

La *apropiación* de la evaluación por las administraciones¹

Saville Kushner

Resumen: *En los últimos treinta años venimos asistiendo a una progresiva “internalización” de la evaluación de programas y políticas en las administraciones públicas. Esto se ha producido bien porque las administraciones mismas han asumido la función de evaluación o porque éstas han estipulado cómo hacer las evaluaciones y con qué objetivos. Este hecho plantea preguntas importantes para el ciudadano y para la democracia. En este artículo se examinan las consecuencias de esta “apropiación” de la evaluación por parte de la administración a través de dos ejemplos: uno tomado del ámbito de la cooperación internacional y otro extraído del encargo que una agencia gubernamental hizo para la evaluación de uno de sus programas.*

Palabras clave: *Independencia de la evaluación, evaluación de programas, política.*

1. A MODO DE INTRODUCCIÓN

En su trabajo sobre “la inversión del ejecutivo en evaluación”, MacDonald (1981) planteaba la cuestión de qué es lo que gana la administración con la contratación de evaluaciones, si son actividades potencialmente peligrosas para la burocracia política. Los observadores profesionales (los evaluadores), que a menudo dan mucho valor a su independencia, obtienen acceso directo a lo que a menudo son fuentes confidenciales de información y además tienen capacidad de publicarlo. Por supuesto, los evaluadores proporcionan lo que puede ser información útil, pero dicha información no siempre compensa los riesgos políticos. Sin embargo, lo que es más importante, como dice MacDonald, es que sólo el racionalista piensa que un programa social es un medio para lograr fines políticos (lo que convierte la evaluación en algo útil en el sentido democrático y en el sentido formativo). A menudo, la realidad es que el programa suele ser un fin en sí mismo –con fines políticos sin consecuencias más allá

1 Revisión científica de la traducción a cargo de Ester García Sánchez.

de su fecha de finalización-, en cuyo caso la evaluación es sólo “simbólica... un elemento esencial de la decoración de cualquier ministerio que se precie”.

Aquí tenemos, pues, las razones principales por las que, para MacDonald, las administraciones subvencionan las evaluaciones; el reto psicológico, al leer esto, es mantener un sólido equilibrio entre el cinismo y el realismo:

- No hay opción (legislación, órdenes, presión, fondos especiales).
- La única alternativa puede ser un nivel *no* aceptable de trabajo común y consulta con los *stakeholders* que, a menudo, tienen agendas incompatibles.
- Mantiene a los escépticos –críticos potenciales o reales– ocupados en tareas “inofensivas”.
- La presencia de evaluadores especializados evita, o al menos pospone, una evaluación más exhaustiva por parte de evaluadores no especializados.
- Hace ganar tiempo: tiempo para que el gobierno disipe el interés que hizo necesaria la inversión en el programa.
- El gobierno puede requerir, sobre todo en los primeros momentos, un cierto nivel de implantación del programa y la presencia de evaluadores contendrá el comportamiento subversivo del colectivo que ha recibido los fondos.
- Es una señal para los escépticos de que el gobierno quiere conocer la verdad, se preocupa por la eficacia y está abierto a las evidencias.
- Con toda seguridad, resultará inofensiva gracias a la combinación de debilidades epistemológicas, impopularidad social y al propio interés de los evaluadores.
- La mejor razón de todas: alguna otra entidad puede financiar una evaluación de su programa.

Continúa mostrando cómo la administración puede protegerse frente a los riesgos de encargar una evaluación (incluso, por ejemplo, “retrasar la evaluación porque aumenta su necesidad...”). Aquí MacDonald no está argumentando que la evaluación no deba servir a la administración, sino que no debería servir *exclusivamente* a la administración y que mientras esté prestando el servicio, sería bueno que el evaluador conociera las intenciones que hay tras su contratación.

En la época en la que MacDonald escribió esto, la evaluación era una actividad que se estaba extendiendo pero que aún estaba poco desarrollada y se llevaba a cabo sobre todo en las universidades y por parte de académicos independientes. Este au-

tor se refería a la evaluación *externa*. Desde entonces, ha habido una tendencia a nivel mundial para la internalización de la evaluación, para incorporarla a las administraciones como una actividad propia. Según el análisis de MacDonald, esto permite controlar todos los riesgos y centrarse en los beneficios, independientemente de dónde se enmarcan. Hay dos vías principales de internalización: la de los organismos que cuentan con sus propios evaluadores y la de aquellos organismos que encargan evaluaciones externas a través de un contrato en el que se establecen unas determinadas condiciones a partir de sus necesidades de gestión de la información.

