

FICHA 2

¿CÓMO SE HACEN LAS EVALUACIONES EDUCATIVAS?

Pedro Ravela, 2008

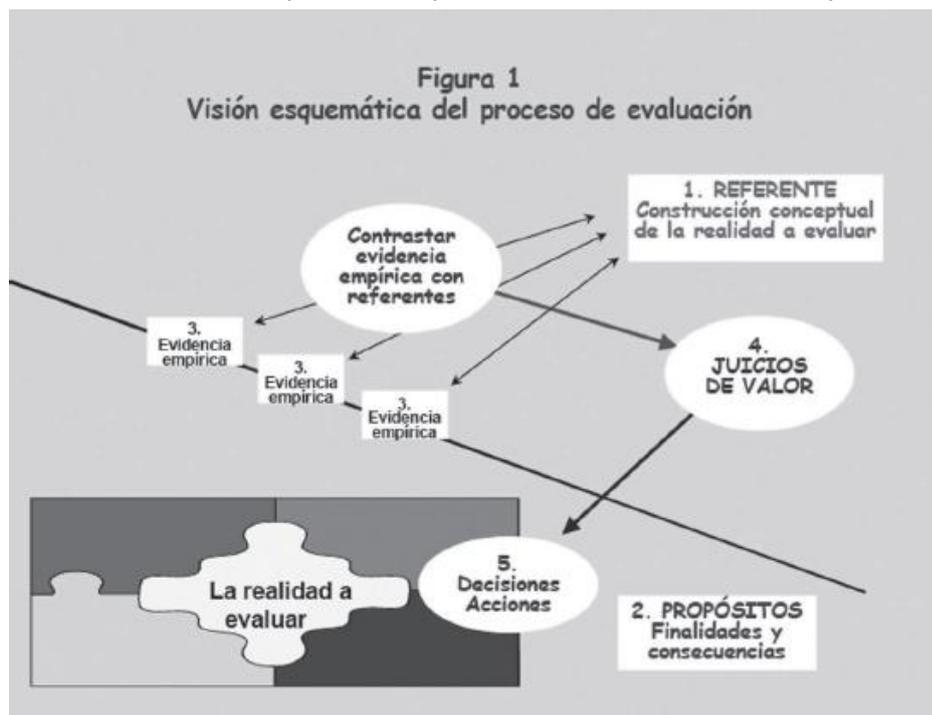
Los elementos básicos del proceso de evaluación

El propósito de esta Ficha es explicar al lector cinco elementos centrales que subyacen a todo proceso de evaluación:

1. la selección de la realidad a evaluar y la “construcción conceptual” de la misma;
2. la definición de los propósitos de la evaluación;
3. la producción de evidencia empírica (información, datos acerca de la realidad evaluada);
4. la formulación de juicios de valor sobre la realidad evaluada;
5. la toma de decisiones o acciones que transformen dicha realidad.

Comprender el significado de estos cinco elementos es fundamental para entender cómo funcionan las evaluaciones, conocer sus posibilidades y limitaciones y leer de manera más inteligente sus resultados.

Los cinco elementos centrales aparecen representados en forma de esquema en la Figura 1.



1. La selección y “construcción conceptual” de la realidad a evaluar

El primer paso obvio de cualquier evaluación es decidir qué se quiere evaluar. En principio esto parece sencillo: queremos evaluar el desempeño de los docentes, la calidad de una universidad o si los estudiantes han aprendido a leer.

Sin embargo, cualquiera de las expresiones anteriores involucra aspectos de la realidad cuyo significado puede ser definido de diversas maneras. Existen diversas perspectivas y posibilidades para definir “desempeño docente”, “calidad universitaria” o “capacidad de lectura”.

Un error en muchas evaluaciones es dar por supuesto que la realidad está al alcance de la mano, que hay algo predefinido inequívocamente que es “el buen desempeño docente”, “la capacidad de lectura” o “la calidad de una institución universitaria” y que, por lo tanto, el problema de la evaluación se limita a diseñar instrumentos y recoger información.

Por el contrario, cualquier “realidad” debe ser definida, es decir, requiere de una “construcción conceptual” que involucra conocimientos, visiones y valores acerca de dicha realidad.

La realidad no es algo de lo que podamos tener conocimiento directo. Es construida por los seres humanos y puede ser percibida y conceptualizada de diversas maneras. Por esta razón, en la Figura 1 está representada con un “rompecabezas” y “separada” de nuestra percepción por una línea negra, que representa la “opacidad” de la realidad.

