

## UN MATRIMONIO CONVENIENTE: DISEÑO DE INVESTIGACIÓN CUALITATIVA Y ESTÁNDARES PARA LA EVALUACIÓN DE PROGRAMAS

by/por

[Article record](#)

[About authors](#)

[HTML format](#)

Margaret D. LeCompte ([lecompte@spot.colorado.edu](mailto:lecompte@spot.colorado.edu))

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

### Abstract

This article provide a qualitative researcher's perspective on the 1993 Program Evaluation Standards. These new Standards are of critical importance for evaluators because they establish agreed-upons for proper behavior within the evaluation community.

Three aspects of the Standards are most salient to qualitative researchers: the concern for the human aspects of evaluation; issues of validity and meaning; and a broadened definition of legitimate data collection and analysis techniques.

It has taken several decades for the utility of match-making between qualitative and quantitative research designs to be recognized by the evaluation community. The 1993 Program Evaluation Standards update evaluation procedures and legitimate a sensible integration of the wide range of research procedures available to program evaluators.

### Keywords

Evaluation, program evaluation, research design

### Resumen

Este artículo proporciona una valoración, desde la perspectiva de la investigación cualitativa, sobre los *Standards* de 1993 para la evaluación de programas.

En este sentido tres son los aspectos de los *Standards* que más aportan a los investigadores cualitativos: a) la preocupación por los aspectos humanos de la evaluación; b) los temas de la validez y el significado; y c) una definición más amplia de las técnicas legitimadas de recogida y análisis de datos.

Han sido necesarias varias décadas para que la comunidad evaluativa reconozca la utilidad del enlace entre los diseños cualitativos y cuantitativos. Los *Program Evaluation Standards* actualizan los procedimientos de evaluación, a la vez que legitima una integración sensata de la amplia variedad de procedimientos disponibles para los evaluadores de programas.

### Descriptoros

Evaluación, evaluación de programas diseño de investigación

Mi tarea en este artículo es proporcionar una perspectiva como investigadora cualitativa sobre los *Standards* de 1993 para la evaluación de programas. Estos nuevos *Standards* son de una importancia fundamental para los evaluadores ya que establecen las bases de acuerdo para una actuación adecuada en el seno de la comunidad evaluativa. Además, el documento es gratificante para investigadores y evaluadores cualitativos, como lo es para mí misma, porque incorpora un reconocimiento oficial de lo que, en mi opinión, ha ocurrido en la educación y las ciencias sociales a lo largo de estos últimos cincuenta años: un

alejamiento desde la exclusiva dependencia de las normas positivistas por las que medir la calidad de la investigación. Desarrolladas, como lo fueron, desde modelos de laboratorio imposibles de realizar en la complicada realidad del trabajo de campo, estos criterios fueron una constante fuente de frustración y de baja autoestima para los evaluadores de campo. Reconociendo la subjetividad de los evaluadores y participantes, los *Standards* muestran efectivamente la noción de que puede darse algo así como una evaluación "objetiva", mientras que al mismo tiempo proporciona pautas para comple-

tar y adecuar la documentación de todas las perspectivas dentro de un proyecto dado. Los nuevos *Standards* han sido de un inmenso alivio para una de mis alumnas de postgrado que había utilizado una versión anterior. Después de haberle dado una copia de los mismos ella dijo: "es maravilloso que ésto se haya hecho. Finalmente nos trae a los noventa y quizás... es realmente una visión holística. Nos permite hablar con todos los patrocinadores, utilizar todo tipo de estrategias de recogida de datos y considerar todas las posibilidades".

Quizás mis sentimientos se pueden expresar más claramente en el contexto de mi bautismo de fuego como evaluadora de programas. Esto ocurrió durante un trabajo de tres años (1968-1971) con un equipo de evaluación asignado al Woodlawn Experimental Schools Project (WESP), uno de los primeros experimentos de descentralización de centros (Lecompte, 1969). El objeto de evaluación era un centro de enseñanza primaria en la zona sur de Chicago, cuyos alumnos tenían un récord poco envidiable de fracasos. Los participantes en el WESP fueron el Comité de Educación de Chicago, la Universidad de Chicago, cuyo campus rodeaba Woodlawn donde el colegio experimental estaba localizado, y la Organización Woodlawn, una institución de la comunidad fundada por Saul Alinsky y dedicada a enfrentarse con las intenciones de la Universidad de Chicago de apropiarse de buena parte del suelo de Woodlawn. WESP comenzó en 1968 y fue fundado por el Título III de ESEA (Elementary and Secondary Education Act). Como tal, el proyecto debía proporcionar a la agencia financiadora informes con los resultados de la evaluación. Los diseños de evaluación requeridos para tales proyectos eran cuasi-experimentales; las apreciaciones de la eficacia de los proyectos se basaron en la comparación pre y post-test de los datos de los alumnos, normalmente midiendo aspectos tales como la atención y el rendimiento relacionado con el programa (ver LeCompte, 1972). Sin embargo, el director de nuestro equipo de evaluación, el Dr. Robert Rippey, tenía una perspectiva algo más amplia de cómo evaluar los cambios sociales. El encargo hecho a WESP era

favorecer todo lo que fuera posible la participación de la comunidad; se había partido de la hipótesis de que restaurando el control del colegio desde los estamentos más bajos, contando con los profesores de mayor nivel, y los administradores y con un comité más cercano a los intereses y necesidades de la comunidad, se incrementaría el rendimiento del alumnado. Rippey estaba, no obstante, preocupado por la escasa conexión entre las metas de WESP y los resultados de los alumnos tal y como se medían en los tests estandarizados. Con la ausencia de acciones pedagógicas, "cómo se puede esperar un impacto en el aprendizaje del alumno?". Es más, la propia naturaleza del centro, de la comunidad, del equipo directivo del mismo, y de las relaciones con la universidad, hacían problemático cualquier éxito a la hora de alcanzar los objetivos, y sin tener en cuenta lo directamente que podían estar conectados con el rendimiento del alumnado.