Lo que pretendo con este artículo es analizar dos de los efectos de esta internalización: uno, en el campo de la cooperación al desarrollo², en el que veremos cómo la preocupación de los gobiernos por la *gestión basada en resultados* “aplata” el compromiso democrático y participativo durante la evaluación; otro, más cercano, es un caso de contratación de una evaluación por el gobierno del Reino Unido, en el que vemos cómo la internalización de la evaluación ha conducido a un marco muy rígido de cláusulas que han predeterminado la evaluación y limitado sus objetivos, haciendo que la evaluación fuera menos eficaz y democrática. Claramente, lo que me preocupa es la utilidad de la evaluación para la democracia. Sin embargo, primero necesitamos entender la evaluación en los términos más amplios y democráticos de sus primeros años. Últimamente, a medida que la evaluación se ha ido internalizando como un elemento para la gestión y la reforma en las administraciones públicas, sus distintas formas y enfoques se han ido restringiendo, cada vez más, a la medición *post hoc* de la productividad de los programas. Esta situación oculta el hecho de que los 25 años transcurridos desde la aparición de la evaluación de programas como una práctica y una disciplina diferenciada han visto cómo la evaluación adquiría facetas muy variadas, útiles y definidas. Esto se refleja en la amplia gama de denominaciones³ que se han usado para los distintos enfoques, cada una de ellos (enumerados a continuación) con unos rasgos característicos precisos⁴:

- Evaluación democrática (MacDonald)
- Evaluación libre de objetivos (Scriven)
- Evaluación comprensiva (Stake)
- Evaluación iluminativa (Parlett and Hamilton)
- Evaluador como científico público (Cronbach)
- Evaluación naturalista (Guba)
- *Empowerment evaluation* (Fetterman)
- Modelo CIPP - Contexto, Input, Proceso, Producto (Stufflebeam)
- *Dialogic evaluation* (Greene)
- *Affirmative evaluation* (Mertens)

2 En el original *international development*.

3 En la revisión científica se ha optado por dejar en su original en inglés aquellas denominaciones que no cuentan en castellano con una traducción suficientemente acuñada.

4 Véanse House (1980), para una clasificación de los enfoques de evaluación y un análisis de las tendencias en el surgimiento de la evaluación, y Alkin (2004) sobre el “árbol de la evolución” de la evaluación.

- Evaluación para la utilización (Patton)
- Evaluación personalizada (Kushner)
- Evaluación democrática deliberativa (House)

Cabe destacar, de paso, que la mayoría de estos autores hablaban de *evaluación educativa* y uno de los ejemplos que expondré a continuación pertenece a dicho campo. No es casual. En gran parte, fue en el campo de la educación, concretamente en el de la enseñanza y el aprendizaje⁵, donde se pusieron los cimientos de la evaluación de programas. Esto se debió, en cierta medida, a que éste fue el laboratorio en el que el gobierno hizo los primeros encargos de evaluación (Norris, 1990; House, 1993)

Para apreciar cómo se ha concebido la evaluación fuera de las “limitadas” definiciones dadas por las administraciones, Cronbach *et al.* (1980), en el famoso prefacio de su importantísima obra sobre evaluación de programas, establecieron 95 “tesis” o principios sobre evaluación, unos postulados valiosos y fundamentales, todo un manifiesto. Fueron escritas en la misma época en la que MacDonald se preguntaba por el uso de la evaluación por la burocracia y una inquietud similar subyace en el trabajo de Cronbach. Él hablaba del evaluador como “científico público”. El primero de estos 95 principios dice, simplemente: “La evaluación es el medio por el cual la sociedad aprende sobre sí misma”.

El sentido de esto es que cada evaluación de un programa es un estudio de caso sobre la sociedad, en la medida en que todos los programas son una representación, en miniatura, de la sociedad en la que se desarrollan. Tienen estructuras sociales de poder, una cultura, funciones y rituales, una economía y una teoría de la justicia. Tanto si el evaluador lo reconoce como si no, tanto si lo contempla su contrato como si no, es lo que sucede y ello plantea muchos interrogantes, que después han sido explorados en profundidad, sobre el carácter democrático de los programas y de su evaluación; por ejemplo, ¿quién tiene derecho a disponer de información sobre los programas y sobre su funcionamiento?

Con ésta y otras preguntas similares en mente analizaré a continuación mis dos ejemplos del impacto de la internalización de la evaluación. El primero tiene que ver con la situación de la evaluación en la cooperación al desarrollo y comienza con una historia personal.

