

La Categoría Precisión en la Meta-evaluación: Aspectos Prácticos y Teóricos en un Nuevo Enfoque¹

Ana Carolina Letichevsky
Marley Maria B. R. Vellasco
Ricardo Tanscheit
Reinaldo Castro Souza

RESUMEN

Este artículo trata de la categoría precisión de una verdadera evaluación en el ámbito de la evaluación y de la meta evaluación (evaluación de la evaluación). Presenta, inicialmente, la conceptualización moderna de la evaluación y de la meta evaluación para, entonces, discutir el papel de la meta evaluación de asegurar la calidad de un proceso evaluativo, en sus diferentes enfoques, a partir de las cuatro categorías de una verdadera evaluación (utilidad, viabilidad, ética y precisión). La énfasis del artículo está en la categoría precisión: (i) su importancia en una meta evaluación, en todas las fases o etapas de una evaluación, con sus avances y sus limitaciones y; (ii) su aplica-

Ana Carolina Letichevsky

Doutoranda em Métodos de Apoio à Decisão do Depto. de Engenharia Elétrica da PUC-RJ
Assessora Técnica Estatística da Fundação Cesgranrio
Professora da PUC-RJ
anacarolina@cesgranrio.org.br

Marley Maria B. R. Vellasco

Doutora pela University College London, Inglaterra
Professora do Depto. de Engenharia Elétrica da PUC-RJ
marley@ele.puc-rio.br

Ricardo Tanscheit

Doutor pela Queen Mary College, Inglaterra
Professor do Depto. de Engenharia Elétrica da PUC-RJ
ricardo@ele.puc-rio.br

Reinaldo Castro Souza

Doutor pela University of Warwick, Inglaterra
Professor do Depto. de Engenharia Elétrica da PUC-RJ
reinaldo@ele.puc-rio.br

ción en la propia meta evaluación que, como proceso evaluativo, también debe respetar los cuatro patrones de una verdadera evaluación. En este contexto se presentan aspectos teóricos y prácticos que merecen atención en la negociación, en el diseño, en la implantación y en la dirección de una meta evaluación cuando se discute, incluso, el paradigma del tratamiento de los datos de una meta evaluación a través de la lógica clásica y no de la lógica *fuzzy*. También se presenta una propuesta de metodología para realizar meta evaluación con base en la lógica *fuzzy*.

Palabras-clave: Meta evaluación. Evaluación. Categoría precisión. Lógica *fuzzy*.

¹ Este artículo es una adaptación del artículo presentado en la Primera Conferencia de la Red de Seguimiento, Evaluación y Sistematización en América Latina y el Caribe (ReLAC), realizada en octubre de 2004.

RESUMO

A Categoria Precisão na Avaliação e na Meta-avaliação: Aspectos Práticos e Teóricos em um Novo Enfoque

Este artigo trata da categoria precisão de uma verdadeira avaliação no âmbito da avaliação e da meta-avaliação (avaliação da avaliação). Apresenta, inicialmente, a conceituação moderna da avaliação e da meta-avaliação para, então, discutir o papel da meta-avaliação de assegurar à qualidade de um processo avaliativo, nas suas diferentes abordagens, a partir das quatro categorias de uma verdadeira avaliação (utilidade, viabilidade, ética e precisão). A ênfase do artigo está na categoria precisão: (i) a sua importância em uma meta-avaliação, em todas as fases ou etapas de uma avaliação, com seus avanços e suas limitações e; (ii) a sua aplicação na própria meta-avaliação que, enquanto processo avaliativo, também deve respeitar os quatro padrões de uma verdadeira avaliação. Neste contexto são apresentados aspectos teóricos e práticos que merecem atenção na negociação, no desenho, na implantação e na condução de uma meta-avaliação quando se discute, inclusive, o paradigma do tratamento dos dados de uma meta-avaliação através da lógica clássica e não da lógica fuzzy.

Palavras-chave: Meta-avaliação. Categoria precisão. Avaliação. Lógica fuzzy.