Conociendo que WESP y el proyecto Ocean-Hill-Brownsville de New York, con él relacionado, eran inicios radicales en cuanto a reforma escolar, Rippey estableció un diseño de evaluación cuyo objetivo primordial fue documentar qué ocurría durante el desarrollo del proyecto [1]. Datos como éstos permitirían al equipo de evaluación juzgar no sólo los resultados finales del proyecto, sino también el por qué ocurrieron tal y como se sucedieron. El equipo de investigación era tan poco ortodoxo como el diseño de la investigación; Rippey nombró como asistentes de la investigación a tres estudiantes graduados, uno en antropología, otro en sociología de la educación, curriculum e instrucción, y otro en ciencias políticas que había crecido en el vecindario de WESP. Los estudiantes éramos los "documentalistas"; ahora probablemente nos denominaríamos etnógrafos. Nuestra tarea era movernos dentro y fuera del centro, asistir a reuniones relacionadas con él y con el personal de la comunidad, y "escribir una historia" de todos los acontecimientos que observamos. La evaluación de WESP prefijó un interés en la así llamada evaluación educativa cualitativa; lo más cercano que estaba a la evaluación educativa tradicional era el uso de protocolos observacio-

nales o listas de control para evaluar a los profesores y la ya mencionada recogida de datos sobre el rendimiento de los estudiantes obtenidos en el pre y post-test. Para la mayoría, los métodos de investigación estaban sacados de las técnicas de trabajo de campo de la sociología y la antropología (ver, por ejemplo, McCall y Simmons, 1969). Estos "métodos cualitativos" eran, de hecho, las únicas orientaciones disponibles al alcance de todos.

### ¿Qué es la investigación cualitativa?

La investigación cualitativa se define de forma poco precisa como una categoría de diseños de investigación que extraen descripciones a partir de observaciones que adoptan la forma de entrevistas, narraciones, notas de campo, grabaciones, transcripciones de audio y vídeo cassettes, registros escritos de todo tipo, fotografías o películas y artefactos. Sólo desde mediados de los 80 se han considerado respetables en los círculos educativos. La mayor parte de los estudios cualitativos están preocupados por el **contexto** de los acontecimientos, y centran su indagación en aquellos contextos en los que los seres humanos se implican e interesan, evalúan y experimentan **directamente** (Dewey, 1934; 1938). Esto es lo que significa calidad: lo real, más que lo abstracto; lo global y concreto, más que lo disgregado y cuantificado. Es más, la investigación cualitativa investiga contextos que son naturales, o tomados tal y como se encuentran, más que reconstruidos o modificados por el investigador (Sherman y Webb, 1988). Esta es la clase de diseño de investigación que Rippey quería realizar.

Robert Rippey centró su evaluación en la quintaesencia de la preguntas cualitativas: "¿Qué está pasando en este escenario?", y "¿qué significa para los participantes?" (Erickson, 1986). Estas preguntas han sido utilizadas en investigaciones denominadas de distintas formas: **interpretativas** (Erickson, 1986), **naturalistas** (Lincoln y Guba, 1985), **fenomenológicas** (Wilson, 1997), y **descriptivas** (Wolcott, 1980). Estos descriptores enfatizan la importancia de los constructos de los participantes, o los significados que los sujetos de la investigación asignan a sus accio-

nes, el contexto del estudio, la relación entre el investigador y los que están siendo estudiados, los métodos para la recogida de datos, los tipos de evidencias aducidas en apoyo de las afirmaciones realizadas, y los métodos y la importancia del análisis utilizado. Son, tal y como ya he argumentado en otros momentos (LeCompte, 1990; LeCompte y Preissle, 1993), más exactos y precisos que el término **cualitativo**, que simplemente sugiere que los investigadores cualitativos no cuantifican, miden, o cuentan algo, que realmente no es el caso.

### Modelos tradicionales en la investigación evaluativa

Robert Rippey tuvo que inventarse su propio diseño de evaluación porque la bibliografía existente en evaluación educativa estaba dominada por acercamientos analíticos sistemáticos empleados habitualmente en la economía y la industria (Rivlin, 1971). Estas perspectivas enfatizaban la medición de variables fácilmente escalables y cuantificables que generaban datos muy fiables; se ajustaban al dominio prevalente en la investigación educativa ejercido por el positivismo y los modelos experimentales. Mientras que apuntaban temas de causalidad que eran de interés para los planificadores de los programas, los datos que proporcionaban eran insatisfactorios porque no podían explicar por qué los programas tenían el impacto -o la falta del mismo- que indicaban los resultados. Además, la validez de muchos de los datos era cuestionada por los prácticos, quienes encontraban que las preguntas respondidas y las interpretaciones dadas eran incongruentes con sus propias experiencias y creencias (Guttentag, 1977; Helfgot, 1974; House, 1979; LeCompte, 1972). También el paradigma experimental era inadecuado para mostrar las complejidades del cambio institucional e instruccional. A principios de los setenta estaba claro que ni la experiencia de los alumnos y profesores, ni los factores relacionados con el centro que afectan a los resultados de los alumnos, podían consignarse adecuadamente con los procedimientos metodológicos empiricistas (Goetz y LeCompte, 1984:28). Por el contrario, eran necesarios

estudios en profundidad a pequeña escala utilizando observaciones participantes y no participantes durante un largo período de tiempo. Únicamente este tipo de estudios podrían ayudar a los investigadores a entender los procesos de cambio, los obstáculos que lo impiden y los factores que lo hacen exitoso.

