobre el tema; el Ford Teaching Project (1975) constituye la fuente principal al respecto⁸⁴. En realidad, la mayor parte de los autores exponen el tema desde el punto de vista del investigador externo (p. ej.: SIMONS, 1977). Como señalamos antes, al referirnos a la realización de entrevistas con los alumnos a cargo del profesor, la cuestión fundamental consiste en la adopción de un nuevo rol. Como en la observación directa, los profesores deben quedarse en segundo plano, aunque atentos, de manera que los alumnos puedan hablar con libertad. En otras palabras, deben asumir el papel de aprendices, no de profesores. Es más, es preciso que se diga a los alumnos que su profesor adopta este nuevo rol, aunque les pueda parecer algo confuso o difícil de creer⁸⁵.

Todo lo dicho hasta aquí se aplica también a las escuelas; sin embargo, a continuación presentamos algunos aspectos adicionales mencionados por el Ford Teaching Project (y otros)⁸⁶:

1. Los alumnos necesitan que se les ayude a superar sus reticencias⁸⁷
2. Hay que mantener la informalidad y evitar remitirse continuamente a papeles o notas.
3. Las preguntas deben plantearse en el nivel adecuado, pero también con discreción, es decir, evitando las preguntas directas.
4. Hay que estar atentos al aburrimiento de los alumnos, aunque sin malinterpretar las señales no verbales (p. ej., encogerse de hombros); pueden indicar que los alumnos tienen dificultad para expresarse.

El muestreo de los alumnos constituye un problema particular en las escuelas y no siempre en relación con quienes son escogidos, sino con quien efectúa el muestreo. La selección de alumnos que entrevistar realizada por sus profesores constituye una posible fuente de inclinaciones tendenciosas.

Un enfoque de la entrevista que ya mencionamos antes es el de la que se realiza con grupos. En las escuelas, es importante porque los alumnos experimentan la clase tanto en calidad de individuos como de grupos. ADELMAN y WALKER (1975) recomiendan entrevistar a grupos de dos o tres alumnos durante veinte minutos o menos. El equipo del *Ford Teaching Project* utilizaba lo que denominaba “entrevista múltiple”, que supone la participación de un agente externo para grabar y observar la clase, entrevistando al profesor antes y después de la hora lectiva, a los alumnos y, por último, al profesor y los alumnos juntos. Para los docentes de mayor categoría también son útiles los diálogos en grupo. Por ejemplo, si están realizando entrevistas con compañeros de categoría inferior, pueden encargarse de realizar una entrevista de grupo, lo que ayudará a reducir la importancia de su propio rol. De igual manera, si han de ser entrevistados, su inclusión en un grupo puede obviar el problema de las preguntas directas (que recibirían respuestas “oficiales”)⁸⁸. Por supuesto, las entrevistas de grupo presentan dificultades, no sólo de transcripción, sino también problemas de validez, así como la posibilidad de que algunos individuos del grupo no compartan los valores de otros⁸⁹.

POLLARO (1982b) cuenta los problemas que tuvo al realizar entrevistas, como profesor investigador, como,

⁸⁴ Procedente de otro campo, está la obra clásica de PIAGET (1929) utilizando el método clínico. El capítulo introductorio (“Problemas y métodos”) es útil como orientación para entrevistar a niños.

⁸⁵ SIMONS (1977, pág. 113) se refiere al problema que se le plantea a un observador externo de mantener el equilibrio entre el bando de los profesores y el de los alumnos, ¡Es dudoso que los docentes que entrevistan a sus propios alumnos sean capaces de realizar esa alternancia!

⁸⁶ SIMONS (1977, págs. 111-116) ofrece orientaciones muy importantes para los investigadores externos que, en su mayoría, son igualmente útiles para los profesores, sobre las cuales nos basamos,

⁸⁷ Esto se relaciona con el papel que adopta el profesor y con la atmósfera normal del aula.

como en el caso de la observación directa, el cambio de rol es más fácil cuando la atmósfera normal del aula es informal (SIMONS, 1977, pág. 114).

⁸⁸ DALTON (1964) realizó investigaciones en las fábricas en que trabajó y sólo mantuvo entrevistas formales con los directivos para conseguir información sobre la “línea oficial”, o para ¡recoger lapsi linguae! (Agradecemos a Martín HAMMERSLEY que nos hiciera llegar ésta y otras comunicaciones relativas a ambientes no educativos.)

⁸⁹ Estas notas sobre las entrevistas de grupo con el profesorado son de SIMONS (1980a)

por ejemplo, tener que hacerlas generalmente a la hora de comer. Una estrategia interesante utilizada por él consistía en hacer participar a los niños en la investigación, constituyendo un equipo entrevistador. Les dio unas orientaciones generales sobre lo que quería saber, pidiéndoles que grabaran sus entrevistas con otros alumnos. Más avanzado el estudio, se presentaba cuando se estaban realizando algunas entrevistas, utilizando de este modo al alumno entrevistador para romper el hielo. Esto le permitía abordar de manera más específica lo que le interesaba para su investigación. El equipo del Ford Teaching Project utilizó una técnica similar, de manera que los mismos niños presidieran grupos de discusión, cuyos diálogos se grababan sin que ningún adulto estuviera presente (ELLIOTT y WALKER, 1973). En un estudio sobre las perspectivas de los niños respecto a la interacción en clase, B. DAVIES (1980) describe una entrevista conversacional (o cara a cara) en la que entrevistador y entrevistado descubrían sus mutuos marcos conceptuales. El propio marco de DAVIES consistía en parte en comprender los marcos de los niños. Estos ayudaban al investigador en esta tarea, rebobinando, por ejemplo, la conversación grabada y llamando la atención sobre lo que consideraban importante. RICH (1968) facilita una guía general para entrevistar a niños que, aunque se sitúa en la tradición de la entrevista en el trabajo social y el consejo (p. ej., para evaluación), se dirige, no obstante, a los inexpertos. En contraste con la mayoría de los textos, trata específicamente del trabajo con niños y jóvenes.

Validez y fiabilidad de las entrevistas

Tanto los entrevistadores como los entrevistados constituyen fuentes de inclinaciones tendenciosas: los primeros, mediante las preguntas que hacen o el papel y la presencia que se les asigna desde fuera; los segundos, a través de su concepción de la entrevista, su recuerdo del hecho, su habilidad para responder y su motivación para tomar parte en la entrevista.⁹⁰ Es obvio que las advertencias indicadas antes sobre las preguntas y la orientación que se imprima a la entrevista ayudarán a mejorar su validez y fiabilidad. Como en el caso de las observaciones directas, la estructura favorece la fiabilidad y puede reducir el sesgo del entrevistador. Las entrevistas no estructuradas exigen la reflexividad. Por ejemplo, tal como indicamos, es preciso incluir detalles del contexto (incluyendo el estado de ánimo del entrevistador) al lado de cualesquiera inferencias que se hagan. Si se duda de la fiabilidad de un relato como en el caso de los recuerdos confusos respecto de un acontecimiento, dicho relato ha de considerarse como una percepción subjetiva del hecho y, en consecuencia, como un tipo de dato diferente. Con respecto a las entrevistas de grupo, en las que los sujetos pueden dar respuestas distintas de las que darían en una entrevista individual, a causa de la presión o apoyo del grupo de compañeros⁹¹, puede adoptarse un enfoque similar. Por tanto, no se considera más fiable un tipo de entrevista que otro, sino que ambos proporcionan datos de carácter diferente.

Cuestionarios

Aunque las destrezas precisas para elaborar cuestionarios sean distintas de las de tipo interpersonal necesarias para las entrevistas, no son menos exigentes.

Muchos cuestionarios, incluidos los elaborados por investigadores universitarios, están mal contruidos, siendo de validez y fiabilidad dudosas. Cuando decimos esto, no nos referimos a los “cuestionarios” diseñados para medir rasgos psicológicos o algún constructo sociológico como, por ejemplo, las medidas de la “curiosidad”, “dogmatismo del profesor” (COHEN, 1976, págs. 30 y 158, respectivamente) o del “ambiente de clase” (FRASER, 1981)⁹². Estos instrumentos suelen estar sujetos a procedimientos de validación. Sin embargo, su atención a problemas específicos de investigación, frecuentemente los hace inadecuados para la evaluación⁹³. El principal esfuerzo ha de orientarse a la construcción del cuestionario y de ahí los consejos a los que antes nos referimos sobre los tipos de preguntas, su redacción y sucesión en los formularios de entrevistas.⁹⁴ Como en el caso de las entrevistas estructuradas, pueden utilizarse diversas respuestas especificadas de antemano:

⁹⁰ SMITH (1981, págs. 182-184), al explicar con detalle los inconvenientes de los métodos de encuesta, se refiere a estos tipos de sesgos. BOYD y WESTFALL (1970) hablan del sesgo del entrevistador en el enfoque de encuesta, y DEAN y WHYTE (1958) se refieren al sesgo del informador (entrevistado) desde el punto de vista etnográfico (de igual significación para la entrevista de encuesta).

⁹¹ BECKER (1958) sugiere este enfoque para ocuparse de informes individuales sospechosos y entrevistas de grupo. Presenta también una lista de preguntas para comprobar la validez de las respuestas de los entrevistados.

⁹² Estas técnicas se integran en la medida de actitudes que revisaremos en el apartado sobre “Resultados del currículum”. Por ejemplo, OPPENHEIM (1966) los estudia en el mismo texto.

⁹³ FRASER (1981) se manifiesta a favor del uso del *individualised Classroom Environment Questionnaire* a efectos de evaluación, en especial para comparar las discrepancias entre las percepciones de los alumnos y del profesor respecto a ciertos aspectos del ambiente del aula. No obstante, sólo pudo encontrar un estudio que utilizase esas medidas ambientales y en relación con la enseñanza Superior.

⁹⁴ MOSER y KALTON (1971, Capítulo 13) estudian los principios generales de redacción de preguntas en encuestas de estudios sociales y revisan la bibliografía pertinente. BYNNER (Open University 1979, Bloque 4) hace una exposición introductoria y presenta algunas ilustraciones útiles de cuestionarios de encuesta. OPPENHEIM (1966) es un texto típico que orienta sobre el diseño, redacción y tipos de respuestas.

dicotómicas; de opción múltiple; escalas (p. ej., escalas de tipo Likert: “muy de acuerdo”, “de acuerdo”, “neutral”, “en desacuerdo”, “en total desacuerdo”), y clasificaciones⁹⁵. Estas respuestas pueden combinarse con otras, cortas o libres (sin limitación de espacio) a preguntas abiertas. No obstante, para facilitar el análisis, debe haber más preguntas cerradas que abiertas. El contraste con las entrevistas se produce en relación con el diseño. Éste es el (pobre) sustituto de la simpatía en la entrevista, como antes dijimos. YOUNGMAN (sin fecha) da ciertas indicaciones útiles a este respecto, incluyendo observaciones sobre los encabezados, para señalar los temas de las preguntas, y la atención que hay que prestar al aspecto, por ejemplo, el espaciado, la tipografía, la separación entre las instrucciones y las preguntas, la colocación coherente de las respuestas. Para quienes pueden disponer de instrumentos reprográficos, la fotorreducción constituye un buen método para mejorar el aspecto. Es obvio que la longitud es un elemento importante del diseño, debiendo mantenerse el equilibrio entre el deseo de obtener “información interesante” y el efecto que se cause sobre la motivación de quienes han de responder. YOUNGMAN (sin fecha, pág. 22) sugiere dejar las páginas sin numerar, para reducir el impacto sobre la motivación! Todo ello se aplica a los cuestionarios formales, pero no cabe duda de que también hay sitio para ejercicios del tipo: “Escribe media página de comentarios de interés sobre las clases de inglés que he impartido este año” (Tony EVANS, TES, 23 de abril de 1982), que pueden proporcionar perspectivas reveladoras.

A diferencia de las entrevistas, los cuestionarios pueden ser examinados por terceros, así como también pueden probarse con potenciales destinatarios de los mismos para estudiar las preguntas y el diseño general. Esto ofrece una forma de validación, al tiempo que imprime cierta orientación al análisis. En realidad, el problema del análisis debe tenerse muy presente al construir un cuestionario; una característica de esta técnica es su utilidad cuando se presenta un problema claro de evaluación que puede traducirse a un conjunto de preguntas. Las categorías de las respuestas deben ser exclusivas y exhaustivas, dándoseles códigos numéricos que faciliten el análisis de rutina. Pueden tabularse las respuestas a cada pregunta y, si conviene, establecerse tabulaciones cruzadas para buscar relaciones⁹⁶. Las repuestas a las preguntas abiertas pueden someterse a análisis de contenido, con la posibilidad de contabilizar frecuencias.

Diarios de alumnos

Al referirnos a la observación directa, hablamos de los diarios de los profesores como una forma de anotaciones, dando por supuesto que las hacía quien realizaba la evaluación. Sin embargo, pedir a todos los profesores, o a una muestra de alumnos, que elaboren un diario constituye una forma de cuestionario diseñada para reconstruir los hechos que el evaluador no puede observar, como las actividades que el profesor o el alumno ejecutan fuera de la escuela (ya hemos expuesto un ejemplo de diario del profesor). SHIPMAN (1979, pág. 107) considera que esta técnica es una forma de evaluar las actividades extracurriculares, sobre todo la participación en las mismas. Sugiere que los alumnos hagan un diario en el que anoten en qué clubes y asociaciones participan y cuándo. Esta actividad pueden realizarla durante un período de dos o más semanas, en dos trimestres del año. Es obvio que se trata de un diario muy selectivo. En cambio, COHEN (1976, págs. 85-90) muestra cómo los investigadores han recogido información a través de diarios no específicos, pidiendo a los alumnos que anotasen en ellos todo lo que hicieran en el tiempo extraescolar. En un caso, se utilizó este procedimiento para estudiar el tiempo dedicado a ver la televisión (recabando información espontánea); en otro, se empleó para conocer las actividades de tipo general desarrolladas por los alumnos. En este último caso (Universidad de Birmingham: Survey of Children's Habits), se pidió a los alumnos que anotasen con quién realizaban las actividades y dónde, se codificaron éstas mediante 92 categorías.

Los diarios de las actividades escolares de los alumnos pueden revestir la forma de cuestionarios, con categorías especificadas de antemano, como “atender el dictado del profesor, copiar de la pizarra, escuchar a otros alumnos, etc.” (ROBERSTON, 1980, pág. 97). En realidad, esto se convierte en un inventario de observación y, una vez fijadas las categorías, puede encargarse a una muestra de alumnos que lo cumplimenten durante las clases del día, por ejemplo. Constituye, pues, una alternativa, parcial al menos, respecto al estudio de seguimiento permanente que expusimos antes (*Sir Frank Markham School*, 1981, 2ª fase).

Medidas no obstrusivas

Todas las técnicas de observación directa e indirecta expuestas hasta ahora provocan diversos grados de reactividad que exigen tomar ciertas iniciativas para reducirla u optar por una postura reflexiva respecto a sus

⁹⁵ YOUNGMAN (sin fecha) y HOOK (1981) dan detalles y ejemplos de estos formatos de respuesta, así como orientaciones sobre la redacción de las preguntas.

⁹⁶ El Ford Teaching Project (1975, Unidad 2, Team Based Action Research) presenta un ejemplo de elaboración, estudio piloto, revisión y análisis de un cuestionario. En el análisis se incluye la tabulación cruzada sucesiva de una pregunta en relación con otras varias.

efectos. Las medidas no obstrusivas constituyen un intento de efectuar observaciones no reactivas, por lo que se basan en datos recogidos sin que los participantes lo sepan. El texto clásico sobre estas medidas (WEBB y cols., 1966) tiene en cuenta tres tipos básicos:

señales físicas, archivos y observación simple. Es más adecuado ocuparse de esta última en el apartado dedicado a la observación directa, aunque los tipos de situaciones a los que se refieren WEBB y cols. (p. ej.: un investigador que pasee arriba y abajo por Regent Street anotando las conversaciones de la gente) requieren un grado de anonimato imposible en las escuelas. A un profesor se le *reconocería* como participante (de ahí el relieve de las ideas sobre la observación participante) y un extraño que se dedicara a observar *aparacería* como tal extraño. La observación sería, por tanto, obstrusiva y reactiva. A continuación, presentaremos ejemplos de los otros dos tipos de medidas⁹⁷ extraídos de BOUCHARD (1976), HOOK (1981) y SHIPMAN (1979)⁹⁸

Señales físicas

Entran en esta categoría: el deterioro o suciedad en los libros de la biblioteca que indique la cantidad y la clase de uso de los mismos; pintadas que muestren prejuicios (o preocupaciones sexuales, i en los servicios!); ventanas rotas o desperfectos en el mobiliario, que indiquen el grado de orgullo de pertenecer a la escuela o la moral de los alumnos⁹⁹ los avisos de los tableros de anuncios de la sala de profesores, que indiquen el nivel de comunicación escrita; los carteles, cuadros, etc., distribuidos por la escuela (poniendo de manifiesto valores, p. ej.: cuadros de honor, obras artísticas) y por las clases (mostrando las actitudes de los profesores).

Archivos

En este apartado se incluyen las actas de reuniones, informes de comisiones, informes al profesorado, cartas a los padres, el prospecto de la escuela, los informes escolares anuales, ficheros de la escuela y de la LEA, expedientes (como los registros personales de los alumnos), cuestionarios de caridad, cartas elogiosas y de quejas, controles de asistencia, registros de elección de centro de los padres.

A partir de la relación anterior, es fácil comprobar que algunos de estos ítems permiten obtener medidas cuantitativas, p. ej.: su frecuencia, y otros requieren análisis de contenidos, como en el caso de los documentos. Como indica SHIPMAN (1979, págs. 120-121), se plantean dos problemas: ¿qué indican las medidas? y ¿qué criterios se aplican para juzgar el nivel? Al referirnos a las “señales físicas”, señalábamos ya lo que podían indicar las medidas; pero no dejan de ser conjeturas, por lo que es preciso utilizarlas junto con otros métodos¹⁰⁰. Cualquier medida del número de sillas rotas por trimestre, por ejemplo, no tiene un significado absoluto (¡salvo en términos financieros!) y el criterio para juzgar debe extraerse de la comparación con otras escuelas, con otras aulas o en el tiempo. A diferencia de la investigación en ciencias sociales a la que se refiere BOUCHARD (1976), la mayor parte de las exposiciones sobre medidas no obstrusivas permanece sepultada en estudios, al lado de otras clases de medidas, o se queda en el nivel de la simple sugerencia (p. ej.: MILLER, 1978)- En consecuencia, no hay mucho sobre orientaciones concretas, sobre todo en relación con las cuestiones de validez y fiabilidad¹⁰¹.

El análisis de contenido

Engloba esta expresión los tipos de análisis realizados en diversas circunstancias. Tiene su origen en los estudios sobre medios de comunicación de masas en los que, por ejemplo, se utiliza para determinar las tendencias de los periódicos en el tratamiento de los candidatos a elecciones o para analizar la propaganda en tiempo de guerra. HOLSTI (1968, pág. 601) lo define como una “técnica para hacer inferencias mediante la identificación sistemática y objetiva de características especificadas de los mensajes”. Aunque constituye de por sí una técnica de investigación que utilizar cuando no se dispone de ninguna otra (p. ej., respecto a documentos), se aplica cada vez más a datos obtenidos por otros métodos (ya la hemos mencionado en

⁹⁷ Éstas pueden subdividirse (véanse WEBB y cols., 1966, y BOUGHARD, 1976). No obstante, estas subdivisiones no parecen adecuadas para el medio escolar y los autores no dan muchas orientaciones sobre sus posibles usos.

⁹⁸ MILLES (1978) presenta una extensa lista de medidas, pero muchas de ellas son más bien categorías para la observación directa, por ejemplo: número de situaciones en las que los profesores sonríen (chistes, etc.).

⁹⁹ SHIPMAN (1979) presenta una lista de indicadores de este tipo, entre los que se encuentran: número de ventanas reparadas por trimestre; número de sillas rotas o pupitres estropeados.

¹⁰⁰ BOUCHARD (1976) hace hincapié en este aspecto en el inventario recopilado a partir de diversos estudios de investigación, presentando las medidas y variables indicadas por ellos.

¹⁰¹ WEBB y cols. (1966) tienen muy en cuenta estas cuestiones cuando exponen todos los tipos de medidas no obstrusivas, aunque sus ejemplos no corresponden a situaciones educativas.

relación con el análisis de entrevistas y cuestionarios). Por ejemplo, puede utilizarse sobre datos no estructurados:

1. Documentos de centros escolares: materiales curriculares, actas de reuniones, informes o cartas, contenidos de tableros de anuncios.
2. Transcripciones de entrevistas.
3. Respuestas abiertas a cuestionarios.

La definición de HOLSTI indica que el análisis se realiza utilizando categorías para codificar los datos. Como en la observación estructurada, las categorías que deben tener relieve teórico, han de definirse con claridad. Por ejemplo tiene sentido contar el número de personas que hablan en una reunión (como recogen las actas), salvo que se tome como indicador de “participación”, por ejemplo. El análisis no tiene que ser cuantitativo (aunque es una cuestión debatida en la bibliografía¹⁰²) porque no es esencial contar la frecuencia de las categorías. Puede bastar con la aparición o no de una categoría concreta.

La técnica del análisis de contenido se utiliza también de manera habitual para tratar las respuestas abiertas a los cuestionarios. La tarea consiste en reducir de alguna manera significativa los datos, aceptando que se trata de efectuar un sondeo de opiniones, o de cualquier otra cosa. El primer paso consiste en leer las respuestas y anotar las categorías que parezcan relevantes. Esa importancia se deriva de lo que pretenda esclarecer la pregunta. Por ejemplo, en su nivel más sencillo, quizá sea importante tomar nota del punto de vista a favor o en contra de un tema o lección concreto. Tras extraer una lista de categorías, el siguiente paso consiste en intentar definir cada una de ellas, seguido de la codificación de cada respuesta. Es posible que alguna no se ajuste a la distribución categorial o que haya que subdividir una categoría. En otras palabras, puede que haga falta revisar las categorías, lo que llevará consigo la recodificación de las respuestas. En consecuencia, para codificarlas satisfactoriamente, es preciso ir una y otra vez de los datos a las categorías y viceversa.¹⁰³ Este tipo de técnica es sistemático en la medida en que se procura obtener una categorización fiable y válida de los datos. La validez se deriva del relieve de las categorías y de la precisión de su definición, con las inevitables transacciones entre ambas. Es perfectamente aceptable la consideración de los significados latentes (o sea, ocultos) como, por ejemplo, los sesgos sexistas que aparezcan en una respuesta. En este caso, se hace una inferencia desde la categoría hacia una idea general, o sea, la inclinación tendenciosa. No obstante, la categoría, que debe indicar cómo se reconoce la tendencia o sesgo, ha de estar clara. Conviene que alguien más codifique los datos utilizando las categorías elaboradas.