ABSTRACT

The Category Precision in the Evaluation and Meta-evaluation: Practical and Theoretical Aspects

This article deals with the category precision of a true evaluation in the scope of evaluation and meta-evaluation (evaluation of evaluation). First, it presents the modern conceptualization of evaluation and meta-evaluation and then discusses the role of meta-evaluation to ensure the quality of an evaluation process, in its different approaches, from the four categories of a true evaluation (utility, viability, ethics and precision). The emphasis of the article is on the category precision: (1) its importance in a meta-evaluation, in all the phases or stages of an evaluation, with its advances and its limitations and; (2) its application in the meta-evaluation itself which, while an evaluative process, must also follow the four patterns of a true evaluation. In this sense, practical and theoretical aspects which merit attention in the negotiation, the design, the implementation and the conduction of a meta-evaluation are presented, when the paradigm of the treatment of the data of a meta-evaluation, through the classical logic and not the fuzzy logic, is discussed.

Keywords: Meta-evaluation. Category precision. Evaluation. Fuzzy logic.

Introducción

Actualmente la evaluación se entiende como un proceso por el cual se emite un juicio de valor respecto al foco u objeto que está siendo evaluado; el ser humano es

capaz de evaluar personas, grupos, instituciones, organizaciones, programas, sistemas, procesos y una serie de otros aspectos. Siempre que se emite un juicio de valor es necesario hacer uso de uno o más criterios, aunque estos no sean explícitos o claros, como ocurre, frecuentemente, en evaluaciones informales. Estas fueron realizadas durante siglos, antes de la creación de procedimientos formales para dirigir una evaluación, y todavía se vienen utilizando hasta hoy, a causa de que "nacemos evaluadores, pero no necesariamente buenos evaluadores" (STAKE, 2004).

Desde principios del siglo XX la evaluación se viene realizando por medio de técnicas, procedimientos y métodos sistematizados que permiten la dirección de un proceso evaluativo formal. A lo largo de un siglo de estudios esta evolucionó y pasó a ser entendida como un proceso documentado de recogida sistemática de datos e información precisa y relevante para responder a cuestiones que surgían *a priori* explicitando los criterios de excelencia que permiten la emisión de un juicio de valor con relación (i) al mérito de su foco de atención (en términos de calidad interna de sus componentes y de su funcionamiento) y (ii) a la relevancia (que se refiere a sus efectos y a su resultado, en términos de su impacto), con el intuito de fornecer subsidios para la mejora del foco evaluativo y para la tomada de decisiones pertinentes (PENNA FIRME, 1998).

La evaluación de un proceso evaluativo se denomina meta evaluación y debe ser realizada con base en las cuatro categorías que añaden los treinta patrones de una verdadera evaluación (JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION, 1994). La meta evaluación puede ser dirigida de diferentes for-

mas, pero frecuentemente se utiliza la lista de chequeo (del inglés *checklist*) como instrumento de recopilación de datos. Y, como en cualquier proceso que implique recopilación de datos, la calidad de información y la adecuación del tratamiento de esta información merecen especial atención. En este sentido las listas de chequeo son instrumentos importantes para auxiliar la dirección de una meta evaluación, pero es preciso estar atento para la utilización de formas adecuadas para la recopilación de información que subsidien su registro, así como para el tratamiento adecuado de los datos recopilados por las mismas.

Meta evaluación

La preocupación en crear patrones (principios obtenidos a través de consenso entre personas involucradas en la práctica de la evaluación que, logren, garantizarán la calidad de un evaluación) para realización de procesos evaluativos es antigua y, tal vez, sea tan antigua como la preocupación con la evaluación. Es una tarea ardua no apenas por la dificultad técnica a ella inherente, sino principalmente, por la dificultad de sensibilizar, movilizar y obtener consenso junto a los diferentes interesados pertinentes, de forma que se produzca un buen trabajo técnicamente y que se acepte por quien realiza la evaluación, por quien es evaluado, por quien encarga la evaluación y por quien la utiliza. En los años 60, algunos evaluadores de renombre como Scriven, Stake y Stufflebeam comenzaron a discutir procedimientos y criterios formales de meta evaluación, sugiriendo lo que serían evaluaciones buenas y malas. Sin embargo, con el surgimiento de diferentes listas de criterios para juzgar la calidad de los procesos evaluativos los que encargaron evaluación, los que son evalua-

dos y los que evalúan comienzan a cuestionarse la lista que deben utilizar. Ya por entonces la utilización de testes en la evaluación de programas educativos era una cuestión que generaba discusión y demandaba una atención especial. Sin embargo fue tan sólo en 1975 cuando se formó un comité, compuesto inicialmente por integrantes de la Asociación Americana de Psicología (*American Psychological Association*), de la Asociación Americana de Pesquisa Educativa (*American Educational Research Association*), y del Consejo Nacional de Medidas en Educación (*National Council on Measurement Education*) de los Estados Unidos y que luego incorporó profesionales de otras organizaciones formando el *Joint Committee on Standards for Educational Evaluation* permitiendo que, en 1981, se generase la publicación *Standards for Evaluations of Educational Programs, Projects and Materials* (JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION, 1981).