2. PRIMER EJEMPLO: “DENTRO DE LA TIENDA DE CAMPAÑA” DE LA EVALUACIÓN DEL DESARROLLO

Cuando me estaban entrevistando para el puesto de responsable regional de control y evaluación de UNICEF que después llegaría a ocupar, me encontré metido

5 En el original *schooling*.

en una teleconferencia con cuatro países: Brasil, Panamá, Nicaragua y Estados Unidos. Por algún motivo parecía importante preguntar con qué país estaba hablando en cada momento. La pregunta más penetrante vino de Nueva York. “Interesante currículum vitae –muy creativo-, qué extenso. Pero durante 25 años ha estado usted trabajando, desde fuera de las administraciones públicas, en el análisis de lo que sucede dentro de ellas; ahora quiere trabajar dentro de la maquinaria burocrática más grande del mundo, para analizar lo que sucede fuera. ¿Cómo logrará ese cambio de mentalidad?”. Simplemente respondí que no tenía la menor idea, pero que era consciente del reto –que se trataba de una cuestión práctica– y que tendría que resolver ese problema cuando me enfrentara a él. Valoraron mi sinceridad.

Sin embargo, no respondí a todas las implicaciones de la sabia pregunta que luego tuve que encarar. La verdadera pregunta era: ¿qué sucede cuando la evaluación entra en la “tienda de campaña”?

Naciones Unidas es una organización muy fragmentada. Tiene muchos organismos que diversifican sus esfuerzos, por ejemplo, entre la ayuda humanitaria, el desarrollo cultural, el SIDA, los derechos del niño, el cambio climático, la reducción de la pobreza, las migraciones y la producción y distribución de alimentos; mientras, los que habitan en el hemisferio sur se ven afectados por todas estas cuestiones a la vez. Cambio climático, problemas de acceso a la escuela, migración, escasez de alimentos, enfermedades; se pueden padecer todos a la vez. No es de extrañar que exista una gran presión dentro de la ONU –y entre sus donantes– para la armonización, la normalización y la integración. Es lógico que el programa estrella de reforma se llame *One UN*.

Dentro de los propios organismos existe también una fragmentación similar. En mi oficina regional había expertos en pediatría, en educación, en políticas públicas, en protección de la infancia y en otros temas -y desarrollaban sus actividades de forma independiente. De hecho, había cierta reticencia a integrarlas. “¿Me han contratado como pediatra!” me contestó, protestando, uno de estos asesores cuando intenté sugerirle que todos ellos, en conjunto, debían gestionar el cambio. Era una respuesta comprensible pero disfuncional, teniendo en cuenta que los problemas sociales están interrelacionados.

Quizás, lo que produce una mayor sensación de fragmentación es la “geografía”; recordemos esa teleconferencia. La sede central de UNICEF en Nueva York, por ejemplo, está a una enorme distancia de sus siete oficinas regionales, a lo que se suma la distancia entre éstas y sus más de cien oficinas en distintos países; y de éstas a las miles de oficinas de campo en los centros de operaciones. La cuestión que se le plantea a la organización es cómo interpretar esas distancias. Puede hacerse de dos modos: como un problema de comprensión (aprender de la diversidad) o como un problema de control (disminuir o atenuar las diferencias). Es la típica cuestión del centro y la periferia. La sede central tiene dos preocupaciones: cómo aprender de las operaciones que llevan a cabo sobre el terreno y cómo asegurarse de que su política

se sigue de una manera coherente, de modo que la voz de UNICEF sea también coherente. Ambas cuestiones son reales e ineludibles y ambas, en mayor o menor medida, se asumen. De hecho, UNICEF es conocida como la agencia más descentralizada dentro de la ONU, en la que los representantes por países gozan de una autonomía mayor que la de, por ejemplo, sus colegas del PNUD –la mayor agencia dentro de la ONU– que gestiona los problemas geográficos de un modo más jerarquizado. La ONU (sí es que se puede hablar de una sola ONU) es, en términos de gestión, un organismo extremadamente heterogéneo y singular.

Sin embargo, en la actualidad se está volviendo al centralismo, auspiciado por un proceso de reforma en la ONU. De hecho, al igual que las demás agencias de la ONU, UNICEF debe cumplir con un nuevo sistema de gestión basado en resultados, una versión del *New Public Management* (resultados previamente especificados dentro de un sistema que exige responsabilidades pero dota a los responsables de escasa capacidad de decisión –*low-trust accountability*– y que asigna los recursos en función del rendimiento). Este sistema, como bien sabemos los que trabajamos en el sector público en el hemisferio norte, es un método para centralizar el poder y la autoridad en las organizaciones. Cuando me marché de UNICEF, la sede central estaba poniendo en marcha una nueva política que cambiaba el papel de los (poderosos) Directores Regionales de asesores e impulsores a asesores y *supervisores*: se esperaba que fueran más exigentes con determinado tipo de resultados y que fueran más sensibles a las comparaciones en cuanto al avance en el logro de los Objetivos de Desarrollo del Milenio. Se mantiene la autonomía, pero supeditada a la preocupación por garantizar una mayor homogeneidad. En mi opinión, la balanza se está inclinando (por las presiones de los países donantes) a favor de una agenda globalizada (igualdad de género, violencia infantil, medidas de supervivencia infantil, gripe aviar) y a costa de la diversidad de las operaciones en los distintos países.