Cuando estudiemos el análisis del material curricular (“El curriculum oficial”), veremos algunos ejemplos de análisis documental. Otros casos de utilización del análisis del contenido como técnica de elección corresponden al análisis de transcripciones de entrevistas. Por ejemplo, podría utilizarse para descubrir si un alumno ha conseguido cierto nivel de comprensión de la “conservación del volumen”¹⁰⁴ Asimismo, puede utilizarse para analizar los errores que cometen los alumnos en sus trabajos escritos, por ejemplo, de ortografía o, en un nivel más elevado, en cuanto a concepciones equivocadas. En relación con este último ejemplo, no es esencial el recuento del número de faltas, sino sólo la comprobación de su existencia (aparición/no aparición). En consecuencia, el proceso general requiere formular un problema, definir categorías, escoger una unidad para categorizar (unidades de contexto), cómo una oración, pasaje o libro, y algún sistema de enumeración (KRIPPENDORFF, 1981). Este último aspecto se refiere al recuento de frecuencias, duración del tiempo empleado, espacio ocupado, intensidad (o sentimiento) y aparición o no aparición. Así, por ejemplo, la inclinación política tendenciosa puede juzgarse por la cantidad de espacio concedido a determinado punto de vista, por el número de veces que se menciona un partido político, por el tiempo que se le concede en televisión o por el trato que reciben los miembros del partido en cuestión (favorable o desfavorable). En este campo, el texto por excelencia es el de HOLSTI (1969), aunque HOLSTI (1968) constituye un buen resumen. Por desgracia, la mayoría de los ejemplos proceden de estudios de comunicación y pocos del campo de la enseñanza¹⁰⁵ aunque la legibilidad (un elemento de estos estudios) sea considerable (cf. “El curriculum oficial”).

Resultados del curriculum

¹⁰² HOLSTI (1968, pág. 595) revisa brevemente los argumentos.

¹⁰³ Es obvio que, si las categorías pueden definirse con claridad de antemano, debe utilizarse un formato de respuesta preespecificado, salvo que sea difícil darles una redacción o la categorización se refiera a elementos lingüísticos (p. ej.: vocabulario) o a significados latentes (p. ej.: sesgo sexual).

¹⁰⁴ Podría juzgarse en la misma entrevista, pero esto depende de la habilidad del entrevistador con su método clínico” y de la complejidad del pensamiento del alumno. Por ejemplo, PERRY (1970) intentó identificar la etapa evolutiva de los alumnos en sus puntos de vista sobre el conocimiento (p. ej. puntos de vista absolutos o relativistas) a partir de las transcripciones de las entrevistas. Identificó ciertas formulaciones características y entrenó a jueces para que clasificasen el pensamiento de cada estudiante (es decir, una transcripción completa), de acuerdo con una categoría de entre varias (es un ejemplo del uso cualitativo del análisis de contenido).

¹⁰⁵ Pero véase en CAIRNS (1982) un ejemplo del análisis del contenido de materiales curriculares.

Introducción

El Capítulo VII demuestra la importancia concedida a la medida de los resultados en cuanto aspecto de la evaluación y su predominio como modelo de evaluación¹⁰⁶. Dada su importancia, sorprende un poco que los materiales de las pruebas, dirigidos a los profesores y relativos a la valoración, contengan muy pocas orientaciones sobre la evaluación del curriculum. Es frecuente que los autores incluyan la evaluación curricular como uno de los objetivos importantes del proceso de valoración, junto con la “selección”, por ejemplo, pero no dan muchas orientaciones concretas respecto a cómo utilizar la valoración a efectos de evaluación. La mayoría de los textos que hemos consultado reconocen la naturaleza polivalente, en cuanto a sus objetivos, de la información valorativa, y un importante aspecto de ella consiste en el “diagnóstico”¹⁰⁷. En los Estados Unidos, es frecuente que el diagnóstico esté incluido en el concepto de aprendizaje para el dominio de una materia (p. ej.: GRONLUND, 1981; BOOM, HASTINGS y MADAUS, 1971; BLOCK, 1971). Se basa esto en la idea de que el dominio se logra mediante pequeñas unidades de aprendizaje. La valoración (assessment) se utiliza para diagnosticar si el alumno ha logrado dominar una unidad. Si es así, pasa a la siguiente; si no, se le proporciona ayuda. (Volveremos sobre este tema más adelante.) En Gran Bretaña, también preocupa el diagnóstico, pero no se relaciona con ideas de dominio. Por ejemplo, HARLEN, DARWIN y MURPHY (1977) y MUNCEY y WILLIAMS (1981) se ocupan de la adaptación del curriculum a los individuos, a través de la valoración de lo que han aprendido los alumnos. Esto supone que la evaluación (evaluation) forma parte del proceso de aprendizaje, porque el profesor supervisa constantemente a los alumnos, adaptando en consecuencia la enseñanza. Aunque éste es un tipo importante de evaluación, parece más adecuado considerarlo, como hace SHIPMAN (1980), como un aspecto de la organización del aprendizaje. Además, se basa en un enfoque individualizado del aprendizaje, mientras que nuestra preocupación principal se refiere a elementos mayores de la enseñanza y el aprendizaje, y a grupos de alumnos más que a individuos.¹⁰⁸ Por tanto, ¿qué diferencia hay entre utilizar la información valorativa para la evaluación y la valoración de los individuos? Uno de los cometidos habituales de la valoración consiste en clasificar a los individuos, por ejemplo, en un aula. Esa clasificación puede ser útil en ciertas circunstancias, pero, en general, al evaluar queremos saber lo que conoce un individuo o aula, de manera que, si surgen problemas, pueda mejorarse la enseñanza. Como veremos más adelante, esto requiere que se interpreten de distinta forma las puntuaciones o grados obtenidos en la valoración.¹⁰⁹ De ahí la diferencia fundamental entre referir la valoración a una norma o hacerlo en relación con un dominio o criterio. Al clasificar a un sujeto, se compara el conocimiento o comprensión de ese individuo con un grupo (p. ej., el grupo de clase o el de edad), importando menos *lo que* ese individuo sabe en realidad; en esto consiste la referencia a una norma. Cuando se trata de manifestar lo que sabe el alumno, la referencia que se utiliza es el dominio. Preferimos la expresión “referencia al dominio” a la más corriente de “referencia al criterio” porque esta última está asociada específicamente con el aprendizaje para el dominio de una materia y con la tarea de conseguir que los alumnos alcancen un criterio determinado (p. ej., el 80% de sumas correctas en una prueba de matemáticas)¹¹⁰. Hay circunstancias, que consideraremos más adelante, en las que ambos tipos de información resultan útiles para la evaluación¹¹¹. Entonces, ¿adónde pueden dirigirse las escuelas para encontrar orientaciones sobre la utilización de las

¹⁰⁶ DEALE (1975), que ha escrito el mejor texto para profesores de enseñanza secundaria del Reino Unido, pone la evaluación como uno de sus ejemplos introductorios sobre los usos de la valoración del alumnado (pág. 15). Asimismo, incluye la evaluación de los materiales curriculares como objetivo (pág. 20). No obstante, sólo se ocupa directamente de éstos en tres páginas, al final del libro (pág. 160-162), dedicando específicamente sólo cuatro líneas a la valoración de un grupo (el libro se centra en la valoración de un sujeto). Del mismo modo, SATTERLY (1981), en un libro británico más reciente, cita la evaluación como el principal objetivo de la valoración de alumnos (pág. 6), pero no trata la cuestión al hablar de las técnicas. Refiriéndose a la distinta utilización de los términos evaluation y assessment en los Estados Unidos, GRONLUND (1981), por ejemplo, habla del uso de assessment para mejorar el aprendizaje y la instrucción (pág. 8-9), pero sólo se refiere a la evaluación sumativa en las páginas 497-498.

¹⁰⁷ MACINTOSH y HALE (1976) plantear seis propósitos: diagnóstico, evaluación, orientación, clasificación, selección y previsión.

¹⁰⁸ Según las acepciones utilizadas en este libro, “assessment” se refiere más a los individuos que “evaluation” SHIPMAN (1983) no está de acuerdo con esta distinción, y lo discutiremos más adelante, aunque indirectamente, cuando consideremos la “comparación ipsativa”. véase también la exposición sobre los registros de rendimiento en el capítulo IV

¹⁰⁹ LEWY (1977) y EGGLESTON (1974) se encuentran entre los pocos autores que se ocupan directamente de los problemas derivados de utilizar la información de la valoración formativa para la del curriculum, aunque LEWY lo hace en el contexto de un proyecto curricular o de un organismo central encargado del curriculum y no de una escuela.

¹¹⁰ Por ejemplo. SATTERLY (1981) describe pruebas referidas a criterio en relación con el aprendizaje para dominar algo y las objeciones que se esgrimen en su contra, aunque él las considera objeciones contra el aprendizaje para el dominio de una materia. HAMBLETON y cols. (1973) indican que la expresión correcta es “referido a dominio”, aunque no conviene hacer cambios terminológicos porque pueden llevar a confusión.

¹¹¹ Autores como EGGLESTON (1974) asumen que la referencia a norma no es adecuada para la evaluación y suele pensarse que los profesores no quieren el tipo de comparaciones que supone este tipo de referencias, sin embargo, SHIPMAN (1979) dice que los profesores quieren comparar, por ejemplo, a sus alumnos con normas nacionales, confirman esto último los datos derivados del estudio del University of London Institute of Education del uso que hacen de las pruebas las LEA y las escuelas.

medidas de resultados? En primer lugar, estamos de acuerdo con SHIPMAN (1979, pág. 19) respecto a que es probable que muchos aspectos técnicos y cuantitativos de la valoración del alumnado carezcan de utilidad y que la exposición de métodos de valoración es excesiva y supera con creces lo que buscan los profesores a efectos de evaluación. Por tanto, aunque la utilización de las medidas de resultados en la evaluación sólo será tan buena como lo sean los instrumentos de medida utilizados, creemos que, en general, sólo conviene considerar métodos nuevos de valoración para mejorar la de los individuos (a efectos de selección, diagnóstico, etc.). Su uso en la evaluación es una añadidura y los revisaremos desde este punto de vista. Como SHIPMAN (1979), creemos que es conveniente utilizar la información valorativa que se haya obtenido ya en la escuela y revisaremos lo que la bibliografía dice al respecto. No obstante, dada la dificultad de obtener una buena información valorativa, contemplaremos también la que puede obtenerse fuera de la escuela, o sea, a partir de las pruebas normalizadas y de los exámenes oficiales¹¹².

El procedimiento de valoración: diseño y construcción¹¹³

La mayor parte de las orientaciones sobre el diseño y la construcción de procedimientos de valoración se dirigen a los profesores de enseñanza secundaria y, sobre todo en Gran Bretaña, aquéllas incluyen orientaciones para elaborar exámenes (p. ej., DEALE, 1975). Sobre la valoración de los alumnos en la enseñanza primaria, no hay gran cosa, quizá porque en ella se utilizan métodos de enseñanza informales e individualizados y pruebas normalizadas.¹¹⁴ Los autores que insisten en los objetivos múltiples de la valoración también hacen hincapié en ajustarla a esos objetivos, aunque sólo exponen los usos de la valoración en términos generales, es decir, no relacionan los objetivos, los métodos y las formas de analizar los resultados (p. ej., DEALE, 1975). Las orientaciones para el diseño y la elaboración se estructuran en una serie de fases:

1. Elaborar una formulación de objetivos, preferiblemente en términos conductuales (o sea, indicando la conducta que deben mostrar los alumnos).
2. Crear una formulación que describa el tema o el área de contenidos completa.
3. Escoger el método de valoración (p. ej.: prueba escrita, examen oral, actividad práctica).
4. Extraer ítems (preguntas, etc.) de los objetivos, para comprobarlos.
5. Si el procedimiento de valoración del alumnado no consiste en una prueba “objetiva”¹¹⁵, elaborar criterios e instrucciones de calificación.
6. Proponer un procedimiento o prueba de valoración que refleje en conjunto el tema (o sea, que constituya una muestra del mismo), aspecto que dependerá del número de ítem y su tipo, y de la elección del alumno.
7. Probar las preguntas antes de utilizarlas para la valoración del alumnado, cosa que, en la escuela, no suele ser fácil.

Este procedimiento pone de manifiesto la necesidad de definir de manera operacional lo que se pretende medir (p. ej., si queremos hallar el tamaño, hay que definirlo operacionalmente en relación con una propiedad concreta, por ejemplo, la longitud, con el fin de medirlo).

El establecimiento de objetivos que abarquen la totalidad del tema (segunda fase) requiere alguna forma de describir el campo que medir. Suelen reconocerse tres campos: el cognitivo, referido al pensamiento; el afectivo, relacionado con los sentimientos, y el psicomotor, que se refiere a las destrezas manipulativas, etc. BLOOM (1956) preparó un volumen que contenía una taxonomía de objetivos pertenecientes al campo cognitivo, organizados en niveles jerárquicos, desde el recuerdo de los conocimientos”, en el más bajo, hasta la “aplicación” y la “evaluación” en los superiores. Los objetivos se expresan en términos independientes de cualquier contenido concreto, pudiendo utilizarse, en consecuencia, con cualquier tema o materia¹¹⁶. Esta taxonomía fue elaborada por grupos de profesores y contrasta con el enfoque de GAGNÉ (1970), que se basa en una teoría psicológica del aprendizaje. También es jerárquica, con conexiones estímulo-respuesta en su nivel más bajo y de resolución de problemas en el superior. Aunque esta jerarquía de aprendizajes tiene una

¹¹² como SHIPMAN (1979), utilizamos la distinción entre valoración con referencia interna y con referencia externa. Sin embargo, no nos ocuparemos de lo que él llama programa de valoración en marcha” dentro de nuestra categoría de referencia interna. Ya hemos dicho que ese tipo de valoración tiene que ver con la organización del aprendizaje.

¹¹³ utilizamos la expresión “procedimiento de valoración” para abarcar todos los tipos, desde las pruebas formales a los ejercicios de clase. La bibliografía estadounidense utiliza la palabra “testen relación con todo tipo de valoración, pero creemos que las connotaciones de este término inducen a confusión.

¹¹⁴ Pero véanse HABLEN, DARWIN y MURPHY (1977), en relación con *Gran Bretaña* y GOODWIN y DISCOLL (1930), para los Estados Unidos.

¹¹⁵ Una prueba “objetiva” está compuesta por ítems de respuesta no ambigua, p. ej.: sí o no. La corrección es objetiva, pero no la prueba.

¹¹⁶ EBEL (1972) presenta una jerarquía similar, partiendo de “comprensión de terminología” y terminando con capacidad de hacer juicios de evaluación”.

base empírica, GAGNÉ reconoce que puede modificarse a medida que se desarrolle la teoría. GERLACH y SULLIVAN (1969) también se opone a BLOOM al insistir en que sólo se utilicen enunciados relativos a conductas manifiestas (la taxonomía de BLOOM pretende describir procesos mentales). EGGLESTON (1974, págs. 246-247) recomienda un enfoque de compromiso y supone niveles de actuación respecto a la 'comprensión': en el nivel inferior se sitúan las operaciones de "iniciar, duplicar y repetir", y en el más elevado, "crear, descubrir, reorganizar", etc. Es evidente que la diversidad de formas de describir los objetivos supone que ninguna se destaca como más autorizada que otra, de modo que deben utilizarse como herramientas que ayuden a describir el campo que nos ocupe¹¹⁷.

La descripción de los objetivos, establecidos de alguna de las formas señaladas, lleva a la especificación del procedimiento de valoración en el que se "sitúe" cada ítem, pregunta o actividad para comprobar un objetivo concreto. También es posible analizar retrospectivamente un procedimiento de valoración para determinar los objetivos que evalúa. En la práctica, es probable que se utilicen tanto la planificación prospectiva como el análisis retrospectivo.

La elección de un método de valoración (tercera fase) requiere tener en cuenta el modo de presentación de las preguntas y actividades, y la forma de respuesta del alumno. Así, a los niños pequeños se les puede presentar un estímulo visual con una pregunta oral, mientras la persona que administra la prueba anota "sí" o "no" en la hoja de respuestas. Los textos estadounidenses se limitan a las pruebas escritas (p. ej.: GRONLUND, 1981; TENBRINK, 1974)¹¹⁸ éstas se dividen en pruebas objetivas (es decir, que se puntúan de forma objetiva) y pruebas de ensayo (o preguntas de respuesta libre). Las pruebas objetivas incluyen preguntas que van desde las de bajo nivel, de "sí o no", pasando por las de opción múltiple, hasta los ejercicios interpretativos de alto nivel que exigen utilizar los datos para descubrir relaciones y evaluar supuestos (véase GRONLUND, 1981). Las preguntas de ensayo o de respuesta libre pueden dividirse entre las de respuesta corta y de respuesta larga (cualquier longitud) y en ellas el alumno elige la forma de responder de acuerdo con las instrucciones que van desde las muy estructuradas hasta las abiertas. La característica que distingue las pruebas objetivas de las de ensayo consiste en que en las primeras el alumno escoge la respuesta, mientras que en las segundas, da la respuesta. La característica concomitante para el evaluador es que, comparadas con las de ensayo, las pruebas objetivas exigen que se dedique mucha atención a su preparación y poco esfuerzo para su corrección. La facilidad de calificación las hace muy útiles para valorar a grupos grandes de alumnos. Sin embargo, también las pruebas de ensayo requieren una preparación cuidadosa, sobre todo en cuanto a los criterios de calificación. Otras ventajas e inconvenientes que deben sopesarse consisten en la dificultad de elaborar una prueba objetiva y su elevada fiabilidad, en comparación con la facilidad de construcción de las pruebas de ensayo, aunque su fiabilidad sea reducida¹¹⁹. Generalmente, el argumento que se utiliza para justificar el uso de las pruebas de ensayo es su capacidad de valorar los aprendizajes de alto nivel, aunque INGENKAMP (1977) duda de que existan pruebas que justifiquen esa afirmación. Además, las pruebas objetivas pueden utilizarse también para comprobar los aprendizajes de alto nivel en ciertas áreas temáticas¹²⁰. Una ventaja de usar pruebas objetivas para evaluación consiste en su relativa facilidad de analizar las opciones equivocadas de los alumnos y de representarlas en valores numéricos. Sin embargo, su capacidad de previsión de errores es limitada, mientras que las pruebas de ensayo no sólo pueden poner de manifiesto las concepciones erróneas, sino los conocimientos e ideas que tienen los alumnos no conocidos por los profesores. Asimismo, aunque la utilización del análisis de contenido de pruebas largas de ensayo lleva tiempo, pueden proporcionar una medida satisfactoria de los errores cometidos por los alumnos.

Otros métodos de valoración del alumnado son menos corrientes, pero, ante las críticas del HMI (p. ej.: DES, 1979a) por las cantidades excesivas de trabajos escritos, la valoración de estudiantes con pruebas orales y auditivas, de producción y de recepción, está cobrando importancia en Gran Bretaña. Aunque esta valoración puede comprobar destrezas muy especiales (generativas, es decir, de habla, y receptivas, o de escucha), INGENKAMP (1977) dice que, en especial, el interés de la valoración con pruebas orales de producción es dudoso a causa de las dificultades de realización de la prueba oral. Sin embargo, estas dificultades (relativas por ejemplo, a la fiabilidad) son menos importantes en la evaluación y sería conveniente que se evaluaran los intentos de introducción o de incremento del desarrollo de destrezas de producción y recepción orales¹²¹. Los

¹¹⁷ consideraremos más adelante los objetivos no cognitivos, aunque no los del campo psicomotor, aunque éstos son muy adecuados para los niños pequeños (véase, por ejemplo, GOODWIN y DRISCOLL 1980). En el subapartado sobre los enfoques cualitativos, mencionamos también una tipología alternativa relacionada con las interacciones de aprendizaje del alumno.

¹¹⁸ Podemos decir que, en Gran Bretaña y en el nivel secundario, al menos, la "Industria" de los exámenes oficiales ha sido responsable de grandes innovaciones en cuanto a métodos de valoración. MACINTOSH (1974) expone un conjunto de tales innovaciones.

¹¹⁹ Incluso los problemas de matemáticas que permiten una respuesta libre" adolecen de baja habilidad (INGENKAMP, 1977).

¹²⁰ La Open University lo utiliza para su valoración en el nivel de estudiantes no graduados.

¹²¹ MILLER (DEALE, 1975 *Open University*, 182a, Bloque 4) elaboró un sistema para valorar la discusión en grupo que ayuda también a evaluar este nuevo elemento del *curriculum*.

trabajos prácticos y de proyecto se valoran en pleno proceso de enseñanza y aprendizaje y, a diferencia de la mayoría de los métodos, permiten cierta valoración del propio proceso de aprendizaje, sobre todo cuando se utilizan listas de comprobación del tipo mencionado en el apartado sobre "Observación directa".

La validez y la fiabilidad son cuestiones fundamentales en el diseño de un procedimiento de valoración del alumnado. Cuando se escoge un método de valoración para que se ajuste a los objetivos concretos que se pretenden medir, la forma más obvia de validez que hay que tener en cuenta en primer lugar es la *validez aparente*. Por ejemplo, si se trata de evaluar destrezas prácticas, la evaluación basada en respuestas escritas tendrá poca validez aparente. Sin embargo, la forma más importante de validez es la de *contenido*. Esta es la validez que se comprueba cuando se establece una especificación porque se pretende comprobar que el procedimiento de valoración refleja el contenido y los objetivos que se quieren medir. El principal problema radica en el muestreo de contenidos y objetivos, cuestión sobre la que volveremos más adelante. Como indicamos antes, la fiabilidad tiene menor importancia en el contexto de la evaluación curricular (LEWY, 1977, pág. 218) porque interesan más las puntuaciones de grupo. Más adelante, consideraremos las técnicas para medir la fiabilidad, aunque es difícil que estén al alcance de la mayoría de los centros. No obstante, interesa reducir los sesgos derivados de tres fuentes principales:

1. En relación con el alumno, efectuando las mediciones más de una vez.
2. En relación con la prueba, muestreando los ítems para que reflejen los objetivos (más de un ítem por objetivo).
3. En relación con la calificación, estructurando el ítem o la calificación y disponiendo de más de un elemento calificador.

Los ejemplos publicados de construcción de procedimientos de valoración del alumnado están inevitablemente limitados a estudios de investigación, pero hay tres que merecen la consideración de los profesores: en primer lugar, el trabajo de los investigadores del *Observational Research and Classroom Learning Evaluation* (ORACLE) de valoración del estudio y otras destrezas (GALTON y SIMON, 1980). Estos investigadores se ocuparon de preparar valoraciones de rendimiento que reflejaran mejor la actividad normal de clase que las pruebas normalizadas. Este caso reviste especial interés porque muestra la planificación prospectiva y el análisis retrospectivo en acción. Para valorar las destrezas de estudio, el equipo trabajó con profesores para elaborar ejercicios que, a juicio de éstos, reflejaban las técnicas de estudio que se practicaban en sus aulas. Los analizaron después para ver qué medía cada uno (análisis retrospectivo). A este análisis descriptivo le siguió una valoración de prueba en relación con todos los criterios, que llevaron a cabo los profesores y, por último, se efectuó una valoración en relación con los criterios más adecuados. Con el fin de valorar las habilidades no contempladas por las pruebas convencionales (por ejemplo, la capacidad de escucha, la originalidad, la adquisición de información por otros métodos diferentes de la lectura), los profesores prepararon actividades estructuradas, fijando, en primer lugar, los objetivos que se proponían, elaborando los criterios de conducta a partir de estos objetivos y diseñando las tareas que los evaluarán (planificación prospectiva).