Más tarde se formó un nuevo comité, formado por profesionales de evaluación y representantes de diferentes grupos de interés de la comunidad más amplia (interesados e involucrados de alguna forma en procesos evaluativos), que revisó los patrones desarrollados específicamente para el contexto escolar haciendo una adaptación para que estos pasasen a atender las evaluaciones de programas proyectos e instituciones, de cualquier naturaleza, desarrollando un trabajo que combinó algunos de los patrones que ya existían y añadió nuevos patrones. Este trabajo propició un conjunto de treinta patrones universales que deben, siempre que sean aplicables, estar presentes en la dirección de procesos evaluativos

realizados en cualquier área del conocimiento. De esta forma, actualmente, los patrones ya son aplicados para evaluación: educativa, de programas sociales, empresarial e institucional, entre otras. Estos patrones son organizados en los cuatro atributos característicos (o categorías) que el comité cree que son necesarios y suficientes para que se tenga una verdadera evaluación, conforme define el *Joint Committee on Standards for Educational Evaluation* (1994, grifo nosso)²:

Utilidad: el patrón utilidad pretende asegurar que una evaluación supla las necesidades de información práctica de los usuarios pretendidos.

Viabilidad: el patrón viabilidad pretende asegurar que una evaluación sea realista, prudente, diplomática y sencilla.

Ética: el patrón ética pretende asegurar que una evaluación sea dirigida legalmente, éticamente y con el debido respeto al bienestar de los involucrados, así como de aquellos afectados por sus resultados.

Precisión: el patrón precisión pretende asegurar que la evaluación lo revele y transmita, técnicamente, informaciones adecuadas sobre las características que expresan el mérito o la relevancia del foco de atención.

Por consiguiente, el *Joint Committee on Standards for Educational Evaluation* entiende que una evaluación no se debe hacer (i) si esta no es útil para una determinada audiencia o público lo que se define en términos de identificación de interesados, credibilidad del evaluador, selección y escopo de la información, identifi-

² Los patrones están siendo actualizados este año y pronto será divulgada una versión actualizada de los mismos

cación de valores, claridad en el informe, presentación (a tiempo), de los resultados, divulgación e impacto de la evaluación; (ii) si no fuese viable llevarla a cabo de forma realista, prudente y diplomática, lo que se define en términos de procedimientos prácticos, viabilidad política y de costo beneficio y (iii) si no fuese posible asegurar que su dirección será legal y ética, con el debido respeto al bienestar de los involucrados, lo que se define en términos de acuerdos formales y contratos, derechos humanos de los individuos, relaciones humanas, verificación y aferición completa y justa, transparencia de procedimientos y de resultados, solución de conflictos de interés y responsabilidad fiscal. El orden de listado de estos atributos no es accidental. El primero es la utilidad, pues si una evaluación no fuese útil, esta no debe ser realizada. Caso este criterio sea atendido se observarán las condiciones para atender a los atributos de viabilidad y de ética y solamente si estos tres atributos fuesen atendidos los evaluadores precisan (y deben) dedicarse a la cuestión más técnica de buscar la precisión en la colecta y en el tratamiento de los datos, que se define en términos de documentación del programa, análisis contextual, descripción de propósitos y procedimientos, información confiable, información válida, información consistente, información sistemática, análisis de información cuantitativa, análisis de información cualitativa, conclusiones justificables, informe imparcial y meta evaluación.

“Cualquier proceso evaluativo será hasta cierto punto tendencioso” (WORTHEN; SANDERS; FITZPATRICK, 1987). Cuando se traza un proceso evaluativo varios aspectos, tales como, cuestiones evaluativas, métodos y técnicas de recopilación de datos y quién serán los respondientes, deben ser, desde el inicio, negociados con quien está encargado la evaluación y con representantes de los que se-

rán evaluados. Estas decisiones, de alguna forma, afectarán los resultados del proceso evaluativo. Evaluadores y clientes precisan estar conscientes de las tendencias del proceso evaluativo, buscando, siempre que sea posible, minimizarlas, y, cuando esto no sea posible, relatarlas. Es justamente para minimizar el bias y garantizar la calidad de un proceso evaluativo en todas sus etapas (decisión sobre la realización de la evaluación, definición del propósito de la evaluación, diseño de la evaluación, recopilación de información, análisis de información, elaboración del presupuesto, elaboración del contrato, coordinación, formación del equipo) que son dirigidas meta evaluaciones.

