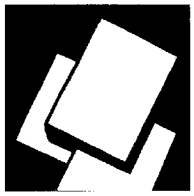
Revista de Psicología del Trabajo y de las Organizaciones - 2001
Volumen 17, n.º 1 - Págs. 21-43



Artículo

Metaevaluación de un sistema de evaluación integral de la calidad

E. REBOLLOSO[1]
B. FERNÁNDEZ RAMÍREZ[2]
P. CANTÓN
C. POZO

Unidad de Calidad e Investigación de Evaluación. Área de Psicología Social.
Universidad de Almería[3]

RESUMEN

El presente estudio describe la metaevaluación de un sistema de Evaluación Integral de la Calidad, desarrollado para valorar la calidad de los servicios de administración universitaria. La evaluación sigue un procedimiento activo y participativo, comprometido con los valores de negociación, consenso y responsividad. En consecuencia, la metaevaluación se centra en la perspectiva de los clientes o participantes en el proceso, con el fin de aproximar el marco de la guía EFQM a las necesidades y perspectivas del personal de administración y servicios. En concreto, se utilizan los criterios de comprensión, claridad, relevancia, utilidad y esfuerzo. Las limitaciones muestrales se superan aplicando una combinación de pruebas no paramétricas en el análisis de los datos. Los resultados muestran los niveles de concordancia alcanzados entre evaluadores y participantes, sugieren cierta redundancia entre los criterios de metaevaluación y aconsejan la revisión y mejora de algunos factores de calidad incluidos en la guía original de autoevaluación. Se señala la importancia del contexto y del liderazgo como elementos fundamentales para comprender y mejorar el Servicio que ha sido evaluado, así como la escasa trascendencia de los factores de impacto social y gestión de recursos.

ABSTRACT

This paper describes the meta-evaluation of a total quality management evaluation system aimed at rating the quality of college management services. Evaluation follows an active and

[1] e-mail: erebollo@ualm.es
[2] e-mail: bfernan@ualm.es
[3] Dirección: La Cañada de San Urbano, s/n. 04120 Almería. Tel.: 950 01 55 72.

participative procedure, committed with values such as negotiation, consensus and responsiveness. Hence, meta-evaluation focuses in customer -or participant- view in order to adapt the frame of EFQM guide to the needs and goals of administration and services personnel.

Particularly, criteria of understanding, clearness, relevance, utility and effort are used. A combination of non-parametric tests balances sample restrictions. Results show the agreement between assessors and participants and some redundancy in meta-evaluation criteria, and suggest the need of a revision and improvement of a number of quality factors in the original self-evaluation guide. Results emphasize the relevance of context and leadership as key factors for the understanding and improvement of services being assessed. They also show the limited relevance of social impact and resource management factors.

PALABRAS CLAVE

Metaevaluación. Gestión Integral de la Calidad (TQM). Modelo EFQM. Administración de servicios.

KEY WORDS

Meta-evaluation, Total Quality Management, EFQM Model, Services Management.

INTRODUCCIÓN

Estamos acostumbrados a escuchar las excelencias de los sistemas de calidad de cara al incremento de la productividad, reducción de errores y aumento de la satisfacción de usuarios. Sin embargo, todos somos conscientes de que también representan un suculento negocio para determinadas organizaciones nacionales e internacionales. Es evidente, por tanto, la necesidad de realizar estudios específicos para demostrar la realidad de su eficacia y la viabilidad de su implementación (Choi y Ebock, 1998; de Cock y Hipkin, 1997; Westphal, Gulati y Shortell, 1997). Con este trabajo pretendemos contribuir a clarificar esta cuestión a partir del concepto y la práctica de la metaevaluación, aplicada a la revisión del proceso evaluativo de un sistema de calidad recientemente implantado en un Servicio de administración universitaria (Rebolloso, Fernández Ramírez, Pozo y Peñín, 1999; Rebolloso, Fernández Ramírez, Pozo y Cantón, 2001). En concreto, se utilizan los criterios de comprensión, claridad, relevancia, utilidad y esfuerzo, como un modo de determinar la valía o el éxito de la guía de autoevaluación elaborada a tal efecto, toda vez que estos criterios inciden en la potencialidad de la evaluación para ser utilizada directamente

en la mejora de servicios. Los resultados del estudio dan pie a señalar, además, la trascendencia y viabilidad de los distintos factores incluidos en el sistema de calidad (EFQM, 1999).

CRITERIOS DE METAEVALUACIÓN

Entendemos por «evaluación» la actividad científica de determinar el mérito o valor de un objeto (Rebolloso y Rebolloso, 1998; Stufflebeam y Shinkfield, 1987) a partir de una serie de criterios de valor previamente determinados (Scriven, 1996; Weiss, 1990). Por ejemplo, podemos ponernos de acuerdo con un grupo de clientes en que lo que interesa es un servicio *rápido y sin errores*; a partir de estos dos criterios podemos evaluar la calidad de distintas opciones de servicio, y llegar a la conclusión de que una de ellas es mejor que las demás. En segundo lugar, entendemos por «metaevaluación» la aplicación de este procedimiento al propio proceso de evaluación (Chelimsky, 1983; Rebolloso, 1994a); la cuestión será determinar qué criterios de valor parecen más adecuados, y valorar que el proceso ha sido bueno desde el punto de vista de dichos criterios. Finalmente, en el marco de este trabajo se realiza la metaevaluación de un sistema de evaluación integral de la calidad implantado en un Servicio de administración universitaria (Rebolloso, Fernández Ramírez, Cantón y Pozo, 2000); es decir, se quiere averiguar si el modelo de evaluación implantado es suficientemente bueno desde determinado conjunto de criterios, con vistas a mejorarlo en posteriores aplicaciones en contextos organizacionales.

Los evaluadores tienen la obligación profesional de analizar el proceso y las decisiones que han tomado a lo largo de la evaluación (Chelimsky, 1983; Rebolloso, 1994b; Scriven, 1969; Stufflebeam, 1975). La metaevaluación (o evaluación de la propia evaluación) cumple los tres propósitos más habituales (Chelimsky, 1997; Stufflebeam y Shinkfield, 1987): a) *formativo*, persiguiendo el aprendizaje de los errores, la corrección de las conclusiones y la mejora de futuros planes de evaluación; b) *sumativo*, si se requiere demostrar la adecuación y eficacia del proceso evaluativo; y c) *de conocimiento*, si se pretende mejorar la teoría y la práctica de la disciplina.

Entre los marcos teóricos disponibles para la realización de metaevaluaciones, describiremos a continuación tres posibilidades: los requisitos de validez asociados a la práctica científica, los estándares desarrollados por las asociaciones profesionales y los procedimientos propuestos por diversos autores de forma independiente (M. Scriven y E. Chelimsky, en nuestro caso).

Variaciones sobre la perspectiva científica tradicional

Los evaluadores que se adscriben a un modelo de investigación cuantitativa y experimental, defienden los estándares científicos «clásicos», incluyendo la necesidad de que el diseño de investigación establezca relaciones causales válidas (validez interna y externa), mediante el uso de procedimientos válidos (validez de conclusión estadística y de constructo).

Desde posiciones críticas al paradigma convencional se argumenta que es más importante lograr la validez externa, al menos en la evaluación de programas, donde la utilidad y el impacto de los resultados evaluativos son críticos para lograr la

supervivencia de la disciplina (Cronbach, 1982; Cronbach y asoc., 1980; Stake, 1975). Por otra parte, la validez, al igual que la noción de «objetividad», tiene que ver con el acuerdo que se logra en torno a qué se entiende por calidad en una investigación, por lo que se torna más bien un asunto de intersubjetividad o consenso dentro de la comunidad científica (Fernández Dols, 1990; Glass y Ellett, 1980; Kaplan, 1964; Ziman, 1981). La crítica se ha acentuado con el debate sobre los procedimientos naturalistas (cualitativos, construccionistas) de evaluación. Los evaluadores de cuarta generación proponen dos tipos de criterio: la *confianza* [trustworthiness] y la *autenticidad* (Guba y Lincoln, 1989; Lincoln y Guba, 1986). La diferencia radica en considerar la investigación como una construcción social, en que la validez depende de la negociación y el consenso entre evaluadores, comunidad científica y audiencias del programa (Guba, 1967; Seoane, 1996).