Un síntoma de este cambio es el llamamiento a la “armonización” y a la “estandarización”. Hay tres razones principales para ello, dos de ellas son obvias y otra es más sutil. Una de las obvias es que a menudo hay maneras alternativas de medir las “patologías” sociales y es importante ponerse de acuerdo sobre cuál se utiliza. La desnutrición a menudo se analiza de dos formas: retraso en el crecimiento y atrofia. El retraso en el crecimiento se calcula midiendo la estatura con respecto a la edad; la atrofia se calcula midiendo el peso con respecto a la estatura. Estos indicadores arrojan unos resultados muy distintos. Por supuesto, pueden utilizarse en conjunto (y debería hacerse así porque son indicadores que se modulan mutuamente), pero a veces también se usan como una opción para llegar a las conclusiones políticas más convenientes según las circunstancias. Para evitarlo, se fomenta la “armonización” de los indicadores como vía para controlar la politización de la información. El otro ejemplo más publicitado y notorio es la incidencia del VIH/SIDA en una población determinada. En el Reino Unido, por ejemplo, hace mucho tiempo que dejamos de hacer extrapolaciones de las estadísticas obtenidas de las pruebas realizadas a mujeres en clínicas obstétricas, una muestra “cautiva” muy útil. La falta de fiabilidad se debe a razones obvias (por ejemplo, estábamos contabilizando únicamente a las mujeres

sexualmente activas y a mujeres que se presentaban voluntariamente) y las estadísticas antes estaban muy infladas. En África, estas metodologías y las estimaciones “infladas” sobre la prevalencia de la enfermedad en esos países persistieron durante muchos años; ¿recuerdan las cifras que se manejaban de un tercio de la población infectada? Había resistencia a cambiar la metodología, entre otras cosas porque las elevadas tasas de incidencia de la enfermedad estimulaban un flujo mayor de fondos “excepcionales”⁶.

La otra razón obvia para la armonización es que hay una necesidad impuesta por los burócratas de comparar unos países con otros, al igual que las distintas zonas o regiones del planeta. Esto forma parte del sistema global de *accountability* que ha propiciado la aparición de un fenómeno histórico único: los Objetivos de Desarrollo del Milenio. Independientemente de que existen realidades globales que superan contextos y culturas, el hecho es que la supervisión por parte de la ONU de determinadas responsabilidades a nivel global –en derechos humanos, cambio climático, conflictos violentos, etc., por dar algunos ejemplos– ha supuesto un impulso para el progreso moral y la reducción del caos (Black, 2008).

La razón sutil es la que más me preocupa y tiene que ver con el “control administrativo”. Los que trabajan en las administraciones no pueden permitirse los mismos lujos que nosotros, los evaluadores, de documentarse, teorizar y debatir sobre la complejidad de éste o aquel problema: tienen que trabajar con ese problema y resolverlo; ése fue el principal cambio de mentalidad que experimenté. Y además han de resolverlo en un contexto en el que deben rendir cuentas a los donantes por los impactos (atribuibles y mensurables). Hay que lograr la cuadratura del círculo; el deseo de guardar las “formas” dentro del proceso –*input*, proceso causal, *output*, impacto– es abrumador.

A veces esas “formas” se corresponden con la realidad, sobre todo cuando los problemas se pueden tratar con soluciones técnicas: porcentaje de niños vacunados, número de nacimientos registrados, cobertura con mosquiteras impregnadas de repelente, ingesta de sal yodada, proximidad a agua potable; todos los ejemplos anteriores tienen un impacto más o menos atribuible y mensurable sobre la salud y los derechos de las personas. Hay muchas otras cosas en el ámbito de la cooperación al desarrollo que no lo tienen, aún más desde que empezó a producirse un nuevo cambio que nos ha llevado de las operaciones de campo a la defensa de determinadas políticas en las que los impactos son dudosos y, en el mejor de los ca-