El segundo caso publicado de utilidad es el presentado por WOOD y NAPTHALI (1975). Utilizaron técnicas de rejilla (véase la nota 60) para determinar los constructos utilizados por los profesores para valorar a sus alumnos. Reunieron después todos los constructos de diversos docentes y les pidieron que los clasificasen, con el fin de descubrir los más utilizados. Concluyeron que había que traducir esos constructos a conducta observable. El tercer ejemplo, tomado de BLACK y DOCKRELL (1980), hacía precisamente esto, aunque en el campo de la afectividad. Pidieron a un grupo de profesores que pensaran en una valoración del "esfuerzo", por ejemplo y definieran operacionalmente puntos en la escala de medida (en realidad, varias definiciones aceptadas para cada punto) en términos de conductas específicas.¹²² Creemos importante reiterar nuestra apreciación inicial de que la evaluación sola no justifica el trabajo que supone la elaboración cuidadosa de procedimientos de valoración.¹²³ Para la evaluación no hace falta tanta minuciosidad porque se refiere más a grupos que a individuos.

Los fundamentos de la comparación

Hemos indicado ya que el tipo de referencia que se utiliza en las medidas (norma o dominio) influye en la forma de interpretación de las puntuaciones. Así mismo, afecta al diseño del procedimiento de evaluación.¹²⁴ Ambos tipos de medida requieren las mismas fases básicas de diseño y construcción mencionadas antes, pero las medidas referidas a una norma exigen la selección empírica de los ítems o preguntas (el modelo de Rasch

¹²² Éste tiene mucho en común con HARLEN, DARWIN y MURPHY (1977, aunque ellos abarcan tanto el dominio cognitivo como el afectivo).

¹²³ Debe preocupar el hecho de que los profesores no tengan tiempo para elaborar minuciosamente procedimientos de valoración que aplicar a su alumnado.

¹²⁴ ROWNTREE (1977, pag. 185) niega la existencia de cosas como las pruebas referidas a normas o a criterios, porque la distinción importante consiste en la forma de interpretar las puntuaciones obtenidas en las pruebas y no en su diseño. No estamos de acuerdo con esta interpretación, como HAMBLETON y cols. (1978), pero véase CHRISTIE y FORREST (1981, pag.56).

constituía a base de un procedimiento de este tipo, expuesto en el Capítulo IV). A continuación, examinaremos las diferencias, considerando los distintos supuestos y procedimientos adoptados por cada clase de medida.

Medidas referidas a una norma

El supuesto básico correspondiente a esta referencia consiste en una medida de naturaleza unidimensional (“rasgo”) correspondiente a algún logro general o específico. Por ejemplo, la “habilidad matemática” puede constituir un rasgo, del mismo modo que la “habilidad para hacer álgebra”¹²⁵ Este rasgo se representa mediante una puntuación única y cada ítem de la prueba contribuye con una pequeña porción de información a esta medida. Así, aunque la primera fase debe consistir en garantizar la validez de contenido de la prueba en sí misma, esto es insuficiente; deben considerarse también las puntuaciones alcanzadas por los alumnos. Si cada ítem de la prueba contribuye a la medida de este rasgo unitario, las puntuaciones de los alumnos en un ítem deben correlacionarse con sus puntuaciones totales en la prueba, lo que significa también que las puntuaciones en cada ítem pueden sumarse porque miden lo mismo y que la puntuación total será más fiable, es decir, menos sujeta a error.¹²⁶ Esa correlación garantiza la posibilidad de diferenciar a los alumnos si los ítems puntuados se correlacionan con la puntuación total, o sea, sus puntuaciones se distribuyen de manera que la probabilidad de que haya alumnos que consigan la misma puntuación sea baja. En consecuencia, esta correlación recibe el nombre de *índice de discriminación*. Por otra parte, un ítem debe presentar una dificultad moderada o, lo que es lo mismo, la mitad de la población debería dar una respuesta acertada y la otra mitad errónea; la medida correspondiente a esta característica constituye el *índice de facilidad*, indicación del porcentaje de alumnos que responden correctamente.¹²⁷ Estos supuestos sirven de fundamento a las recomendaciones sobre la forma de tratar las puntuaciones que hacen los textos dirigidos a los profesores (p. ej.: DEALE, 1975; GRONLUND, 1981; SATTERLY, 1981), aunque aconsejan siempre técnicas simplificadas. Los supuestos citados significan que una prueba se define en relación con el grupo que la realiza, y estos supuestos se utilizan para establecer su *validez de constructo*, es decir que, en este caso, mide un rasgo simple, verificado por medios empíricos. En el contexto de esta referencia, la puntuación individual de un alumno sólo tiene sentido en relación con un grupo y, a poder ser, un grupo grande, que sea representativo de la población de todos los niños de 11 años.¹²⁸ Otro supuesto consiste en que la habilidad (rasgo) se distribuye normalmente entre toda la población en cuestión (p. ej., los niños de 11 años).¹²⁹ Cuando la muestra es grande o seleccionada de manera representativa o aleatoria, este supuesto es relativamente seguro. En consecuencia, tras diseñar y elaborar una prueba, se ensaya y se calculan los índices de facilidad y discriminación para cada ítem. Si estos índices se hallan entre límites razonables, el ítem en cuestión se acepta; si no, se rechaza.¹³⁰

Medidas referidas al dominio

Las medidas basadas en esta referencia sitúan en primer lugar la validez de contenido: ante todo, se consideran apropiados los ítems si prueban un objetivo o dominio determinado. En consecuencia, no podemos presuponer cómo se distribuirán las puntuaciones. Por ejemplo, si la enseñanza es buena, los alumnos alcanzarán, en generar, los objetivos; los ítems no discriminarán entre alumnos y serán fáciles (índice de facilidad alto; índice de discriminación bajo o negativo). HAMBLETON y cola (1978) dicen que fiarlo todo a la validez de contenido presenta problemas:

1. El dominio que se prueba puede no estar bien especificado y, en todo caso, esta especificación supone una materia muy estructurada, como las matemáticas.¹³¹
2. Quizá los ítems no constituyan una muestra aleatoria (o representativa) de todos los ítems posibles que

¹²⁵ Véase en CHOPIN (1981) una breve exposición de esta visión de la medida, y en CHRISTIE y FORREST (1981, pág. 48-52) una revisión del pensamiento actual al respecto- Téngase en cuenta también que, en la mayor parte de estas exposiciones, se da por supuesto que el procedimiento de valoración es una “prueba” y que los ítems de las pruebas “objetivas” se utilizan con respuestas correctas y erróneas. Seguiremos analizando esta cuestión.

¹²⁶ La teoría clásica de pruebas supone que la puntuación de cada ítem es una estimación de la “puntuación verdadera” del alumno (puntuación sin error) en el rasgo que se mide.

¹²⁷ NUTTALL y WILLMOTT (1972, págs. 19-20) exponen la justificación matemática, basada en la teoría clásica de pruebas.

¹²⁸ Al decir que un alumno tiene una capacidad matemática superior a la media” respecto a los alumnos de la clase 1 B de la escuela X, sólo se consigue una información limitada, aunque eso es lo que aportan muchos informes escolares.

¹²⁹ La distribución normal es una distribución teórica con propiedades matemáticas conocidas que describe aproximadamente la distribución de las propiedades físicas p. ej., la altura y las medidas mentales en una población.

¹³⁰ Los límites aceptables suelen ser: índice de facilidad: entre 30 y 70%; índice de discriminación: mayor que +0,3 (esta última es una medida de correlación; los números negativos indican una correlación negativa).

¹³¹ HAMBLETON y cols. (1978) dicen que incluso los objetivos conductuales suscitan dudas; ¡solían ser mas criticados por ser demasiado específicos!

midan el cumplimiento de los objetivos.

3. Los ítems se relacionan con aspectos específicos de alguna habilidad y, en relación con la habilidad matemática, por ejemplo, sería más adecuado un perfil de puntuaciones que una puntuación global. (INGENKAMP, 1977; PILLINER, 1979, págs. 38-39).

HAMBLETON y cols. (1978), partiendo de la bibliografía estadounidense, e INGENKAMP (1977), basándose en la europea, señalan varias formas de paliar los dos primeros problemas: 1) pedir a expertos que clasifiquen los ítems como representativos del dominio; 2) pedir a expertos que relacionen los ítems con objetivos (los desacuerdos reflejarán una validez de contenido baja); 3) pedir a grupos de redactores de ítems que elaboren ítems paralelos correspondientes a los mismos objetivos. HAMBLETON y cols. (1978) afirman también que es fundamental utilizar pruebas empíricas de la validez de constructo de las *puntuaciones* (y no sólo de la validez de los *ítems*). En el caso de las pruebas de criterio relacionadas con el aprendizaje para el dominio de una materia, es difícil utilizar el análisis convencional de ítems porque la mayoría de los alumnos los cumplimentan de forma correcta (índice de facilidad alto); en consecuencia, no se produce discriminación (índice de discriminación bajo) y la varianza de las puntuaciones es baja. Se sugiere, no obstante, que se utilice el análisis de ítems para comprobar los que no sirvan. Los ítems que deben medir el mismo objetivo deben estar correlacionados; si no es así, no estaría garantizada la validez de constructo de los ítems de *este grupo* (la referencia a la norma es un procedimiento dependiente del grupo). En consecuencia, llegamos a un punto en el que la diferencia entre las dos pruebas disminuye con respecto a la forma de construir las.¹³² El supuesto de la falta de varianza de las puntuaciones no se aplica necesariamente a la evaluación (p. ej., cuando la enseñanza obtiene resultados mixtos), de ahí que las técnicas analíticas de la medida con referencia a la norma pueden aplicarse para comprobar los ítems erróneos.

HAMBLETON y cols. (1978) creen que la teoría y las orientaciones prácticas que existen son suficientes para elaborar buenas pruebas referidas al dominio, aunque INGENKAMP (1977), al revisar la situación europea, señala que, a pesar de la alta estima que en general se les profesa, hay pocos ejemplos de este tipo de pruebas.¹³³ El advenimiento del GOSE engrosará el archivo sobre las valoraciones referidas al dominio. Con independencia de las circunstancias que afecten a los creadores profesionales de pruebas, la tecnología que requiere su elaboración está fuera del alcance de las escuelas, tanto en términos de tiempo como de apoyo estadístico.¹³⁴ Es más, pocas escuelas estarán dispuestas a conformarse con pruebas de tipo objetivo, y no es aplicable la misma tecnología a todos los tipos. Cuando se hace así (p. ej., en los exámenes de ensayo referidos a una norma), la estadística que se debe utilizar es aún más compleja (véase NUTTALL y WILLMOTT 1972). Más adelante, revisaremos cómo pueden utilizarse las pruebas publicadas.

Comparación ipsativa

Antes de pasar adelante, conviene decir algo de la comparación ipsativa: concepto expresado en lenguaje técnico que ha surgido no hace mucho, asociado en gran parte con la elaboración de perfiles de los alumnos. Derivado del término latino *ipse*, que puede traducirse por “sí mismo”, se refiere al individuo, constituyendo la base de comparación el rendimiento diferencial del mismo sujeto en relación con distintos criterios o dominios, o el rendimiento diferencial del mismo individuo respecto al mismo criterio en el tiempo. En el primer caso, se presenta la actuación en una serie de áreas de manera que puedan compararse los aspectos positivos y negativos. Los “perfiles” de los alumnos, que se representan a veces en forma de diagrama de barras, pueden elaborarse a partir de las puntuaciones obtenidas en conjuntos de pruebas normalizadas. Por ejemplo, los resultados de los Richmond Tests of Basic Skills (HIERONYMUS, LINDQUIST y FRANCE, 1974), que están constituidos por una batería de once pruebas, cada una de las cuales está diseñada para valorar la actuación en un dominio diferente, pueden representarse de ese modo. (CLIFT, ABEL y SIMPSON, 1982, manifiestan sus dudas respecto a que valoren dominios independientes, como explicaremos a continuación.) De este modo, la comparación entre las actuaciones en distintos dominios es fácil, requiriendo sólo examinar las distintas puntuaciones. Más difícil resulta utilizar la información en forma no estadística con fines semejantes, lo que se pone de manifiesto en perfiles de tipo muy diferente, relacionados con las iniciativas sobre registros de rendimiento. (Véase una breve descripción en el Capítulo y.) Pueden abarcar muchas áreas de experiencia curricular y extracurricular del alumno, pero sólo pueden registrarse los aspectos positivos y, con frecuencia, a modo de informes no normalizados. La mayor parte de los diagnósticos de los aspectos positivos y negativos se realizan en forma oral en sesiones de revisión entre profesor y alumno; no obstante, debe quedar clara

¹³² En esta situación, la postura de ROWNTREE (1977) respecto a la inexistencia de diferencias entre pruebas referidas a normas y las referidas a criterios tiene cierta justificación (véase la nota 124). Sin embargo, las pruebas referidas a dominio (criterio) son más específicas y tienen más ítems por cada objetivo.

¹³³ SUMMER y ROBERTSON (1977) revisan lo existente en Gran Bretaña.

¹³⁴ GOWER (NIXON, 1981) participó en la elaboración de alguna prueba, pero admite que lleva mucho tiempo. Aunque en la introducción a la descripción de sus experiencias se refiere a la utilización de las pruebas para evaluar el enfoque de un tema, parece haber elaborado las pruebas en el marco de un enfoque orientado al aprendizaje para el dominio de alguna materia.

constancia escrita de los aspectos positivos, aunque la identificación de las áreas de relativa debilidad requiera “leer entre líneas”. El perfil de City and Guilds of London Institute (CGLI) (adoptado en algunos esquemas de perfil TVEI) permite sombrear un área de acuerdo con una escala en la que figuran descripciones cualitativas de los distintos niveles de cada uno de los catorce criterios. El perfil puede representarse también mediante McCORMICK, 1987, pueden verse algunos ejemplos).

En el segundo sentido, la comparación ipsativa se realiza cuando se valora a los individuos, o se valoran ellos mismos, en relación con normas fijadas de acuerdo con su propia actuación anterior. Es una característica de los sistemas de valoración graduados o por etapas (incorporados a veces a los registros de rendimiento como, por ejemplo, en los componentes “G” de inglés, matemáticas, ciencias y lenguas modernas del Oxford Certificate of Educational Achievement) que, mediante la concepción de los niveles de rendimiento en un dominio, constituyen el fundamento de comparación de la actuación actual con la anterior.

ROWNTREE (1977) diferencia la valoración de los individuos en relación con su actuación anterior puntuaciones en cada uno de los criterios, indicando los niveles de actuación alcanzados (en como base de comparación independiente y distinta, aunque las dos formas descritas de comparación ipsativa pueden considerarse reductibles en último término a modos de valoración referidos a norma y a criterio o dominio, porque, en algún momento, la actuación se define en relación con determinados criterios, como en el caso del CGLI citado, o con la actuación de los demás alumnos.

Más difícil resulta descubrir la base de referencia de los componentes abiertos en los registros de rendimiento, sobre todo en el registro personal. En algunos casos, parece que los criterios se generan con referencia al individuo más que al grupo. Este aspecto se refleja mejor en aquellos esquemas cuyos enunciados de rendimiento se extraen de relatos de experiencia, relativamente no estructurados, recopilados por los alumnos, a menudo en forma de diarios. Como la selección de lo anotado depende en gran medida del alumno, los criterios “pertenecen” al individuo y pueden guardar poca relación con los utilizados por los demás. En este sentido, son genuinamente ipsativos. No obstante, en la práctica, la mayor parte de los enunciados sumativos de rendimiento se redactan en diálogo con los profesores, quienes orientarán la selección sobre la base de un conjunto implícito o explícito de criterios, más o menos aplicable a todos los grupos de alumnos. En realidad, las pruebas preliminares, no publicadas aún, de la evaluación PRAISE (*Pilot Records of Achievement in Schools Evaluation*) de esquemas piloto indica que a la mayoría de los profesores les resultan poco manejables las narraciones libres, cambiando sin dilación a informes estructurados de los alumnos, diseñados de manera que faciliten información sobre criterios especificados de antemano y referidos al grupo. (Ambos autores pertenecemos al equipo del PRAISE.)

Es interesante señalar que, al investigar de qué formas los profesores y alumnos hacen valoraciones de perfiles, el equipo de evaluación del PRAISE ha conseguido algunas pruebas de la resistencia a la desaparición de los hábitos de referencia a una norma. Aunque es patente que participan en la tarea de juzgar la actuación en relación con un criterio, algunos profesores, y alumnos en el contexto de la autoevaluación, hacen comparaciones tácitas con otros miembros del grupo de alumnos, sobre todo cuando han de indicarse los niveles de actuación respecto a un criterio. En otras palabras, hay una tendencia a clasificar a los alumnos, y a clasificarse ellos mismos, en relación con sus compañeros antes de juzgar si han alcanzado un criterio o un nivel especificado. Con frecuencia, esto es una consecuencia del establecimiento de criterios amplios que abarcan un dominio extenso y requieren, por tanto, una valoración impresionista y holística. (Por ejemplo, la “destreza” en relación con CDT*.)

En resumen, quizá el único caso de comparación ipsativa diferente de las comparaciones referidas a normas o dominios se produzca en circunstancias en las que los mismos criterios pueden derivarse, y de hecho así lo hacen, de forma ipsativa, es decir, sin especificarse de antemano en relación con un dominio generalizado, sino que se formulan a partir de los datos de la actuación y la experiencia individuales. Es posible que esta forma resulte especialmente adecuada en áreas del *currículum* como la de artes creativas, en las que, para los profesores es difícil e incluso inaceptable especificar de antemano unos resultados, aunque estarían dispuestos a analizar lo conseguido, o a ayudar a los alumnos a hacerlo, en retrospectiva. En tales casos, es muy probable que los enunciados sobre rendimiento y, por tanto, la base de comparación, se refieran mucho más a cada individuo que a un grupo. Ante esto, las formas ipsativas de valoración, con referencia estricta al individuo, son aún menos útiles que las relativas a un grupo en el contexto de la evaluación curricular. No obstante, en la medida en que quienes experimentan el *currículum* son los individuos, los profesores harán bien en prestar atención a todas las consecuencias que se deriven de las pautas individuales de rendimiento para el *currículum* y la enseñanza, ya se conciben éstas en relación con el grupo o con los mismos individuos. Pero esto nos devuelve al problema señalado en el Capítulo V, que supone que el profesor responde en el nivel del alumno individual adaptando su enseñanza a sus necesidades.

El uso de la información rutinaria sobre valoración del alumnado¹³⁵

* CDTI: *Craft, Design and Technology*. (N. del R.)

¹³⁵ Queremos agradecer la ayuda que nos prestó Desmond NUTTALL con algunas ideas que exponemos en este apartado, pero, como es habitual, aceptamos toda la responsabilidad por los defectos que puedan encontrarse.

Dadas las dificultades existentes; antes señaladas, para elaborar procedimientos de valoración, las escuelas tienen que basarse en su sistema interno de valoración, aunque utilicen los principios generales indicados. Es importante la idea de mejorar la validez de contenido de las valoraciones que se realizan mediante especificaciones claras, por ejemplo, a través del análisis retrospectivo, y de asegurar la validez aparente seleccionando un método de valoración adecuado. Por ejemplo, los compañeros podrían actuar en calidad de expertos para juzgarla correspondencia de los ítems con los objetivos, o acordar descripciones conductuales para las escalas de clasificación. En tal caso, centrándose especialmente en la validez de contenido, podría emprenderse la elaboración de una prueba referida a dominio. (Si fuera posible ensayar la prueba antes de utilizarla en la evaluación, la estadística serviría para ayudar a eliminar ítems inadecuados. De lo contrario, pueden excluirse las puntuaciones de dichos ítems de las generales de la prueba; véase en KIBBLEWHITE 1981, un enfoque gráfico.)

Volvamos, por tanto, a los tipos de análisis que pueden hacerse o al uso que se puede dar a los procedimientos de valoración referidos a un dominio, en el nivel del aula y en relación con una destreza determinada. La tarea básica consiste en comprobar si la clase consigue los objetivos medidos en la valoración del alumnado. En otras palabras, ¿alcanza la mayoría de los alumnos los objetivos?; ¿consiguen todos los objetivos principales?; ¿hay algún grupo concreto que no logra alcanzarlos (p. ej., niños o niñas, grupos étnicos concretos, grupos de ciertas clases sociales)? Puede analizarse también el procedimiento de valoración para determinar errores o concepciones equivocadas corrientes: en relación con las preguntas de ensayo, mediante el análisis del contenido y la elaboración de una lista de errores; contabilizando éstos en las pruebas objetivas (o sea, los ítems con bajos índices de facilidad)¹³⁶. La acción que se emprenda sobre la base de este tipo de análisis dependerá del momento en que se realice la evaluación. Habrá que hacer algo con los individuos afectados (función diagnóstica), pero quizá haga falta reconsiderar también el enfoque de la enseñanza.¹³⁷

La función diagnóstica, tanto en el aprendizaje individual como en la enseñanza, constituye uno de los objetivos fundamentales de muchos registros de rendimiento y de otros esquemas de elaboración de perfiles. Resalta especialmente en aquellos sistemas que incluyen una valoración sobre la base de criterios especificados a partir de los objetivos de enseñanza correspondientes a ciertos módulos curriculares o unidades de trabajo. Se parte de la base de que las valoraciones efectuadas pueden utilizarse como fundamento para diagnosticar los puntos fuertes y débiles, estableciendo a continuación las metas a las que se orienten las mejoras futuras. El establecimiento de esas metas se refiere generalmente a objetivos de los alumnos, pero suele suponerse también que el mismo proceso ayudará a las escuelas a descubrir los méritos y las debilidades del *currículum*, la enseñanza y la organización (DES, 1984d, pág. 3). En efecto, el equipo de evaluación de PRAISE tiene pruebas de que las escuelas consideran que los registros de experiencia y rendimiento de los alumnos constituyen un medio de comprobación de lo enseñado, reconociendo que hay cosas, como las experiencias que los alumnos tienen fuera de la escuela, que salen del marco del control curricular. Lo mismo puede decirse de las cualidades personales, que aparecen en muchos sistemas de registro, aunque en manos de la escuela está la posibilidad de propiciar oportunidades curriculares, en donde se demuestran esas cualidades personales.

Por supuesto, en relación con el rendimiento de los alumnos, siempre existe el peligro de atribuirlo directamente al *currículum* y la enseñanza. También pueden ser decisivas las condiciones antecedentes, como la extracción social y la experiencia educativa previa, o las influencias contextuales, como la presión del grupo de compañeros, el estado de salud del individuo e, incluso, el tiempo meteorológico.