De la misma forma que se recomienda que las evaluaciones se realicen en las perspectivas formativa y sumativa (SCRIVEN, 1967), las meta evaluaciones también se deben dirigir en estas dos perspectivas, que, de cierta forma, se complementan. La meta evaluación formativa se dirige a lo largo del proceso evaluativo. Lo ideal es que su inicio coincida con el de la evaluación. La principal finalidad de la meta evaluación formativa es fornecer al equipo responsable por dirigir el proceso evaluativo que está siendo sometido a la meta evaluación información útil para mejorar este proceso mientras este aun está en curso. . La meta evaluación sumativa se realiza al término del proceso evaluativo, buscando respuestas conclusivas sobre su mérito y su relevancia para los contratantes, usuarios y otros interesados en el proceso evaluativo, y, sobretudo, dando credibilidad al proceso y a los resultados finales que este genere. En otras palabras, mientras la meta evaluación formativa tiene el papel de mejorar el proceso evaluativo a lo largo de su dirección, la sumativa tiene el papel de informar a los involucrados y a la comunidad más amplia, y de contribuir para la mejora de procesos futuros.

Discutir procedimientos de meta evaluación es discutir la calidad del proceso evaluativo siendo, por tanto, fundamental considerar también los principios que un evaluador debe seguir para dirigir, a la luz de los criterios de una verdadera evaluación, una evaluación de calidad. La Asociación Americana de Evaluación creó, en 1992, una fuerza-tarea para desarrollar principios universales orientadores para los evaluadores. El trabajo de esta fuerza-tarea comenzó a través de la revisión, por cada uno de sus miembros, de documentos similares y relevantes elaborados por otras asociaciones profesionales el diseño elaborado por cada miembro circulaba por los otros miembros que hacían críticas y sugerencias y entonces era compilado por el coordinador del grupo. Tras incorporar las sugerencias y considerar las discusiones, en 1994, la propuesta fue sometida a los miembros de la referida asociación que aprobaron, a través de voto, el documento con cinco principios universales que deben estar presentes en evaluadores responsables por todo y cualquier tipo de evaluación (SHADISH et al., 1995, grifo nosso), a saber:

Indagación Sistemática: *evaluadores dirigen, sistemáticamente, indagación de datos sobre lo que quiera que sea que esté siendo evaluado.*

Competencia: *evaluadores ejercen actuación competente frente a los interesados.*

Integridad / honestidad: *evaluadores aseguran la honestidad y la integridad de todo el proceso evaluativo.*

Respeto por las personas: *evaluadores respetan la seguridad, dignidad y autoestima d los respondientes, de los participantes del programa, de los clientes y de otros intere-*

sados con los cuales él interactúe.

Responsabilidad por el bienestar general y público: *evaluadores articulan y tienen en consideración la diversidad de intereses y de valores que pueden estar relacionados al bienestar general y público.*

Una meta evaluación puede ser dirigida de variadas formas, adoptando diferentes instrumentales, pero la utilización de listas de chequeo para indicar, a lo largo y al término del proceso evaluativo, el cumplimiento de los treinta patrones de una verdadera evaluación ha sido un procedimiento adoptado por muchos evaluadores y centros de evaluación y ha generado resultados satisfactorios. En la práctica, cuando se transforma una lista de chequeo en un instrumental de recopilación de datos, se crea un conjunto de instrumentos donde cada ítem es una afirmativa con opciones de respuestas cerradas. Las listas de chequeo pueden ser de dos tipos: (i) en cada ítem el respondiente debe indicar la intensidad de su veracidad o de su ocurrencia, o aun informar que aquella afirmativa no se aplica, cuando sea el caso; (ii) en cada ítem el respondiente debe simplemente indicar si la afirmativa es verdadera o no. En el caso específico de la meta evaluación tales afirmativas procuran investigar la presencia de cada uno de los treinta patrones de una verdadera evaluación en el proceso evaluativo foco de la meta evaluación. Las listas de chequeo constituyen un instrumento eficiente, en un formato amigable, para compartir las lecciones aprendidas en la práctica (STUFFLEBEAM, 1994). Las listas de chequeo son muy útiles porque, en el mundo complejo de la evaluación, donde es necesario estar atento a los más variados aspectos, es preciso asegurar que ciertos criterios se están respetando.