En el campo de la Evaluación, donde la dimensión política y la necesaria acomodación de los valores en juego (Cronbach y asoc., 1980; Weiss, 1990) han hecho tomar conciencia de la importancia del proceso de negociación y búsqueda de consensos, los diseños deben ser lo suficientemente flexibles para adaptarse a las necesidades de los responsables de la organización y optimizar las oportunidades de investigación en un ambiente políticamente cambiante (Conner, Altman y Jackson, 1984; Rebolloso, 1990). Un método de trabajo respondente y participativo parece el modo más adecuado para despertar el interés de las personas implicadas por conocer y utilizar los resultados de la evaluación (Patton, 1994, 1997; Rebolloso, 1994a; Rebolloso, Fernández Ramírez, Cantón y Pozo, 2000). Además, ningún evaluador asume el papel de estar en posesión de ninguna «verdad» incuestionable que deba ser aceptada por los demás. Más bien, sirve a los intereses de audiencias y personas concretas, por lo que debe compaginar los criterios metodológicos de calidad con la obligación de clarificar y responder a sus perspectivas, incertidumbres o necesidades (Rebolloso y Rebolloso, 1998).

Chen (1988, 1990) ha propuesto una síntesis de criterios de calidad para la evaluación, teniendo en cuenta la perspectiva metodológica tradicional y las críticas aparecidas entre los evaluadores de corte constructivista. El autor propone cuatro criterios: (a) *objetividad*, que se asemeja a un criterio de fiabilidad o confirmación (ausencia de sesgos subjetivos; Scriven, 1991), alcanzada por consenso intersubjetivo; (b) *confianza*, que implica que la evidencia proporcionada por la evaluación es convincente o de confianza para el uso de los interesados; (c) *generalizabilidad*, que se pregunta si los resultados de la evaluación pueden ser aplicados a circunstancias o problemas futuros de interés para las audiencias; y (d) responsividad, que implica que la evaluación debe ser relevante y útil para las necesidades e intereses de todas las audiencias (de cara a producir acciones y cambios que redunden en la mejora social de sus miembros), incluyendo valores como la oportunidad [timeliness], la relevancia y la amplitud de consecuencias.

Los procedimientos metaevaluativos

Las normas profesionales recogen una serie de prescripciones y sugerencias sobre los requisitos que debe cumplir el evaluador en el desarrollo de su actividad (Evaluation Research Society [ERS], 1981; Joint Committee on Standards for Educa-

tional Evaluation [JCS],1981, 1988; Stufflebeam y Shinkfield, 1987). Dentro del amplio conjunto de normas existentes, Rebolloso (1994b) recomienda el uso de los estándares JCS, atendiendo a su mayor amplitud de miras y a la posibilidad de ser aplicados en múltiples contextos de evaluación. Las normas JCS se agrupan en cuatro conjuntos: (a) normas de viabilidad, que exigen que la evaluación sea fácil de implementar, eficiente en el uso del tiempo y los recursos, dotada con recursos apropiados y viable desde cierto número de perspectivas diferentes; (b) normas de honradez, que requieren que la evaluación atienda principios legales y éticos, el respeto a la intimidad (privacidad), la libertad de información o la protección de las personas; (c) normas de precisión, que tratan de asegurar la exactitud de los resultados y las conclusiones de la evaluación, evitando cualquier tipo de sesgo; y finalmente, (d) normas de utilidad, que persiguen ejecutar evaluaciones que sean informativas, oportunas e influyentes, ayudando a las audiencias a juzgar o hacer algo mejor de lo que normalmente lo hicieron en el pasado.

En cuanto a la realización de metaevaluaciones, Cronbach y asoc. (1980) recomiendan un procedimiento de revisión por pares, supuesto que la honestidad y el equilibrio de la evaluación se incrementará por revisiones críticas de la actuación de los evaluadores y patrocinadores. Las revisiones por pares suelen formar parte de los sistemas de acreditación al uso, por ejemplo, el propuesto por la European Foundation for Quality Management (EFQM) o los que siguen los sistemas de evaluación universitaria en varios países europeos (Consejo de Universidades, 1998; van Vught y otros, 1993). No obstante, Trow (1999; véase también Cook, 1984) critica las «picardías» que se utilizan para preparar y dar una imagen excelente a los evaluadores externos que visitan la institución, tales como la contratación de última hora de profesores con currícula deslumbrantes, los ensayos exhaustivos de la visita, la intervención de consultores que asesoran sobre el modo de presentarse, y otras por el estilo.

Scriven (1969) opina que pueden usarse diferentes procedimientos metaevaluativos, incluyendo las denominadas listas de control o las normas profesionales. La lista de control de Scriven (el multimodelo de evaluación) se compone de veintiún puntos en forma de preguntas que inquieren sobre distintas decisiones y tareas usuales dentro del proceso evaluativo (Stufflebeam y Shinkfield, 1987). El evaluador puede repasar la lista muchas veces a lo largo de la evaluación, analizando si cada cuestión ha sido tratada de forma correcta y si necesita volverse atrás para asegurar la validez del resultado final.

Con un marcado carácter pragmático, Chelimsky (1983) ha propuesto una síntesis de criterios de calidad, distinguiendo entre la adecuación técnica y la utilidad de la evaluación. La adecuación técnica depende de las restricciones del trabajo (tipo de información requerida, fondos disponibles, momento en que se necesitan los resultados), y se define como la diferencia de coste entre un diseño y otro. Se puede determinar si el diseño es adecuado desde el punto de vista de las necesidades del usuario. La adecuación técnica es relativa, y requiere un equilibrio entre varios elementos: adecuación del diseño, adecuación en términos de ejecución, existencia de opciones viables y ausencia de errores conceptuales.

La *utilidad* implica cuatro componentes objetivamente analizables: la relevancia de los datos respecto a una necesidad infor-

mativa, la oportunidad de los resultados de la evaluación, la presentación del informe (entendida como organización lógica de la información que resulta coherente y fácilmente comprensible) y el impacto real del informe (en términos del uso demostrable o de su influencia en la legislación o en la toma de decisiones). Todas las medidas de la utilidad persiguen reforzar la ejecución de negociaciones específicas con el cliente y el seguimiento de la satisfacción y el uso.

En conclusión, los evaluadores, sobre todo después de la llamada crisis de la utilización (Cook, 1997; Patton, 1997; Shadish, Cook y Leviton, 1991), han tratado de extender la perspectiva tradicional de la calidad científica, basada en el concepto de validez, para incluir la preocupación por que las evaluaciones logren un impacto evidente en la mejora de los problemas y condiciones sociales y organizacionales. Aunque la objetividad es necesaria para ganar en credibilidad ante sus clientes, no es suficiente si el esfuerzo empleado no resulta útil para los interesados en el proceso evaluativo.

A pesar de trabajar desde una perspectiva sistemática (es decir, tratando de conjuntar múltiples cuestiones, criterios, actividades, etc., dentro del proceso de evaluación; Chen, 1990; Rebolloso, 1994a; Rossi y Freeman, 1989), no todos los criterios pueden ser satisfechos con igual seguridad en una misma investigación. Idealmente, se debe llegar a un acuerdo sobre cuáles serán maximizados, evaluando diferentes teorías en una secuencia temporal, o balanceando todos dentro de la teoría del programa, funcionando con valores científicos y políticos ajustados (Chen, 1990). En nuestro caso, dada la novedad de los procesos evaluativos en el contexto de la gestión universitaria, parece interesante enfatizar criterios que aseguren que la evaluación gana adeptos y es reputada por su utilidad y ventajas en la mejora de los servicios. En consecuencia, destacaremos algunos criterios encaminados a aumentar el valor de la evaluación para los intereses de los responsables organizacionales.

OBJETIVOS DE LA INVESTIGACIÓN

El objetivo primario era revisar el proceso seguido en la evaluación de un Servicio de administración universitaria (Rebolloso, Fernández Ramírez, Cantón y Pozo, 2000; Rebolloso, Fernández Ramírez, Pozo y Cantón, 2001; Rebolloso, Fernández Ramírez, Pozo y Peñín, 1999). Se trabajó con un modelo de evaluación activa y participativa que pretende, manteniendo los estándares tradicionales de objetividad y validez, optimizar las decisiones para ganar en responsividad o capacidad de adaptación a las necesidades de información de los participantes en el proceso.