Los investigadores educativos denominan a esto "método de estudio de caso" (Stake, 1978), mientras que investigadores más orientados a las ciencias sociales se centran en observaciones participantes y le asignan el nombre de "evaluación etnográfica" a tales procedimientos. Las agencias financiadoras comienzan rápidamente a exigir que el "componente etnográfico" se incluyera en cualquier diseño de evaluación. El uso del término etnográfico, en lugar de un término más general y apropiado como "cualitativo" o "proceso", puede haberse potenciado por el hecho de que los componentes etnográficos eran llevados a término por antropólogos y sociólogos, porque, tal y como Robert Rippey había aprendido, eran los únicos individuos formados en observación participante e investigaciones de campo de larga duración. Además, en tanto que las primeras evaluaciones de este tipo fueron de hecho diseñadas por antropólogos, tendieron a reseñar los estudios realizados en antropología cultural. Entre las primeras llamadas evaluaciones etnográficas nos encontramos con los estudios del National Institute of Education sobre las escuelas de integración racial, realizados bajo la supervisión de un antropólogo, Murray Wax (1979), y apoyado por un equipo de jóvenes antropólogos.

A pesar de su fuerza, la investigación etnográfica no estuvo exenta de dificultades. Los estudios a menudo duraban mucho, incluían demasiados datos y generaban pocos resultados satisfactorios para los que debían tomar decisiones. Como remedio, muchos metodólogos se decantaron por la combinación de estrategias cuantitativas y cualitativas (Jaeger, 1988; Goetz y LeCompte, 1991; Reichardt y Cook, 1979), las primeras para satisfacer a los que toman decisiones deseosos de una apreciación del impacto y las últimas para los planificadores y prácticos que realmente querían comprender cuáles eran

las consecuencias de su trabajo. A mediados de los ochenta, algunos investigadores evaluativos comienzan a proclamar una nueva forma de evaluación de proceso, formativa, naturalista (Lincoln y Guba, 1985), etnográfica (Fetterman y Pitman, 1986) o cualitativa (LeCompte y Goetz, 1982). Sin embargo no ha sido fácil que los fundamentos y orientaciones de estos modelos se legitimaran y se consideraran estándares en la evaluación de programas, veinticinco años después del pionero diseño de evaluación de Robert Rippey.

## Orientaciones para la evaluación cualitativa

Algunas características de la investigación cualitativa son de especial importancia para los evaluadores cualitativos. Ya he mencionado anteriormente lo relativo a la validez y utilidad de resultados para una audiencia variada y la focalización en el proceso o los eventos formativos, así como la recogida de datos sobre el producto o los resultados sumativos. La postura positivista de los evaluadores también se ha alterado. Como participantes de la evaluación, más que como sujetos de la misma, aquellos que son objeto de estudio adquieren una nueva identidad que supone un papel más activo a la hora de negociar los objetivos de la evaluación. Del mismo modo también cambia el papel del investigador, que pasa de un diseño detallado, con una postura objetiva vis-a-vis y una ejecución del estudio, hacia una posición más centrada en las subjetividades del proyecto. Esto no significa que el evaluador se vuelva un participante activo en los trabajos del proyecto objeto de evaluación, significa que él o ella no pueden seguir manteniéndose en una "presencia ausente". Al igual que los etnógrafos, los evaluadores cualitativos son responsables de su comportamiento y su postura ética dentro de las difíciles y estrechas interrelaciones que caracterizan la vida de las instituciones que evalúan (véase el código ético de la Sociedad para Antropología Aplicada, por ejemplo; también Deyhle, Hess y LeCompte, 1992). Al igual que los médicos deben considerar la admonición de "no hacer daño". Como investigadores en la acción o colaborado-

res, a menudo se encuentran con tensiones entre las constricciones de la apreciación y la medición más sólida y los asuntos a tratar por los electores competentes en sus lugares de evaluación. Especialmente si adoptan una postura de defensa de los etnógrafos post-modernos (Anderson, 1989; Gitlin y Smyth, 1989), estas tensiones resultarán aspectos destacados y substantivos de su trabajo que complican tanto los aspectos éticos como el desarrollo de las tareas de evaluación.

### ¿Cómo debería diseñarse una evaluación cualitativa?

Una concepción equívoca muy común acerca de la investigación cualitativa en general, y de la etnográfica en particular, es que carece de una teoría y de una estructura. El etnógrafo simplemente "merodea" (LeCompte y Preissle, 1993: 113), registrando todo lo que pasa. La recogida de datos no está guiada *a priori* por unos interrogantes de investigación, un marco conceptual, o teoría, y el análisis se ciñe a contar historias o a la explicitación de buenas afirmaciones hechas por los informantes para al final ofrecer una presentación que no sea cuantitativa. De hecho, la ejecución actual de una etnografía eficaz puede que no varíe mucho de este estereotipo.

Los plazos de tiempo que las agencias requieren de los evaluadores, las preguntas específicas de interés para los electores, y las múltiples audiencias que necesitan acceder a los resultados de la evaluación, especialmente cuando se utilizan los métodos cualitativos y etnográficos, todos intervienen para que el evaluador piense cuidadosamente *de antemano* qué tipo de preguntas constituirán el centro de la evaluación, qué fuentes de datos responderán las cuestiones, dónde y de quién pueden obtenerse esos datos, cómo se manipularán, dividirán y se presentarán. Sin esta planificación por adelantado se perderá

mucho tiempo explorando metas que no llevan a ninguna parte y que son tangenciales a la evaluación, o recogiendo datos que no responden a las preguntas planteadas por los clientes. Como cualquier investigador, el evaluador debe primero decidir cuáles serán las preguntas fundamentales que deben tener una respuesta. Los evaluadores, no obstante, deben negociar las preguntas de la investigación con los patrocinadores. Después pueden tomarse algunas libertades a la hora de planificar un conjunto de técnicas de recogida de datos que sea lo más variado, rico y eficiente posible. Una ventaja del uso de métodos cualitativos es que casi todo puede definirse como un dato potencialmente útil. Un problema es que resulta atractivo recoger cualquier conjunto de datos, pero los evaluadores deben - como enfatiza elocuentemente los *Standards*- conservar con cautela sus recursos de tiempo, personal y dinero. Esto requiere una nota de advertencia: no siempre es fácil recoger los datos más importantes de una forma rápida y eficaz; no obstante, los datos que son recogidos fácilmente pueden no ser lo suficientemente ricos como para contestar a preguntas significativas.