La afirmación de que pueden analizarse las valoraciones a efectos de diagnóstico se basa en el supuesto previo de que los procedimientos de valoración son fiables, o sea, que no existen sesgos en el reconocimiento del rendimiento o en la asignación de calificaciones o grados en las pruebas. DEALE (1975, págs. 36-38) indica diversas formas de reducir los sesgos en las calificaciones rutinarias que, en el contexto del proceso de registro del rendimiento, pueden aplicarse a las simples valoraciones del cumplimiento o incumplimiento de un criterio o de la calidad o logro demostrados. Las sugerencias de DEALE se relacionan con las dos fuentes principales de sesgos:

1. Sesgo intracalificador: elaborando una respuesta paradigmática o distinguiendo con toda claridad los criterios (p. ej., argumento del ensayo o hechos abarcados). Si ninguna de estas posibilidades resulta adecuada, puede

¹³⁶ En relación con las preguntas de opción múltiple, pueden analizarse los distractores para descubrir los errores cometidos. Los distractores son las respuestas alternativas, plausibles, pero incorrectas, que se presentan. Se escogen con la idea de que los alumnos que cometen ciertos tipos de errores los seleccionen: por tanto, no sólo apodan una medida de error, sino que facilitan la elaboración de hipótesis sobre las razones de la elección.

¹³⁷ Sin embargo, no podemos ser optimistas sobre lo que pueda lograrse utilizando medidas de resultados. HARLEN (1973, pág. 31), por ejemplo, desarrolló unas muy cuidadas para ayudar a evaluar los materiales de un proyecto curricular. En relación con los resultados obtenidos, dijo: "Aunque, en general, los resultados de la prueba fueron satisfactorios, tuvieron pocas consecuencias para la nueva redacción". No obstante, reconocía que sirvieron de orientación sobre las partes más débiles de los materiales, aunque hubo que recabar de otras fuentes información sobre las razones de esas debilidades.

utilizarse la calificación de acuerdo con la impresión.

2. Sesgos intercalificadores: normalizando la calificación (p. ej., dialogando sobre los criterios, calificando muestras y poniéndose de acuerdo sobre el nivel mínimo); contando con otra opinión respecto a respuestas difíciles de calificar; pidiendo a otra persona (p. ej., el director del departamento) que actúe como moderadora para muestrear las respuestas y comprobar la uniformidad de nivel entre calificadores.

Cuando hay que efectuar el análisis de un curso completo (siempre en el marco de una materia o dominio), en vez de hacerlo con una sola aula, hay que poner especial cuidado en reducir el sesgo intercalificador que puede surgir cuando se califica una tarea o prueba común. En este nivel, es posible comparar clases pertenecientes al mismo curso. Si las clases son comparables (p. ej., se trata de aulas heterogéneas en cuanto a la capacidad de los alumnos que las integran o comparten un nivel homogéneo de capacidad de los alumnos), pueden determinarse los puntos fuertes y débiles de cada una (respecto a ciertos objetivos). Los profesores que imparten clases en el curso en cuestión pueden, por tanto, discutir las etapas apropiadas para solucionar los problemas que se detecten. Es obvio que también puede contemplarse cómo se desenvuelven otros grupos (p. ej., chicos y chicas), tanto en el conjunto del curso como en cada clase¹³⁸.

Hasta aquí, hemos dado por supuesto que la información obtenida se relaciona con determinados objetivos (referencia al dominio). Sin embargo, quizá no se tenga suficiente certeza respecto a lo que mide en detalle el procedimiento de valoración del alumnado; a lo mejor sólo se tiene constancia de que mide la capacidad matemática, por ejemplo. En este caso, conviene utilizar medidas referidas a la norma, siempre que los ítems se ajusten a los límites antes expuestos (véase la nota 130). En tal caso, pueden revisarse las puntuaciones totales de los alumnos o, más en concreto, las puntuaciones medias de los subgrupos (p.ej.: chicos y chicas), para ver si se desvían de lo previsto. Si las clases que pertenecen al mismo curso son comparables en cuanto a la capacidad de sus alumnos, las puntuaciones obtenidas a partir de procedimientos de valoración distintos, pero equivalentes, pueden normalizarse, de manera que las puntuaciones medias y la dispersión de puntuaciones sean las mismas en cada clase.¹³⁹ Aunque es obvio que no pueden efectuarse comparaciones entre aulas (porque todas las medias son iguales), es posible comparar los subgrupos que forman parte del curso (chicos y chicas, clases sociales, etc.). Aun en ese caso, se presentan dificultades porque, por ejemplo, en general se prevé que los alumnos de clase social baja se desenvuelvan peor que el grupo en su totalidad. Si, por otra parte, se hacen esfuerzos especiales para compensar estas previsiones, las puntuaciones normalizadas medias (véase la nota 139) pondrán de manifiesto el éxito o el fracaso de estas medidas.

No parece que pueda aceptarse sin más el supuesto de comparabilidad de los procedimientos de valoración utilizados por diversos profesores que desarrollen sus actividades de forma paralela, aun tratándose de la misma asignatura¹⁴⁰. Una alternativa consiste en utilizar una prueba común (para un mismo curso) para elaborar una escala de normalización de los demás procedimientos de valoración (p. ej., las tareas para casa o los ejercicios de clase), de manera que puedan efectuarse comparaciones directas¹⁴¹. Esto permitiría la comparación de la actuación de cada aula (de un mismo curso) respecto a distintos elementos de la materia (p. ej., los trabajos prácticos), para ver, por ejemplo, la manera de comparar los enfoques utilizados por cada uno de los profesores. No obstante, este procedimiento requiere una correlación elevada entre la prueba común y el procedimiento de valoración sometido a normalización.

Nos hemos ocupado sólo de las comparaciones dentro de una misma materia o dominio. **SHIPMAN** (1979, pág. 27) indica que pueden compulsarse diversas materias extrayendo las puntuaciones percentiles (véase la nota 139) correspondientes a cada materia de todos los alumnos de un aula¹⁴². **INGENKAMP** (1977, págs. 28-30) manifiesta sus dudas respecto a este supuesto, desde el punto de vista del rigor variable de las calificaciones, que se supone compensará la normalización. **CHRISTIE y FORREST** (1981) dan un paso más, diciendo que no hay razón para suponer que, en el terreno de los exámenes oficiales, puedan compararse las distintas asignaturas. Nosotros pensamos que ya se dan por supuestas muchas cosas en relación con las comparaciones dentro de la misma asignatura y que pretender la comparabilidad entre asignaturas en un nivel grupal es ir demasiado lejos. El grupo más adecuado para utilizar este tipo de referencia a normas es el

¹³⁸ Por ejemplo, podemos suponer que los niños de clase social más baja [p. ej., según la escala del Registrar General/—Registrador general. Es la persona responsable oficial del gobierno británico de los censos de población (Nota del R.)—] actúan igualmente bien en todas las clases. Si su conducta difiere según aulas, podrían realizarse otras investigaciones. Es obvio que esa opción no carece de problemas por sus consecuencias, pues el resultado podría depender también de la enseñanza de cada profesor, sobre todo si los subgrupos considerados se categorizasen sobre la base del origen étnico.

¹³⁹ En general, los percentiles dan una medida ordinal en una clasificación y las puntuaciones normales proporcionan una medida de diferencia respecto a la puntuación media. Los percentiles no pueden sumarse ni introducirse en cómputo alguno, mientras que las puntuaciones *z* sí pueden.

¹⁴⁰ Suponiendo que no haga falta desarrollar ningún paso especial para garantizar que sean comparables.

¹⁴¹ **DEALE** (1975, págs. 133-139) presenta dos técnicas con este fin, una de las cuales supone elaborar "mapas de puntuaciones" y la otra, combinar los órdenes de clasificación.

¹⁴² **DEALE** (1975, pág. 130) confía también hasta cierto punto en la estabilidad del grupo de clase, por lo que supone que la actuación del conjunto del grupo será en general semejante en todas las asignaturas. No obstante, como indicamos antes, sólo resulta sensible en la comparación de subgrupos pertenecientes a la misma clase, porque todas las asignaturas reciben la misma media a través del proceso de normalización (ose reducen a los mismos órdenes de clasificación, en el caso de los percentiles), o sea, no existen diferencias entre asignaturas en el nivel de grupo.

curso¹⁴³, y parece excesivo pretender normalizar las puntuaciones de todas las materias de cada aula en relación con una norma común de curso.

La valoración del alumnado basada en referencias externas

Como hemos comentado, la valoración interna plantea problemas en la obtención de una norma con la que comparar la actuación de los alumnos. Es más, la referencia a una norma de clase o de curso se basa en suposiciones de dudoso valor respecto a la distribución de las capacidades. La valoración referida al exterior del centro, con una norma basada en una muestra más grande de la población (p. ej., todos los niños y niñas de 11 años), ofrece, dentro de ciertos límites, un nivel estable. La información bruta obtenida, por regla general, una puntuación global (de capacidad matemática, por ejemplo), puede constituir un indicador útil si se tiene en cuenta la información derivada de la valoración interna. De este modo, las pruebas de rendimiento estandarizadas y referidas a una norma constituyen una posible referencia externa.

A la escuela o al profesor le compete seleccionar una prueba adecuada. JACKSON (1974) constituye la fuente de orientación más importante sobre pruebas normalizadas para los profesores británicos¹⁴⁴. La consideración fundamental se refiere a la validez de contenido de la prueba, o sea: ¿se ajusta a los objetivos de la enseñanza? Esto exige un análisis detallado y NUTTALL (Open University 1981a, Bloque 4) advierte su disconformidad a dar por buena, sin más, la apreciación del creador de la prueba respecto a lo que ensaya, ya que, sin otros elementos de juicio, la validez de contenido puede ser específica del contexto. Así mismo, sugiere que se tenga en cuenta la validez aparente, planteándose si los alumnos están acostumbrados a la situación de prueba. Otro aspecto que considerar a la hora de escoger una prueba es la información que facilita el manual de la misma; en concreto:

1. Validez: JACKSON (1974) dice que son importantes la validez predictiva y la concurrente, pero NUTTALL (Open University, 1981 a, Bloque 4, pág. 50) no está de acuerdo con esta apreciación.
2. Fiabilidad.
3. Condiciones de estandarización, o sea, cómo se realizó, cómo se relaciona con los alumnos a quienes se piensa administrar y en qué fecha se efectuó la estandarización.¹⁴⁵

Ya hemos señalado que existen pocas pruebas referidas a criterio, pero, dado que son más específicas, es preciso examinar con todo cuidado su validez de contenido.

Como dijimos al principio de este apartado, no abundan las orientaciones sobre el uso de la información derivada de la valoración del alumnado a efectos de evaluación curricular, lo que se refleja en la falta de ejemplos de la utilización de pruebas normalizadas en las escuelas para este fin.¹⁴⁶ Esos ejemplos son importantes porque, como señala NUTTALL (Open UNIVERSITY 1981a, Bloque 4, página 49) en el contexto de las pruebas de Richmond, los autores de pruebas afirman que son útiles sin facilitar datos empíricos de tales aplicaciones. Un estudio de casos de una escuela que utilizaba los Richmond Tests of Basic Skills puso de manifiesto algunos problemas a este respecto (*Open University* 1982, Estudio de casos). Esta escuela pasó la batería de pruebas a todo un curso a su ingreso en el centro y un año más tarde, de nuevo (la prueba presentaba normas relacionadas con la edad). Se trataba de ver si al cabo del año, el curso había progresado más de lo previsto. El conjunto de pruebas facilita una puntuación global y once puntuaciones parciales relativas a áreas como vocabulario, capacidades lingüísticas, técnicas de trabajo intelectual y habilidades matemáticas¹⁴⁷. Sin embargo, se descubrió que las puntuaciones obtenidas en las subpruebas de la batería estaban muy correlacionadas, por lo que parecía que un único factor subyacía a la puntuación global, lo que cuestionaba la justificación de las puntuaciones parciales en cuanto indicadores de habilidades independientes. Sin embargo, la puntuación total sólo indicaba tendencias generales, facilitando solamente una información en bruto. Pueden compararse los mismos grupos en años diferentes y distintos cursos, dando por supuesta la comparabilidad entre grupos, o sea, la misma distribución de las habilidades. Por ejemplo, SPARSHOTT

¹⁴³ Esto es así porque es más probable que los supuestos respecto a la distribución normal de capacidad sean válidos para un mismo grupo de edad; sólo uno de cada 400 grupos de clase cumple con este supuesto (INGENKAMP, 1977, pag. 28).

¹⁴⁴ Los textos estadounidenses típicos sobre evaluación, como TENBRINK (1974), contienen en gran medida la misma información, aunque sin los detalles relativos a pruebas concretas. GRONLUND (1981) facilita esta última información. En relación con la infancia, GOODWIN y DRISCOLL (1980) dan una información muy detallada sobre una muestra de pruebas de rendimiento. NUTTALL (Open University, 1981 a) analiza una batería concreta de pruebas (Richmond Tests of Basic Skills) poniendo de manifiesto el tipo de preguntas que deberían plantearse en relación con las pruebas.

¹⁴⁵ El estudio de investigación del London University Instituto of Education sobre las pruebas de las LEA y las escuelas pone de manifiesto que muchas administraciones locales utilizan pruebas normalizadas unos diez años antes. Aun en el caso de que existan normas más recientes, el programa de pruebas de las LEA abarca una muestra mayor que la utilizada por los autores de las pruebas.

¹⁴⁶ O'DONNELL (1981) sólo encontró en un condado inglés cinco escuelas, entre varios centenares, que mencionaran la evaluación curricular como un motivo para utilizarlas pruebas.

¹⁴⁷ Los autores de la prueba no dicen que sus puntuaciones parciales reflejen habilidades mutuamente excluyentes, aunque no sería una suposición irracional, por lo que parece que la prueba carece de validez de constructo.

(FAIBROTHER, 1980) utilizó una prueba de habilidad cognitiva para mostrar cómo variaban los distintos departamentos de una escuela, pero, a partir de los detalles no está clara su validez.

Todo esto sirve para hacer hincapié en la necesidad de tomar las puntuaciones globales de las pruebas sólo como indicadores; si los resultados que arrojan son imprevistos, hará falta obtener más información para explicarlo. SHIPMAN (1979) advierte a los profesores que busquen otras causas de cambio como, por ejemplo, la variación de los alumnos de nuevo ingreso o, incluso, un cambio de ámbito nacional en los niveles de "lectura, por ejemplo. No obstante, también dice que deben pasarse por alto los cambios de pequeña magnitud a causa de la poca fiabilidad de las puntuaciones de las pruebas y la forma de contabilizar a los ausentes.

Las puntuaciones del grupo (aula, etc.) en cada objetivo en una prueba referida a criterio (dominio) pueden utilizarse también para ver si una clase o curso se desenvuelve peor que lo previsto. Esto supone una forma de referencia a norma que puede combinarse con juicios sobre la comparabilidad de los objetivos y de los ítems de la prueba con la enseñanza.

Aunque en el Capítulo IV planteamos nuestras dudas respecto al uso de los resultados de los exámenes oficiales como medio de evaluación, es cierto que pueden desempeñar un papel, facilitando una referencia externa utilizable en la escuela. El principal problema de interpretación de estos resultados, que señalamos en el Capítulo IV, consiste en la necesidad de tener en cuenta las características de los alumnos de cada escuela como, por ejemplo, su clase social y anterior experiencia escolar. Además, la normativa de la escuela sobre la presentación de sus alumnos a los exámenes públicos influirá en los resultados (SHIPMAN, 1979). Por ejemplo, una escuela puede dejar al arbitrio de cada uno la presentación a esos exámenes, con lo que el número de alumnos será muy alto, aunque también lo será la tasa de suspensos; o bien, puede dejar examinarse sólo a aquellos alumnos de los que tiene garantía suficiente de que aprobarán, con lo que la tasa de aprobados será elevada (bajo número de suspensos). Cada una de estas alternativas refleja objetivos diferentes de las escuelas, que deben ponerlos en claro antes de utilizar los resultados a efectos de evaluación curricular. Por ejemplo, las restricciones en la presentación de alumnos a los exámenes sólo beneficia a los más capaces, mientras que la libertad de presentación da la oportunidad de aprobar a un mayor número. El problema inicial consiste en obtener los datos en forma utilizable, tabulándolos después para su análisis. SHIPMAN (1979) propone comenzar con unas pocas tabulaciones simples, pasando a otras más complejas, a medida que hagan falta. Asimismo, sugiere una tabulación básica (pág. 71) que proporcione la siguiente información de cada alumno: cociente de razonamiento verbal (CRV) (Verbal Reasoning Quotient, VRQ) y edad lectora, y la calificación obtenida en cada asignatura en el GCE o OSE. John GRAY (Open University, 1982, "Making More Sense Out of Examination Results", pág. 14) presenta también una tabla sencilla para demostrar la relación entre la actuación inicial en una prueba normalizada (como el CRV) y los resultados de los exámenes de quinto curso. Los alumnos se clasifican en cinco grupos de edad y los resultados de los exámenes se presentan por separado en relación con el GCE y con el CSE, así como en una puntuación global combinada¹⁴⁸.

Pueden hacerse diversas comparaciones. En primer lugar, puede calcularse la puntuación total de cada grupo de edad, comparándola con la de años anteriores. A este respecto, son importantes las tendencias, aunque será preciso que los resultados de años distintos indiquen actuaciones "típicas". Dichas tendencias pueden representarse en forma gráfica (Open University, 1982, "Making More Sense Out of Examination Results", págs. 14-15). En segundo lugar, pueden efectuarse comparaciones entre materias, aunque el reducido número de alumnos que se presentan a algunas asignaturas, las diferencias entre chicos y chicas y la falta de comparabilidad de asignaturas (y quizá de las normas de ingreso de los departamentos de las asignaturas) las hagan sospechosas. Según SHIPMAN (1979) es más adecuado comparar cada asignatura consigo misma a lo largo del tiempo. Como en el caso de la evaluación interna, es posible considerar diversos subgrupos en cada curso (p. ej., quinto), teniendo en cuenta los grupos étnicos minoritarios y los de clase social (SHIPMAN, 1979). No obstante, si esos subgrupos han de dividirse también en grupos de pruebas (p. ej., en relación con el CRV), es probable que la proporción de alumnos de cada subgrupo sea muy pequeña¹⁴⁹. Por último, es posible comparar escuelas, aunque esto requiere recabar datos de la LEA (véanse: John GRAY, *Open University* 1982, "Making More Sense Out of Examination Results", págs. 15-20; GRAY, MCPHERSON y RAFFE, 1982; PLEWIS y cols., 1981; SHIPMAN, 1979, págs. 91 -100).

Como dijimos antes, estos resultados sólo pueden considerarse como indicadores y, como toda la información sobre valoración del alumnado, no dicen nada a las escuelas sobre lo que no va bien. El diagnóstico debe provenir de los profesores. con la ayuda de terceros cuando convenga teniendo que resolver los problemas que se produzcan. No obstante, el análisis de los resultados de los exámenes puede convertirse en un catalizador

¹⁴⁸ Puede hallarse una puntuación total combinada suponiendo cierta equivalencia en cada una de las dos escalas. A continuación presentamos algunas posibilidades apuntadas por SHIPMAN (1979, pág. 85) y por John GRAY (Open university, 1982, pág. 12).

Nivel "O" GCE	A	B	C	D	E	(o D E) U
CSE		1	2	3	4	5 U
Punt., total	8	7	6	5	4	3 2 1 (Shipman)
Punt. total	7	6	5	4	3	2 1 O(Gray)

(PLEWIS y cols., 1981, pág. 41, dan otras posibilidades).

SHIPMAN sugiere también una puntuación total positiva y negativa que resulta útil a los profesores preocupados por reducir el fracaso al mínimo (pág. 86).

¹⁴⁹ como alternativa, pueden reducirse todos a una única franja de ingreso, de manera que habrá pocos alumnos en esa franja con quienes puedan compararse.

Enfoques cualitativos e idiográficos de los resultados del aprendizaje

La preocupación por lo que no marcha bien cuando las puntuaciones de los alumnos son inferiores a lo previsto nos devuelve a nuestra exposición anterior sobre la organización del aprendizaje. Si los profesores están preparados para enseñar escuchando (EASLEY y ZWOYER, sin fecha) y observando, es probable que consigan entender de algún modo el pensamiento del niño, lo que ayudará directamente a los alumnos y mejorará en general su enseñanza. Dijimos antes que esto no constituía el núcleo fundamental de este apartado, sin embargo, abre una vía alternativa. Hasta aquí nos hemos centrado en los resultados del aprendizaje; pero, ¿qué podemos decir sobre sus procesos? Nuestra revisión sobre la observación en el aula sólo se ha ocupado de ellos en términos globales: reflejando las preocupaciones que aparecen en la bibliografía respecto a los procesos del aula o del medio del aprendizaje (PARLETT y HAMILTON, 1972). La exposición sobre los resultados del aprendizaje constituye, en parte, un intento de corregir este desequilibrio. Sin embargo, como afirman KEMMIS, ATRIN y WRIGHT (1977, capítulo 7), esto se basa en una visión del rendimiento que hace hincapié en las asignaturas y en las conductas productos. Dicho con toda crudeza, valoramos las estructuras cognitivas de los alumnos (porque ése es el producto del aprendizaje) valorando hasta qué punto se ajustan a las estructuras de las asignaturas, dejando de lado lo que KEMMIS, ATKIN y WRIGHT denominan el “conocimiento íntimo” del alumno. No podemos ocuparnos aquí de su argumentación¹⁵⁰, pero, de acuerdo con quienes insisten en la observación directa del aprendizaje, que ya hemos mencionado, ésta hace hincapié en la estructura cognitiva del alumno y en cómo puede desarrollarse. Un enfoque de esta cuestión parte de la base de observar muy de cerca cómo interactúa el alumno con la tarea de aprendizaje con el fin de hacer inferencias sobre su estructura cognitiva. Otro consiste en la técnica de entrevista clínica utilizada por PIAGET (1929), a la que ya nos hemos referido¹⁵¹. Una consideración más detenida de estos enfoques exige el examen de las medidas cualitativas del aprendizaje. Los requisitos de este examen son exactamente los mismos de los métodos de valoración del alumnado que revisamos antes, aunque expresados en esta ocasión en términos de aprendizaje y no de la asignatura. Por tanto, se precisan:

1. Una descripción del aprendizaje.

2. Técnicas para evaluar el aprendizaje.

En relación con la primera, las propuestas de KEMMIS, ATKIN y WRIGHT (1977) resultan más interesantes que susceptibles de aplicación inmediata, sobre todo cuando la puesta en práctica de sus ideas se realiza en el aprendizaje asistido por ordenador y, en consecuencia, no se imita a las escuelas. No obstante, señalan la importancia de insistir en el aprendizaje del alumno, algo que algunos evaluadores “iluminados” perdieron de vista al hacer hincapié en el medio de aprendizaje. KEMMIS, ATKIN y WRIGHT (1977) aplicaron sus ideas en la evaluación del National Development Programme in Computer Assisted Learning, elaborando una tipología de las interacciones entre estudiante y ordenador. Describieron a continuación diversos proyectos de aprendizajes asistidos por ordenador, según el tipo de interacciones a que se traducían. Aunque defendían un enfoque idiográfico de esa evaluación, poniendo ejemplos de casos de interacciones especialmente satisfactorios (o insatisfactorios), en vez de resúmenes cuantitativos, utilizaban cuasi-estadísticos para describir los programas. Como es lógico, el uso del ordenador no sólo significa que todas las interacciones han de estar definidas con claridad y ser observables (mediante los datos introducidos en el ordenador y los resultados de salida del mismo), sino que el ordenador puede almacenarlos. Por supuesto, en una clase convencional, la situación es muy distinta.