En otras palabras, apostamos por desarrollar una herramienta de evaluación que resulte directamente útil para los intereses de la organización, con el objetivo de servir de ayuda en la toma de decisiones sobre la mejora de servicios y lograr una actitud positiva hacia el trabajo evaluativo. La razón fundamental de esta elección radica en el convencimiento de que el fracaso de muchas evaluaciones ha sido no saber aproximarse a las auténticas necesidades de información de sus clientes (Cronbach y asoc., 1980; Patton, 1997; Sonnichsen, 2000).

Concretamente, analizaremos si el proceso es eficiente (esfuerzo), suficientemente informativo (relevancia, utilidad), y si

no hay problemas para comprender las cuestiones que plantea la guía de autoevaluación (comprensión, claridad). No se pretende contrastar una teórica estructura de análisis metaevaluativo, por cuanto hemos optado por seleccionar sólo una parte de los posibles criterios disponibles. Fundamentalmente, la elección se debe a razones de tipo pragmático, teniendo en cuenta las limitaciones de los participantes (sólo se les pregunta lo que pueden contestar con facilidad y sin requerir un esfuerzo elevado) y que el objetivo básico es descubrir cómo mejorar la guía para que se comprenda mejor, requiera menos esfuerzo y resulte más relevante y útil.

La situación peculiar de nuestros servicios de administración, respecto de la posible generalización de estrategias de Gestión Integral de la Calidad (GIC; véase Rebolloso, 1999), requiere especialmente actividades metaevaluativas. Las guías originales fueron diseñadas para ser aplicadas en el contexto amplio de los Servicios sociales y la Administración pública (Consejo de Universidades, 1998; EFQM, 1996/1997; MAP, 1999). En general, carecemos de claves teóricas para comprender adecuadamente las implicaciones de los nuevos sistemas de gestión en el ambiente de la administración universitaria, y necesitamos desarrollar investigaciones dirigidas a ampliar, recortar, afinar o modificar las cuestiones que se plantean en el sistema de gestión y en el procedimiento de evaluación.

En definitiva, con la metaevaluación se pretende, por una parte, modificar la guía de autoevaluación, incluyendo criterios que optimicen el esfuerzo, la comprensión y la utilidad de la misma, y en segundo lugar, comparar la perspectiva de los distintos participantes en el proceso de evaluación.

MÉTODO

Muestra

El proyecto contó con la participación de un grupo de cinco miembros del personal de administración, constituídos en Comisión de Calidad del Servicio, y de dos técnicos pertenecientes a la Unidad de Calidad e Investigación de Evaluación (UCIE), departamento responsable de las actividades de evaluación y la implantación de sistemas de calidad en nuestra Universidad, que coordinaron las actividades de autoevaluación. Evidentemente, se trata de una muestra de conveniencia, al modo de un estudio de caso en el que han colaborado todas las personas implicadas en el proyecto marco. No se plantean cuestiones de representatividad, sino que se aprovecha la opinión de todo el personal disponible. Esta medida no resta validez a ninguno de ellos como informantes clave que pueden opinar, desde la experiencia directa, sobre las debilidades y fortalezas del procedimiento y la guía de autoevaluación.

Instrumentos

La Guía de Autoevaluación para los servicios de administración universitaria está compuesta por un conjunto de once factores de calidad[1], básicamente derivados del modelo de gestión de la EFQM (1996/97, 1999). Cada uno de ellos se subdivide en

[1] En las guías de evaluación al uso se denomina *criterios* a cada una de las grandes áreas de gestión que se consideran relevantes para el análisis de la calidad. En los modelos de evaluación se denomina igualmente *criterios* a las variables o medidas que indican el valor del objeto. Para evitar confusiones en la exposición, hemos preferido utilizar el concepto en este segundo sentido, reservando a los primeros el nombre de *factores de calidad*.

un conjunto de subcriterios o actividades particulares que se suponen asociadas con prácticas excelentes de gestión organizacional. Los once factores están ordenados del siguiente modo (Rebolloso, Fernández Ramírez, Pozo y Cantón, 2001):

1. *Evolución y contexto actual del Servicio.* Se describen y analizan los grandes hitos que han marcado la trayectoria del Servicio y que determinan, en cierto modo, su situación actual.

2. *Liderazgo.* Hace referencia al grado de compromiso visible de los dirigentes y responsables de la organización y del Servicio en la creación y uso de valores y estrategias de gestión de la calidad.

3. *Política y estrategia.* Se considera cómo el Servicio define y formula su política estratégica, traduciéndola en planes y acciones concretas que incorporan los principios y herramientas de la calidad.

4. *Gestión del personal.* Se evalúa cómo se promueve la participación del personal del Servicio en la consecución de mejoras dentro del mismo.

5. *Recursos.* Se refiere a los procedimientos de gestión de los recursos principales del Servicio (económicos, instalaciones, tecnológicos, etc.), para que estén orientados hacia los objetivos de calidad.

6. *Procesos.* Se trata de examinar cómo el Servicio gestiona sus procesos clave y los procesos de soporte, es decir, cómo se identifican, se revisan y se corrigen, si es necesario, para asegurar la mejora continua.

7. *Satisfacción de los usuarios.* Se valora cómo se comprende y se avanza en la satisfacción de los usuarios a través de medidas directas (encuestas de opinión) e indirectas (estadísticas sobre quejas).

8. *Satisfacción del personal.* Se estudia cómo se comprende y se avanza en la satisfacción del personal, a través de medidas directas (encuestas de opinión) e indirectas (absentismo, niveles de formación, promoción interna, reconocimiento de esfuerzos, rotación, etc.).

9. *Impacto social.* Se examinan las actividades del Servicio que tienen consecuencias directas e inmediatas para la comunidad local y regional, y para la sociedad en general.

10. *Resultados generales.* Se discuten los resultados de las acciones bajo responsabilidad del Servicio y que son considerados más significativos de su actividad. También se estudian los resultados económicos (recursos propios conseguidos, ejecución del presupuesto, reducción de costes de los servicios, etc.).

11. *Metaevaluación.* Finalmente, se realiza una valoración crítica del proceso de evaluación y el análisis de las sugerencias de mejora.

El instrumento utilizado para realizar la metaevaluación fue un cuestionario dirigido a analizar cada uno de los diez primeros factores, así como el procedimiento global de trabajo. En concreto, cada factor se analizaba mediante dos tipos de cuestiones. Con un formato abierto, se preguntaba sobre los puntos fuertes y débiles de cada factor general, y se solicitaban sugerencias para mejorar sus preguntas y contenidos. A continuación, con un formato cerrado de respuesta (escalas tipo Likert de cinco puntos, desde totalmente en desacuerdo hasta totalmente de acuerdo), cada uno de los subfactores pertenecientes a cada factor general (cuarenta y uno

en total) era valorado según cinco criterios de metaevaluación (ver infra).

En resumen, el protocolo de metaevaluación está compuesto por cuarenta y un subfactores de calidad (valoración mediante respuesta cerrada), que se agrupan en diez factores generales (valoración mediante respuesta abierta).

Como los posibles criterios de metaevaluación son variados y no pueden ser fácilmente satisfechos al mismo tiempo (Chen, 1990), interesa seleccionar los que mejor respondan al punto de vista del personal de administración, y que puedan producir una información más rica, teniendo en cuenta que el objetivo de adaptar la guía de autoevaluación a la realidad de los servicios universitarios puede ser mejor alcanzado si se atiende a los problemas, dudas y necesidades que planteen los implicados en el proceso. A pesar de las deficiencias teóricas de las guías, ya existe suficiente investigación previa (Oakland, 1993; Rebolloso, 1999) y un amplio uso extendido en distinto tipo de organizaciones, como para garantizar mínimamente los requisitos científicos de validez y fiabilidad, y centrarnos en otro tipo de consideraciones de interés.

Los cinco criterios elegidos se identifican con el concepto de utilidad propuesto por Chelimsky (1983), y con las normas de precisión, utilidad y viabilidad del JCS (1981, 1989), con el propósito de mejorar el procedimiento de evaluación, aumentando su carácter informativo y su capacidad de influencia en el sistema. En concreto, se ha partido de la clasificación de Chen (1990), especialmente en relación con los estándares de objetividad y responsividad:

(a) *Comprensión.* El concepto que trata de reflejarse se comprende fácilmente.

(b) *Claridad.* Está enunciado con claridad.