6 Véanse, por ejemplo: (i) M. Cheng, “AIDS cases drop, but bad data to blame”, *The Washington Post*, 20 de noviembre de 2007; (ii) J. Chin (2007), *The AIDS Pandemic: the Collision of Epidemiology with Political Correctness*, Oxford: Radcliffe; (iii) J.R. Glynn, A. Buve, M. Carael, R. Musonda, M. Kahindo, I. Macculey, M. Hawken, F. Tembo, L. Zekeng (2000), “A new method of estimating HIV prevalence in men using antenatal clinic sentinel surveillance”, ponencia presentada en la *Conferencia Internacional sobre el SIDA*, 9-14 de julio de 2000; 13: abstract núm. ThOrC669; (iv) C.A. McGarrigle, A. Cliffe, A.J. Copas, C.H. Mercer, D. DeAngelis, K.A. Fenton, B.G. Evans, A.M. Johnson y O.N. Gill (2006), “Estimating adult HIV prevalence in the UK in 2003: the direct method of estimation”, *British Medical Journal*, 82(3): 78-86.

sos, *inferidos* y contextualizados. En este caso, la evaluación como análisis de contextos es esencial, al igual que lo es utilizar esa evaluación para poner nuevamente en contacto a los políticos con la realidad de cada comunidad. Cuanto más nos alejamos de las soluciones técnicas en los problemas muy específicos (como la vacunación), más nos adentramos en un territorio incierto en cuanto a los resultados, respecto a lo que significan para los distintos *stakeholders*, cuáles son sus causas, lo que representan en términos de calidad y, lo que es más importante, hasta qué punto dependen de los contextos de la actuación. “Al final, la evaluación precisa de la cadena de hechos y datos y de las interpretaciones de quienes tienen experiencia en el campo correspondiente, es insustituible”, afirmaba Deaton (2008) al cuestionarse la enorme difusión de los enfoques experimentales que pueden ser útiles para intervenciones muy concretas, pero que no logran reflejar la complejidad de los programas sociales.

En estos casos, necesitamos la evaluación para lo que con excesiva frecuencia se consideran funciones accesorias: aprendizaje, intercambio de información, descripción, deliberación (debate) o para dejar claro cuál es el contexto. Cuando las causas no están claras o son difíciles de averiguar necesitamos la evaluación para conocer las contingencias; no se trata tanto de saber cómo A da lugar a B sino más bien de saber cómo A y B tienen una relación dinámica entre sí, pero también con X, Y y Z. Picciotto (2007) defiende que la evaluación ha de situar las intervenciones en cooperación al desarrollo dentro de su contexto político y económico, sacarlas del enfoque miope que se centra en la propia intervención y sus objetivos.

La dificultad que se nos plantea es que la apropiación de la evaluación por las administraciones tiende a que se devalúen estas cosas en favor de funciones supuestamente cruciales y por las que el burócrata es responsable: la medición de resultados. Como me comentó en cierta ocasión un alto funcionario de UNICEF, “el seguimiento nos dice si vamos bien encaminados; la evaluación nos dice si hemos llegado”. Suena bien, pero es un sofisma. Dado que los Objetivos Globales del Milenio definen puntos de destino universales, podemos ver la lógica que subyace al deseo de armonizar y estandarizar la evaluación. A esto se suma el deseo de incrementar y mantener niveles de competencia y calidad en la práctica evaluadora, defender los logros en el desarrollo de capacidades que no han sido fáciles de alcanzar, consagrándolos como estándares o normas e insistiendo en el compromiso de aprender de “los mejores”.

El problema es la tensión entre la estandarización y el contexto: lo primero niega lo segundo. La urgencia para asegurar que se ponen en práctica las políticas y se aseguran los impactos tiende a desplazar las funciones democráticas de la evaluación: estimular la crítica y el debate social; hacer que los programas de intervención sean transparentes; facilitar el intercambio de información; dar voz a la ciudadanía; proporcionar datos públicos como base para el diálogo entre la sociedad y la administración. Estas funciones sólo se pueden realizar en su contexto, *estas* personas en *este* lugar hablando entre ellas de sus preocupaciones en *este* preciso momento. Como he

sugerido en algún otro momento, la evaluación democrática fue concebida a imagen de la política municipal, local (Kushner, 2009). En el nivel municipal las personas se conocen, hablan cara a cara, cuentan con pocos recursos para desarrollar una ciencia muy sofisticada, la rendición de cuentas se basa en una conciencia de las relaciones humanas, la política se basa en el conocimiento del lugar y en las preocupaciones de sus ciudadanos. Todo esto se ve amenazado por la armonización. La estandarización es enemiga del contexto.

No es que esto no se reconozca; desde luego en UNICEF lo saben. Hay un intenso debate sobre la democracia y la evaluación. El problema es que debe verse en paralelo al compromiso ineludible con la gestión basada en resultados -y conste que se hacen meritorios esfuerzos por reconciliarlas, por compaginar ambas exigencias (Franklin, 2008). Ciertamente son casi irreconciliables y simplemente ponen de relieve la dificultad de la tarea de los burócratas, que tienen que hacer juegos malabares con “ortodoxias” rivales. Es injusto para ellos y para los ciudadanos. La democracia no debe ser vista como responsabilidad de las administraciones.