### Desarrollando un Plan

Una estrategia útil para asegurar unos diseños de evaluación que puedan ser operativos y eficaces es construir una matriz de recogida de datos. Esta matriz resulta también de utilidad para los clientes; les permite entender el alcance final del proyecto e incluso sugerir alternativas o fuentes adicionales de datos. La figura 1 muestra una matriz que se utilizó para la evaluación de un programa de relación entre la familia y el centro, tras la jornada escolar, en alumnos indios americanos de enseñanza primaria

**Figura 1. Una matriz de planificación para los procedimientos de recogida de datos**

¿Qué necesito conocer?	¿Qué datos responderán a esta cuestión?	¿De qué fuentes deben obtenerse los datos?	¿Quién es el responsable de contactar con las fuentes y recoger los datos?
¿Se está realizando el programa tal y como se concibió?	<ul style="list-style-type: none"> <li>* Documento de propuesta d proyecto</li> <li>* Observaciones del aula</li> <li>* Entrevistas con personal de proyecto</li> <li>* Análisis de materiales curriculares</li> </ul>	<ul style="list-style-type: none"> <li>* Coordinación de recursos</li> <li>* Orientadores familiares</li> <li>* Director del proyecto</li> </ul>	<ul style="list-style-type: none"> <li>* La coordinación de recursos establece el guión de la entrevista para sí misma; el personal; el director del proyecto aporta el material curricular y establece los intervalos de tiempos de observación</li> </ul>
¿Qué cambios, si hay alguno, se han realizado sobre el proyecto original?	<ul style="list-style-type: none"> <li>* Entrevistas con el personal del proyecto y la Coordinación de los recursos</li> </ul>		<ul style="list-style-type: none"> <li>* Recogidos a lo largo de las entrevistas del personal</li> </ul>
¿Está aportando la zona educativa el nivel de apoyo que prometieron?	<ul style="list-style-type: none"> <li>* Comparación de la propuesta común con el actual presupuesto del proyecto y las contribuciones desinteresadas</li> </ul>	<ul style="list-style-type: none"> <li>* Director del proyecto</li> <li>* Coordinador de zona</li> <li>* Director de programas especiales de la zona</li> </ul>	<ul style="list-style-type: none"> <li>* Con la coordinación de los recursos contacta B. Fowler, secretaria de la oficina de la zona, para establecer el guión de la entrevista con el personal de la oficina del distrito</li> <li>* El evaluador también entrevista a director del proyecto</li> </ul>
¿Qué logros y obstáculos ha encontrado el personal en la realización del programa?	<ul style="list-style-type: none"> <li>* Entrevistas con el personal del proyecto</li> </ul>	<ul style="list-style-type: none"> <li>* Orientadores familiares</li> <li>* Director del proyecto</li> </ul>	<ul style="list-style-type: none"> <li>* El director del proyecto establece las entrevistas con los orientadores familiares</li> </ul>
¿Con qué ánimo y espíritu de colaboración afronta el personal la puesta en marcha del programa?	<ul style="list-style-type: none"> <li>* Entrevistas con el personal del proyecto</li> </ul>	<ul style="list-style-type: none"> <li>* Orientadores familiares</li> <li>* Director del proyecto</li> <li>* Coordinación de los recursos</li> </ul>	<ul style="list-style-type: none"> <li>* Recogida a lo largo de las entrevistas con el personal</li> </ul>
¿Es bien recibido el programa por el personal de los centros?	<ul style="list-style-type: none"> <li>* Entrevistas a directores de centros</li> <li>* Entrevistas con el orientador escolar asignado al centro</li> </ul>		<ul style="list-style-type: none"> <li>* La Coordinación de los recursos establece las horas de las entrevistas para los directores de los centros y los orientadores escolares</li> </ul>
¿El personal del proyecto está guardando los registros especificados de los alumnos tal y como se especificó?	<ul style="list-style-type: none"> <li>* Registros anecdóticos</li> </ul>	<ul style="list-style-type: none"> <li>* Orientadores familiares</li> </ul>	<ul style="list-style-type: none"> <li>* El evaluador controla los ficheros guardados por los orientadores familiares en los centros</li> </ul>
¿Se están realizando con regularidad las visitas a las familias?	<ul style="list-style-type: none"> <li>* Anotaciones sobre las visitas a las familias</li> <li>* Observación participante</li> </ul>	<ul style="list-style-type: none"> <li>* Orientadores familiares</li> </ul>	<ul style="list-style-type: none"> <li>* El evaluador acompaña al orientador en una muestra de visitas a las familias</li> </ul>
¿Qué influencia está teniendo el programa sobre los alumnos?	<ul style="list-style-type: none"> <li>* Datos sobre el desarrollo de los alumnos</li> </ul>	<ul style="list-style-type: none"> <li>* Registros anecdóticos (desarrollados sólo para los chicos objeto del programa)</li> </ul>	<ul style="list-style-type: none"> <li>* El director del proyecto recoge y analiza continuamente estos formularios para todos los alumnos y se los pasa al evaluador</li> </ul>
	<ul style="list-style-type: none"> <li>* Datos sobre el rendimiento de los alumnos</li> </ul>	<ul style="list-style-type: none"> <li>* Puntuaciones en el Test Gates-McGrinitie</li> <li>* Metropolitan Readiness Test (sólo preescolar)</li> <li>* Home Language Survey (todos los alumnos)</li> <li>* Arizona State Assessment Tests (en 3º)</li> <li>* Arizona State Assessment Tests, administrados localmente (cursos 1º a 3º)</li> <li>* Iowa Test of Basic Skills (cursos 3º y 4º)</li> <li>* Diagnóstico individual de lectura (cursos 2º-3º)</li> </ul>	<ul style="list-style-type: none"> <li>* El director del proyecto obtendrá de los jefes de estudio las puntuaciones de los tests administrados en la zona, y los tests administrados a nivel de cada centro de los secretarios de los mismos, seleccionará los alumnos concretos y calculará las puntuaciones individuales y de grupo. También obtendrá puntuaciones de grupo de otros alumnos para utilizarlas como comparación. Estas puntuaciones se suministrarán al evaluador.</li> </ul>
	<ul style="list-style-type: none"> <li>* Datos sobre las actitudes de los profesores de aula y los directores</li> </ul>	<ul style="list-style-type: none"> <li>Entrevistas con los profesores de aula; directores</li> </ul>	<ul style="list-style-type: none"> <li>* El evaluador establecerá y realizará las entrevistas seleccionadas; el director del proyecto distribuirá recordatorios para cumplimentar los cuestionarios no respondidos</li> </ul>