A la construcción conceptual de la realidad que queremos evaluar se la denomina “referente”. El “referente” siempre tiene una connotación valorativa, porque expresa lo deseable o lo que se desea alcanzar. Elaborar y explicitar el “referente” es la primera tarea central de toda evaluación.

Los Recuadros 1 y 2 ejemplifican dos referentes distintos para la evaluación del desempeño docente.

En el Recuadro 1 se incluyen las pautas empleadas en Uruguay para la evaluación del desempeño de los docentes por parte de los Inspectores, de acuerdo a lo estipulado en el Estatuto del Funcionario Docente. Esta pauta, junto con un formulario que incluye un espacio para que el Inspector registre algunas observaciones o comentarios sobre cada uno de los aspectos anteriores, es todo con lo que cuenta el Inspector para evaluar a cada docente, a través de una única visita al aula para observar una clase.

Como se puede apreciar, se trata de una definición conceptual absolutamente general y poco precisa. ¿Qué significa, por ejemplo, “posibilidades de desarrollo del trabajo creativo”?

El Inspector debe asignar al docente un puntaje entre 0 y 100. La escala tiene juicios de valor asociados a distintos puntajes (por ejemplo, un puntaje entre 90 y 100 significa “excelente”), pero no existen pautas explícitas que definan qué significa ser “excelente”, por ejemplo, en la “capacidad técnico pedagógica”. En cierto modo, se da por supuesto que dicha expresión tiene una única interpretación posible que es conocida y compartida por docentes e Inspectores. Como ello no ocurre, los resultados de las evaluaciones que se realizan dependen fuertemente de la subjetividad de cada Inspector.

En el Recuadro 2 se incluye una parte del referente elaborado por el “*Educational Testing Service*” (ETS) para la evaluación de docentes. Es un referente mucho más elaborado, preciso y explícito, que recoge las investigaciones y enfoques más recientes en relación a la profesión docente y la práctica de la enseñanza.

Este referente se basa en cuatro grandes dimensiones (véase la primera parte del Recuadro 2), cada una de las cuales es luego explicitada a través de entre cuatro y seis criterios, que a su vez explicitan el significado de la dimensión. En la segunda parte del Recuadro 2 se incluyen los criterios correspondientes a las dimensiones 'A' y 'B'.

Pero además de establecer criterios para cada dimensión, cada uno de los 19 criterios resultantes está explicado en un documento mediante tres o cuatro párrafos que los describen y sustentan en la investigación y literatura más recientes en relación al desempeño docente.

Como resulta obvio, un dispositivo de evaluación desarrollado sobre este último referente tiene mucho mayor interés y utilidad que el incluido en el Recuadro 1: el docente tiene más claro sobre qué bases está siendo evaluado, es más fácil construir instrumentos para relevar información cuanto más explícito esté el referente, es más fácil también entrenar a los evaluadores para que las evaluaciones de distintos docentes sean menos subjetivas y más equitativas.

Recuadro 1	Pautas para el Informe de Inspección empleado en Uruguay
	<p style="text-align: center;">Juicios sobre la aptitud docente</p> <ol style="list-style-type: none"> 1. "Capacidad técnico pedagógica" 2. "Conducción del proceso de enseñanza-aprendizaje" 3. "Orientación dada al curso; planeamiento y desarrollo del mismo" 4. "Aprendizaje realizado por los alumnos y capacidad para seguir aprendiendo" 5. "Clima de trabajo, cooperación e iniciativa" 6. "Respeto al alumno y promoción de su capacidad de autodeterminación" 7. "Posibilidades de desarrollo del trabajo creativo"
Recuadro 2	El "referente" de desempeño docente empleado por el Modelo de Evaluación PRAXIS¹
	<p style="text-align: center;">DIMENSIONES</p> <p>Organizar los contenidos para el aprendizaje de los alumnos Crear un entorno para el aprendizaje de los alumnos Enseñar para que los alumnos aprendan Profesionalismo Docente</p> <p style="text-align: center;">CRITERIOS</p> <p>A.1. Familiarizarse con los aspectos relevantes de las experiencias y conocimientos previos de los alumnos A.2. Formular metas claras de aprendizaje que sean apropiadas para los alumnos A.3. Demostrar comprensión de los vínculos entre el contenido que se aprendió anteriormente, el contenido actual y el contenido que se trabajará más adelante A.4. Crear o seleccionar métodos de enseñanza, actividades de aprendizaje y materiales u otras fuentes de Instrucción apropiados para los alumnos y para las metas de aprendizaje A.5. Crear o seleccionar estrategias de evaluación apropiadas para los alumnos y para las metas de aprendizaje</p> <p>B.1. Crear un clima que promueva la equidad B.2. Establecer y mantener una relación interpersonal adecuada con los alumnos B.3. Comunicar expectativas de aprendizaje desafiantes a cada alumno B.4. Establecer y mantener normas coherentes de comportamiento en clase B.5. Hacer que el entorno físico sea lo más seguro y conducente al aprendizaje que sea posible</p> <p><small>1) Fuente: DWYER, C.; "Evaluación de los maestros". En: ALVAREZ, B. y RUIZ-CASARES, M. (editores), 1997; Evaluación y reforma educativa. Opciones de política; Capítulo 7, pp.187-222. PREAL, Santiago de Chile.</small></p>