Sin embargo, el principio de observación directa y cercana no es menos importante en la clase convencional y ARMSTRONG (1981) pone un ejemplo de ese enfoque. (Algunas observaciones suyas no tienen lugar en clase.) Estuvo observando a los niños durante un período considerable, tomando notas detalladas a diario que, junto con muestras de sus escritos, pinturas, etc., constituyeron un estudio del incremento de su conocimiento individual [NIXON, 1981, presenta un estudio de uno de estos niños]. ARMSTRONG prosiguió esta investigación, aunque como observador participante en el aula de otro profesor (de una escuela primaria), percatándose de las dificultades que encierra hacer esto en circunstancias normales. No obstante, su trabajo

¹⁵⁰ Tampoco estamos cualificados para evaluar su nivel en psicología cognitiva.

¹⁵¹ Hasta donde sabemos, se ha prestado poca atención a esto a efectos de evaluación. COOD (1981) hace una descripción muy general respecto a su uso. DAHLGREN y MARTON (1978) ponen un ejemplo de la utilización de las entrevistas para observar las estructuras de conocimiento de los estudiantes en la enseñanza superior, aunque MARTON ha ido adoptando un enfoque fenomenológico. Existe un creciente interés por investigar el aprendizaje de los estudiantes en el sector postescolar, pero se ha aplicado muy poco al mundo de la evaluación. En un nivel más práctico JAMES (1979) utilizó las entrevistas para examinar la “comprensión social de los alumnos en la evaluación de una asignatura sobre ciencias sociales, pero, como en el caso de DAHLGREN y MARTON no usó la técnica de entrevista clínica.

pone de manifiesto la necesidad de la investigación idiográfica, porque una persona sólo puede estudiar a unos pocos alumnos en un mismo período de tiempo. En este sentido, su enfoque difiere de los que hacen hincapié en la organización del aprendizaje, a quienes nos hemos referido. ARMSTRONG elaboró su punto de vista respecto a la evolución del alumno mediante un estudio minucioso y detallado, en vez de prestar atención a la actividad del profesor, como ocurre en muchas evaluaciones cualitativas (véase el Capítulo XI).

Entre otros ejemplos recientes de investigaciones sobre los resultados del aprendizaje utilizando enfoques cualitativos está el detallado trabajo de los investigadores sobre el aprendizaje de las ciencias en niños. El *Children's Learning in Solence Project* (CLISP), que tuvo como base la *Leeds University*, analizó en primer lugar las respuestas erróneas en ejercicios escritos de la APU (DRIVER, 1982 y 1983), estudiando después las concepciones erróneas de los niños en relación con un conjunto reducido de conceptos clave (p. ej.: nutrición de las plantas, energía, constitución de la materia —partículas—) pidiendo directamente a los alumnos que explicasen tales cuestiones. Hacia la misma época, diversos investigadores de la *Surrey University* de Nueva Zelanda se ocupaban de temas similares de un modo semejante (p. ej.: POPE y GILBERT, 1983; OSBORNE, BELL y GILBERT, 1983; WATTS y ZYLBERSTEIN, 1981). A todos ellos les interesaban las estructuras cognitivas de los alumnos y su ajuste con la teoría aceptada sobre la estructura de la ciencia y la enseñanza de la misma. Concluyeron que “lo que aprenden los niños a partir de la interacción con un fenómeno... no sólo depende de lo que abstraen de la situación, sino también de las construcciones mentales que aportan a la misma” (MILLAR y DRIVER, 1987).

Basándose en la perspectiva constructivista de la obra de psicólogos cognitivos como PIAGET, BRUNER y KELLY, interpretaron las pruebas de que disponían como indicadores de que las concepciones científicas erróneas de los niños se derivan de conjuntos de ideas, esquemas o representaciones mentales internas antecedentes que utilizan para interpretar e interactuar con situaciones nuevas. Los investigadores juzgaron que el ajuste de aquéllas con la forma que adopta la enseñanza de las ciencias no es bueno; tanto el antiguo enfoque respecto a los contenidos como las nuevas orientaciones centradas en los procesos suponen que la mente del niño es una *tabula rasa* en la que basta con inscribir el saber. La segunda fase del proyecto CLISP ha intentado abordar las consecuencias pedagógicas de estas conclusiones, ayudando a los profesores a desarrollar estrategias para “partir de donde se encuentran los niños”, pero tendiendo hacia una mejor comprensión científica. El enfoque es esencialmente psicológico.

En este ejemplo, la investigación básica fue dirigida por investigadores profesionales que utilizaron técnicas muy complejas, incluyendo entrevistas clínicas (PIAGET, 1929) y técnicas de rejilla (FRANSELLA y BANNISTER, 1977). A pesar de sus ventajas, en términos de tiempo, destrezas y apoyo, sus investigaciones estaban restringidas, por necesidad, a pequeñas muestras de alumnos y a algunos conceptos científicos. La posibilidad de que los profesores emprendan investigaciones similares es, por tanto, remota, a pesar de su valor potencial para planificar el *currículum* y la enseñanza. Las necesarias destrezas de investigación son considerables, pero el tiempo y las ocasiones para ponerlas en práctica aún son más difíciles de encontrar. Es posible que los profesores que organizan el aprendizaje con una orientación individual sean capaces de encontrar tiempo para hablar con los alumnos con cierto detenimiento; otra posibilidad sería aprovechar las sesiones de revisión individual que constituyen un elemento clave de los esquemas de registros de rendimiento.

La formulación de la normativa sobre registros de rendimiento del DES (1984d) hacía hincapié en que las escuelas tenían que modificar su funcionamiento interno con el fin de promover un diálogo “adecuado” entre los alumnos y todos los profesores que tuviesen contacto con ellos. En principio, se preveía que los procesos de registro diesen oportunidad a todos los profesores para revisar el aprendizaje de cada alumno con el fin de diagnosticar las dificultades y fijar los objetivos futuros, así como cumplir el requisito de recopilar un registro de experiencias y rendimiento. En la práctica, al menos hasta el momento de redactar este libro, cuando muchos esquemas de registros de rendimiento se encuentran aún en sus balbuceos, las sesiones de revisión individual en las que participen todos los profesores siguen siendo un objetivo a largo plazo. Las disposiciones formales internas que permitan la revisión se limitan en muchos casos a la formación de tutores, quienes están en situación de explorar el progreso general de los alumnos, pero no de discutir resultados específicos de aprendizaje en cada asignatura, que saldría fuera de su competencia (PRAISE, 1987). No cabe duda de que el diálogo sobre los aprendizajes específicos se realizara en las sesiones de revisión y registro, pero, hasta ahora, no tenemos noticia de que ningún profesor haya efectuado un análisis profundo de las estructuras cognitivas de los alumnos que pueda compararse con los ejemplos presentados.

Otras medidas de resultados

Hemos dedicado la mayor parte de este apartado a la información de valoración y, más en concreto, a la valoración del aprendizaje cognitivo. Aunque esto se justifica por su importancia en el interior de las escuelas, conviene considerar también otras medidas. Por tanto, revisaremos ahora la valoración y la medida no cognitivas, así como otras dos medidas de resultados: “permanencia” y “destino de quienes se van”. En el aprendizaje no cognitivo se incluyen aspectos como sentimientos, emociones, consciencia, valores, creencias y actitudes. Algunos de ellos pueden investigarse utilizando los métodos que señalamos en “Observación

indirecta', mediante entrevistas y cuestionarios, por ejemplo. Aquí nos ocupamos de los más duraderos, como las actitudes, que pueden, por tanto, medirse.¹⁵² Las interrelaciones de los aspectos no cognitivos son complejas. Por ejemplo, OPPENHEIM (1966, págs. 109-110) describe una serie de niveles, situando las creencias en el más superficial y las actitudes, valores y personalidad en niveles sucesivamente más profundos. Las interconexiones se producen tanto en cada nivel como entre niveles. Esta complejidad supone que no basta con una sola pregunta de un cuestionario para penetrar en la actitud subyacente. Aunque los aspectos no cognitivos son duraderos, también se aprenden, por lo que está justificado que la escuela los mida. Sin embargo, los profesores deben tener claras las conexiones entre las actitudes en cuestión y las formas en que la enseñanza las trata (BROWN, S., 1979). Por supuesto, existe un debate sobre si deben valorarse los resultados no cognitivos del aprendizaje. No obstante, dado que la evaluación curricular se ocupa del grupo y de la enseñanza, y no de los individuos, creemos que las cuestiones éticas al respecto son menos discutibles.

Sin embargo, para las escuelas, el problema fundamental consiste en *cómo* valorar estos aspectos del aprendizaje. Muchas técnicas utilizadas (p. ej., las escalas Likert, Thurstone y Guttman¹⁵³) constituyen más bien técnicas de investigación que requieren recursos para su construcción y validación. INGENKAMP (1977) es muy mordaz respecto a los juicios de los profesores en este campo:

Sin embargo, en la actualidad, nos hallamos en la incómoda situación en la que los profesores actúan como aficionados al emitir juicios en un campo cuyos objetivos suelen considerarse los más importantes. (Pág. 81.)

Considera necesario hacer operacionales los conceptos que evaluar, de manera que puedan expresarse en términos de conducta manifiesta. Como hemos señalado ya, BLACK y DOCKRELL (1980) demuestran que los profesores pueden hacerlo, poniendo el ejemplo del "esfuerzo", una característica habitual en la valoración no cognitiva de la escuela.¹⁵⁴

Las otras dos medidas de resultados de las que nos ocuparemos aquí: la "permanencia" y el "destino de quienes se van", las considera SHIPMAN (1979) como formas de evaluar los "aspectos más generales de la vida escolar". La "permanencia" puede considerarse como un voto de confianza en la escuela, aunque en estos tiempos de elevado desempleo juvenil (que puede obligar a los alumnos a volver a la escuela) y de programas gubernamentales de formación juvenil (que puede estimularlos a dejarla antes), la validez de la "permanencia" como indicador de "confianza" no está muy clara. Sin embargo, quizá convenga calcular los datos y SHIPMAN (1979, págs. 114-115) indica dos proporciones: sexto inferior/quinto del año anterior; sexto superior/quinto de dos años antes. Estas proporciones pueden compararse con las estadísticas nacionales. El "destino de los que se van" constituye por supuesto, una auténtica medida de resultados (1) y, si se utilizan las categorías del *Department of Employment*, los destinos de quienes abandonan la escuela pueden compararse con las estadísticas nacionales¹⁵⁵. Es obvio que esas medidas no representan una evaluación de la escuela en sí, sino de las normas de la escuela respecto a la ayuda que brinde a sus alumnos para encontrar trabajo¹⁵⁶. Una escuela utilizó una medida muy útil consistente en descubrir el destino de sus antiguos alumnos un año después de dejar el centro.

Conclusión

Sabemos que, al utilizar en este apartado mucho tiempo en dos tipos de medidas con referencia externa, es decir, referidas a normas ya dominios (criterios), nos hemos ocupado menos de las autorreferidas. Sin embargo, creemos que éstas son más importantes para la organización del aprendizaje individual que para la evaluación escolar, que pensamos está más relacionada con los grupos¹⁵⁷. Además, no hemos mencionado de manera explícita la evaluación del Mode III, que combina normas internas y externas. Lo hemos omitido por dos

¹⁵² OPPENHEIM (1966, pág. 105) define la actitud como "la tendencia a actuar o reaccionar de una manera determinada cuando alguien se enfrenta a determinados estímulos". Si no tuviese un carácter duradero, sería imposible medirla con alguna fiabilidad.

¹⁵³ OPENHEIM (1956) y THOMAS (sin fecha) los describen con cierto detalle y LEWIS (1974) hace una exposición más general de los mismos.

¹⁵⁴ WOOD y NAPHTALI (1975) recomiendan también la representación conductual de esas evaluaciones que, en el área no cognitiva categorizan en los aspectos motivacionales y afectivos. (En esta área existe una gran diversidad de categorías; véase también S. BROWN, 1979.)

¹⁵⁵ Así lo indica SHIPMAN (1979, pág. 119). Las categorías son: aprendizajes orientados al empleo; trabajos que generan cualificaciones profesionales reconocidas; empleos burocráticos; otros empleos (con o sin formación planificada).

¹⁵⁶ una escuela que conocemos bien entró en contacto con todos los empleadores locales para ver si podían garantizar puestos de trabajo para todos los alumnos que salieran de la escuela.

¹⁵⁷ Quizá el establecimiento de perspectivas comunes sobre el progreso individual sea una contradicción in terminis en todo caso, resultaría extremadamente difícil, como indicamos antes (nota 108) nadie comparte este punto de vista

razones. En primer lugar, las solicitudes del Mode III constituyen una pequeña parte de las peticiones de exámenes (TORRANCE, 1982). En segundo lugar, como el mismo TORRANCE (1982), lo consideramos muy importante para el desarrollo profesional, por lo que nos ocuparemos de él al considerar el «curriculum oficial». Se trata, por supuesto, de algo más que un instrumento para el desarrollo profesional, puesto que reúne enseñanza, aprendizaje y valoración; sin embargo, en el contexto de la evaluación curricular, creemos que su principal contribución se refiere al desarrollo profesional. Es más, NUTTALL (1982b) se muestra pesimista respecto a su futuro en cualquier esquema combinado 16-plus (en los que se combinarán los exámenes GCE y CSE).

Nuestros intentos de dar sentido al campo de la valoración del alumnado se traducen en dos reflexiones. En primer lugar; nos preguntamos por la ética de la afirmación, que suscriben muchos autores, de que uno de los objetivos importantes de la valoración del alumnado es la evaluación del curriculum, sin decir, al mismo tiempo, cómo puede realizarse.¹⁵⁸ Por ejemplo, nadie explica qué importancia tienen para la evaluación curricular los supuestos pertinentes para la calificación. En segundo lugar a la vista de las dificultades de utilizar la información de la valoración del alumnado, nos preguntamos por qué se considera ésta una técnica evidente. Después de todo, muchos evaluadores curriculares con dedicación completa iban a hacer grandes esfuerzos para construir instrumentos de valoración de los alumnos y es obvio que no se sienten muy felices con el uso de los resultados (véase la nota 137).

El curriculum oficial

Introducción

En el Capítulo III, estudiamos el tipo de revisión curricular que se defiende como forma de evaluación intrínseca que desarrollen las escuelas. Aunque terminábamos con una nota pesimista, dudando de la significación que pudiera revestir la implantación de dicha revisión (a causa de las limitaciones impuestas desde fuera, los derechos de los profesores para realizarla y su preparación para hacerlo), revisamos aquí las técnicas que podrían utilizarse. Teniendo presente el limitado margen de maniobra de que disponen los profesores, sostenemos que la revisión curricular constituye una práctica importante para el desarrollo del profesorado.¹⁵⁹ Aunque los docentes tienen muy poca experiencia sobre el conjunto del *curriculum* (aspecto al que ya aludimos en el Capítulo III), con independencia de que lleve o no, al menos en principio, a cambios espectaculares en el *curriculum* oficial, constituye una tarea fundamental. Por tanto, en este apartado, estudiaremos la revisión y planificación del *curriculum* (los procesos y clases de análisis que suponen), así como el análisis de los materiales curriculares. Al presentarlo como argumento que abarca un “campo en disputa”, carece de las características de una revisión, por lo que no se parece a otros apartados del capítulo.

Revisión y planificación del curriculum

Hemos combinado la “revisión” y la “planificación” porque las formulaciones procedentes del DES, de la HMI e, incluso, del *Munn Report* se refieren a ambos aspectos sin distinción. La “revisión” queda caracterizada por las siguientes citas:

Cada escuela deberá analizar sus objetivos, ponerlos por escrito y evaluar con regularidad en qué medida cumple el *curriculum* esos objetivos, tanto en el conjunto de la escuela como en relación con cada alumno.

(DES, 1981b, pág. 5.)

Sólo si la escuela formula con claridad los objetivos y un conjunto de criterios o lista de comprobación... será capaz de asegurarse... de que su concepto de la enseñanza secundaria es esencialmente bueno para todos los alumnos.

(DES, 1980b, pág. 13.)

Esta utilización de los objetivos o de las “áreas de experiencia” como lista de comprobación o conjunto de criterios se combina también con ideas de planificación:

Es una lista de comprobación [de áreas de experiencia].., para el análisis y la estructuración del *curriculum*.

(DES, 1977a, pág. 6.)

¹⁵⁸ SHIPMAN (1979 y 1983) es uno de los pocos autores que ha tratado esta cuestión desde el punto de vista de la escuela, aunque sólo en los términos más generales.

¹⁵⁹ WESTON (1979, pág. 225) señala que el diálogo sobre el *curriculum* es importante para ampliar la visión de los profesores sobre los problemas curriculares y como forma valiosa de formación permanente.

Cuando las escuelas planean su programa detallado de trabajo han de poder medir la adecuación de estos programas en relación con objetivos más específicos..

(DES, 1960b, pág. 3.)

El uso de los objetivos, etc., como criterios para la revisión y para la planificación es diferente, aunque, por supuesto, no constituyan procesos sin relación mutua. No obstante, las citas anteriores muestran que, tanto el DES como la HMI, los agrupan. Es obvio que lo que dicen estos documentos es menos importante que los tipos de actividades que estimulan y orientan, por lo que a continuación nos ocuparemos de ejemplos concretos de actividades de revisión.

El primero está tomado del esquema LEA-HMI al que nos referimos en el Capítulo III. Trataremos, en concreto, de la actividad realizada en el ámbito de la administración local de Cheshire, dado que es el mejor documentado e incluye una evaluación independiente. Esta administración pidió a sus escuelas que prestaran atención a cuatro áreas (una más que las propuestas en principio por la

HMI): educación del individuo; preparación para el mundo laboral; preparación respecto a la sociedad, y relaciones personales. Teniendo en cuenta estas cuatro áreas, la administración elaboró instrumentos que sirvieran de ayuda en el proceso de revisión (Cheshire Education Committee, 1981). Al no disponer aquí de espacio suficiente para examinar con detalle todos estos instrumentos, hemos hecho una selección. El elemento central del ejercicio, indicado en el Capítulo III, consistía en un análisis de las asignaturas mediante un modelo tridimensional respecto a la educación y el individuo (conceptos, destrezas y actitudes; progreso en esos tres aspectos durante los cinco años de escolaridad; relación entre las asignaturas y el currículum global). El instrumento El se ocupaba de las dos primeras dimensiones, pidiendo a cada departamento de asignatura que hiciera una lista de: sus metas; sus objetivos, en términos de conceptos, destrezas y actitudes y contenidos de conocimiento; su método de enseñanza, evaluación de los alumnos y evaluación curricular. Las listas resultantes se comprobaron en la escuela, analizándose después mediante determinadas estrategias como la siguiente:

Objetivos conceptuales

- 1) ¿se estudian los mismos conceptos en más de una asignatura durante el mismo o diferentes cursos?
¿Hasta qué punto están correlacionados? Acción: conjuntar las materias apropiadas.

(Cheshire Education Committee 1981, pág. 29.)

La tercera dimensión se investigó por medio de la clasificación de cada materia por el departamento y la posterior baremación de la contribución de su materia a las ocho áreas de experiencia propuestas en Curriculum 11-16 (DES, 1977a).¹⁶⁰ Se recomendó a las escuelas que analizaran sus resultados exponiendo las baremaciones en matrices que presentaran Las relaciones mayores o menores de cada asignatura con cada una de las áreas de experiencia. Esto llevaría a una "acción" que supondría la reunión de los departamentos para estudiar cómo correlacionar o reforzar sus esfuerzos.

La 'preparación para la sociedad' se examinó en dos partes: en primer lugar el marco general (normativa escolar; aspectos estructurales, p. ej.: consejo escolar; actividades educativas, p. ej.: festivos; participación comunitaria; relaciones con padres y governors), y en segundo, Las aportaciones de cada asignatura. El primer instrumento (S1) se utilizó en forma de cuestionario con respuestas abiertas y cortas¹⁶¹ 164, que se envió a los consejos de dirección, a los coordinadores de curso y a los cabezas de familia. Entre las cuestiones planteadas para el análisis estaba: "¿La cantidad y equilibrio de las actividades de extensión educativa satisfacen con eficacia las necesidades de los alumnos?" (*Cheshire Education Commillee*, 1981, pág. 57).

De las descripciones anteriores se desprende que la forma recomendada de análisis requería el uso de los instrumentos (listas de comprobación) en cuanto listas de criterios respecto a los que juzgar el *currículum* oficial¹⁶². Constituía también un importante elemento de recogida de datos, tanto en el nivel de escuela como en el de la LEA. Por desgracia, los detalles del trabajo realizado por otras administraciones ocales son insuficientes para juzgar si éste era el enfoque principal; no obstante, la revisión del conjunto del ejercicio (DES, 1981a, págs. 21-22) revela que el departamento de lengua de una escuela de Nottinghamshire utilizó determinadas metas como punto de partida para elaborar objetivos y unidades de aprendizaje.

Conviene tener en cuenta también el enfoque adoptado por las escuelas que no formaron parte del ejercicio HMI-LEA, aunque se inspiraron en los documentos de la HMI y del DES. La primera fase de La evaluación de

¹⁶⁰ Las escuelas recibieron un "Supplementary" to Curriculum 11-16' que analizaba brevemente cada una de las ocho áreas.

¹⁶¹ En el informe (Cheshire Education Commillee 1981) no se reprodujo el segundo instrumento (S2), por lo que no podemos comentar su formato.

¹⁶² Las listas de objetivos elaborados por los departamentos de las distintas asignaturas se *utilizaron* de forma similar, como indica el extracto sobre "objetivos conceptuales".

la *Sir Frank Markham School* (véase el Capítulo V) fue similar al ejercicio de Cheshire y utilizó ocho áreas y Listas de conceptos, etc., como criterios para juzgar su *curriculum*. Sin embargo, la combinación de este tipo de actividad con el estudio del *curriculum* tal como lo experimentaban los alumnos proporciona mayor utilidad al estudio, aunque algunas de las cuestiones planteadas (en la tercera fase) sobre el equilibrio de las ocho áreas, por ejemplo, son semejantes a las de la reevaluación curricular de Cheshire.