La formación de los equipos de evaluación y de meta evaluación es una fase decisiva para la dirección y el éxito de un proceso evaluativo. Lo ideal es montar un equipo de evaluación Inter. Disciplinar que involucre tanto a evaluadores como a profesionales del área conocimiento del foco de la evaluación, que agregue competencias diversificadas y complementares. Existen varias formas de montar el equipo de meta evaluación, que difiere, principalmente, de acuerdo con el tamaño del proceso evaluativo que será sometido a la meta evaluación y con el papel de la meta evaluación. En evaluaciones menores, la meta evaluación formativa se puede hacer por un evaluador, con experiencia en meta evaluación, que interactuará como un consultor con el equipo responsable por la evaluación. Conforme aumenta el tamaño y la complejidad del proceso evaluativo, se puede aumentar el número de consultores con este papel o montar un equipo específico para meta evaluación. En un caso u otro lo fundamental es que haya mucha interacción entre los responsables por la evaluación y por la meta evaluación formativa. El equipo responsable por la meta evaluación sumativa debe ser formado por uno o más evaluadores cuya credibilidad sea reconocida.

Como es la calidad de la evaluación que va a garantizar, en un último análisis, o grande desafío que los evaluadores enfrentan actualmente san el éxito de la evaluación de la calidad, la capacitación de profesionales para el ejercicio de la evaluación, la expansión del conocimiento en el área para atingir, democráticamente, todos los involucrados en el desarrollo y en el perfeccionamiento de la calidad del foco de la evaluación (PENNA FIRME; LETICHEVSKY, 2002).

La importancia de la meta evaluación en la categoría PRECISIÓN

Como ya se ha dicho una evaluación sólo debe llevarse a cabo si esta es útil, ética y viable y, de esta forma una meta evaluación formativa debe verificar aun en la negociación del proceso evaluativo que estos tres patrones sean atendidos. En el diseño del proceso evaluativo propiamente dicho es preciso dedicarse a la cuestión más técnica de asegurar la precisión cuando es importante estar atento para: (a) calidad de los instrumentos de recopilación de datos, (b) calidad de la información recopiladas, (c) elección y aplicación adecuada de técnicas y modelos de tratamiento de los datos y, (d) transmisión adecuada de los resultados obtenidos dejando claro su alcance y sus limitaciones. Así, la meta evaluación también tiene un papel decisivo a lo que se refiere precisión auxiliando los evaluadores en la elección de las técnicas más adecuadas así como en su correcta aplicación y utilización. Este es un punto crucial de la meta evaluación ya que es frecuente observar procesos evaluativos que fracasan o presentan resultados que no ayudan a entender de forma adecuada la realidad en virtud de fallos en estos aspectos.

Con relación a la calidad de los instrumentos de recopilación de datos es preciso asegurar que estos sean adecuados para obtener datos referentes a lo que realmente se pretende recopilar. La validez de los instrumentos de recopilación de datos puede ser realizada de diferentes formas. En el caso de instrumentos que recopilan información cualitativa esta validez puede ser realizada con el auxilio de especialistas del área o a través de triangulación (confrontación entre diferentes evaluadores, técnicas e instrumentos) mientras que en el caso de informa-

ción cuantitativa se recomienda la aplicación de un Análisis Factorial Confirmatorio (BOLLEN, 1998) que es una técnica de reducción de dimensionalidad de los datos así como el Análisis Factorial Exploratorio, más conocido y utilizado con mayor frecuencia. La principal diferencia es que el Análisis Factorial Confirmatorio, se realiza a partir de la aplicación de un modelo de sistemas de ecuaciones estructurales y, por tanto, se supone *a priori* un modelo teórico relacionando variables latentes (o no observables) a las variables observables, mientras que en el Análisis Factorial Exploratorio toda y cualquier variable latente puede estar influenciando las variables observables, ya que no se conocen cuantos y cuales son los factores latentes antes de procesar el análisis. Es justamente esta diferencia que vuelve la primera confirmatoria y la segunda eminentemente exploratoria. Otra diferencia importante es que en el Análisis Factorial Confirmatorio los errores también son modelados³ y pueden (o no) estar correlacionados, mientras que en el Análisis Factorial Exploratorio existe la suposición de que los errores no pueden estar correlacionados (lo que no siempre es verdad) y lo que se observa es una vasta utilización de esta técnica sin realizar ningún tipo de teste para verificar si los errores, de hecho, no son correlacionados. Lo ideal es utilizar diferentes tipos de instrumentos de recopilación de datos contemplando, preferentemente, información cuantitativa y cualitativa.