Lo mismo ocurre en cualquier otro aspecto de la realidad que se quiera evaluar. Por ejemplo, hay múltiples modos de definir capacidad de lectura, los que pueden ser más o menos generales o específicos y pueden o no recoger los conocimientos más actualizados en el campo del aprendizaje del lenguaje.

En las Fichas 6 y 7 este tema será abordado nuevamente, a través de ejemplos sobre cómo distintas evaluaciones estandarizadas de lectura definen de manera distinta el aprendizaje de los alumnos y utilizan actividades de prueba muy distintas. Por este motivo, los datos que entregan esas evaluaciones tienen sentidos diferentes.

2. La definición de los propósitos de la evaluación

El segundo elemento clave para una buena evaluación es definir claramente los propósitos de la misma: las preguntas que busca responder, el tipo de consecuencias que tendrá la evaluación, los usos que se pretende dar a sus resultados, quiénes harán uso de los mismos. Estos aspectos fueron tratados en la Ficha 1.

Lo que interesa destacar en esta Ficha es que la definición de los propósitos y consecuencias de una evaluación tiene importancia decisiva para el resto del diseño de la misma.

Por ejemplo, si el propósito de una evaluación es calificar a los docentes, y esta calificación tendrá consecuencias para sus carreras funcionales, se requiere una gran confiabilidad y precisión en los puntajes que se les asignen, porque diferencias de puntajes muy pequeñas entre los individuos pueden determinar que unos accedan a un cargo y otros no. En este caso, la precisión de los puntajes es fundamental para que el proceso sea equitativo y para que el sistema educativo se asegure de promover a los profesionales más competentes.

En cambio, si la evaluación tiene como finalidad principal orientar a los docentes para mejorar su trabajo como forma de desarrollo profesional, la asignación de puntajes precisos pasa a tener importancia secundaria, dado que no está en juego la carrera de los individuos.

De la misma manera, una evaluación de alumnos cuyos resultados tendrán consecuencias “fuertes”, tales como decidir si aprueban o no un curso, debería restringirse a aquello que fue realmente enseñado durante el curso.

En cambio, si se trata de una evaluación estandarizada de carácter diagnóstico que no tendrá consecuencias para los estudiantes, es legítimo evaluar aspectos relevantes, aun cuando no hayan sido enseñados, justamente para entregar al cuerpo docente una señal en el sentido de que dichos aspectos deberían estar siendo enseñados.

Otro aspecto importante es el relacionado con los destinatarios de la evaluación. Definir con claridad desde el comienzo quiénes utilizarán los resultados de la evaluación, y para qué propósitos, también tiene implicancias importantes para el diseño de la evaluación.

Será necesario asegurarse que el tipo de información y los juicios de valor que la evaluación produzca sean apropiados y comprensibles para los destinatarios.

3. La producción de evidencia empírica

No es raro que cuando se piensa en evaluar algo, la primera tarea que se emprenda sea el diseño de instrumentos. Sin embargo, de acuerdo a lo analizado en los apartados anteriores, recién después de haber definido con claridad el referente y los propósitos de la evaluación tiene sentido trabajar en el diseño de instrumentos para recoger información.

El problema central en este terreno es reconocer que no tenemos acceso directo a la realidad. Ante un profesor, no podemos, saber directamente si como profesional es bueno, regular o malo. Del mismo modo, no tenemos acceso directo a la mente y el corazón de los alumnos. Por lo tanto, necesitamos construir “mediaciones” que nos permitan aproximarnos a la realidad que queremos evaluar.