3. SEGUNDO EJEMPLO: “PERSONAS SENSATAS” EN CIRCUNSTANCIAS INSENSATAS

El siguiente ejemplo viene del Reino Unido y tiene que ver con el encargo de una evaluación que hizo un organismo gubernamental. A principios de los años 70 se publicó en Reino Unido el Informe Rothschild (1971), en el que se establecían las condiciones para la contratación de proyectos de investigación por parte del gobierno. El Informe definía la relación entre el gobierno y el investigador como la de un “cliente y un prestador de servicios”. El cliente (gobierno) establece las condiciones, mientras que el prestador del servicio (investigador) cumple con dichas condiciones a tenor de unas cláusulas contractuales. Básicamente, definía el encargo de evaluación como una transacción económica, como “condiciones de contratación” (Pettigrew, 1994). Ésta fue la causa principal de la internalización gradual de la investigación y la evaluación. Promovía el concepto de “cliente informado”, es decir, el burócrata con suficientes conocimientos de investigación y evaluación como para poder especificar exactamente cómo deben llevarse a cabo.

El cambio subsiguiente, y bastante rápido, de la contratación externa de la evaluación a su internalización, quedó ampliamente documentado en su momento por evaluadores preocupados por el control de la investigación académica (Pettigrew, 1994; Norris, 1995; Simons, 1995). Más recientemente, la preocupación por el control de la información pública ha sido absorbida por otras preocupaciones más generales sobre lo que ha dado en llamarse *The New Public Management* -la contratación por resultados y la centralización del control (Kushner y Norris, 2007)-, incluso al nivel de los cambios sociales estructurales (Powers, 1999; O’Neill, 2002). Estos debates rara vez afloran en la evaluación del desarrollo: es una cuestión democrática que se ignora en aras de la supuesta necesidad urgente de acción.

El texto que incluyo a continuación es parte de un memorándum que escribí y, que posteriormente dio lugar a que la *British Educational Research Association* revisara los consejos que daba a sus miembros a la hora de negociar sus contratos con organismos oficiales. Esto muestra el tipo de control sobre el conocimiento que surge cuando la administración asume la responsabilidad de definir las condiciones en las que aquél se produce. Trata de las condiciones y los detalles concretos de una evaluación establecidos por un organismo gubernamental responsable de una cuestión educativa fundamental:

Para los que buscan colaborar en el desarrollo pedagógico y educativo, las oportunidades de trabajar con la [Agencia] son de máxima importancia. Sin embargo, las condiciones contractuales de financiación son tan onerosas como para resultar inaceptables. Dichas condiciones suscitan graves dudas sobre la independencia y la base ética de los estudios encargados por la Agencia, pero lo que es aún más preocupante, amenazan con comprometer la calidad de la información que se traslada a la propia Agencia. Tratemos cada uno de ellos:

(1) Condiciones de financiación: la [Agencia] impone tales exigencias en la elaboración de los informes sobre los que a menudo son proyectos a corto plazo, que comprometen nuestra capacidad de realizar estudios rigurosos y de proporcionar informaciones correctas. En los requisitos de un concurso reciente para un contrato de diez meses de duración se exigían informes de progreso quincenales, un “diálogo constante” y la presentación de seis informes parciales/finales junto con un número indeterminado de reuniones para hablar sobre el contrato y tres reuniones adicionales para hablar sobre los borradores del informe final, además de un número no especificado de reuniones con el denominado “comité de seguimiento”. Al parecer, los “prestadores del servicio” también deben presentar y negociar un “plan de salida” al mes de comenzar el estudio.

(2) Ética e independencia: la Agencia impone unos niveles inusualmente elevados –incluso para las agencias oficiales– de control y revisión de los informes que están sujetos a procedimientos de calidad de la [Agencia], incluso por su “tono”. Resulta sorprendente que insistan en inspeccionar y examinar la correspondencia con terceros (externos) además de supervisar los instrumentos de investigación y los informes. Los informes parciales tienen que contar con el visto bueno de la Agencia antes de convertirse en informes finales. La [Agencia] se reserva el derecho de decidir a quiénes se les envían los informes (‘audiencias’). La [Agencia] no sólo insiste en quedarse con los derechos de propiedad intelectual –en el mejor de los casos, por la confusión entre “comprador” y “entidad financiadora”, y en el peor de los casos porque abusan de sus privilegios – sino que también insiste en la propiedad de datos personales y confidenciales que, por contrato, deben ser enviados a la Agencia. Esto sólo puede provocar que se supriman los comentarios críticos de profesores y profesores-educadores a quienes se les aconsejaría que se abstuviesen de aportar puntos de vista que quizás pudieran estar cuestionando la labor de la [Agencia], por ejemplo, privando así a la Agencia de una valiosa fuente de opinión. Aparte de todo esto, no está claro que la Agencia pueda “hacerse con la propiedad” de los datos personales de ninguna persona.