En el ejercicio HMI-LEA, la HMI recibió informes en los que se decía que se había previsto el uso de otros modelos, al lado de la lista de comprobación de las ocho áreas (DES, 1981a, pág. 68). Esto supone que las escuelas podrían no haber tenido en cuenta otras cuestiones importantes. La aceptación incondicional de las escuelas de la lista de comprobación puede ser también un síntoma de la dejadez respecto a los problemas fundamentales en los que se basaba. Los informes sobre la evidencia de lo ocurrido y de la percepción de los profesores acerca de ello (*Cheshire Education Committee*, 1981; DES, 1981a; *North West Educational Management Centre*, 1981; WRIGLEY, 1981) señalan tres aspectos importantes y relacionados de este ejercicio.

1. La importancia de estimular la consciencia

De los datos se desprende la impresión sobresaliente de que constituye un ejercicio apropiado y estimulante que generó importantes debates y discusiones sobre cada uno de los aspectos considerados (p. ej.: análisis de materia; ocho áreas; el mundo laboral). Puso de manifiesto áreas sobre las que el profesorado no había meditado mucho y ayudó a correlacionar los esfuerzos de los departamentos de cada materia.¹⁶³ No se produjeron desacuerdos en relación con la profundidad del análisis llevado a cabo, pero no cabe duda de que incrementó la consciencia, tanto de lo que hacían los compañeros¹⁶⁴, como de la naturaleza de las cuestiones que supone la revisión curricular. Es interesante señalar que aumentó la solidaridad, sobre todo dentro de los departamentos.

2. El uso de los instrumentos del tipo de las listas de comprobación

Gran parte del debate respecto a si el análisis fue lo bastante penetrante se centra en torno a los instrumentos utilizados en Cheshire. Por ejemplo, la clasificación de las materias en relación con las ocho áreas de experiencia se consideró carente de sentido; su significación radica en el debate que rodeó la tarea de ponerse de acuerdo en la clasificación. En realidad, el significado de las "áreas" constituyó un importante tema de debate, lo que puso de manifiesto una confusión y falta de acuerdo considerables. Se pensaba que, en general, las listas de comprobación y los cuestionarios proporcionaban información demasiado detallada, por lo que tenían un valor limitado en sí mismos.

3. Los resultados de la revisión

Algunos profesores pensaban que se trataba de un ejercicio artificial, que no reflejaba la actividad de clase ni las cuestiones de importancia, y la proporción de modificaciones curriculares de las que se ha informado parece discreta. Sin embargo, la LEA se benefició de la creación de un modelo curricular sobre el que puede basarse la normativa futura sobre la distribución del personal docente¹⁶⁵. Para las escuelas, la *misma* actividad de discusión resultó importante. En las conclusiones del "informe oficial" de la administración Local se insiste en ello, como muestra la siguiente cita:

Ahora podemos reconocer que el proceso de reevaluación ha llevado a un programa de formación permanente a gran escala... Para una escuela, esta forma de re-valoración es esencialmente un proceso de autoevaluación: no se propone dar respuestas a la gran cantidad de cuestiones que pueden surgir. Su mayor valor consiste en que proporciona un centro de atención para el profesorado.

(*Cheshire Education, Committee*, 1981, pág. 21.)

Por tanto, el mensaje más importante que se deriva de las prácticas en donde los objetivos o listas de comprobación de áreas de experiencia, etc., se utilizan como criterios para juzgar el *curriculum* consiste en que

¹⁶³ El estudio del *North West Educational Management Centre* (1981) caracteriza esto diciendo que el análisis de las asignaturas ayuda a los departamentos a organizar y justificar sus intenciones, así como a recordarles las intenciones perdidas en los contenidos.

¹⁶⁴ La importancia de aumentar el conocimiento del trabajo de sus colegas constituía también un objetivo relevante para los profesores que participaron, a iniciativa de la escuela, en una revisión curricular en la que un equipo formado por miembros de distintos claustros revisaron el trabajo de uno concreto (*Open University 1982*, Estudio de caso 2).

¹⁶⁵ El hecho de que WRIGLEY (1981) uno de los profesores que participaron en la experiencia, sólo esperase que la revisión diera algún resultado que le permitiera proteger el fundamento del *curriculum* en el restrictivo clima financiero del momento, da la medida en que este modelo dominaba los resultados.

proporcionan una valiosa oportunidad para que los profesores estudien cuestiones que habían olvidado hasta ese momento¹⁶⁶.

Sobre los datos y la discusión del uso de metas y objetivos para la *planificación* curricular, el otro elemento señalado en las publicaciones de la HMI y del DES, se ha trabajado mucho y aquí sólo nos basaremos de manera selectiva en la bibliografía¹⁶⁷. La principal crítica contra el enfoque de ‘metas y objetivos’ es que no presta suficiente atención al origen de las metas y objetivos (problema semejante al puesto de relieve por las escuelas, en el ejercicio HMI-LEA, que prestaba poca atención al fundamento de las ocho áreas de experiencia”). MACDONALD-ROSS (1973), en una de las críticas más completas de los objetivos conductuales como fundamento de la planificación del *currículum*, hace objeto de su primera crítica a su mismo origen. Considera tres tipos de respuestas a la pregunta por la procedencia de los objetivos: hay quienes evitan la cuestión¹⁶⁸; están los partidarios de la «línea dura» que dicen que los objetivos se derivan del análisis de la conducta, y también los partidarios de la «línea flexible», que dicen que los objetivos provienen de la consideración de la «sociedad», el «alumno» y la «asignatura», pero que no existen reglas para su derivación. MACDONALD-ROSS afirma que la postura de los partidarios de la «línea dura» es insostenible cuando se aplica a la enseñanza (p. ej., no se puede analizar la actuación de un doctor en economía para extraer objetivos económicos). La postura de los partidarios de la «línea flexible», parecida a la de la HMI y del DES¹⁶⁹, puede sostenerse sin dificultad, pero, como enfoque práctico, no proporciona muchas orientaciones sobre la forma de extraer los objetivos o de cómo proceder a partir de ellos para diseñar el *currículum*¹⁷⁰.

El uso de metas y objetivos como *prescripción* para a planificación no constituye la única dificultad, sino que WALKER (1975) dice que no *describen* lo que en realidad sucede cuando se planea. En su estudio sobre el proyecto curricular, descubrió que la planificación arranca de lo que él denominó *plataforma* de ideas:

“un sistema compartido de creencias y principios de trabajo” (pág. 108), y que es “metas” no predominaban en los contenidos de tales plataformas. En consecuencia, esto hace dudar de las “metas” como punto de partida de la revisión y como orientación para la planificación. A partir de los de las revisiones curriculares, hemos demostrado que, cuando se parte de formulaciones de objetivos o de listas de comprobación de ‘áreas de experiencia’, queda mucho por elaborar y explorar. SCHWAB (1973) decía que, en el caso de los objetivos, esto era esencial, para poner de manifiesto la riqueza de los valores encerrados en ellos.

El tratamiento de los valores plantea otro problema, ya señalado en el Capítulo III. El ejercicio HMI-LEA busca el consenso. Por ejemplo, los instrumentos utilizados en Cheshire piden que se llegue a un acuerdo para clasificar la aportación de una asignatura a las ocho áreas. Es probable que esta postura dejase de lado la importancia de las cuestiones de valor¹⁷¹ y el conflicto resultante¹⁷². LEWY (1977) suscita el problema del consenso en el contexto del resumen de los juicios de los “expertos” (p. ej., los profesores). Expone una técnica importante para considerar los juicios de valor de un grupo de personas: la técnica Delphi. Aunque se desarrolló para ayudar a los grupos a conseguir el consenso, lo hace de manera que procura preservar los puntos de vista divergentes. Evita los acuerdos falsos pidiendo a los individuos que, por ejemplo, elaboren por escrito una lista de objetivos para una escuela. Las respuestas se tabulan y se reparte un resumen. Cada individuo clasifica entonces los objetivos de acuerdo con sus prioridades personales; se resumen de nuevo y se distribuyen, poniendo de manifiesto la clasificación del grupo y la de cada individuo. Después, se les invita a que reconsideren sus clasificaciones. Cuando estas clasificaciones se apartan del consenso del grupo, se les invita a que expongan sus razones. El método se ha utilizado para descubrir los puntos de vista respecto a los objetivos que debe tener una *University School of Education* (CYPHERT y GANT, 1973). En relación con este caso, se dijo que la técnica podría utilizarse para moldear la opinión, aunque permite que los puntos de vista de las minorías queden reflejados, creando, por tanto, conciencia de ello. CYPHERT y GANT seleccionaron un

¹⁶⁶ O'HARE (1981) llega a una conclusión similar a partir de una encuesta realizada en las escuelas de Lancashire que participaron en el estudio de las orientaciones curriculares.

¹⁶⁷ En el capítulo III, mencionamos la reserva de LAWTON (1980) respecto al modelo de planificación que estaba perpetuando el HMI, pero los datos derivados de las experiencias de revisión no muestran que se estuviera realizando de ese modo, o sea, que la planificación basada en metas y objetivos no constituye la actividad predominante.

¹⁶⁸ En realidad, esto es lo que hicieron las escuelas que participaron en la revisión del HMI-LEA cuando aceptaron las ocho áreas sin discusión.

¹⁶⁹ En parte, porque no insisten en que los objetivos tengan que ser conductuales, es decir, especificados en términos de conductas finales de los alumnos.

¹⁷⁰ MACDONALD-ROSS (1973) considera también otras cuestiones, como el uso de objetivos conductuales para diseñar programas de enseñanza. Sin embargo, como enfoque, no sale mejor parado en el escrutinio que utiliza. Hay otras críticas más filosóficas que dicen que la planificación por objetivos se basa en una concepción errónea de la naturaleza de la planificación (véase SOCKETT, en *Open University*, 1976, unidades 16-18).

¹⁷¹ El capítulo III presentaba el punto de vista de REID acerca de que los enfoques que, desde el primer momento incluyen metas y objetivos, tratan de adoptar soluciones de procedimiento no adecuadas para los problemas curriculares.

¹⁷² Hay diferencias de opinión sobre la posibilidad de reconciliar puntos de vista ideológicos opuestos. En el capítulo III presentamos el enfoque de DARLING (1978), que dudaba de ello, pero LAWTON (1973, pag.30) afirma que un *curriculum* equilibrado puede derivarse de puntos de vista opuestos. Muchos autores que adoptan la perspectiva del consenso no han intentado analizar las incompatibilidades de las posturas en conflicto, dando por supuesto que cabe la reconciliación. WHITTY (1981) dice que el enfoque de consenso adoptado por LAWTON oscurece la naturaleza de los problemas curriculares que pretende solucionar.

grupo distinto y, al utilizar “cuestionarios”, pudieron escoger uno geográficamente disperso. RASP (1974) informa también de un estudio que utilizó esta técnica para averiguar la importancia relativa de ciertos objetivos (véase también HUNKINS, 1980, págs. 172-176, y BROOKS, 1979).

La técnica del ‘grupo nominal’ constituye un enfoque similar, aunque las personas consultadas se reúnen para realizar el proceso. Esto supone, en primer lugar, que los miembros del grupo elaboren, de forma individual, una lista de objetivos (sin consultas), por ejemplo. Se recopila después una lista general que los recoja todos, haciéndose las aclaraciones pertinentes al grupo. Los individuos clasifican estos objetivos (no hace falta que sean todos), mostrándose a continuación una lista general en la que aparezcan las ‘votaciones’. Puede haber una segunda vuelta de votaciones. HEGARTY (1977) afirma que ambas técnicas (Delphi y “Grupo nominal”) muestran la misma eficacia; la técnica Delphi requiere más trabajo de administración, pero ahorra tiempo a los participantes.

Sin embargo, debemos ser cautos respecto a lo que puede esperarse de cualquier técnica de revisión si los profesores estudian el *curriculum* sometidos a diversas presiones externas. Esta presiones pueden hacer ineficaz y, durante cierto tiempo, improbable una reevaluación fundamental.¹⁷³ Esto no quiere decir que los alumnos no se beneficien de que sus profesores exploren las cuestiones subyacentes al *curriculum*, pero este beneficio se deriva de una mayor conciencia adquirida por los profesores como, por ejemplo, en la experiencia de Cheshire y por tanto, de la formación y desarrollo de los profesores.¹⁷⁴ Si la naturaleza de la revisión curricular consiste precisamente en esto, podemos preguntarnos: ¿son adecuados los métodos defendidos por la HMI y el DES? Aunque reconocemos la importancia de las prácticas de revisión acometidas hasta ahora, de lo dicho se deduce que no. Dicha inadecuación no sólo surge del procedimiento adoptado, sino también de los “puntos de partida” o tipos de análisis empleados. En el Capítulo III ya advertimos esto en relación con *The School Curriculum* (DES, 1981 b), comparándolo, por ejemplo, con el *Munn Report* (SED, 1977a). Los documentos publicados por la HMI y el DES estimulando a las escuelas a revisar sus *curricula*, también las invita a considerar más de un tipo de análisis como herramientas de trabajo en la planificación curricular (DES, 1981b pág. 6). Se hace esta invitación como si todos los tipos de análisis fuesen equiparables, cosa que sometimos ya a examen en el Capítulo III. Aun en el caso de que se alcanzase un consenso respecto al *curriculum*, como afirma el *Munn Report* (SED, 1977^a), no puede hacerse en el nivel de cada escuela, salvo que quienes participen reconozcan los valores y supuestos subyacentes que sostienen ellos mismos y los demás. El hecho de partir de una lista de ocho áreas no estimula tal reconocimiento.

Un elemento que forma parte de cualquier consideración que se haga en la escuela de los problemas que subyacen al *curriculum* debe consistir en la adquisición de cierto conocimiento de las concepciones contrapuestas del *curriculum*, como se ofrece en las asignaturas curriculares¹⁷⁵, aunque reconocemos que esto seguirá constituyendo la preocupación de unos pocos¹⁷⁶. Para el profesorado de una escuela sigue siendo factible, en general, adoptar formas de trabajo que aumenten al máximo la preocupación por estas cuestiones básicas. Un comienzo modesto podría consistir en la consideración de los exámenes del *Mode III* que se realizan en las escuelas. Aunque, como ya dijimos (Capítulo VI), se trata de una actividad minoritaria (y quizá amenazada de extinción), da lugar a que los profesores participen en el pensamiento curricular, al menos en el nivel departamental (TORRANCE, 1982). No obstante, dejando de lado los detalles de esta posibilidad, nos ocuparemos a continuación de las formas más generales de tratar los problemas básicos.

Deliberación y negociación

Como indicamos antes, se considera que la deliberación constituye el enfoque necesario para explorar las cuestiones que subyacen al *curriculum* de manera que conduzcan a una acción práctica. Es más, se configura como un medio para la acción que ninguno de los enfoques de la revisión del *curriculum* hasta aquí considerados parece capaz de ofrecer. Constituye una actividad ordenada, y el análisis de las deliberaciones de un proyecto curricular que hace WALKER (1975) proporciona una visión muy útil del proceso. WALKER analizó

¹⁷³ Tenemos que manifestar nuestro escepticismo sobre el papel del profesor que encierra la visión del desarrollo curricular basado en la escuela de SKILBECK (*Open University* 1976, unidades 14-16, págs. 90-102), aunque, en la época en que se escribió (1975), quizá fuera difícil prever el clima que presidiría los últimos años setenta y los ochenta. El pequeño estudio de P. TAYLOR y cols. (1974) a principios de los setenta sentó las bases del pesimismo.

¹⁷⁴ WESTON (1979, pág. 259) lo considera importante aun en el contexto limitado que supone la innovación de superar las barreras entre asignaturas de un mismo departamento de ciencias. REY NOLDS y SKIBECK (1976) consideran que la participación en la planificación del *curriculum* es un medio para aclarar a propia postura de valor del profesor.

¹⁷⁵ HORTON (*Open University*, 1983b, unidad 7) adopta esta postura cuando estudia cuatro modelos curriculares: liberal, dual, comunitario y elegido por el alumno. Admite que deben considerarse como orientaciones que, de ser mejor conocidas, ampliarían las opciones disponibles para la planificación curricular. Esto nos llevaría a considerar la bibliografía sobre los estudios curriculares, en general, a la que nos sería imposible conceder la atención que merece, por lo que no nos ocuparemos de ella.

¹⁷⁶ Sería deseable que el profesorado de categoría superior lo considerara como una de sus responsabilidades, poniéndose al día en la bibliografía correspondiente a este campo.

transcripciones de la discusión del equipo del proyecto y descubrió tres estratos. Aquí nos centraremos en el segundo y tercero. Los movimientos deliberativos se producen en el segundo estrato y son de seis tipos:

1. Formulaciones de problemas: identificación de problemas.
2. Propuestas de líneas de acción: qué puede hacer el grupo del proyecto.
3. Argumentos en pro de propuestas o problemas concretos.
4. Argumentos en contra de propuestas o problemas concretos.
5. Ejemplos: ilustraciones o abstracciones presentadas en plan neutral.
6. Aclaraciones: modificaciones de los pasos anteriores.

Los tres tipos de mayor interés son los problemas, las propuestas y los argumentos. WALKER optó por explorar la naturaleza de los argumentos y, en concreto, el origen de los datos utilizados (el tercer estrato). De nuevo, identificó diversas categorías, susceptibles de agrupación por parejas:

1. Observacionales o judiciales: ¿el ítem de datos citado constituye una observación concreta, o es un juicio basado en la experiencia, en la sabiduría convencional o en la especulación?
 2. Externas o internas: si este ítem de datos es observacional, ¿proviene del proyecto, de sus miembros o de sus reuniones, o los datos proceden del medio exterior al proyecto?
 3. De primera mano o tomado de informes: los datos observacionales de este ítem, ¿fueron obtenidos por algún miembro del proyecto o recibidos de alguien no perteneciente al mismo, tomados de algún libro o de otra fuente externa?
 4. Obtenidos a propósito o por casualidad este ítem de datos observacionales, ¿fue objeto de búsqueda de uno o más miembros del proyecto u obtenido de manera fortuita?
- (WALKER, 1975, págs. 113-114¹⁷⁷)

Aplicando estas categorías a las transcripciones de los debates para determinar la proporción de uso de distintas fuentes de datos, descubrió:

que algo más de la mitad de los *argumentos* se basaban en juicios, en vez de fundarse en datos; que más o menos el mismo número de *argumentos* juzgados como observacionales surgió tanto del interior como del exterior del proyecto; que la inmensa mayoría de las observaciones concretas en las que se basaban los *argumentos* se realizaron por miembros del proyecto, y que la mayor parte de las observaciones se produjeron por casualidad y no a propósito (la cursiva es del autor).

(WALKER, 1975, págs. 121 -122.)

WALKER señala que el carácter *deseable* o no de estas proporciones es opinable. Es obvio que, si la discusión se centraba en los problemas que experimentaban los alumnos en una clase determinada, la falta de datos observacionales, o el hecho de que éstos hubiesen sido recogidos sólo por casualidad, constituiría un motivo de preocupación.

La lección para los profesores, que, como se indicó en el Capítulo V, se basan considerablemente en la discusión, consiste en que ésta puede ser ordenada. Ante esto, quizá sea interesante reflexionar sobre su naturaleza. La reflexión respecto a la cantidad de propuestas de acción consideradas, el número y fuerza de los argumentos en pro y en contra, el grado de juicio presente en los argumentos y el tipo de observaciones realizadas, por ejemplo, pueden ayudar a los profesores a perfeccionar sus deliberaciones. La autoconciencia respecto de este proceso parece artificial, pero no más que la tentativa de perfeccionar otras técnicas de evaluación, como los cuestionarios. Por desgracia, en este campo no hay muchas experiencias documentadas, pero sería interesante grabar y analizar algunas discusiones y que los miembros del profesorado que participaran reflexionasen sobre la naturaleza de la deliberación. Asimismo, podrían formalizarse las discusiones de manera que reflejaran los movimientos del segundo estrato de WALKER; esto garantizaría, por ejemplo que se considerara un número adecuado de propuestas de acción.¹⁷⁸

Por supuesto, todo ello supone que la forma de operar de los grupos de profesores en la escuela es semejante a la de los equipos de proyectos curriculares. WESTON (1979) dice que las escuelas no se parecen a los proyectos curriculares y que la negociación constituye una descripción mejor del pensamiento curricular. Propone un modelo sencillo de negociación en tres etapas:

1. Definición: definición y exposición de problemas.
2. Generación y configuración de propuestas: p. ej., elaboración de planes escritos o materiales.

¹⁷⁷ WALKER tiene en cuenta también quiénes son los objetos de los datos (es decir, internos, p. ej.: profesorado, materiales curriculares: externos, p. ej.: alumnos, escuelas, sociedad), pero éstos son específicos de contexto y serían diferentes si las discusiones estuvieran a cargo de los profesores de una escuela. No obstante, la división interno-externo es útil.

¹⁷⁸ Es posible que una escuela primaria, con un número reducido de profesores que constituyan un grupo unido, se parezca más a un grupo de proyecto curricular que una gran escuela secundaria como la estudiada por WESTON.

3.Consolidación: p. ej., consideración de Las consecuencias de las propuestas.

Cada una de estas etapas se solapa con las demás, de manera que, por ejemplo, mientras se exponen los problemas, se van configurando las propuestas. No obstante, éstas no deben formularse antes de efectuar una mínima consideración de los problemas.

La forma de tomar decisiones y, por tanto, los problemas de gestión relativos a la viabilidad ya la preocupación por los efectos, son características comunes de los debates que tienen lugar en las escuelas. A diferencia del proyecto de WALKER, las escuelas suelen tener una existencia permanente, una responsabilidad directa respecto a las decisiones y experimentan un ciclo de acontecimientos. (La vida de un proyecto puede considerarse lineal.) Quienes participan en una discusión (los directores de departamentos, por ejemplo) no suelen adoptar un papel neutral, sino que representan intereses (los de sus departamentos¹⁷⁹). Según nuestra opinión, la investigación de WESTON es importante y digna de estudio por su visión de las negociaciones curriculares¹⁸⁰. El propósito que se persigue al caracterizar estas negociaciones no sólo consiste en describirlas, sino en estimular una utilización más intencionada de las mismas. Un elemento clave de la idea de negociación de WESTON es la apreciación¹⁸¹ que recuerda la noción de deliberación de WALKER:

“es un proceso de argumentación razonada a la luz de toda la experiencia y la información disponibles que puedan aportar los participantes” (WESTON, 1979, página 195). Es más, ‘en sí mismo, constituye un proceso educativo que conduce a una nueva comprensión de quienes toman parte en él’ (páginas 195-196). Esto nos devuelve a nuestra advertencia anterior de que una preocupación fundamental de la revisión curricular debe ser el incremento de la comprensión de los profesores¹⁸². La consideración de “los problemas, los argumentos y las decisiones sobre cursos de acción como preocupaciones fundamentales de la revisión curricular, contrasta radicalmente con el diseño curricular basado tanto en el enfoque de los partidarios de la “línea dura” (cuyas prescripciones son difíciles de llevar a la práctica) como en el de los partidarios de la “línea flexible” (que se traduce en un ejercicio más bien vago y no siempre conectado con los problemas curriculares reales¹⁸³). Al final de este apartado volveremos brevemente sobre las ideas de la deliberación.¹⁸⁴

El análisis de los materiales curriculares

Los materiales curriculares de cualquier clase (fichas de trabajo, libros de texto, unidades didácticas completas) son una manifestación del *currículum* oficial. Implícita o explícitamente, engloban muchas decisiones curriculares, aunque no se hayan tomado de forma consciente. A pesar de los diferentes modos de utilizar los materiales, a juicio de los alumnos siguen siendo una formulación de las pretensiones curriculares. Como tales, merece la pena examinarlos. En un primer nivel, constituyen una prueba documental de intenciones, pero también una ‘agenda’ o estímulo para la acción en el aula. En consecuencia, ocupan una posición intermedia entre las formulaciones inactivas de propósitos y la actividad de clase. El examen de los materiales curriculares puede contemplarse de tres formas: como evaluación intrínseca; como la primera fase de una evaluación empírica, y como herramienta para la deliberación. A continuación, revisaremos cada una de ellas, centrándonos en el análisis de los materiales como evaluación intrínseca.