En cuanto a la calidad de la información existen dos aspectos que precisan ser observados: (i) calidad y adecuación de las fuentes de información y; (ii) tratamiento adecuado de las bases de datos. Al escoger las fuentes de infor-

mación es importante asegurar (cuando sea el caso) que los diferentes grupos de informantes posibles sobre el foco evaluativo sean contemplados. Por otro lado, tras recopilar los datos antes de calcular los indicadores es fundamental tratar la base de datos expurgando la información que no reflejen el trazo latente que se quiere medir. Así en el caso de instrumentos que recopilen datos cuantitativos, deben ser excluidos los que presentan respuestas siempre con el mismo patrón, entregados en blanco o con señales de protesta, así como cualquier instrumento cuyos datos no reflejen una información útil de lo que se quiere medir. Cuando la información es cualitativa este problema se puede solucionar a través de triangulación de datos.

Al escoger la técnica más adecuada de modelar y analizar los datos es importante tener claro que cuestión(es) evaluativa(s) se pretende responder ya que una técnica adecuada para buscar la respuesta de una determinada cuestión puede no ser adecuada para responder otra. Por ejemplo, si una evaluación de larga escala de desempeño de alumnos se dirige con el objetivo de determinar en que escuelas están los mejores alumnos es posible trabajar directamente con los scores (resultados) de los alumnos. Por otro lado, si lo que se pretende es identificar cuales son las escuelas más eficaces⁴ es preciso considerar qué alumnos presentan algunas diferencias y traen históricos de vida diferenciadas, tanto por lo que se refiere a la educación formal como a conocimientos generales. Los niveles socioeconómicos de las familias son variados, los conocimientos preexistentes de los alumnos son

³ En el Análisis Factorial Confirmatorio el modelo establecido *a priori* – que considera la existencia de factores latentes (x o h) influenciando las variables observables (x) y, teniendo en cuenta los errores de medida (d o e), así como la asociación entre estos errores y las variables latentes (f) – es ajustado en el intuito de minimizar los residuos calculados a través de la diferencia entre las matrices de variancia y covariancia observada y estimada.

⁴Una escuela eficaz es aquella que contribuye para el progreso escolar de los alumnos, aunque se haga un control sobre el conocimiento traído por los alumnos antes de entrar en la escuela (MORTIMORE, 1991).

diferentes. Así, alumnos con nivel socio económico más alto, o alumnos que tengan mayor gama de conocimientos preexistentes, tenderían a tener mejor desempeño. Las interacciones entre el alumno y el ambiente escolar también interfieren en su desempeño y también deben ser incorporadas a los modelos.

Es en este contexto que surge la necesidad de aislar los efectos que independen de la escuela, o sea, que no están bajo el control de sus gestores, profesores, equipo pedagógico y de apoyo - y ni pueden llegar a estar -, de aquellos que dependen exclusivamente de los que dirigen o proceso de enseñanza. En este sentido, normalmente se pretende aislar, en especial, los efectos de nivel socioeconómico de los alumnos y de las escuelas que de alguna forma han impacto sobre sus desempeños. Así, lo que se desea es medir el valor agregado (GOLDSTEIN et al., 1993; CREEMERS, 1994; YANG et al., 1999b) por la escuela. Los métodos tradicionales de estudio de relación causa y efecto envuelven modelos de regresión hecha en un único nivel, donde una variable dependiente es explicada por un conjunto de variables independientes más un error. En este caso específico, la variable dependiente es la competencia del alumno (estimada por su desempeño en testes de contenido). Surge, sin embargo, la cuestión: ¿Es posible aceptar la validez de estos modelos ignorando las relaciones existentes entre diferentes niveles jerárquicos y la forma cómo estas relaciones impactan los resultados del estudio? La respuesta es no y estudios estadísticos confirman esta respuesta que puede ser obtenida hasta de forma intuitiva. Al analizar cuestiones multiniveles a través de modelos de un

único nivel, es posible que se cometan errores y en este caso, por tanto, es preciso utilizar modelos multiniveles (RABASH, 1999) que es la metodología de análisis indicada⁵ para datos con estructura jerárquica y patrón complejo de variabilidad. A la semejanza de lo que ocurre en escuelas, el mismo cuidado debe ser utilizado al evaluar el desempeño de sectores o directorias dentro de una empresa o el impacto obtenido en los beneficiarios de programas sociales con objetivos similares.