(c) *Relevancia.* Es clave para comprender el funcionamiento del Servicio.

(d) *Utilidad.* Es útil para mejorar el Servicio.

(e) *Esfuerzo.* Nivel de esfuerzo que ha requerido contestar la cuestión.

Con los dos primeros (comprensión y claridad) se persigue obtener información para mejorar la objetividad que sugiere la guía como instrumento de recogida de datos, en el sentido de que aseguran que los participantes comprenden las cuestiones que se preguntan y no perciben la existencia de sesgos. Los dos siguientes (relevancia y utilidad) contribuyen a incrementar el carácter respondente de la guía, aumentando su proximidad y adecuación a las necesidades e intereses del personal profesional implicado. Cualquiera de estos objetivos ayudará a potenciar el uso de los resultados de la evaluación para la mejora del clima y de las actividades de calidad de la organización.

El último criterio (esfuerzo) ha sido incluido en virtud de los problemas al respecto que señalaron en repetidas ocasiones los participantes en la evaluación. El nivel de trabajo del personal de administración y servicios es elevado en determinadas etapas del año y se debe evitar que la evaluación suponga una carga añadida que resulte excesiva, pues son de esperar reacciones en contra o efectos de desmotivación (Fabi y Pons, 1995; Peiró y González Romá, 1993). El esfuerzo tiene que ver, de forma indirecta, con las normas de viabilidad (valor de eficiencia); en concreto, con el objetivo de que la evaluación sea tan fácil de aplicar como sea posible (JCS, 1988).

Procedimiento

El objetivo inicial de esta investigación era diseñar una guía de autoevaluación adaptada al contexto peculiar de la gestión y administración universitaria. Partiendo de un primer borrador de trabajo, se colaboró con diversos responsables de gestión (gerente, jefes de servicio, personal de administración y servicios), desarrollando un proyecto anual de evaluación de la calidad de servicios.

El proceso de autoevaluación requirió veintiuna sesiones de trabajo durante un periodo aproximado de cinco meses (Rebolloso, Fernández Ramírez, Pozo y Peñín, 1999). En una primera reunión, el equipo de autoevaluación discutía sobre las preguntas y procedimientos que se seguirían para cumplimentar el primer factor de calidad de la guía. Durante los días siguientes, los miembros del Servicio se hacían responsables de coordinar a sus compañeros y recoger la información requerida. Finalmente, en una nueva reunión, se discutía sobre los resultados entregados y se preparaba la redacción para el informe final. En todo momento, los dos evaluadores miembros de la UCIE defendían una actitud de negociación, buscando consensos sobre cualquier tema planteado, incluyendo el procedimiento de trabajo, la estructura de la guía, la formulación de las preguntas o la redacción final del informe de autoevaluación (Rebolloso, 1994a; Rebolloso, Fernández Ramírez, Cantón y Pozo, 2000).

La metaevaluación se planteó como una tarea a ser desarrollada una vez finalizado el proceso, aunque la mejora del proceso se perseguía en cualquier momento, al adoptar el equipo técnico una actitud respondente y negociadora frente a los participantes (Guba y Lincoln, 1989; Rebolloso, 1994a; Rebolloso, Fernández Ramírez, Cantón y Pozo, 2000). Retrasando la metaevaluación se evitaba sobrecargar a los miembros de la Comisión, toda vez que no habían sido aliviados en absoluto de sus responsabilidades en la marcha normal del Servicio.

El cuestionario de metaevaluación fue presentado en la última sesión de trabajo, discutiendo con los participantes sobre su conveniencia, formato, relevancia e interés para la mejora del sistema de evaluación. Su cumplimentación supuso un esfuerzo importante, pues exigía volver a pensar en cada fase del proceso que se estaba cerrando, con el problema añadido de que algún factor o área de análisis fuera mal recordado o hubiera cambiado de significado con la perspectiva adquirida después de hacerse una idea global de la autoevaluación. Para vencer las reticencias de los participantes y superar las dificultades inherentes, se redujo la tarea allí donde no fuera imprescindible para averiguar la impresión general que tenían sobre cada factor de calidad. De este modo, sólo se preguntó sobre los subfactores de la guía de autoevaluación, y no sobre cada una de las preguntas específicas de que se componían. No obstante, a pesar de la actitud negociadora del equipo técnico, dos participantes no devolvieron el cuestionario.

Análisis

En primer lugar, se estima el nivel de concordancia entre los respondentes de la metaevaluación, en conjunto y por separado (miembros de la Comisión de Calidad del Servicio y equipo técnico), comparando sus respuestas directas a lo largo del cuestionario. Se emplean el coeficiente de

concordancia *W* de Kendall y la prueba *U* de Mann-Whitney. El coeficiente de Kendall es un contraste no paramétrico sobre la hipótesis de que varias muestras relacionadas proceden de la misma población, utilizado generalmente para medir el acuerdo entre evaluadores. La *W* de Kendall varía entre 0 (no hay acuerdo) y 1 (acuerdo completo). La significación estadística se calcula mediante χ^2. La prueba de Mann-Whitney es el equivalente no paramétrico de la prueba t, que contrasta si dos muestras independientes proceden de la misma población. La significación estadística se calcula por aproximación a la curva normal. Además, se tiene en cuenta el valor alfa de Cronbach, que puede ser interpretado también como una medida de la correlación entre respondentes.

Aunque en el texto incluyamos los resultados de ambas pruebas, lo cierto es que parece discutible otorgar a los respondentes el papel de jueces, en la medida que no interesa tanto su grado de acuerdo como su opinión acerca de la calidad de cada uno de los subfactores de la guía. La prueba de Kendall mide concordancia, mientras que el valor alfa mide covariación, razón por la cual atenderemos preferentemente a este último en los análisis correspondientes.

Posteriormente, se analiza la homogeneidad entre los criterios de metaevaluación, para determinar si son independientes, o si ofrecen información redundante. Se utiliza la traspuesta de la matriz original de datos, de modo que se puedan correlacionar los criterios de metaevaluación. Somos conscientes de que este procedimiento introduce un sesgo de diferencias individuales, en el sentido de que se consideran independientes las distintas respuestas de un mismo respondente que utiliza repetidas veces el mismo conjunto de criterios para valorar una serie de cuestiones, con el riesgo de aumentar, por tanto, los niveles de covariación. Este sesgo debería tenerse en cuenta al discutir la potencia de este análisis. No obstante, se elige el coeficiente rho de Spearman, por ser el más adecuado para realizar cálculos con datos que no cumplen algún supuesto paramétrico. Por último, se usa la prueba U de Mann- Whitney para comprobar el uso diferencial de los criterios de metaevaluación por parte de los grupos de miembros del Servicio y del equipo técnico.

Finalmente, los datos metaevaluativos se resumen siguiendo un análisis descriptivo (media y desviación típica), para estudiar la comprensión, claridad, relevancia, utilidad y esfuerzo requerido en cada una de las áreas de gestión o factores de calidad.

RESULTADOS

Concordancia entre los participantes

Según el coeficiente de Kendall (W =.354; χ^2 = 69.004, p =.002), la concordancia indica un nivel de acuerdo del 35.4%, demasiado bajo para considerar que los siete informantes mantienen, en general, idénticas opiniones en la metaevaluación. No obstante, si se tiene en cuenta la puntuación alfa de Cronbach (α =.729), el nivel de covariación puede ser considerado aceptable para analizar cada criterio de calidad con los datos promedio. Estos valores resultarían suficientes para continuar con la descripción de los datos, aunque plantean la cuestión de si las diferencias de opinión son mayores entre los miembros de la Comisión de Calidad y del equipo técnico, y en qué contenidos de la guía existe mayor desa-

cuerdo. En concreto, el grado de concordancia entre los cinco miembros de la Comisión (W =.406; χ^2 = 63.325, p =.008) es algo mayor que el que se encuentra entre el total de participantes, a pesar de que el coeficiente de Cronbach (α =.491) disminuya sensiblemente. Por su parte, no es posible calcular el coeficiente de concordancia entre los dos miembros del equipo técnico, si bien el nivel de covariación entre ambos es mayor si se expresa a través del alfa de Cronbach (α =.670).