A estas mediaciones se las denomina en forma genérica como “evidencia empírica”, entendiéndose por tal toda pieza de información que muestra un aspecto de la realidad que nos interesa conocer y evaluar.

Por ejemplo, si queremos evaluar el desempeño de un docente, necesitaremos recoger información sobre aquellos aspectos que hayamos definido como relevantes en nuestro referente: cómo planifica sus clases, cómo explica los temas a sus alumnos, qué actividades les propone que ellos realicen, en qué grado se esfuerza para que los alumnos se interesen en los temas, etc.

Los “instrumentos” de una evaluación son todos aquellos dispositivos contruidos para recoger evidencia empírica en forma sistemática sobre los aspectos relevantes de la realidad a evaluar.

Los instrumentos para la recolección de evidencia empírica en las evaluaciones educativas **2** pueden ser muy variados:

- pruebas escritas de diverso tipo (de ensayo, de respuesta construida, de opción múltiple, etc.);
- pruebas prácticas como, por ejemplo, dar una clase o conducir una reunión de profesores;
- registros de observación de diverso tipo, como por ejemplo, de las actividades de los alumnos;
- carpetas con trabajos producidos por los alumnos o los profesores (denominados “portafolios” desde hace algún tiempo en la literatura anglosajona);
- encuestas de opinión de padres, alumnos o profesores.

La amplitud y diversidad de las fuentes de recolección de evidencia empírica es enorme.

La enumeración anterior de instrumentos es solamente ilustrativa.

Cuanto más variada sea la diversidad de fuentes de información que se utilice en una evaluación, más sólida será la misma. Sin embargo, al mismo tiempo hay restricciones de tiempo y costos que hacen necesario limitar la cantidad de instrumentos.

Un aspecto muy importante que será tratado con detalle en la Ficha 4 es cuidar la adecuada relación y coherencia entre el referente de la evaluación y los contenidos de los instrumentos de recolección de evidencia empírica. Un problema común en muchas evaluaciones es que se define de una manera la “realidad” a evaluar, pero los instrumentos recogen información sobre otros aspectos no contemplados en el referente. Esto, así como otros errores en el diseño de las evaluaciones y en la interpretación y uso de sus resultados, genera un problema de “validez” en la evaluación.

Otro problema, siempre presente, deriva del hecho de que la realidad es inabarcable, por lo que inevitablemente nuestra información tendrá limitaciones que no hay más remedio que aceptar.

Por ello es importante evitar los simplismos, tanto el de creer que un número puede definir la calidad de una persona o de una institución, como el simplismo opuesto: creer que, como hay limitaciones en la información, solo es válida la observación directa y el juicio subjetivo, con la consecuente descalificación genérica de las evaluaciones sistemáticas.

De lo antes expuesto se desprende que la selección de los instrumentos a emplear en una evaluación debe ser coherente con el referente y con los propósitos de la evaluación, así como con cierta dosis de sentido común y realismo en cuanto a los recursos, el tiempo y la viabilidad práctica de la evaluación.

Siempre es necesario realizar un balance entre la cobertura más exhaustiva posible de los diversos aspectos de la realidad que se quiere evaluar, la inversión (principalmente de tiempo y de dinero) y el tipo de consecuencias que la evaluación tendrá.

Por ejemplo, puede ocurrir que la evaluación sea tan exhaustiva que el tiempo necesario para el procesamiento de la información y la producción de resultados tenga como consecuencia que éstos lleguen tarde y que la evaluación no tenga impacto porque sus resultados no llegaron en el momento oportuno.

Esto ocurre en las evaluaciones nacionales de carácter estandarizado: muchas veces cuando se presentan los resultados ha pasado tanto tiempo desde que se realizó la evaluación, que los resultados ya no son pertinentes para la toma de decisiones.

Asimismo, los profesores constantemente necesitan hacer en el aula el siguiente balance: si evalúan con demasiada frecuencia, probablemente no tengan tiempo para corregir las evaluaciones que proponen y para hacer devoluciones significativas a sus alumnos.

4. La formulación de juicios de valor

La esencia de la evaluación es establecer un juicio de valor. Este surge principalmente de contrastar la evidencia empírica con el referente para formular valoraciones sobre la realidad.

Ejemplos de juicios de valor son: “este alumno tiene un desempeño satisfactorio en Matemática”, “este alumno tiene dificultades en el uso de la puntuación en la producción escrita”, “este maestro es excelente enseñando Ciencias”. En estos casos se trata de juicios de valor absolutos, formulados en relación al referente de la evaluación. Se los suele denominar como “referidos a un criterio”.