La matriz está organizada alrededor de preguntas de evaluación; junto a cada pregunta hay columnas explicando qué datos serán necesarios para contestarlas, de qué fuentes se pueden obtener los datos, quién puede establecer los procedimientos para la recogida de datos, y quién es el responsable de recogerlos. La matriz puede asimismo incluir tanto las tentativas en cuanto a la temporalización de la recogida de datos, como las estrategias de análisis, incluyendo la codificación recursiva y continua, el desarrollo de diagramas y la construcción de dominios, tipologías y taxonomías.

Desarrollar un plan como éste y corroborarlo con los clientes, puede ayudar al evaluador a determinar qué datos van a ser "agradables de recoger" a partir de las fuentes originalmente identificadas y cuáles pueden no ser accesibles. Por tanto la matriz puede servir tanto como de contrato con los patrocinadores como de plan de ejecución para los propios evaluadores.

### **Una puntualización en torno a la contratación del equipo**

Robert Rippey también se aseguró de que el personal de la evaluación que había contratado se sintiese cómodo con los procedimientos cualitativos escogidos, por lo que seleccionó individuos cuya formación disciplinaria fuese compatible con los modelos de interacción cara a cara con frecuencia demandados por la observación participante. Aunque no fuera una norma no contar como evaluadores potencialmente cualitativos a aquellas personas cuya formación estuviera limitada a diseños experimentales, los evaluadores debían eliminar a los individuos que pensaran que la observación de campo o los datos de una entrevista son inferiores a las puntuaciones de los tests, o que se sienten incómodos hablando e interaccionando socialmente con extraños, incluso preguntándoles cualquier cuestión profesional o personal algo más embarazosa que el diseño de evaluación pudiera tener en cuenta.

### **Analizando los datos y contando la historia**

Una vez que se han recogido los datos, deben ser organizados y colocados de forma que

los patrocinadores puedan entenderlos y usarlos. El análisis comienza con el proceso de "poner en orden" (Romagnano, 1991) por lo que los datos recogidos son, en primer lugar, categorizados, limpiados, agrupados y después comprobados con la matriz de datos. Esto permite saber al evaluador si le falta algo o aún le resta algo por recoger, y le permite entender aquellas alteraciones en el esquema original de ejecución que las exigencias del trabajo de campo hayan provocado.

A continuación el evaluador intenta darle sentido a los datos. Este proceso a menudo se ve facilitado mediante el intento de "contar la historia" de lo que se ha encontrado. Los resultados de la evaluación no son historias, tan sólo son los resultados de un trabajo duro que tiene en su base una investigación. Sin embargo, la advertencia de Harry Wolcott a sentarse simplemente, aún cuando estás escribiendo mal y no tienes ánimos para hacerlo, y escribir la historia sin ninguna referencia a las notas (Wolcott, 1990), es una buena manera de crear el marco de trabajo inicial sobre el cual unir el análisis, independientemente del propósito al que sirvan los resultados, evaluación o investigación básica. La historia -o marco de trabajo inicial- se une entonces una vez más con la matriz de datos ya revisada para determinar dónde encaja definitivamente cada conjunto de datos. El marco de trabajo requiere alteraciones y, algunas veces, revisiones mayores, así como procesos de análisis más finos, y hasta el punto que los datos contradigan las primeras impresiones del investigador. Las cuestiones inicialmente indicadas en la matriz pueden ser irrelevantes, mientras que otras no contempladas llegan a resultar de especial importancia. No obstante, el esquema de trabajo inicial es una forma de comenzar con lo que algunas veces se presenta como una pila de información aparentemente sin ninguna relación. Es también una buena manera de identificar los tipos de productos que el evaluador deberá producir al terminar la evaluación. Estos productos, por supuesto, serán predicados sobre la base de las necesidades de las distintas audiencias o patrocinadores que querrán acceder a los resultados.

## ¿Qué dicen los estándares de evaluación de programas a los investigadores cualitativos?

Entiendo que hay tres aspectos de los *Standards* que son los que más importancia tienen para los investigadores cualitativos: la preocupación por los aspectos humanos de la evaluación; las cuestiones de validez y significado; y una definición más amplia de las técnicas de recogida y análisis de datos que se consideran legítimas.

### La parte humana de la evaluación de programas

Uno de los aspectos más beneficiosos de *Program Evaluation Standards* de 1993 es que los temas y diseños meramente técnicos no son considerados nunca más como supremos. Por el contrario, los temas humanos de la interacción, los intereses encubiertos, los éticos -aquellos aspectos implicados en la complicada existencia del mundo real en la línea de fuego- son de especial significado en la evaluación de la evaluación. Utilizo la metáfora "línea de fuego" a conciencia después de haber sido advertida por un directivo de AERA (American Educational Research Association) que tales metáforas eran inapropiadas y demasiado numerosas en la investigación educativa. Mi interpretación es que tal crítica puede únicamente haber sido hecha por alguien cuya investigación esté primordialmente localizada en una biblioteca. Mi propia experiencia ha sido completamente diferente; fuimos literalmente tiroteados mientras trabajamos en el vecindario de WESP, y me he sentido a punto de ser tiroteada, siempre metafóricamente, en otras muchas tareas de evaluación.