Para determinar las diferencias entre grupos se utiliza la prueba U de Mann-Whitney, comparando las valoraciones que se realizan en cada subfactor de la guía, promediando las respuestas directas correspondientes. En ningún caso la comparación es significativa, oscilando entre U = 5 (z =.000, p = 1.000) y U = 0 (z = -1.879, p =.060).

Para complementar estos análisis, la tabla 1 recoge los niveles de concordancia

Tabla 1
Concordancia entre los participantes

	Contexto		**Liderazgo**		**Política y estrategia**		**Gestión del personal**		**Recursos**	
Todos	N	7	N	7	N	7	N	6	N	6
	W	.381	W	.036	W	.242	W	.146	W	.519
	χ^2	2.667	χ^2	.500	χ^2	10.150	χ^2	3.495	χ^2	9 340
	gl.	1	gl.	2	gl.	6	gl.	4	gl.	3
	Sig.	.102	Sig.	.779	Sig.	.118	Sig.	.479	Sig.	.025
Miembros del	N	5	N	5	N	5	N	4	N	4
Servicio	W	.360	W	.050	W	.188	W	.266	W	.711
	χ^2	1.800	χ^2	.500	χ^2	5.626	χ2	4.263	χ^2	8.526
	gl.	1	gl.	2	gl.	6	gl.	4	gl.	3
	Sig.	.180	Sig.	.779	Sig.	.466	Sig.	.372	Sig.	.036
	Procesos		**Satisf. del usuario**		**Satisf. del personal**		**Impacto social**		**Resultados generales**	
Todos los jueces	N	6	N	7	N	7	N	6	N	7
	W	.341	W	.640	W	.669	W	.167	W	.143
	χ^2	14 333	χ^2	8.957	χ^2	9 360	χ^2	2 000	χ^2	1 000
	gl.	7	gl.	2	gl.	2	gl.	2	gl.	1
	Sig.	.046	Sig.	.011	Sig.	.009	Sig.	.368	Sig.	.317
Miembros del	N	4	N	5	N	5	N	5	N	5
Servicio	W	.577	W	.612	W	.744	W	.224	W	.067
	χ^2	16.157	χ^2	6.118	χ^2	7444	χ^2	2.235	χ^2	.333
	gl.	7	gl.	2	gl.	2	gl.	2	gl.	1
	Sig.	.024	Sig.	.047	Sig.	.024	Sig.	.327	Sig.	.564

de la muestra total de participantes y de los miembros de la Comisión, en cada factor de la guía (entre los dos miembros del equipo técnico no pudieron ser calculados). En ella se aprecia que los mayores acuerdos se obtuvieron en los apartados de Recursos, Procesos, Satisfacción del usuario y Satisfacción del personal, únicos en que se encuentran valores estadísticamente significativos.

Relaciones entre los cinco criterios de metaevaluación

Como ya se comentó, para determinar las relaciones empíricas se utiliza la traspuesta de la matriz original de datos, a partir de la cual se calcula la matriz de correlaciones (tabla 2).

Los resultados sugieren que podrían haberse utilizado tres criterios metaevaluativos, teniendo en cuenta la agrupación de las variables *comprensión-claridad* (r=.848, p <.000) y *relevancia-utilidad* (r = .725, p <.000). Entre ambos grupos existe una relación positiva, indicando que la mayor comprensión y claridad de cada tópico evaluado en la guía se asocia con una valoración de mayor relevancia y utilidad parta la mejora del Servicio. Respecto a la variable esfuerzo, está estrechamente relacionada con los valores de comprensión-claridad. Es decir, que el esfuerzo requerido es independiente de que las áreas de gestión incluidas en la guía sean más o menos relevantes para comprender y mejorar el Servicio, y se relaciona, más bien, con las dificultades surgidas durante el proceso de evaluación para alcanzar acuerdos sobre el significado de algunas cuestiones de la guía.

Metaevaluación de los factores de la Guía de Autoevaluación

En primer lugar, comprobaremos la utilización diferencial que hacen los dos grupos de participantes de los cinco criterios de metaevaluación. Con este fin, hemos promediado sus respuestas dentro de cada factor de calidad (contexto, liderazgo, política y estrategia, etc.), para comparar las valoraciones mediante la prueba *U* de Mann-Withney. En general, los valores *U* resultantes señalan tendencias no significativas (estos datos no se han incluido), sugiriendo que los dos miembros del equipo técnico tienden a valorar los contenidos de la guía como más comprensibles, claros, relevantes y útiles para la mejorea del Servicio. Igualmente, afirman en todos los casos que la tarea requiere menor esfuerzo

Tabla 2
Coeficientes de correlación entre los criterios de metaevaluación[a]

	Comprensión	Claridad	Relevancia	Utilidad
Claridad	.848**	1.000		
Relevancia	.323**	.387**		
Utilidad	.287**	.304	.725**	1.000
Esfuerzo	-.442	-.449	-.066	-.076

[a] Rho de Spearman
** p<.000

del que valoran los miembros de la Comisión de Calidad del Servicio. Sin embargo, ninguno de los resultados alcanza significatividad estadística, así que consideramos las respuestas de todos los participantes como similares, de cara a la realización del siguiente análisis.

En la tabla 3 se recoge la valoración promedio de cada factor de calidad (contexto, liderazgo, etc.), incluyendo la opinión de todos los participantes en el estudio. Para apoyarnos en la interpretación de estos resultados se utiliza la prueba de Friedman, realizando comparaciones entre los factores, para determinar estadísticamente cuáles son los que obtienen mayor o menor puntuación en comprensión, claridad, relevancia, etc. Sólo se comentarán las comparaciones que resultaron significativas.

Distinguiremos los resultados siguiendo los cinco criterios de metaevaluación:

(a) *Comprensión*. Las puntuaciones sobrepasan el valor medio de la escala, lo cual indica que, en general, las preguntas que plantea la guía son de fácil comprensión, a pesar de que algún participante haya manifestado una opinión menos favorable, tal como se aprecia en los factores con menor valor promedio y mayor desviación típica. Las preguntas planteadas dentro en *impacto social* son las peor comprendidas (χ^2=7.000 p=.008, con liderazgo, política y estrategia, gestión del personal, y satisfacción de usuarios; χ^2=6.000 p =.014, con satisfacción del personal). En menor grado, también debe revisarse el contenido de los factores de *procesos* (χ^2=6.000 p=.014, con liderazgo, y con política y estrategia) y de resultados *generales* (χ^2=4 000 p=.046, con satisfacción de usuarios, y satisfacción del personal).

(b) *Claridad*. Las puntuaciones superan el punto medio de respuesta, indicando el nivel general de claridad de la guía. Las comparaciones entre pares muestran que el *impacto social* es la cuestión peor valorada en este criterio (χ^2 =6.000 p =.014, con liderazgo, y gestión del personal; y χ^2 =

Tabla 3
Metaevaluación de los criterios de la Guía

	Comprensión		Claridad		Relevancia		Utilidad		Esfuerzo	
	$\bar{x}$	Sx	$\bar{x}$	Sx	$\bar{x}$	Sx	$\bar{x}$	Sx	$\bar{x}$	Sx
CONTEXTO	3.21	1.219	4.07	.449	4.71	.488	4.21	.636	3.07	1.304
LIDERAZGO	4.52	.572	4.33	.608	4.52	.604	4.71	.488	2.95	1.532
POLÍTICA Y ESTRATEGIA	4.31	514	4.08	.692	4.15	.777	4.19	.661	2.87	1.040
GESTIÓN DEL PERSONAL	4.40	.632	4.31	.620	4.08	.630	4.31	.615	2.71	.958
RECURSOS	4.29	.640	4.04	.781	3.29	.765	3.62	.847	2.80	1.272
PROCESOS	3.66	.640	3.50	.627	4.14	.588	4.12	.607	3.18	.665
SATIS. DE USUARIOS	4 47	766	4.28	1.026	4.38	.678	4.33	.638	2.42	.762
SATIS.DELPERSONAL	4.47	.857	4.47	.978	4.23	1.031	4.14	.813	2.47	1.069
IMPACTO SOCIAL	3.64	.735	3.73	.912	3.66	.577	4.00	.430	2.95	1.161
RESULTADOS CENERALES	3.71	1.216	3.92	1.090	4.00	1.000	4.50	.774	3.92	.607

5.000 p =.025, con política y estrategia, satisfacción de usuarios, y satisfacción del personal), junto con las cuestiones incluidas en el análisis de *procesos* (χ^2 =6.000 p =.014, con liderazgo, y con política y estrategia).