Otras veces, los juicios de valor están basados en comparaciones entre individuos o entre instituciones: “estos alumnos son mejores que aquellos”, “este alumno está dentro del 10% de estudiantes con mejor desempeño”, “éstas son las escuelas con mejores resultados en contextos desfavorecidos”. En estos casos se trata de juicios de valor “relativos”, porque están basados en la comparación entre unidades de análisis. A este tipo de juicios de valor se suele denominar como “normativo” (por referencia a la curva normal, según se explicará en la Ficha 3).

En ocasiones, los juicios de valor son conjuntos de apreciaciones que no se resumen en un puntaje o una categoría, sino que se utilizan en forma global para orientar el mejor desempeño de un individuo. Es el caso de un profesor orientando a un alumno o de un supervisor orientando a un docente. En estos casos puede haber valoraciones que se expresan en el marco de la interacción directa entre evaluador y evaluado, sin necesidad de resumirlas en un juicio taxativo o en un puntaje.

Al análisis de diferentes modos de establecer los juicios de valor está dedicada la Ficha 3.

Lo central a destacar en este punto es que lo más importante en una evaluación no son los datos en sí mismos, sino la valoración de la realidad que es objeto de evaluación, valoración que se construye a partir de la contrastación entre los datos y el referente.

5. La toma de decisiones o acciones que modifiquen la realidad evaluada

La vocación de toda evaluación es tener alguna consecuencia sobre la realidad. Las consecuencias pueden ser de diverso tipo –Ficha 1– pero una evaluación no es tal si no pretende tener algún tipo de consecuencia, al menos una nueva comprensión de la realidad evaluada que ayude a los actores a pensar en nuevos modos de actuar.

Ninguna evaluación se hace simplemente por curiosidad, sin intención de que sus resultados sean empleados de uno u otro modo para fines específicos.

En la Ficha 1 se hacía una distinción entre las evaluaciones con consecuencias formales y explícitas y evaluaciones formativas sin consecuencias formales, pero que modifican la realidad a través del aprendizaje de los individuos, que logran nuevas visiones sobre su propio desempeño, el de sus alumnos o el de las instituciones de las que forman parte.

En este punto es necesario señalar algunas dificultades.

Cuando una evaluación tiene consecuencias “fuertes”, a veces ocurre que se producen efectos no deseados o “efectos perversos”.

Por ejemplo, un examen de finalización de la primaria –con consecuencias para los alumnos– cuyo propósito explícito es garantizar que todos logran niveles de aprendizaje similares, puede tener como efecto “perverso” un incremento de los niveles de repetición y deserción si muchos alumnos no consiguen aprobar.

Otro ejemplo podría ser la introducción de incentivos monetarios a las escuelas en función de sus resultados, cuyo propósito explícito es incentivar a todos a mejorar, pero porque puede tener como efecto “perverso” la desmoralización de los “perdedores” y un incremento de la segmentación interna del sistema educativo. Estos aspectos serán tratados con mayor detalle en la Ficha 11.

Cuando las evaluaciones son de carácter formativo, el tipo de consecuencias que tengan dependerá del grado en que los actores se apropien de los resultados y los utilicen para generar nuevas comprensiones sobre la realidad que les permitan aprender y mejorar su modo de actuar. Que ello efectivamente ocurra requiere que se implementen estrategias y acciones específicas de divulgación de resultados (a este aspecto está destinada la Ficha 12).

Una dificultad central en relación al impacto de los procesos de evaluación en la transformación de la realidad es el divorcio que a veces se produce en la práctica entre evaluadores, tomadores de decisiones y quienes se desempeñan como los directivos y docentes.

Muchas veces los evaluadores consideran que su labor termina cuando producen el informe de la evaluación y que la responsabilidad de tomar decisiones y usar los resultados corresponde a otros. Pero, al mismo tiempo, suele ocurrir que los tomadores de decisiones y quienes están en las escuelas –los educadores– no comprenden adecuadamente los resultados, no se interesan por ellos o no los consideran relevantes.

Por lo tanto, no los utilizan y la evaluación corre el riesgo de tornarse un ejercicio estéril y de perder legitimidad.

Por ejemplo, es bastante común que las autoridades educativas quieran evaluaciones pero luego no utilicen sus resultados en la toma de decisiones ni emprendan las acciones necesarias para revertir los problemas que la evaluación devela.