Los mandatos positivistas de la mayoría de la evaluación e investigación educativas ignoraban en el pasado este conflicto y dejaban al investigador ser al menos sobrecogido, si no manipulado y manejado, por miembros amenazadores y apasionados de grupos con intereses ocultos. Los *Program Evaluation Standards* son un buen antídoto frente a tal ingenuidad. Mientras quizá se sobreenfatice la necesidad de

que el evaluador ayude a que todas las partes lleguen a un consenso -algo que no considero como una parte de las tareas del evaluador- los *Standards* hacen más claro aún lo profundamente que puede meterse en problemas el evaluador por no hacer caso al poder, los asuntos a tratar y las opiniones expresadas sobre los mismos (LeCompte 1994, en prensa).

Otro aspecto beneficioso de los *Standards* es su reconocimiento explícito de la naturaleza humana de la empresa evaluativa. En el estándar U1, "Identificación de los patrocinadores" (p. 49) y el estándar P "Prioridad" (p. 120), los evaluadores deben identificar y establecer relaciones contractuales con todos los patrocinadores interesados. En otra parte del documento, los patrocinadores son caracterizados como un grupo, más que como el grupo más directamente implicado en el proyecto: participantes del programa, evaluadores, y la entidad o agencia que financia la evaluación. Los patrocinadores incluyen "la gama completa de los participantes-objetivo", (P1, p. 122) todos aquéllos implicados y todos aquéllos afectados por los resultados de la evaluación (P, p. 120) y todas las personas afectadas por el programa o la evaluación, y todos aquéllos que tienen legalmente el derecho a tener información derivada de la investigación (P6, p. 157). Esto significa que los evaluadores no pueden definir a los patrocinadores simplemente en términos de aquellas personas cuya inclusión es conveniente o efectiva en términos de costo-eficacia. Los *Standards* urgen incluso a los evaluadores a hacer "esfuerzos especiales ... para promover la inclusión apropiada de grupos o individuos menos poderosos como patrocinadores, tales como grupos minoritarios raciales, culturales o lingüísticos" (p. 49). Tales consideraciones reflejan el consejo dado tradicionalmente a los etnógrafos: no considerar irrelevante o no importante a ningún informante y no interesante a ninguna participación.

Los temas de la revelación también están implicados a la hora de identificar a los patrocinadores. Los *Standards* dirigen todas las revelaciones de los resultados a los patrocinadores

identificados; como si en el caso de los etnógrafos, los evaluadores no pueden asumir por más tiempo que los participantes no leerán sus informes o estarán furiosos si el evaluador intenta esconder información de ciertos grupos o individuos, dándoles una información parcial o filtrada. Los *Standards* buscan institucionalizar las prácticas relacionadas con la identificación y protección de los patrocinadores (p. 128-131), haciéndoles firmar un contrato de acuerdo con las normas vigentes sobre la protección e intervención con las personas objeto de investigación.

### **Temas de Validación en la Evaluación de Programas**

Un aspecto beneficioso de los *Standards* es la descentralización que hace del evaluador con respecto al establecimiento de preguntas significativas de evaluación. Los evaluadores no pueden ya permanecer cómodamente en la certeza de que han preguntado todo lo que es importante investigar en los programas que evalúan. Esto significa que los evaluadores deben estar seguros de que tienen en cuenta las interpretaciones y los significados de los participantes. El Standard U4, "Identificación de valores" dice que cada proyecto incluye a una variedad de participantes que harán -y deben ser capaces de hacerlos- juicios de valor e interpretaciones sobre el informe del evaluador. Los *Program Evaluation Standards* se centran en la cuestión de los múltiples significados presentes en los resultados, o cómo las personas interpretan los datos obtenidos en el informe final. Los evaluadores deben buscar estas interpretaciones múltiples, incluso cuando difieran de las del evaluador o de la agencia contratante, porque muchas de ellas pueden ser válidas. Como en el caso de una buena práctica etnográfica, esto significa tratar con multitud de significados o interpretaciones discrepantes o desaprobatorias, o casos o fuentes de evidencia alternativos (LeCompte y Preissle, 1993), e implica una investigación sistemática y diligente de cuanto más, mejor. Sólo haciendo ésto, se intensifica la fuerza de la validez del informe final.

Los *Standards* también enfatizan la necesidad de que los evaluadores lleven a cabo descripciones exhaustivas características de los programas, de las operaciones, de los procesos de ejecución y de los problemas de los participantes que participan en ese momento en el programa (p. 182) y la conveniencia de solicitar ayuda de los participantes con objeto de asegurarse de que esas descripciones son adecuadas (p. 183). Los etnógrafos, no obstante, es posible que todavía hagan una crítica al tratamiento de la validez interna que se hace en los *Standards*. Aparte de lo relativo al resultado final, los *Standards* tienden a tratar el tema de la validez interna como si careciese de dificultad. Se presta una atención muy considerable al modo en que deben desarrollarse los procedimientos de recogida de datos (p. 206) y a si los que aportan la información dicen la verdad, resultan fiables o recortan la información. Esta cuestión relacionada con la "validez del procedimiento" asume que si las reglas para llevar a cabo el estudio se aplican correctamente, también lo serán los datos recogidos. Sin embargo, los investigadores cualitativos creen que asegurar la validez interna y la de constructo es un asunto bastante complejo. La validez de constructo en medición dice si el método usado o la pregunta hecha correctamente tienen que ver con la información buscada. En el nivel más simplista, los investigadores cualitativos interpretan que esto significa "¿tiene la construcción de las preguntas -y las propias preguntas- sentido para el informante?". Una crítica muy común es que los instrumentos utilizados por los investigadores poseen un lenguaje inaccesible para los informantes. Hacer un buen trabajo de investigación y análisis cualitativo, no obstante, significa que el investigador debe buscar activamente los significados y las interpretaciones que los participantes dan a sus acciones en un proyecto y usarlas para organizar mejor su esquema de análisis (Erickson, 1986).