(c) *Relevancia.* En general, todos los factores de calidad son relevantes para comprender la realidad del Servicio, dadas las altas valoraciones encontradas. Atendiendo al número de diferencias significativas en las comparaciones, los más relevantes son el conocimiento del *contexto* previo y actual del Servicio (χ^2 =6.000 p =.014, con recursos, e impacto social; χ^2 =5.000 p =.025, con procesos; y χ^2 =3.570 p =.059, con política y estrategia) y el apartado de *liderazgo* (χ^2 =7.00, p =.008, con política y estrategia; χ^2 =6.000 p =.014, con recursos; y χ^2 =5.000 p =.025, con impacto social). En el polo opuesto, los factores de *impacto social* (χ^2 =6.000 p =.014, con contexto del Servicio, y con satisfacción de usuarios; χ^2 =5.000 p =.025, con liderazgo), y de *gestión de recursos* (χ^2 =6.000 p =.014, con contexto del Servicio, liderazgo, y gestión del personal) son las menos relevantes.

(d) *Utilidad.* Las altas puntuaciones que reciben, en general, todos los factores de la guía, señalan su importancia potencial para la mejora del Servicio. Según el resultado de la comparación entre pares, ninguno de ellos es considerado como menos útil que los demás. Únicamente la puntuación de *liderazgo* se destaca con diferencias significativas en seis casos (χ^2 =6.000, p =.014, con política y estrategia, y recursos; χ^2 =5.000, p =.025, con satisfacción de usuarios, y satisfacción del personal; y χ^2 =4.000, p =.046, con gestión del personal, y procesos). Este dato ofrece un índice de la importancia que se otorga al papel de los responsables del Servicio en la mejora de la gestión.

(e) *Esfuerzo.* Todas las tareas relacionadas con el análisis de cada factor de gestión han supuesto un nivel de esfuerzo considerable, atendiendo a que las puntuaciones rondan en torno al punto medio de respuesta, y ninguna se aproxima al valor 1, que indicaría un nivel mínimo de esfuerzo. En concreto, el análisis de las cuestiones relacionadas con la *satisfacción de usuarios* es el que menor esfuerzo parece haber necesitado (χ^2 =5.000, p =.025, con procesos; y χ^2 =3.570, p =.059, con política y estrategia). Al contrario, el análisis de los *resultados generales* es la tarea que más empeño ha requerido (χ^2 =7.000, p =.008, con satisfacción de usuarios; χ^2 =6.000, p =.014, con política y estrategia, y satisfacción del personal; χ^2 =5.000, p =.025, con impacto social; y χ^2 =3.570, p =.059, con gestión del personal).

CONCLUSIONES

Tal como cualquier otro proceso, la evaluación necesita ser metaevaluada, con el fin de analizar los posibles problemas y descubrir soluciones para su mejora (Rossi y Freeman, 1989; Scriven, 1969, 1983; Stufflebeam y Shinkfield, 1987). No encontramos al respecto métodos convincentes en las guías de autoevaluación al uso, más allá de breves menciones sobre la conveniencia de revisar el proceso o la sugerencia de proceder mediante revisiones por pares (rechazada tajantemente por Cook - 1984- y por Trow -1999-). En consecuencia, se diseñó un cuestionario de metaevaluación compuesto por un conjunto de criterios dirigidos a conocer mejor la perspectiva de los participantes sobre el proceso de evaluación que finalizaba, y que ha tenido resultados satis-

factorios, permitiendo detectar ciertos problemas del proceso evaluativo, así como la necesidad reformular de algunos factores específicos de la guía.

Concordancia entre los participantes

Como han mostrado los resultados, existe cierta concordancia entre las valoraciones de los siete participantes en el estudio, aunque dista mucho de ser plenamente satisfactoria. En este punto se plantea la cuestión de qué puede haber ocurrido durante el proceso de autoevaluación para que el equipo de trabajo no haya logrado un nivel mayor de acuerdo. A pesar de haber colaborado juntos en el proceso, enfatizando el hecho de que se discutía cada una de las cuestiones a evaluar hasta alcanzar un acuerdo sobre su significado, es evidente que las valoraciones posteriores de los miembros de la Comisión de Calidad del Servicio difieren de las emitidas por los miembros del equipo técnico de la UCIE. El resultado es comprensible, supuesto que, de partida, los mapas conceptuales y las perspectivas de valor sobre lo que significa una gestión de calidad son diferentes en ambos grupos. Ello da una idea de las dificultades que los responsables de la administración universitaria pueden tener en la comprensión y aceptación de un modelo integral de calidad. En nuestro caso, los valores menores de acuerdo se dieron en los criterios de contexto, liderazgo, política y estrategia, gestión del personal, impacto social y resultados generales.

Relaciones entre los criterios de metaevaluación

La distribución de resultados indica cierta redundancia entre los criterios de metaevaluación empleados. La primera agrupación (comprensión-claridad) designaría criterios próximos al valor de objetividad, o ausencia de sesgos, y la segunda (relevancia-utilidad) al valor de responsividad (Chen, 1990). Ambos conjuntos están relacionados, indicando que la mayor comprensión y claridad de las cuestiones tiene que ver con una percepción de que resultan más relevantes y útiles para la mejora del Servicio. Además, el nivel de esfuerzo no está relacionado con variaciones en la relevancia y utilidad percibida, sino con que se formulen los contenidos de la guía de forma clara y comprensible para los participantes.

No estará de más recordar que la elección de los criterios de metaevaluación no respondía a una estructura teórica que pudiera ser validada con estos datos, por lo que las inferencias que se realicen son arriesgadas. Los presentes resultados indican, de modo tentativo, la importancia que puede tener la correcta formulación de las cuestiones de evaluación en el aumento de la responsividad, es decir, que la actitud negociadora debe complementarse con la calidad técnica de la evaluación, al menos en lo que concierne a la objetividad percibida que se gana cuando los participantes comprenden con facilidad lo que se les plantea (Chelimsky, 1983; Rebolloso, 1990).

Metaevaluación de los criterios de la Guia de Autoevaluación

En general, podemos considerar que los factores incluidos en la guía han resultado comprensibles, claros, relevantes y útiles, aunque el esfuerzo requerido para su análisis haya sido mayor de lo deseado. Por supuesto, queda un amplio margen para introducir múltiples correcciones y mejoras, pero en ningún caso parece que deban

ser rechazados o modificados por completo. A continuación señalaremos algunas observaciones recogidas durante el proceso previo de autoevaluación, con el objetivo de ilustrar y comprender mejor estos resultados. Téngase en cuenta también que los comentarios sirven para explicar la situación específica del Servicio evaluado, y que su capacidad de generalización está mermada por las propias limitaciones del estudio.

1. En primer lugar, se deben analizar los problemas de comprensión y claridad de tres de los factores de la guía:

(a) *Impacto social.* El concepto de comunidad local, utilizado para identificar un conjunto de beneficiarios externos del Servicio, es especialmente amplio, y se necesitan estudios que determinen los sectores o grupos sociales relevantes para la Universidad.

(b) *Procesos.* Quizá los problemas de comprensión se derivan de que el modelo de calidad representa un nuevo modo de gestionar las cosas de siempre. El sistema implica un cambio en las prácticas asentadas, en favor de las que preconiza un sistema de Gestión Integral de la Calidad (GIC; Rebolloso, 1999) -planificación exhaustiva mediante un método de investigación acción participativo-.

(c) *Resultados generales.* Los resultados del Servicio no llegaron a ser valorados, aduciendo que, por su escasa antigüedad, los datos recogidos podrían no reflejar con validez la realidad de los esfuerzos realizados durante el último año. Por otra parte, el Servicio no disponía de un sistema de evaluación del rendimiento, y tampoco la guía de autoevaluación facilita procedimientos para su creación.

2. En segundo lugar, deben destacarse, por su mayor relevancia y utilidad, los factores de *contexto del Servicio y liderazgo,* y por su menor relevancia, los de *impacto social y recursos.*

(a) *Contexto del Servicio.* La reciente creación del Servicio explica la importancia del contexto previo y de las decisiones políticas que se han tomado. Su estructura y funciones han sido definidas en este periodo, y muchas de ellas sólo se entienden cuando se presta atención a la situación anterior o a la Sección donde previamente se desarrollaban tareas que ahora están vinculadas orgánicamente a este Servicio. Sin embargo, a pesar de que el contexto sea fundamental para entender el Servicio, no es especialmente más útil para la mejora, como se aprecia en la comparación entre factores.