En el Capítulo VII nos ocupamos ya de la evaluación intrínseca; en el contexto del análisis de los materiales, puede considerarse como un intento de aclarar características importantes, como el tema de los valores. Esto suscita de inmediato la cuestión de cuáles son, en concreto, las características de las que debe ocuparse el análisis y si deben determinarse de antemano. BEN-PERETZ (1981) habla de marcos de referencia”, que

¹⁸⁰ Aunque el informe de la investigación sea largo (casi 300 páginas), pueden señalarse apartados en los que se exponen los principios fundamentales: la negociación y el currículum: páginas 40-44; el capítulo 11: págs. 489-198; caracterización del trabajo de la comisión: págs. 217-226;

reflexiones: págs. 257-260; Capítulo 16: págs. 264-271.

¹⁸¹ Expresión tomada de VICKERS (1965).

¹⁸² Somos conscientes de que pueden acusarnos de dar por supuesta la relación entre desarrollo profesional y cambio, cuestión revisada en el capítulo II, que hemos procurado mantener abierta. Sin embargo, creemos que es posible afirmar que, si los profesores no son conscientes de los problemas que subyacen a las opciones curriculares que puedan hacer, no podrán elegir con prudencia. Esta función educativa puede considerarse también como una primera fase. En todo caso, mantenemos que los problemas curriculares y las decisiones tendientes a su solución deben ocupar un lugar central, de manera que la preocupación por los mecanismos de deliberación y negociación sólo puede constituir una ayuda.

¹⁸³ Por supuesto, es una caricatura!

¹⁸⁴ Aunque hemos expuesto dos técnicas (la Delphi y la de grupo nominal), sabemos que hemos sido selectivos en nuestra exposición sobre la evaluación de las intenciones curriculares. En concreto, hemos pasado por alto las técnicas relacionadas con la distribución temporal. Por ejemplo, ZARPAGA y BATES (1980) describen un estudio informático de viabilidad de diversas formas de llevar a la práctica una determinada filosofía curricular. Se basa en las ideas de DAVIES (1969) y no se ocupa de cuestiones curriculares fundamentales, tratándose, más bien, del “eslabón final de la cadena de planificación curricular” (ZARPAGA y BATES, 1980, pág. 115). En cambio, WILCOX y EUSTACE (1980) afirman que su notación para describir el currículum permite la revisión. La notación genera índices a partir de los cuales puede describirse la agrupación por capacidades y la diferenciación curricular. En tal caso, el currículum puede revisarse en relación con criterios como el “equilibrio” y el “núcleo común”. Reconocen que llene irritaciones: por ejemplo, la notación no se ocupa de lo que está tras la denominación de una materia.

sirven como fuente de categorías para el análisis, y distingue entre marcos internos y externos. Los marcos internos de referencia provienen de los autores de los materiales y pueden constituir el fundamento para su elaboración (p. ej., una lista de objetivos). En este caso, la tarea consistirá en comprobar la coherencia entre el fundamento y los materiales o en cuestionar el fundamento mismo. Cuando los materiales se elaboran en la escuela, los evaluadores pueden ser también los autores. Los marcos externos de referencia pueden ser estrictos o singulares, como los sesgos sexuales y culturales, o globales. Los marcos globales pueden subdividirse en abiertos y cerrados. Un ejemplo de ellos será el que plantee si los materiales se adaptan a un modelo de desarrollo curricular, como, por ejemplo: ¿los materiales tienen una lista de objetivos?; ¿las actividades de aprendizaje se relacionan directamente con los objetivos?; ¿se incluyen las evaluaciones de los alumnos? (TYLER y KLEIN, 1976). En este caso, los criterios para juzgar están incluidos en el esquema, aunque los propios criterios sean controvertidos.

ERAUT, GOAD y SMITH (1975) nos proporcionan un ejemplo de marco de referencia (o esquema) global y abierto. Lo elaboraron para utilizarlo en el análisis de materiales curriculares publicados por editoriales, pero ha sido adaptado para usarlo con materiales elaborados en la escuela (Open University, 1981 a, Bloque 5). Tiene en cuenta cuatro aspectos de los materiales: objetivos (implícitos o explícitos); estrategia curricular (p. ej., actividades de aprendizaje, contenidos, evaluación); adecuación de los materiales (p. ej., lenguaje asequible); viabilidad práctica de utilización (p. ej., durabilidad, coste, etc.). El analista puede optar por evaluar los materiales de acuerdo con uno o varios aspectos; en este sentido, los cuatro aspectos operan como listas de comprobación. En realidad, la primera tarea consiste en establecer las cuestiones que se deben considerar en la evaluación, que pueden surgir al utilizar o inspeccionar los materiales.

Este esquema prevé varias fases tras la identificación de las cuestiones (que puede ampliarse o revisarse en cualquiera de las fases):

1. Recogida de datos pertinentes respecto a las cuestiones; por ejemplo, el análisis de los conceptos utilizados o los tipos de tareas incluidos (análisis documental); o datos sobre la forma de usar los materiales en clase (no es preciso recogerlos mediante observación directa, sino que basta con que el profesor la recuerde).
2. Planteamiento de argumentos a favor y en contra de las distintas posturas (por ejemplo, la sucesión de conceptos enseñados), en relación con cada uno de los cuatro aspectos (objetivos, estrategia, adecuación y viabilidad) identificados en las cuestiones.
3. Elaboración de juicios sobre la base de los argumentos.

McCORMICK (Open University 1981a, Bloque 5) plantea varios ejemplos de análisis de materiales curriculares elaborados en escuelas y publicados así como detalles de los esquemas apropiados¹⁸⁵.

En contraste con muchas técnicas de recogida de datos empíricos presentadas en este capítulo, los métodos de recogida de pruebas son muy diferentes en el caso de análisis de materiales. Una vez más, podemos señalar diversas técnicas:

1. Análisis de metas: elaboración de listas de propósitos y objetivos planteados explícitamente o inferidos, por ejemplo, a partir de las actividades y de la evaluación de los alumnos¹⁸⁶.
2. Análisis de contenido: abarca diversas actividades, por ejemplo, elaboración de listas de conceptos en orden de presentación, construcción de diagramas de conceptos que muestren interrelaciones, análisis de tareas para observar el nivel intelectual que implican, descubrir casos de sesgos en el material¹⁸⁷.
3. Análisis textual: requiere prestar una atención minuciosa al texto y a la naturaleza de los argumentos, ejemplos y conclusiones¹⁸⁸.
3. Legibilidad: es una medida de la dificultad del texto que se utiliza con tanta frecuencia que merece un apartado propio¹⁸⁹.

¹⁸⁵ El Dr. Michael ERAUT, de la Education Area, de la Universidad de Sussex, puede facilitar también ejemplos de análisis de materiales curriculares publicados. LEWY (1977) pone dos ejemplos de esquemas. Uno de ellos se refiere a los estudios sociales y considera los siguientes aspectos: fuentes pictóricas, expresiones y conceptos clave, adecuación de los datos, exactitud de los hechos, tendencias étnicas, niveles y tipos de pensamiento, valores y conflictos con ellos relacionados, formas de alimentar las distintas habilidades. El otro recuerda el marco de TYLER y KLEIN (1976). HULL y ADAMS (1981, págs. 94-96) presentan cuestiones detalladas en relación con los materiales de ciencias, explorando de qué manera se relaciona el estilo del material con los objetivos y el contenido del curso, el uso de los materiales y su viabilidad para los alumnos.

¹⁸⁶ FRASER (1977) presenta un enfoque respecto a este análisis que supone categorizar los objetivos y extraer los índices de frecuencia, aunque no en el contexto del análisis de los materiales curriculares.

¹⁸⁷ véase la exposición de esto en el capítulo sobre el análisis del contenido ("El curriculum observado").

¹⁸⁸ Se trata de un campo muy amplio que abarca el análisis de la retórica y otros instrumentos de presentación. ANDERSON (1981) aplica algunas de estas ideas al análisis de propuestas curriculares. en concreto, las planteadas en diversos proyectos curriculares. Para quienes tengan interés en revisar la bibliografía general relativa a este campo, puede consultarse Open University (1974a).

¹⁸⁹ La bibliografía al respecto es muy amplia: pueden verse, como referencias útiles: Open University (1974b, unidades 1-4); GILLILAND (1972); KLARE (1978); MOYLE (1978); HARRISON (1980).

Aunque gran parte de los datos se derivan del análisis documental, la consideración del modo de utilización de los materiales tiene igual importancia. Los profesores suelen recordar con detalle cómo los usan, lo que resulta especialmente útil cuando abarcan un período largo de enseñanza (p. ej., un trimestre) o se utilizan muchas veces, situaciones que podrían generar una cantidad enorme de datos observacionales directos. Estos utilizan la gran cantidad de pruebas y experiencia con que cuentan los profesores y que creemos debe examinarse antes de emprender ejercicios empíricos elaborados. En realidad, el análisis de los materiales puede considerarse como la primera etapa de un estudio empírico, dada la oportunidad de tener en cuenta nuevos aspectos a partir de su uso cotidiano. El producto principal de tal análisis sería una lista de cuestiones que se deben investigar empíricamente¹⁹⁰.

De nuevo, la deliberación

Al lector avisado no se le habrán escapado las semejanzas existentes entre las fases del análisis de los materiales y los ‘movimientos’ de la deliberación. Ambos se centran en la aclaración de problemas o cuestiones, y ambos utilizan la argumentación. Los argumentos se basan en los datos: en el caso del análisis de materiales, se extraen de los materiales y de su utilización. Se basan, asimismo, de forma sistemática, en la maestría de los profesores y buscan otras fuentes de evidencia y conocimiento.

El esquema propuesto por ERAUT, GOAD y SMITH (1975) preveía que los juicios correspondientes a la evaluación de los materiales se desarrollaran en una fase independiente de decisión. En ésta, se sopesarían los argumentos y la viabilidad, y se tendrían en cuenta las consecuencias de diversos cursos de acción. La semejanza con la deliberación ofrece la posibilidad de utilizar el análisis de los materiales curriculares como instrumento sobre el que basar el proceso de deliberación¹⁹¹. Como es lógico, los materiales curriculares no siempre están en el punto de mira de las discusiones sobre el currículum, pero pueden constituir un buen punto de partida para que los profesores desarrollen y “practiquen” la deliberación. Un grupo de docentes no tiene por qué dedicarse sólo a considerar los objetivos y estrategias implícitas en los materiales, sino también la forma de utilizarlos, relacionando, en consecuencia, sus creencias y valores subyacentes con la práctica de clase. Si la revisión o sustitución de materiales curriculares constituye el núcleo de la discusión, es evidente su relación directa con la acción.

El contexto curricular

introducción

Hasta ahora, en este capítulo, hemos revisado técnicas que pueden utilizarse para evaluar distintos aspectos del currículum: transacciones, resultados e intenciones. Estos aspectos abarcan la actividad fundamental de enseñanza y aprendizaje en la escuela. Pero la enseñanza y el aprendizaje tienen lugar en un contexto escolar que incluye un edificio, una organización y una minisociedad (o conjunto de subgrupos). A su vez, la escuela es una parte de una sociedad mayor. Los evaluadores iluminativos caracterizan este contexto mediante la expresión “medio de aprendizaje” (PARLETT y HAMILTON, 1972), que engloba los factores físicos y los sociales. Por otra parte, los sociólogos se refieren a la cultura’ (o subculturas) de la escuela, aludiendo a las normas compartidas, valores comunes, conocimientos y acuerdos que unen a quienes están en ella (MUSGROVE, 1973, pág. 165). Hay varias razones por las que la evaluación debe tener en cuenta este contexto:

1. La definición amplia de currículum que hemos adoptado nos recuerda un extenso conjunto de aspectos de la vida escolar.
2. Aunque adoptemos una visión más estricta del currículum, otorgando mayor importancia a los resultados, por ejemplo, hay pruebas, aunque discutibles (RUTTER y cols., 1979), de que los factores contextuales (p. ej., la participación del profesorado en las decisiones) influyen en los resultados de los alumnos (p. ej., el éxito académico).

¹⁹⁰ Este enfoque se ha utilizado en una evaluación de un curso de la Open University (P970: “Governing Schools”) en el que las intenciones de los diseñadores del currículum son explícitas porque exponen los principales “mensajes de aprendizaje”. Véase SOMEKH (1980) y Open University (1981. Bloque 5, Apéndice 4), en relación con evaluaciones que combinan el análisis de los materiales curriculares y los métodos de observación en clase.

¹⁹¹ No parece que la obra de WALKER influyera en absoluto en ERAUT, GOAD y SMITH (1975), sino que estos autores llegaron a su formulación de manera independiente. Debemos reconocer también que hemos hecho hincapié en ciertos aspectos del análisis de los materiales que resaltan las semejanzas con la deliberación.

3. En sentido más pragmático, muchas de las listas de comprobación que se utilizan en los esquemas de las LEA (y en la certificación estadounidense) abarcan diversos aspectos de las escuelas y de sus alrededores. Si revisamos las listas de comprobación utilizadas en los esquemas de evaluación de las LEA, observaremos que se ocupan de las condiciones antecedentes observadas¹⁹², empleando la terminología de STAKE (vease el Capítulo VII). Sin embargo, lo hacen así porque quienes elaboraron las listas pensaban que, para mejorar la calidad de la enseñanza, era importante considerar, por ejemplo, el estado de mantenimiento de los edificios escolares¹⁹³. Como estos tipos de listas reflejan el pensamiento colectivo de un conjunto de profesionales, conviene examinar los aspectos de las escuelas que se contemplan en ellas. Hemos seleccionado diversos ítems de una serie de listas que (a nuestro juicio¹⁹⁴) se relacionan con el contexto del aprendizaje. Señalamos aquellas en las que aparece cada ítem.

Clave de las listas:

1. ILEA (1977).
2. JACKSON y HAYTER (1978).
3. NSSE (1978).
4. *Schools Council* (1982).
5. NCA (1980a).

(Las notas se refieren a calificaciones basadas en la categorización de los ítems de las listas de comprobación).

<i>Aspectos que abarcan</i>	<i>Fuente</i>	
Ambiente de	la	escuela ¹⁹⁵
a. limpieza, exposiciones, mantenimiento y ruido		1
b. medios, recursos y financiación	5,3	
Clima de la escuela ¹⁹⁶		
Profesorado	1,3,4	
a. estructuras y responsabilidades	1,2,4	
b. decisiones		
c. desarrollo	1-5	
d. planificación		(curriculum) ¹⁹⁷
Alumnos		
a. funciones internas de la escuela: cómo se les recibe		
servicios de bienestar y psicológicos; absentismo	1,2	
b. características no relacionadas	con	la escuela ¹⁹⁸
p.ej.: origen étnico	3	
270		
<i>Relaciones entre profesores y alumnos</i> ¹⁹⁹	4	
<i>Escuela y comunidad</i>		
a. participación de los padres en la escuela, relaciones entre padres y profesores, uso		
comunitario de los recursos	1-5	
b. composición de la comunidad local	3	

¹⁹² véase a clasificación de las preguntas. etc., de CLIFT (*Open University*, 1982, Bloque 2, Primera parte, pág. 47) en: "Keeping the School Under Review" (ILEA, 1977).

¹⁹³ RUTTER y cols. (1979) opinan que esto no diferencia las medidas de resultados de las escuelas.

¹⁹⁴ puede que, a este respecto, hagamos que parezcan decisiones arbitrarias, pero sólo tratamos de ilustrar el ámbito de temas abarcados.

¹⁹⁵ En contraste con los aspectos que hemos recogido en el clima de la escuela", aquí nos ocupamos de los aspectos físicos.

¹⁹⁶ Es probable que muchas cuestiones o ítems de las listas de comprobación se refieran al clima', por ejemplo premios y castigos' (ILEA, 1977), pero sólo NCA (1980^a) hace explícita referencia a la cuestión.

¹⁹⁷ ILEA (1977) incluye esta cuestión en el epígrafe "autoevaluación departamental".

¹⁹⁸ ILEA (1977) incluye una indicación de "comidas libres", que puede tomarse como indicador de clase social.

¹⁹⁹ NCA (1980^a) las señala indirectamente a través del clima escolar, LEA (1977) supone las relaciones al tratar los aspectos del "Alumno".

Es útil comparar esta lista con la extraída de RuUER y cols. (1979)²⁰⁰, que, a pesar de sus defectos (véase la nota 4 del Capítulo VI), no se elaboró a partir del punto de vista concordante de profesionales, sino de investigaciones anteriores y de correlaciones estadísticas sobre los procesos escolares.

1. Aspectos académicos: tareas para casa y comprobación de las mismas; previsiones del profesor; muestras del trabajo del alumno; utilización de la biblioteca; planificación de las clases del grupo de profesores.
2. Premios y castigos: alabanzas en asambleas; premios deportivos; exposiciones de los trabajos del alumno.
3. Cualidades del alumno: condiciones de trabajo; limpieza de las clases; posibilidades para plantear a los profesores problemas personales.
4. Responsabilidades y participación (de los alumnos): intervención en asambleas escolares; cuestaciones de caridad.
5. Estabilidad de los grupos de trabajo y de amigos: permanencia del grupo de compañeros.
6. Organización del profesorado: planificación en grupo; comprobación del trabajo del profesorado y puntualidad; moral elevada y rutinas de personas que trabajan juntas en armonía; decisiones en el nivel de consejo de dirección con representación de los puntos de vista de los demás profesores.

Además, sobre la base de estudios anteriores, investigaron las relaciones entre sus medidas de resultados y las “características físicas y administrativas” y las “influencias ecológicas”²⁰¹. De éstas, sólo las ‘influencias ecológicas’ diferenciaban significativamente las escuelas de su muestra y, en relación con ellas, sólo resultaba significativo el “equilibrio de nuevos ingresos”. No obstante, establecieron una correlación entre el equilibrio académico (es decir, la proporción de los distintos niveles de capacidad) en una escuela y los resultados académicos, aunque, por ejemplo, sólo tuvieron en cuenta el *efecto de grupo* de una proporción elevada de niños capaces (véase la nota 201).

Nuestro propósito al examinar estos ítems, tanto de las listas de comprobación como del estudio de RUTTER, consiste en demostrar que es más conveniente evaluar el conjunto de aspectos que presenta una escuela que la actividad principal de enseñanza y aprendizaje. El resultado es desalentador y, sobre todo en el caso de las listas de comprobación, hasta cierto punto sin forma²⁰². Esto nos recuerda el ruego de PARSON, al que nos referimos en el Capítulo VII, de que los evaluadores tuviesen más en cuenta el papel de la teoría. Sin embargo, sería erróneo suponer que quienes preparan las listas de comprobación para la evaluación no tienen en cuenta la teoría. Por ejemplo, todas las listas antes mencionadas hacen referencia a las relaciones entre escuelas y padres, aspecto que consideraban importante gran parte de las primeras investigaciones británicas de sociología de la educación. Además, la preocupación por la toma de decisiones surge de las ideas relativas a la naturaleza de las organizaciones y a la forma de gestionarlas. Es fácil que la mayor parte de los profesores posea una serie de conceptos para pensar sobre la escuela como organización, por ejemplo: toma de decisiones; comunicaciones; burocracia, rol, jerarquía, poder; subgrupos de alumnos. Ante la evaluación del contexto curricular, creemos importante conocer la teoría que subyace a algunos de estos conceptos²⁰³. Como señala DAVIES (1973), en relación con el análisis de organización de las instituciones de enseñanza: la especificación de los problemas de investigación se basa en el marco conceptual en el que ha sido concebida” (pág. 253)²⁰⁴. Creemos que esto se aplica de igual modo a los problemas de evaluación porque, con independencia de cómo surjan, su investigación se basa en la forma de concebirlas que, a su vez, se fundamenta en algún marco conceptual (aunque sea incompleto o implícito). Como veremos, en la investigación etnográfica basada en el interaccionismo simbólico, no todas las teorías suponen una especificación rígida de los hechos y aspectos de la escuela (p. ej.: toma de decisiones; roles del profesor) que hayan de tenerse en cuenta.

²⁰⁰ Sobre la base de lo que se correlaciona significativamente (en términos estadísticos) con las medidas de resultados: conducta, delincuencia, rendimiento académico y asistencia. La lista se presenta con sus propios encabezamientos y no con los extraídos de las listas de comprobación.

²⁰¹ ‘características físicas y administrativas’ incluyen los años que tienen los edificios, el número de recintos y la organización interna (tanto de atención al alumno como académica). Las influencias ecológicas se refieren a aspectos externos a la escuela: área de recepción de alumnos, balance de nuevos ingresos y elección de los padres. Se investigaron estos aspectos para ver si se relacionaban con alguna medida general de los procesos de la escuela, demostrándose que, en el caso de las características ecológicas, estas variables carecían de efecto directo sobre los individuos. En otras palabras, aunque la capacidad en el momento del ingreso afecta a los resultados individuales, en esta parte del estudio, RUTTER y sus colaboradores sólo se ocuparon de un efecto de grupo. por ejemplo, de la elevada proporción de niños de baja capacidad en una escuela.

²⁰² La controversia en relación con el estudio de RUTTER (Capítulo vi nota 4. de esta obra) supone que debe tratarse con cierta cautela la forma establecida mediante las seis áreas de procesos internos de la escuela.

²⁰³ Advertencia que uno de nosotros ha hecho en otro lugar (JAMES y EBBUTT 1981, pág. 92) Por supuesto, puede afirmarse que, al considerar a evaluación de la enseñanza y el aprendizaje (intenciones del *currículum*, transacciones y resultados), llene igual importancia la teoría. No obstante, hemos dejado esta cuestión para el Capítulo XI.

²⁰⁴ Con una línea de pensamiento similar, LAMBERT. BULLOCK y MILLHAM (1973, pág. 312) afirman que el análisis de la organización es útil porque es más probable que los descubrimientos objetivos se deriven de un marco teórico válido que de investigadores que no tengan en cuenta los valores’. Por desgracia, no desarrollan la cuestión.