No está claro sobre qué base se arroga la [Agencia] unos niveles tan altos de control sobre los informes que se pagan con dinero público y que frecuentemente contendrán información de gran interés e importancia para los profesionales de la educación y para los padres con hijos en edad escolar. ¿Por qué le “pertenecen” más a la [Agencia] estos informes que a los profesores y a los padres que forman parte del consejo escolar, por ejemplo? En las reuniones, los trabajadores de la Agencia suelen insistir en que la potestad de controlar sólo se utilizará de forma aislada y justificada. Aunque los trabajadores de la [Agencia] son personas sensatas y honradas, la “sensatez” y la “honradez” no describen a un ente jurídico, ni garantizan un comportamiento sensato. Cuando concedemos a los entes jurídicos (no a sus empleados) un control tan férreo sobre el conocimiento, se abre la posibilidad de procesos propios de “sociedades cerradas”. Si la eliminación gradual del derecho público al conocimiento sobre la enseñanza no se parece a un contrato con la [Agencia], ¿a qué otra cosa se podría parecer?

¿Qué vemos en los párrafos anteriores? Desde el punto de vista de la Agencia hay una preocupación por la calidad de la investigación. No es raro, en los últimos años en Reino Unido, que una agencia de este tipo insista en supervisar todos los instrumentos de investigación –protocolos de entrevistas, encuestas– y la explicación que dan es que el rápido incremento en el número de expertos universitarios y consultoras privadas que compiten para conseguir contratos de evaluación (todos buscando ingresos para pagar sueldos) ha llevado a unos niveles insospechados de calidad metodológica. Agencias con procedimientos como los que aquí describimos se ven obligadas a desempeñar un papel de control de calidad.

A esto se añade la necesidad de quienes trabajan en la Agencia de velar por la reputación del organismo, en especial porque ellos también han de rendir cuentas. El compromiso profesional de los cargos públicos no tiene como prioridad la pluralidad de valores, el rigor de la metodología o el derecho de la ciudadanía a la información: su prioridad son las necesidades inmediatas de sus superiores políticos. La burocracia es, necesariamente, la cultura del cumplimiento del servicio.

4. CONCLUSIÓN

Se pueden extraer dos lecciones principales de estas dos historias sobre la internalización de la evaluación. Una, que al evaluador se le niega la oportunidad de diseñar una evaluación comprometida, esto es, en términos “deliberativos y democráticos”. Los primeros teóricos de la evaluación, en su búsqueda de las funciones más apropiadas de ésta, la definían menos en términos de metodología y más como un *espacio* social y político único, dentro del cual se producían unas interacciones que no tenían cabida en ningún otro espacio social. Para ocupar ese espacio (de hecho, un espacio ético) la evaluación debe ser independiente. La independencia debe entenderse no como un privilegio del evaluador como actor político, sino como garantía al ciudadano de que hay un proceso político en marcha (que da cuentas a la sociedad) y al cual tienen acceso. Internalizar la evaluación equivale a denegar el libre acceso a sus resultados.

La otra lección es que la internalización de la evaluación expone al burócrata al escrutinio público y a la rendición de cuentas hasta un punto que, al final, es irracional e insostenible. El burócrata, por definición, no puede rendir cuentas por las decisiones metodológicas ni por la “inspección” metodológica y política que se deriva de su control contractual sobre la evaluación. Esto no le es de mucha ayuda, supone importantes trabas para el evaluador, como ya hemos visto, y niega al ciudadano la oportunidad de exigir responsabilidades a quienes toman las decisiones.

Si la sociedad ha de “aprender sobre sí misma”, deberíamos prestar atención a la relación entre el ciudadano y lo público, ya que las administraciones en estos momentos tienen prácticamente el monopolio de los medios para lograr ese aprendizaje. No fueron concebidas para asumir esa responsabilidad y, lógicamente, no lo llevan bien. Evidentemente, no está en nuestra mano, ni siquiera en democracia, reajustar esa relación, ya que la raíz del problema está en las disposiciones constitucionales, aun cuando la realidad contemporánea hace tiempo que las ha hecho irrelevantes y anticuadas.