(b) *Liderazgo.* Los actuales responsables del Servicio han tomado parte decisiva en la definición de lo que es actualmente dicho Servicio (planificación, definición de puestos, reparto de funciones, asignación de recursos, etc.), y en asuntos como el nivel de participación y autonomía que se otorga al resto del personal. En consecuencia, el liderazgo es una cuestión de especial relevancia para comprender la situación actual. Además, la evaluación del liderazgo es la cuestión de mayor utilidad para el éxito de los planes de mejora, para definir y revisar el sistema de aseguramiento de la calidad, y para animar a todo el personal a que se implique en los esfuerzos para la mejora.

(c) *Impacto social.* Se realizan múltiples actividades dirigidas a la sociedad local, pero su objetivo en realidad es preparar el terreno para que los futuros usuarios internos del Servicio (futuros alumnos, familia-

res, contratantes, etc.) conozcan mejor las posibilidades que éste ofrece y el modo de beneficiarse de ellas. Sin embargo, no existe conciencia del potencial que tiene el Servicio para realizar actividades que apoyen el desarrollo local. En cierto modo, no se considera la comunidad como un receptor de servicios, sino como un intermediario con los futuros usuarios. Quizá ésta sea la razón por la que también se considera reducida la importancia del impacto social en la mejora del Servicio.

(d) *Recursos.* Los servicios universitarios tienen limitada la función de gestión de los recursos, que recae en la Gerencia y el Rectorado. Sólo está en sus manos la posibilidad de demandar recursos de distinto tipo en función de las necesidades internas detectadas; ni siquiera entablan relaciones con los proveedores, quienes contratan directamente con las instancias mencionadas. Por esta razón, la gestión de recursos, tal como se plantea en la guía, resulta ajena a la realidad del Servicio, y es irrelevante para comprender la situación actual y las prácticas de gestión internas. Otra cosa es que parezca útil para la mejora, y hemos visto que no se la considera de modo inferior a otros factores en este punto.

3. En cuanto al esfuerzo requerido, ya durante la autoevaluación se constataron repetidamente los problemas de los participantes para compaginar sus responsabilidades internas con el trabajo de evaluación. Se debe reducir el nivel de esfuerzo en todos los factores de la guía (modificando el procedimiento o los requisitos de trabajo). En concreto, el criterio de *satisfacción de usuarios* fue el que menor esfuerzo exigió, mientras que el más laborioso fue el criterio de *resultados generales.*

(a) *Satisfacción de usuarios.* La UCIE realizó el estudio de satisfacción de los usuarios y del personal profesional (Rebolloso, Fernández Ramírez, Pozo y otros, 2000), y el equipo de autoevaluación se detuvo poco tiempo analizando este apartado, ya que bastaba con describir mínimamente las pocas experiencias previas y con analizar, grosso modo, los resultados del estudio de satisfacción. Evidentemente, en otros casos el esfuerzo requerido dependerá del procedimiento y de la instrumentación que se desarrolle para obtener información sobre los usuarios.

(b) *Resultados generales.* Deben tenerse en cuenta dos problemas. Primero, que se emplearon varias sesiones al comienzo del proceso de autoevaluación para recoger la información sugerida en la guía del Consejo de Universidades (1998); como esta información no llegó finalmente a utilizarse, la percepción de esfuerzo puede haberse incrementado por esta razón. Segundo, que el sistema de indicadores mínimos desarrollado posteriormente requirió recoger, de forma rápida, una información que estaba más o menos accesible en muchos casos, pero que exigió cierto nivel de elaboración y organización, incrementando la percepción del esfuerzo empleado.

LIMITACIONES Y PERSPECTIVA DE FUTURO

Ante todo, debemos reconocer las limitaciones del estudio, teniendo en cuenta sobre todo el tamaño reducido de la muestra y las dificultades que conlleva el análisis de datos. La situación ideal es la realización de la metaevaluación una vez que el sistema de calidad se haya extendido a todas las unidades de la estructura de gestión.

El estudio ha sido un primer ensayo para poner a prueba los instrumentos y el procedimiento de evaluación, sin mayor pretensión que mejorarlos de forma paulatina y reflexiva. Por tanto, las conclusiones no tienen ánimo de generalización. Al contrario, reconocemos que su utilidad principal es mejorar el proceso específico de implantación del sistema de calidad en el contexto de nuestra universidad, sin despreciar la posibilidad de que el estudio resulte ilustrativo para que otros investigadores se animen a la realización de metaevaluaciones como una actividad de autoanálisis crítico dirigida a la mejora de su propio trabajo.

Con todo, durante el proceso de metaevaluación han aparecido dos problemas importantes que evidencian la necesidad de mejorar el procedimiento. En primer lugar, los criterios de metaevaluación podrían reducirse en futuros ensayos, dada la proximidad conceptual y empírica entre los mismos (comprensión-claridad; relevancia-utilidad). En segundo lugar, debería ensayarse la metaevaluación en paralelo, a lo largo de todo el proceso de autoevaluación, con la ayuda, por ejemplo, de algún cuestionario o formulario de registro que se cumplimentara al finalizar cada sesión de evaluación, o al finalizar el análisis de cada factor de calidad. Esta sugerencia se deriva de que resultó problemático el esfuerzo que supuso volver a repasar todos los elementos de la autoevaluación, una vez que ya parecía cercana la finalización del proceso, muy deseada, según los participantes expresaban tras cinco meses de trabajo no recompensado (debemos suponer que por esto algunos no devolvieron el cuestionario).

Independientemente de las realidades políticas y de los intereses creados en el seno de la organización universitaria, estamos obligados a defender un ideal comprometido con la mejora y el fortalecimiento de las personas y las instituciones (Blanco, 1998; Miller, 1969; Rebolloso, 1999; Stake, 1975). Conscientes de la importancia de los cambios que requiere la implantación extensa de un sistema de aseguramiento de la calidad, y de que cualquier cambio es traumático en mayor o menor grado, no podemos sino aconsejar la prudencia, reflexión y habilidad política para buscar el necesario consenso entre los grupos de interés de la institución, e iniciar un lento período de cambios (Rebolloso, 1999).

Este primer ensayo de autoevaluación ha servido para calibrar el procedimiento y los instrumentos de evaluación, así como para realizar una primera aproximación a los sistemas de gestión que se emplean en los Servicios de administración universitaria. No obstante, los resultados ofrecen una visión insuficiente para comprender la realidad de todo el sistema y para aportar sugerencias generalizables en la situación actual. En consecuencia, es necesario continuar con esta labor parcial, extendiendo la evaluación a otros Servicios para ampliar la información disponible sobre el contexto de implantación del sistema de calidad.

Es aconsejable, por último, que las nuevas evaluaciones estén precedidas de un tiempo para discutir y negociar con los Servicios que en un futuro participen, de tal modo que se reduzca el esfuerzo requerido y se realice un acercamiento más válido a la realidad de la organización. En cualquier caso, es evidente la necesidad de realizar metaevaluaciones sistemáticas, que pueden partir de un cuestionario como el empleado en este estudio o ser ampliadas siguiendo las recomendaciones de múltiples autores (Chelimsky, 1983; Chen, 1988,1990; JCS, 1981,1988; Rebolloso, 1994b; Scriven, 1969).

REFERENCIAS BIBLIOGRÁFICAS

Blanco, A. (1998). La coherencia en los compromisos. En I. Martín-Baró, *Psicología de la liberación.* Madrid: Trotta.

Chelimsky, E. (1983). The definition and measurement ofrece evaluation quality as a management tool. *New Directions for Program Evaluation,* 18, 113-126.

Chelimsky, E. (1997). The coming transformations in Evaluation. En E. Chelimsky y W. R. Shadish (Eds.). *Evaluation for the 21st century. A handbook* (pp. 1-26). Thousand Oaks, CA: Sage.

Chen, H. T. (1988). Validity in evaluation research: A critical assessment ofrece current issues. *Policy and Politics,* 16, 1-16.

Chen, H. T. (1990). T*heory-driven evaluations.* Newbury Park, CA: Sage.