Dada la variedad de teorías y, según la opinión de DAVIES (1973), la falta de una teoría de las organizaciones que goce de aceptación generalizada, cualquier exposición de posibles teorías será muy selectiva²⁰⁵. Las teorías que consideraremos se ocupan de los aspectos internos de las escuelas; el contexto curricular externo es más difícil de tratar, aunque también nos ocuparemos brevemente de él más adelante.

Teorías de las escuelas como organizaciones

Vamos a revisar teorías procedentes de dos tradiciones: de gestión y sociológica, pues ambas influyen en la teoría de las organizaciones²⁰⁶. Los enfoques de gestión de las teorías se consideran normativos (DAVIES, 1973, pág. 252) y se centran en la estructura y funciones formales de las organizaciones. Un enfoque de este tipo se ocupa de la decisión y del liderazgo. SILVERMAN (1970, pág. 204) describe la racionalidad de los sistemas de decisión de acuerdo con tres elementos: un conjunto ordenado de preferencias; procedimientos para poner de manifiesto los posibles cursos de acción, y la capacidad de elegir entre las preferencias. Recordará el lector que utilizamos esto mismo para caracterizar el proceso de deliberación expuesto en el apartado dedicado al *Curriculum oficial*. Por tanto, este planteamiento deja de lado el hecho de que quienes “pueblan” la organización aportan motivos y significados diferentes de los del sistema formal. Un ejemplo de esto lo tenemos en la negociación respecto a la categoría de los departamentos, en vez de sobre la planificación del *curriculum*, mencionada por HURMAN (1978) (véase el Capítulo III). COHEN y MANION (1981, pág. 361) identifican la versión del liderazgo y de decisión que llaman modelo B, basada en la tradición de las relaciones humanas en la gestión. El modelo B hace hincapié en la decisión compartida y en pautas horizontales, y no verticales, de autoridad, democracia, participación, armonía y resolución de conflictos²⁰⁷. Este tipo de enfoque se utiliza en el análisis de WESTON (1979) de los problemas de participación y consulta para el *Heads of Department Committee* (véase *Curriculum oficial*).

La consideración del “clima de la organización” también tiene historia en los enfoques de gestión. Los climas escolares pueden caracterizarse, por ejemplo, en términos de liderazgo, autonomía, satisfacción en el trabajo (véase ZAK, 1981) o de insistencia en la producción (la supervisión inmediata de los profesores de parte de los directores), consideración (los intentos de los directores de tratar con humanidad a los docentes) y moral del profesorado (véase COHEN y MANION, 1981). En nuestro caso, dejamos de lado la organización formal para revisar algunos de los sistemas sociales que existen *en la escuela (sistemas informales)*.

La consideración de la escuela como “sistema abierto” proporciona un marco para considerar tales sistemas (o subsistemas) sociales, de los que algunos pueden ser informales (p. ej., las camarillas de profesores). Por ejemplo, puede considerarse que los sistemas abiertos tienen entradas, procesos y salidas. El análisis de los resultados de exámenes de John GRAY (“Resultados del *curricu/unt*”), en el que se medía la eficacia del sistema escolar, utiliza una versión tosca de este marco. Los sistemas reúnen una serie de características²⁰⁸, pero, por ahora, sólo nos ocuparemos de los “tres recordatorios” de áreas desatendidas de DAVIES (1973):

1. Relaciones externas y mantenimiento de límites: este recordatorio llama la atención de las escuelas que tienen que enfrentarse con presiones diversas y mutuamente opuestas procedentes del exterior, bien reconciliándolas, bien elaborando instrumentos que sirvan para aislarlas respecto de algunas de las presiones.
2. Organización formal e informal: este recordatorio nos alerta a tener en cuenta que la relación entre ambas organizaciones es compleja y nos recomienda no dar por supuesto que la organización informal sea un resultado de la organización formal (volveremos a ocuparnos de la organización informal cuando revisemos las teorías sociológicas).
3. Resultados (*output*) y automantenimiento: esto nos recuerda que “gran parte de la actividad que se desarrolla en una organización debe orientar-se a reparar y mantener por cuenta propia en vez de a producir un resultado explícito y oficial”²⁰⁹ (DAVIES, 1973, pág. 262).

²⁰⁵ Véase en Brian DAVIES (1951) una revisión de la bibliografía sobre las escuelas como organizaciones. No obstante, tenemos que advertir a los lectores no acostumbrados a la bibliografía sobre sociología y análisis de sistemas que este artículo es difícil de asimilar. Pueden consultarse resúmenes de la teoría de las organizaciones en: PUGR, HICKSON y HINNINGS (1971). y SILVERMAN (1970).

²⁰⁶ Brian DAVIES (1931, págs. 51-54) identifica tres líneas, la primera de las cuales procede de la administración educativa, mientras que la segunda y la tercera provienen de la sociología (“estudio institucional singular” y “clasificación y selección”). Reconoce que las tres se solapan, cosa que parece cierta en relación con muchas teorías de las organizaciones; por eso hablamos a menudo de “enfoques” y no de teorías”.

²⁰⁷ COHEN y MANION (1981) describen también un modelo profesional que reconoce fas organizaciones colegiales cuyas decisiones se adoptan en los niveles adecuados, de manera que, por ejemplo la autonomía del profesor en el aula no se ve afectada. No obstante, esto deja en pie el problema de la autoridad legal del director escolar.

²⁰⁸ Por ejemplo, mantienen estados estables autorreguladores

²⁰⁹ DAVIES continúa afirmando que es tan reducido el acuerdo sobre los objetivos de las escuelas que es imposible distinguir los dos. Esta descripción de los sistemas sociotécnicos, con adiciones, está tomada de SILVERMAN

El estudio clásico de RICHARDSON (1973) utilizaba la teoría de sistemas como una parte del fundamento teórico de la investigación. Sin embargo, hacía uso de la idea de “sistema sociotécnico” (aunque no empleara el término), que hace hincapié en las interrelaciones de la tecnología (p. ej., características especiales de la asignatura y del método docente), el ambiente de la escuela, los sentimientos de los participantes (p. ej., los objetivos de profesores y alumnos) y la forma de la organización²¹⁰. El problema fundamental radica en la descripción de la tarea primordial de la escuela, cuando se la considera como proceso de conversión, es decir, cuando los niños ingresan (input), experimentan procesos de conversión de crecimiento y aprendizaje (proceso), abandonándola como personas maduras (output) (RICHARDSON, 1973, págs. 17-18). Su argumentación es demasiado compleja para exponerla aquí, pero DAVIES (1973,1981) afirma que este enfoque ofrece la posibilidad de relacionar el sistema formal (que aboca a la tarea) con el informal (sensible²¹¹), así como de centrarse sobre el trabajo que se realiza en las escuelas. Aunque las ideas que se contienen en los enfoques sistémicos parecen muy alejadas de las escuelas “en las que se trabaja aquesse conocen”, debemos recordar que alguna de esas ideas están, sin embargo, presentes en nuestro pensamiento. Por ejemplo, diferenciamos el sistema de la escuela según los grupos de capacidades, grupos de edad, atención personal y sistemas (u organizaciones) académicos, aulas, actividades intra y extraescolares y departamentos de asignaturas²¹². El enfoque sistémico trata de poner algún orden en estos temas y proporciona una forma de pensar sobre las partes (o subsistemas), sin perder de vista el conjunto.

Los enfoques de gestión de la reflexión sobre las escuelas como organizaciones han conducido a métodos de evaluación, aunque no exista una conexión directa entre las teorías mencionadas y los enfoques de evaluación que exponemos a continuación²¹³. En primer lugar, se han elaborado diversas medidas del clima de la organización, que abarcan cuestiones como el control institucional (p. ej., el grado de supervisión de los movimientos de los niños y la medida y capacidad de la escuela para limitar y restringir la conducta de los alumnos), y la percepción que alumno y profesor tienen de su propia conducta y de la del otro. Se basan en instrumentos de tipo cuestionario elaborados y validados con minuciosidad (véase COHEN, 1976, págs. 276-278). En segundo lugar pueden llevarse a cabo definiciones y análisis de roles. Por ejemplo, SHIPMAN (197g, págs. 131-135) presenta un enfoque que requiere una descripción cuidadosa de los distintos papeles desempeñados por los profesores, así como una revisión regular de los mismos. Aparecen de nuevo los instrumentos de tipo cuestionario, utilizando escalas Likert, para definir los roles y las expectativas de rol (COHEN, 1976, páginas 265-273). En tercer lugar, relacionada con la concepción de revisión de roles de SHIPMAN, aparece la evaluación del profesorado. Se refiere ésta a una revisión de cada docente a cargo de uno o varios miembros de la dirección. Abarca aspectos como la actuación del profesor en clase (aspectos positivos y negativos), problemas surgidos en el departamento (a en la escuela, si se trata de una pequeña escuela primaria), necesidades de formación permanente u otras relacionadas con el desarrollo profesional, y planes y previsiones de carrera^{214 215}.

La sociología, que utiliza el concepto de “subcultura” de alumnos y de profesores, refleja también el interés por los sistemas informales dentro de la organización. MUSGROVE (1973) señala también la importancia fundamental que tiene la cuestión del poder para los sociólogos. Se considera que la investigación de las subculturas de la escuela ayuda a los profesores que desean ‘comprender y, quizá, influir en la motivación de [sus] alumnos’ (MUSGROVE, 1973, pág. 168). El interés de los sociólogos por el “poder” les lleva también a ocuparse de cuestiones como:

la estructura de autoridad, la forma y el fundamento de la distribución de tareas, la influencia de los grupos de presión que existen dentro y fuera de la escuela, el tipo de autoridad del director [profesor] y en qué sentido difiere de la autoridad de... [los delegados], la división y grupos de alumnos y profesores basados en la edad y en la antigüedad, la estratificación formal, la vinculación afectiva de los sujetos o el aislamiento o cercanía entre ellos.

²¹⁰ Esta descripción de los sistemas sociotécnicos, con adiciones, esta tomada de SILVERMAN(1970, pág. 109) y no del informe de RICHARDSON. Aunque tengan semejanzas, RICHARDSON se basa explícitamente en un conjunto de teorías.

²¹¹ El sistema o grupo sensible, concepto utilizado por RICHARDSON (1973, p.ej.: págs. 22-23 es el grupo del que los individuos de una organización extraen la fuerza emocional para realizar su “trabajo”, una persona puede pertenecer a más de un grupo sensible, no necesariamente relacionado con la tarea, que puede existir fuera de los límites de la escuela. RICHARDSON pone el ejemplo de un profesor que pertenezca a una asociación científica.

²¹² Están tomados de RICHARDSON (1973, págs. 21-22).

²¹³ ¡No es sorprendente que el enfoque de los sistemas sociotécnicos no haya dado lugar a métodos de evaluación!

²¹⁴ Este tipo de revisión se utilizó, por ejemplo, en la Sir Frank Markham School (véase el capítulo V, de esta obra) y en la Ernulf Community School (JAMES, 1982, págs. 47-48).

²¹⁵ Existen otras medidas para la adopción de decisiones, jerarquías y control, respectivamente [véase COHEN (1976), págs. 284-287, 295-298 y 299-303]. La mayor parte de los métodos de evaluación que arrancan de la tradición de gestión son prescriptivos. Así se definen, por ejemplo, un clima escolar favorable o una estrategia de decisión más sensible.

El estudio de LACEY (1970, págs. 165-169) desarrolla cómo se confiere categoría al profesorado mediante la antigüedad y el efecto de la organización formal e informal del mismo. El estudio de los sistemas informales en las instituciones educativas constituye la preocupación fundamental de LAMBERT y sus colaboradores (1973), en concreto, los sistemas sociales informales de los niños. Definen un sistema informal como “la pauta de normas, valores y relaciones no prescrita en sentido estructural ni normativo por los objetivos oficiales de la organización, pero que puede influir en esos objetivos” (LAMBERT, BULLOCK y MILLHAM, 1973, pág. 298). Esos sistemas pueden ser funcionales u obstaculizadores²¹⁶, pero, según afirman, las variables estructurales de la organización (p. ej., la clasificación) no tienen por qué determinar la sociedad de los alumnos. Descubren tres notas típicas de los sistemas informales que pueden servir para caracterizarlos: el grado de acuerdo en su interior; la fuerza del control informal; la omnipresencia del sistema informal en las áreas y actividades controladas por sus normas informales. WOODS (*Open University*, 1977, Unidad 7-8, págs. 19-21) caracteriza esa investigación como relacionada con la ‘experiencia’ (p. ej., cómo negocian su jornada alumnos y profesores, cómo mantienen sus relaciones y responden a los requerimientos que se les hacen y cómo influye esto en sus vidas y autoconceptos) mas que con el “proceso” (p. ej., los métodos de enseñanza). En consecuencia, se ocupa del *curriculum* oculto. WOODS señala diversos aspectos del *curriculum* oculto que se han investigado: deformaciones del *curriculum* manifiesto (p. ej., ciertos aspectos del *curriculum* que operan como formas de control social); las estrategias de profesores y alumnos (p. ej., alumnos que no participan en el aprendizaje, sino que siguen la ley del mínimo esfuerzo, manifestando su incomodidad); el “marco inalterable del sistema” (p. ej., su tendencia a alienar)²¹⁷. En este campo, se observa un incremento de la bibliografía, incluyendo los trabajos de WOODS (1979, 1980a, 198Gb) y WILLIS (1977). Puede verse una amplia revisión en HAMMERSLEY (1980).

La metodología que adoptan los sociólogos para investigar la “experiencia” es la etnografía, puesto que este enfoque pretende hacer un “retrato de a forma de vida de un grupo humano interactivo” de tipo holístico (*Open University* 1977, Unidad 7-8, pág. 12). Requiere utilizar la observación participante, incluyendo las técnicas adecuadas que señalamos en el apartado sobre la “Observación”, por ejemplo: la observación no estructurada y las entrevistas. (Véase en WILSON, 1977, un resumen de estas técnicas.) Aunque no exista una definición antecedente de lo que sucede en la escuela, y aunque la teoría que explique los significados concretos que alumnos y profesores asignan a su conducta se elabore por inducción (Capítulo VII), los etnógrafos utilizan teorías generales respecto a esa conducta. Uno de estos enfoques teóricos es la “interacción simbólica”. Con frecuencia se considera sinónima de la investigación mediante estudio de casos, pero WOODS (*Open University* 1977, Unidad 7-8, pág. 12) cree que ésta supone estudiar un aspecto o tema más concreto de la vida de profesores y alumnos. STENHOUSE (1978) decía también que la investigación mediante el estudio de casos hace más hincapié en las entrevistas que en la observación.

Al tomar de la gestión y de la sociología las teorías de las escuelas como organizaciones²¹⁸, los profesores se enfrentan a una tarea difícil si quieren utilizarlas de manera que informen sus evaluaciones. Es obvio que no decimos que los profesores tengan que analizar todas las teorías existentes antes de realizar un estudio. Sin embargo, algunas de éstas pueden proporcionar formas útiles de pensar sobre el contexto del *curriculum*; como tales, suponen un avance en relación con las categorías ad hoc. Precisamente por tratarse de teorías elaboradas a menudo por investigadores ajenos a las escuelas, pueden ayudar a los profesores a distanciarse de las suyas²¹⁹. JAMES y EBBUTT (1981) hablan así de su experiencia cuando ejercían como profesores que desarrollaban sus investigaciones en las escuelas en las que daban clase:

A menudo resulta difícil escoger y reflexionar sobre lo que es significativo, dejando aparte el mero “ruido”. Uno de los problemas consiste en aprender a reevaluar o que se ha considerado como mero “ruido” institucional y escucharlo con “oídos nuevos”. En una escuela dotada de un sistema de megafonía, Dave EBBUTT se acostumbró de tal manera a dar órdenes y normas despersonificadas que sus mensajes carecían de todo énfasis especial, hasta que observó que la voz electrónica transmitía una evidencia iluminadora respecto al

²¹⁶ LAMBERT y cols. (1973) describen cuatro formas de relacionar el sistema informal con el formal: de apoyo al formal, manipulador de éste, pasivo respecto al formal y opuesto al mismo.

²¹⁷ Conviene distinguir la idea de sistema informal de la de *curriculum* oculto. Esta última se refiere al aprendizaje y experiencia que se derivan tanto del sistema formal como del informal. Por ejemplo, el sistema formal puede exigir que los alumnos se pongan en pie cuando un profesor entra en clase; “la banda” (sistema) informal puede exigir que se cumplan sus normas. Ambos dejan impresiones y, por ejemplo, influyen en el autoconcepto del alumno. Además, “*curriculum* oculto” suele referirse al *curriculum* de los alumnos, aunque, como señala HARGREAVES (sin fecha), el *curriculum* oculto del profesor es, como mínimo, tan importante como aquél. HARGREAVES facilita una descripción muy accesible de la bibliografía sobre lo que él prefiere llamar el “*paracurriculum*”

²¹⁸ Y somos muy conscientes de o selectivos que hemos sido al respecto.

²¹⁹ En el capítulo IX desarrollaremos a idea de que cierto conocimiento de las teorías existentes puede ayudar a los profesores a “convertir o conocido en sorprendente”.
Dísembodyed despersonificadas (sin hacer visible el emisor). (N.del R.)

currículum oculto. En la misma escuela, también había aceptado sin apercibirlo un tratamiento diferente para los chicos y para las chicas; pero nadie había comprendido las consecuencias de esto, consistentes en la existencia de amplias zonas de la escuela que, en realidad, constituían áreas “prohibidas” para uno u otro sexo. Incluso el salón de actos estaba dividido en dos partes iguales, una masculina y otra femenina, característica evidente, aunque nadie en diez años se había percatado de ella.

(Pág. 90.)

El contexto externo

Ya hemos llamado la atención respecto a cómo la investigación sobre la enseñanza ha informado las listas de comprobación para la evaluación con respecto a las relaciones entre el hogar y la escuela. No obstante, en general, parece mucho menos evidente la ayuda prestada por el trabajo teórico en la consideración del contexto externo del *currículum*. Es posible detectar la influencia sistémica en la lista de comprobación del NSSE (1978), sobre todo con respecto a las características del alumno y de la comunidad como *inputs* y el desarrollo económico de la comunidad como ambiente del sistema al que debe adaptarse el proceso del sistema. No obstante, eso puede quedarse en una simple forma grandilocuente de decir que la escuela debe fijar con claridad sus objetivos y métodos de acuerdo con el tipo de niños que acudan a ella y al futuro que probablemente les toque vivir. Aunque esas listas de comprobación requieren obtener datos estadísticos sobre los alumnos y el ambiente al que accedan²²⁰, no se facilitan orientaciones sobre el uso de esta información. DAVIES (1981, pág. 54) da a entender que los enfoques marxistas, que consideran la enseñanza como reproducción social²²¹, convierten la organización escolar en una “caja negra” y no consiguen relacionar el contexto social de la escuela con su funcionamiento interno²²².

El interés por la escolarización urbana o la enseñanza multitécnica pone de relieve la necesidad de examinar la comunidad que rodea la escuela²²³: sus aspectos físico, político, económico y cultural. Aunque es obvio que un análisis de ese tipo, realizado por los profesores, les ayudaría a comprender mejor el cometido de la escuela, no es habitual que estén claras las relaciones entre el contexto y las operaciones internas de ésta. Por ejemplo, para que el *currículum* se adapte a las necesidades, aspiraciones y condiciones de la comunidad local, es preciso que la revisión curricular cuente con tal análisis. El NSSE (1978, pág. 14) lo plantea así: ‘la cuestión debe ser: ‘¿qué hace falta saber de los jóvenes y de la comunidad que rodea a esta escuela para determinar si ésta satisface las necesidades de la comunidad?’” La idea de desarrollo curricular basado en la escuela de SKILBECK (*Open University*, 1976, Unidad 26) requiere también un análisis de la situación (o sea, un análisis de los factores que estructuran la situación). En ello se incluye un componente externo:

1) Los cambios y expectativas culturales y sociales, incluidas las previsiones de los padres, los requerimientos de los empleadores, los supuestos y valores de la comunidad, las relaciones cambiantes (p. ej., entre adultos y niños) y la ideología.

II) Los requisitos y exigencias del sistema educativo, p. ej.: las formulaciones políticas, los exámenes, las previsiones, demandas o presiones de la administración local, los proyectos curriculares, la investigación educativa (...).

V) El flujo de recursos en la escuela. (Open University 1976, Unidad 26, página 96).

Escribiendo para un curso sobre enseñanza y el medio urbano, GOLBY se ve forzado a admitir: “El resultado del análisis de situación es, por tanto, turbio. Para quienes están comprometidos en el tema, parece algo así como una cuestión de incursiones en experimentos educativos con la comunidad” (Open University, 1978, Bloque VI, Unidad 14, pág. 47). Para qué decir más!

Conclusión

²²⁰ El NSSE (1978) propone la siguiente lista de datos que recabar para el cuerpo del estudio: composición étnica, aptitud escolar, intereses, capacidades destacadas, capacidades disminuidas, percepciones de las necesidades e intenciones educativas y necesidades y oportunidades laborales. El ILEA (1977) dice que las escuelas secundarias deberían recabar detalles acerca de las ocupaciones de sus ex alumnos y de la enseñanza recibida por ellos tras abandonar el centro respectivo. Esto último puede considerarse sin más como una medida de resultados.

²²¹ Los teóricos de la reproducción social y de la correspondencia sostienen que la enseñanza constituye una forma de control social diseñada para reproducir y justificar el orden social existente. BOWLES y GINTIS (1976), quizá los teóricos más conocidos de esta línea, plantean una correspondencia de uno a uno entre la enseñanza y el trabajo que garantiza el mantenimiento de las relaciones sociales de producción en la sociedad capitalista. Un tipo de análisis semejante informa también el estudio de WILLIS (1977).

²²² No obstante, véase en HAMMERSLEY (1980) una revisión de la etnografía del aula que intenta demostrar que existen relaciones entre los modelos de aula y los macroproblemas.

²²³ No pretendemos sugerir que sólo reclamen este tipo de análisis estas áreas de investigación; no son más que ejemplos en donde se ha concentrado el trabajo en cuestión.

La escuela interesada en investigar el contexto curricular interno dispone de diversas teorías que ayudarán a conceptualizar el problema de la evaluación. No así respecto al contexto externo; en este campo, nos quedan las recomendaciones ad hoc de diversas listas de comprobación. Dejando a un lado los problemas del uso de la información, las escuelas se enfrentan aún con los problemas de la recogida de datos y, en caso de necesitar cuantificarlos, con los de análisis estadístico. Cuando la comunidad que rodea la escuela es grande, el muestreo también constituye un problema. Incluso los enfoques etnográficos, en los que el muestreo teórico es importante, los recursos necesarios para elaborar una etnografía de la comunidad local son elevados²²⁴. Después, está la nación: el conjunto de la sociedad! En este nivel, debe dejarse a las escuelas para que utilicen su sentido común respecto a lo importante; como deduce GOLBY: “cada uno a lo suyo”.

²²⁴ JAMES y EBBUTT (1981, pág. 83) indican que, en uno de sus estudios, sus 'primeros pensamientos sobre una explicación [de por qué las niñas 'tropezaban' en las ciencias físicas] habrían supuesto escribir una sociología de toda la localidad'.