Sin embargo, hay medidas que podemos y debemos adoptar como evaluadores para mejorar la situación. Sin duda, las asociaciones profesionales (de psicólogos, profesores, sociólogos y evaluadores) deben entablar un diálogo con las entidades financiadoras oficiales sobre la información pública y las obligaciones del gobierno a este respecto. Pero también deberíamos guiarnos, más de lo que lo hacemos, por los principios no formalizados sobre la ética de la evaluación. Un protocolo ético que regule el acceso a la información y a la publicación de la misma, que establezca las condiciones de confidencialidad y negociación del control de los datos, puede servir de base para llegar a un consenso sobre la delicada cuestión de cómo hacer pública la información. Por ejemplo, una cláusula que estipule que los individuos son dueños de la información sobre su persona o su trabajo, les daría al menos autoridad moral ante cualquier intento de control de un informe de evaluación. Este tipo de acuerdos se pueden negociar con la entidad financiadora con menor presión y mayor transparencia que durante el proceso de encargo y contratación. Los contratos a menudo también cambian los valores y las preferencias de la entidad financiadora y esta clase de acuerdos pueden satisfacer a todos los implicados en la medida en que logran recoger valores comunes.

Los acuerdos éticos tienen una ventaja adicional. Modifican, en cierta medida, la relación económica entre el ciudadano, el evaluador y la administración/entidad financiadora. Las relaciones económicas –como señaló Pettigrew (1994), “condiciones de contratación”– no son una buena vía para las prácticas éticas y las relaciones que se basan en valores. De hecho, recientemente, hemos asistido al estrepitoso fracaso de las relaciones económicas en las que durante más de una década nos convencieron para que nos considerásemos “clientes”, “consumidores”, “recursos humanos” o *stakeholders*, lo que nos condujo a una profunda crisis económica, de la que la cultura del contrato y los sistemas administrativos y regulatorios fallidos tienen mucha culpa. Debiera haber una autoridad moral que apoye el argumento de que las entidades públicas que financian la evaluación y la regulación desempeñen un papel más humilde y comprometido democráticamente.

REFERENCIAS BIBLIOGRÁFICAS Y DOCUMENTALES

- Alkin, M., ed. (2004), *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, Ca: Sage Publications.
- Black, M. (2008), *The No-Nonsense Guide to the United Nations*. Oxford: New Internationalist.
- Cronbach, L.J. et al. (1980), *Towards Reform of Program Evaluation*. San Francisco: Jossey-Bass Publishers.
- Deaton, A. (2008), "Randomisation in the tropics, and the search for the elusive key to economic development", *The Keynes Lecture*, Londres: The British Academy, 9 de octubre.
- Franklin, T. (2008), "Reaching the Millennium Development Goals: Equality and justice as well as results", *Development in Practice*, 18(3): 420-423.
- House, E. (1979), "Technology versus craft: a ten year perspective on innovation", *Journal of Curriculum Studies*, 11(1): 1-15.
- (1980), *Evaluating with Validity*. Thousand Oaks, Ca: Sage.
- (1993), *Professional Evaluation*. Thousand Oaks, Ca: Sage.
- Kushner, S. (2009), "Own goals: evaluation and accountability in millennium development planning", en K.E. Ryan y J.B. Cousins, eds., *Sage International Handbook on Educational Evaluation*. Thousand Oaks: Sage.
- Kushner, S. y Norris, N., eds. (2007), *Dilemmas of Engagement: Evaluation and the New Public Management*. Oxford: Elsevier.
- MacDonald, B. (1976), "Evaluation and the control of education", en D. Tawney, ed., *Curriculum Evaluation Today: Trends and Implications*. Londres: Macmillan.
- (1981), "Mandarins and Lemons. The executive investment in program evaluation", ponencia presentada en el Congreso Anual de la *American Educational Research Association*, Los Angeles, 13-17 abril.
- Marsden, D. (2003), "Rights, culture and contested modernities", conferencia inaugural de la INTRAC 5th International Evaluation Conference, Holanda, abril.
- Norris, N. (1990), *Understanding Educational Evaluation*. Londres: Kogan Page.
- (1995), "Contracts, control and evaluation", *Journal of Education Policy*, 10(3): 271-285.
- O'Neill, O. (2002), *A Question of Trust*. Cambridge: CUP.
- Pettigrew, M. (1994), "Coming to terms with research: the contract business", en B. Halpin y B. Troyna, eds., *Researching Education Policy: Ethical and Methodological Issues*, Londres: Falmer.
- Picciotto, R. (2007), "The new environment for development evaluation", *American Journal of Evaluation*, 28(4): 509-521.
- Powers, M. (1999), *The Audit Society: Rituals of Verification*. Oxford: OUP.
- Rothschild Report (1971), Report of *The organization and management of Government R&D*, en *Cabinet Office. A Framework for Government Research and Development*. Londres: HMSO.
- Simons, H. (1995), 'The politics and ethics of educational research in England', *British Educational Research Journal*, 21(4): 435-449.