Este problema tiene una doble arista. Por un lado, es resultado de que muchas veces prima una lógica político-partidaria y de corto plazo en la toma de decisiones. Al mismo tiempo, obedece a que muchas veces lo que producen los evaluadores no es lo que los tomadores de decisiones necesitan.

La vocación principal de toda evaluación es modificar la realidad, pero la evaluación por sí misma no produce cambios si no hay actores que usen los resultados y tomen decisiones a partir de las valoraciones resultantes de la misma.

En este sentido, es preciso enfatizar que, si bien no son los evaluadores quienes deben tomar decisiones o emprender acciones, sí es su responsabilidad comunicar los resultados de manera apropiada. Esto incluye escribir reportes comprensibles pero, sobre todo, propiciar y participar de instancias de diálogo “cara a cara” con otros actores –políticos, autoridades, docentes, unidades de currículum, formadores de docentes, etc. – ayudando a comprender el significado de los datos y lo que nos dicen sobre la realidad, escuchando las dudas y demandas, ayudando a pensar en términos de alternativas para encarar los problemas.

Síntesis final

Esta segunda Ficha se propuso explicar los cinco elementos centrales de todo proceso de evaluación. Intentó además alertar al lector o usuario de evaluaciones sobre los aspectos clave que es necesario tener presente al diseñar, interpretar o analizar una evaluación. Atender adecuadamente a estos cinco aspectos permitirá mejorar sustancialmente el impacto de las evaluaciones sobre el sistema educativo.

Toda evaluación debe estar basada en un “**referente**” explícito, claro y apropiado. Esto significa partir de una definición conceptual de la realidad a evaluar que recoja las perspectivas más actualizadas y que, al mismo tiempo, esté expresado de manera clara y comprensible. De este modo es posible diseñar mejor los instrumentos para la evaluación y, al mismo tiempo, permitir a los usuarios interpretar con más facilidad los resultados de la evaluación y, por tanto, hacer uso de ellos.

Los “**propósitos y consecuencias**” de las evaluaciones deben estar definidos desde el comienzo. Esto permite, al diseñar la evaluación, tomar decisiones apropiadas en relación a las consecuencias que se espera tenga la evaluación. El hecho de que los propósitos estén claros desde el comienzo facilita a los diversos actores involucrados saber a qué atenerse y qué pueden esperar de la evaluación. La reflexión inicial sobre los propósitos incluye el análisis de la complejidad de los efectos que la evaluación puede generar, de modo de minimizar efectos no deseados.

La “**evidencia empírica**”, si bien siempre será limitada, debe intentar cubrir de manera apropiada la diversidad de aspectos de la realidad a evaluar definidos en el referente. La consistencia entre los instrumentos y el referente hace que la evaluación tenga mayor validez y, por tanto, que sus resultados sean interpretables y utilizables. Al mismo tiempo, el uso e impacto de la evaluación mejoran cuando los instrumentos empleados son suficientemente precisos (teniendo en cuenta las consecuencias que la evaluación tendrá) y no dependen excesivamente de la subjetividad del evaluador, es decir, son confiables. Los conceptos de validez y confiabilidad serán analizados en la Ficha 4.

La esencia de una evaluación es “**establecer un juicio de valor**” acerca de la realidad evaluada. Esta valoración surge principalmente de contrastar la evidencia empírica con el referente (en la Ficha 3 se analiza con más detenimiento cómo se establecen los juicios de valor). El uso apropiado de los resultados de una evaluación requiere tener conciencia de que los datos no “son” la “realidad”, sino una aproximación a ella, por lo cual es muy importante hacer siempre un uso reflexivo y ponderado de los mismos, evitando caer en visiones simplistas. El principal aporte de una evaluación es ayudar a reflexionar y comprender mejor la realidad, con el fin de enriquecer la toma de decisiones. Por el contrario, las visiones simplistas pueden hacer que la evaluación pierda su potencial y conduzca a debates estériles o a decisiones inapropiadas.

El sentido último de toda evaluación es “**provocar cambios en la realidad**”. Para que las modificaciones se produzcan, es fundamental que los evaluadores consideren como parte de su trabajo –y por tanto del diseño de la evaluación– la divulgación apropiada de los resultados a las audiencias que los utilizarán. Esto requiere una preocupación especial por hacer reportes comprensibles pero, además, invertir tiempo en contactos directos con las diferentes audiencias para explicar los resultados y ayudar a la reflexión acerca de las implicancias de los mismos.