Más allá de esto, los temas de la validez implican si las distintas asunciones, esquemas de trabajo, preguntas y temas explorados por el evaluador o investigador tienen o no sentido y guardan relación con la importancia que le con-

ceden los patrocinadores o los participantes en la investigación. Esta fue una de las preocupaciones de Rippey: ha habido muchas evaluaciones en el pasado que han preguntado por cosas que los profesores y miembros de la comunidad creían irrelevantes. Demasiados programas han fallado por partir de criterios que no eran importantes para los prácticos. La evaluación de WESP, con el enorme peso que tuvo la observación participante y la extracción de redes complejas de significados, fue un intento de resolver estos problemas. El énfasis en la mera validez del proceso contenido en los *Program Evaluation Standards*, incluso cuando se usan las comprobaciones con los participantes, son insuficientes para remediarlos.

Un comentario post-moderno podría añadir que una búsqueda de los significados múltiples supone una excesiva preocupación sobre el valor de la verdad o la validez de la respuesta de los informadores, sobre todo si se tiene en cuenta que el sentido que cada informante da a las cosas constituye para él o ella la verdadera realidad. Ninguna historia o conjunto de significados es más o menos válido que otro. Esto no reconforta a los evaluadores -o a cualquier investigador- que se encuentra con el problema de cómo contar una buena historia o presentar un conjunto de resultados y recomendaciones. Los evaluadores pueden, no obstante, tratar las múltiples realidades presentes como aspectos competentes, y luego evaluar los costes relativos y los beneficios para la agencia financiadora adoptando la perspectiva de uno por encima de otro.

### **Una definición amplia de la recogida y análisis de datos**

Los nuevos *Program Evaluation Standards* suponen un avance fundamental para legitimar el conjunto ecléctico de herramientas metodológicas utilizado por los investigadores cualitativos. Los funcionarios de la Oficina de Educación que observaron la evaluación de WESP se quedaron perplejos por el peso dado a la observación participante y al análisis de documentos. Sin embargo, los nuevos *Standards* aceptan estas técnicas así como una amplia gama de

otros procedimientos considerados en alguna ocasión como sospechosos, sesgados o distorsionados, incluyendo "entrevistas estructuradas y no estructuradas, observaciones participantes y no participantes, audiciones, documentos y registros y medidas no obstrusivas de diversos tipos" (*Standards*, p.240). Los *Standards* otorgan explícitamente a los datos cualitativos casi la misma confianza que a los datos cuantitativos, y alertan a los evaluadores contra el "error común" de descartar datos tales como aquéllos que son recogidos a través de entrevistas personales, testimonios y análisis de documentos, que suelen ser considerados como datos distorsionados (p. 202). Estas recomendaciones son verdaderamente tranquilizadoras para aquellos investigadores cualitativos deseosos de realizar una evaluación, ya que son ellos mismos quienes crean el verdadero sesgo y enfoque de su trabajo.

Todas estas técnicas son las que Robert Rippey utilizó para desarrollar lo que ahora los nuevos *Standards* denominan un "análisis de contexto" (A2, p. 189) y describe como fundamental a la hora de realizar "evaluaciones reales que responden a las condiciones en las que el programa se encuentra... y a la (interpretación de) la evaluación" (Ibid.). El contexto es, por supuesto, el fondo de la cuestión de las interacciones humanas; como cualquier interacción humana, los programas educativos están "embebidos en y afectados por el/los contexto/s de una forma difícil de desenmarañar" (*op.cit.*). Los modelos positivistas de evaluación de programas definen los factores contextuales como ruidos necesarios de controlar; sin embargo, la necesidad de la descripción y el análisis exacto de esos factores y efectos contextuales es, creo, también exactamente lo que generó la insatisfacción de los sesenta o setenta con el entonces contemporáneo criterio de una buena evaluación. De forma similar, los *Standards* previenen a los evaluadores para que no caigan en el análisis cualitativo falto de rigor y meramente intuitivo. Se reconoce en ellos que el análisis de datos cualitativos es tan sistemático y trabajoso como los métodos que utilizan los investigadores cuantitativos. También reconoce la natura-

leza especial de los datos cualitativos y del análisis, sugiriendo que los evaluadores cometen un "error común" cuando limitan el análisis de los datos cualitativos a la cuantificación (*Standards*, p. 243) o atienden primordialmente a lo "único" o inusual en lugar de a lo común o a las "características generales" de la vida humana (Ibid). Los antropólogos y sociólogos han considerado desde hace tiempo que las "características generales" de la vida son aquellos patrones normativos frecuentemente no reconocidos que gobiernan la existencia diaria; son éstos, más que los eventos inusuales, únicos o no normativos, los que constituyen los procesos de los programas que causan los efectos que el evaluador desea estudiar.

### Enlace razonable

Robert Rippey se atrevió a finales de los sesenta a lo que los evaluadores defendieron en los setenta y ochenta: crear un matrimonio entre los diseños de investigación cualitativos y cuantitativos y ponerlos al servicio de una evaluación de programas más útil y eficaz. Mientras, han sido necesarias varias décadas para que la comunidad evaluativa reconozca la utilidad de tal enlace, los *Program Evaluation Standards* de 1993 traen procedimientos de evaluación (en términos de mi alumno graduado) "para los noventa y más allá", y legitima una integración lógica, razonable y sensata de la amplia variedad de procedimientos de investigación disponibles para los evaluadores de programas.