Choi, T. Y. y Eboch, K. (1998). The TQM paradox: Relations among TQM practices, plant performance, and customer satisfaction. *Journal ofrece Operations Management,* 17, 59-75.

Conner, R. F., Altman, D.G. y Jackson, C. (1984). 1984: A brave new world for evaluation. En R.F.

Conner, D. G. Altman y C. Jackson (Eds.). *Evaluation Studies Review Annual* (Vol. 9). (pp. 12-22). Beverly Hills, CA: Sage.

Consejo de Universidades (1998). *Plan nacional de evaluación de la calidad de las universidades.* Madrid: Secretaría General del Consejo de Universidades.

Cook, T. D. (1984). Opportunities for evaluation in the next few years. En R.F. Conner, D.G. Altman y C. Jackson (Eds.). *Evaluation Studies Review Annual* (Vol. 9). (pp. 726-751). Beverly Hills, CA: Sage.

Cook, T. D. (1997). Lessons learned in evaluation over the past 25 years. En E. Chelimsky y W.R. Shadish (Eds.). *Evaluation for the 21st century. A handbook* (pp. 30-52). Thousand Oaks, CA: Sage.

Cronbach, L. J. (1982). In praised ofrece uncertainty. En P.H. Rossi (Ed.). *Standards for evaluation practice* (pp. 49-58). San Francisco, CA: Sage.

Cronbach, L. J. y asoc. (1980). *Toward reform ofrece program evaluation.* San-Francisco, CA: Jossey-Bass.

De Cock, C. y Hipkin, I. (1997). TQM and BPR: Beyond the beyond myth. *Journal ofrece Management Studies,* 34, 659-675.

Deming, W.E. (1986). Japan: Quality control and innovation. *Business Week,* 2697, 18-44.

European Foundation for Quality Management (1996/1997). *Premio europeo a la calidad.* Madrid: Club de Gestión de la Calidad.

European Foundation for Quality Management (1999). *Modelo EFQM de excelencia.* Madrid: Club de Gestión de la Calidad.

Evaluation Research Society (1981). *Standards for program evaluation.* Potomac, MD: Evaluation Research Society.

Fabi, B. y Pons, 0. (1995). Los círculos de calidad: la experiencia internacional. *Alta Dirección,* 182, 275-282.

Fernández Dols, J. M. (1990). *Patrones para el diseño de la Psicología social.* Madrid: Morata.

Glass, G. V. y Ellet, F. S. (1980). Evaluation research. *Annual Review ofrece Psychology,* 31, 211-228.

Guba, E. G. (1967). Evaluation and the process ofrece change. En *Notes and worhng papers concerning the administration ofrece programs: Title III ESEA.* Washington DC: U.S. Senate Committee onLabor and Public Welfare.

Guba, E. G. y Lincoln, Y. S. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

Joint Committee on Standards for Educational Evaluation (1981). *Standards for evaluations ofrece educational programs, projects and materials.* New York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation (1988). *The personnel evaluation standards: How to assess systems for evaluating educators.* Newbury Park, CA: Sage.

Kaplan, A. (1964). *The conduct ofrece inquiry.* San Francisco, CA: Chandler.

Lincoln, Y. S. y Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. En D.D. Williams (Ed.). *Naturalistic Evaluation. New Directions for Program Evaluation* (Vol. 30). San Francisco, CA: Jossey-Bass.

Ministerio de Administraciones Públicas (1999). *Guia de autoevaluación para la administración pública.* Madrid: MAP/BOE.

Miller, G. A. (1969). Psychology as a mean ofrece promoting human welfare. *American Psychologist,* 24, 1063-1075.

Oakland, J. S. (1993). *Total quality management. The route to improving performance.* Oxford: Butterworth-Heinemann.

Patton, M. Q. (1994). Developmental evaluation. *Evaluation Practice,* 15, 311-320

Patton, M. Q. (1997). *Utilization-focused evaluation.* Newbury Park, CA: Sage.

Peiró, J. M. y González Romá, V. (1993). *Círculos de calidad.* Madrid: Eudema.

Rebolloso, E. (1990). Algunos apuntes sobre la utilidad de la evaluación en el marco de la toma de decisiones políticas. *En Actas del III Congreso Nacional de* Psicología *Social.* (pp. 567-574). La Coruña: Tórculo.

Rebolloso, E. (1994a). La evaluación del personal profesional: ¿Un modelo de evaluación nuevo o una síntesis de modelos previos? *Revista de Psicología Social, 2,* 231-255.

Rebolloso, E. (1994b). Los standards de evaluación de la Joint Committee. Su utilidad como guía para la práctica profesional en educación y servicios sociales. *Revista de* Psicología *Social aplicada, 4,* 63-91.

Rebolloso, E. (1999). *La evaluación de la calidad como estrategia de supervivencia y futuro de la universidad.* Almería: Universidad de Almería.

Rebolloso, E., Fernández Ramírez, B., Cantón, P. y Pozo, C. (2000). El papel de la investigación cualitativa en la evalua-

ción de los servicios universitarios. *Cuadernos IRC, 4,* 65-82.

Rebolloso, E., Fernández Ramírez, B., Pozo, C. y Cantón, P. (2001). *Estrategias de calidad en la Universidad. Guia para la autoevaluación de los servicios de administración universitaria.* Valencia: Promolibro.

Rebolloso, E., Fernández Ramúez, B., Pozo, C. y Peñín, R. (1999). *Autoevaluación del Servicio de Alumnos -1999. Informe de resultados.* Almería: UCIE.

Rebolloso, E., Fernández Ramírez, B., Pozo, C. y otros (2000). *Gestión Integral de la Calidad. Satisfacción de los usuarios y del personal profesional con los servicios universitarios.* Almería: UCIE.

Rebolloso, E. y Rebolloso, J.R. (1998). Significado y desarrollo actual de la evaluación de programas. En E. Rebolloso (Ed.). *Evaluación de programas. Ambitos de aplicación* (pp. 219-240). Barcelona: Textos Universitarios Sant Jordi.

Rossi, P. H. y Freeman, H. E. (1989). *Evaluación. Un enfoque sistemático para programas sociales.* México: Trillas.

Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report, 2,* 36-38.

Scriven, M. (1983). Evaluation ideologies. En G.F. Madaus, M. Scriven y D.L. Stufflebeam (Eds.). *Evaluation models* (pp. 229-260). Boston, MA: Kluwer-Nijhoff.

Scriven, M. (1991). Beyond formative and summative evaluation. En G.W. McLaughlin y D.C. Philips (Eds.). *Evaluation and education: At quarter century.* Chicago, IL: University ofrece Chicago Press.

Scriven, M. (1996). Types ofrece evaluation and types ofrece evaluators. *Evaluation Practice, 17,* 151-61.

Seoane, J. (1996). El escenario postmoderno de la psicología social. En G. Collier, H. L. Minton y G. Reynolds (Autores). *Escenarios y tendencias de la* psicología *social* (pp. 13-30). Madrid: Tecnos.

Shadish, W. R., Cook, T. D. y Leviton, L. C. (1991). *Foundations of program evaluation.* Newbury Park, CA: Sage.

Sonnichsen, R. C. (2000). High impact internal evaluation: A practitioner's guide to evaluating and consulting inside organizations. Thousand Oaks, CA: Sage.

Stake, R. E. (1975). E*valuating the arts in education: A responsive approach.* Columbus, OH: Merril.

Stufflebeam, D. L. (1975). *Metaevaluation.* Kalamazoo, MI: Evaluation Center, Western Michigan University.

Stufflebeam, D. L. y Shinkfield, A. J. (1987). *Evaluación sistemática.* Barcelona: Paidós.

Trow, M. (1999). Confianza, mercados y rendición de cuentas en la educación superior. *Cuadernos IRC, 2,* 47-63.

VanVught y otros(1993). *Gestión de la calidad y garantia de calidad en la enseñanza superior europea: Métodos y mecanismos.* Bruselas: Oficina de Publicaciones de las Comunidades Europeas.

Weiss, C. H. (1990). I*nvestigación evaluativa.* México: Trillas.

Westphal, J. D, Gulati, R. y Shortell, S. M. (1997). Customization or conformity? An institutional and network perspective on the content and consequences ofrece TQM adoption. *Administrative Science Quarterly, 42,* 366-394.

Ziman, J. (1981). *La credibilidad de la ciencia.* Madrid: Alianza.