### Referencias

Anderson, G.L. (1989). Critical ethnography in education: Its origins, current status and new directions. *Review of Educational Research*, 59(3), 249-270.

Deyhle, D., Hess, G.A., y LeCompte, M.D. (1992). Approaching ethical issues for qualitative researchers in education. En M.D. LeCompte, W. Millroy y J. Preissle. (Eds.) *The handbook of qualitative research in education*. San Diego, CA: Academic Press, pp. 815-861.

Dewey, J. (1934). *Art as experience*. New York: Minton, Balch. Dewey, J. (1938). *Log-*

*ic: The theory of inquiry*. New York: Holt.

Erickson, F. (1986). Qualitative methods in research on teaching. En M.C. Wittrock, (Ed.), *The handbook of research in teaching* (3ª edición). New York: MacMillan, pp. 119-161.

Gitlin, A. y Smyth J. (1989). *Teacher evaluation: Educative alternatives*. Philadelphia, PA: The Falmer Press.

Goetz, J. P. y LeCompte, M.D. (1991). Qualitative research in social studies education. En J.P. Shaver (Ed.), *Handbook of research on social studies teaching and learning, a project of the National Council for the Social Studies*. New York: MacMillan Publishing Co., pp. 56-67.

Goetz J.P. y LeCompte, M.D. (1984). *Ethnography and qualitative design in educational research*. New York: Academic Press.

Gutentag, M. (1971). Evaluation and society. *Personality and Social Psychology Bulletin*, 3, 31-40.

Helfgot, J. (1974). Professional reform organizations and the symbolic representation of the poor. *American Sociological Review*, 38, 475-491.

House, E. (1979). The objectivity fairness and justice of federal evaluation policy as reflected in the Follow-Through evaluation. *Educational Evaluation and Policy Analysis*, 1(1), 28-42.

Jaeger, R.M. (Ed.) (1988). *Complementary methods for research in education*. Washington, D.C.: American Educational Research Association.

Joint Committee on Standards for Educational Evaluation (1993). *The program evaluation standards: How to assess evaluations of educational programs*. United States.

LeCompte, M.D. (1972). The uneasy alliance of community action and research. *School Review*, 79, 125-132.

LeCompte, M.D. (1990). Review of Designing qualitative research, por C. Marshall y G. Rossman, *Qualitative Studies in Education*, 3 (3), 295-298.

- LeCompte, M.D. (en prensa). Some notes on power, agenda and voice: A researcher's personal evolution toward critical collaborative research. En P. McLaren y J.M. Giarelli, (Eds.). *Critical theory and educational research*. Albany, NY: State University of New York Press.
- LeCompte, M.D. y Goetz, J. P. (1982). Ethnographic data collection and analysis in evaluation research. *Educational Evaluation and Policy Analysis*, Fall, 387-400.
- LeCompte, M.D. y Preissle, J. (1993). *Ethnography and qualitative design in educational research*. San Diego: Academic Press.
- Lincoln, Y.S. y Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Reichardt, C.S. y Cook, T.D. (1979). (Eds.). *Qualitative and quantitative methods in evaluation research*. Beverly Hills, CA: Sage.
- Rivlin, A.M. (1971). *Systematic thinking for social action*. Washington, D.C.: The Brookings Institute.
- Romagnano, L. (1991). *Managing the dilemmas of change: A case study of two ninth grade general mathematics teachers*. Tesis doctoral inédita, School of Education, University of Colorado-Boulder.
- Sherman, R.R. y Webb, R.B. (1988). Qualitative research in education: A focus. En Sherman, R.R. y Webb, R.B. (Eds.), *Qualitative Research in Education: Focus and Methods*. New York: The Falmer Press, pp. 2-22.
- Stake, R.E. (1978). The case study method in social inquiry. *Educational Researcher* 7, 5-8.
- Wax, M.L. (1979). *Desegregated schools: An intimate portrait based on five ethnographic studies*. Informe inédito para el National Institute of Education. Social Science Institute, Washington University, St. Louis.
- Wilson, S. (1977). The use of ethnographic techniques in educational research. *Review of Educational Research*, 47, 245-265.
- Wolcott, H.M. (1980). How to look like an anthropologist without really being one. *Practicing Anthropology*, 3(1), 6-17, 56-59.
- Wolcott, H. M. (1990). *Writing up qualitative research*. Newbury Park, CA: Sage.

---

## NOTAS

(Texto traducido del inglés por Ana Corrales Pérez)

### [\*] NOTA DE LA DIRECCIÓN:

Durante la reunión anual de la Asociación Americana de Investigación Educativa (AERA) realizada en la ciudad de Nueva Orleans, se celebró un simposium en el que participaron diversos expertos en evaluación de programas. El objetivo del mismo era presentar diferentes perspectivas sobre la nueva revisión que se ha llevado a cabo de los estándares para la evaluación de programas. En tal contexto, este trabajo presenta la aportación realizada por Margaret LeCompte al citado simposium. Los interesados pueden consultar la nueva versión de los estándares en:

Joint Committee on Standards for Educational Evaluation (1994). *The program evaluation standards: how to assess evaluations of educational programs*. Beverly Hills, CA: Sage.

[1] El proyecto también recogió la típica línea base y los datos del postest de los alumnos, pero no fueron éstos el objeto inicial de la evaluación. De hecho, a los ayudantes de la investigación se les insistió para que aseguraran a los profesores y los administradores del centro que los resultados del proyecto NO serían inmediatos. En efecto, les dijimos que las mejoras de los alumnos no podrían ser evidentes hasta el tercer año del proyecto, o aún después de haber finalizado. Esto último fue lo que en realidad sucedió.

# RELIEVE

**Revista ELectrónica de Investigación y EValuación Educativa**  
*E-Journal of Educational Research, Assessment and Evaluation*

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).