Siempre que se elaboran informes con resultados, los meta evaluadores deben actuar junto a los evaluadores en sentido de garantizar que los resultados están siendo transmitidos de forma clara y lo más próximo posible de la realidad dejando explícito cuales son las limitaciones y los cuidados que se debe tener al utilizar aquellos resultados y, cuando sea el caso, realizando recomendaciones de utilidades pertinentes. Así, en la divulgación de los resultados es preciso dejar claro, por ejemplo, si faltó recopilar información sobre alguna audiencia o si las diferencias presentadas son estadísticamente significativas.

La categoría Precisión de la Meta Evaluación

Tal como ocurre en procesos evaluativos cuando se realiza una meta evaluación lo ideal es agregar varias técnicas de recopilación de datos que contemplen información cuantitativa y cualitativa y diversificar las fuentes de datos de forma que se obtengan información de los diferentes grupos de interés y de las variadas audiencias involucradas.

⁵ Cabe resaltar que existen casos donde, a pesar de que la estructura de la población es jerárquica, no es posible aplicar modelos multiniveles, por no existir diferencias entre los grupos (lo que puede ser verificado a través del coeficiente de intra-correlación, cf. Raudenbush y Bryk, 2002) o debido al diseño de muestra (cuando sea necesario) adoptado en la recopilación de datos (PFEFFERMAN et al., 1998).

Tendrá, así, subsidios para la confrontación de información y para triangulación de datos. Cuando las listas de chequeo son rellenas las informaciones son posteriormente cuantificadas y en este momento también es crucial estar atento para la categoría precisión de la meta evaluación – perciba que como la meta evaluación también es un proceso evaluativo esta debe atender las cuatro categorías de una verdadera evaluación. En este sentido es importante considerar que no siempre es fácil (o incluso posible) para el meta evaluador definir si un patrón es atendido o no, o sea, existen situaciones donde el patrón es parcialmente atingido. En estos casos es necesario crear una escala como, por ejemplo: pobre, regular, bueno, muy bueno y excelente. El meta evaluador debe indicar en que punto de la escala cada patrón está con el auxilio de una lista de chequeo y, posteriormente, se calcula en que punto de la escala cada una de las categorías de una verdadera evaluación está.

El Centro de Evaluación de *Western Michigan University* desarrolló listas de chequeo que se organizan, de acuerdo con su finalidad inicial, en tres categorías, cualquiera que sea, planificación y gerencia, mérito y meta evaluación, que, además de ser utilizadas con éxito en la práctica, encuentran una gran aceptación en el mundo académico. Estos instrumentos son elaborados y sus resultados son tratados con base en la lógica clásica. Sin embargo, responder a estos instrumentos y tratar los datos recopilados con base a la lógica clásica (o bivalente) muchas veces se vuelve una tarea ardua tanto para los respondientes como para los evaluadores principalmente si consideramos que no son raras las solicitudes de evaluaciones de programas cuyo objetivo tie-

ne una imprecisión intrínseca⁶. En la Teoría Clásica de los Conjuntos, la función característica de un conjunto puede asumir apenas dos valores: uno (en el caso de que el elemento pertenezca al conjunto) y cero (en el caso de que el elemento no pertenezca al conjunto). El enfoque tradicional de tratamiento y adaptación de datos, construida a partir de la Teoría Clásica de los Conjuntos, utiliza muchas herramientas matemáticas que presuponen la precisión. Cuando se trabaja con problemas sencillos y bien definidos, este enfoque se adapta y, si está bien aplicada, genera resultados muy satisfactorios. Sin embargo, para problemas complejos, que pueden contener una imprecisión intrínseca, las herramientas matemáticas basadas en la lógica clásica pueden no adaptarse.

Para rellenar y tratar los dados de las listas de chequeo con base en la lógica clásica la frontera entre un punto y otro de la escala es preciso que esté muy claro y definido. Sin embargo no siempre es fácil para el ser humano definir esta frontera que puede tener una naturaleza no muy clara. En la Teoría de los Conjuntos *Fuzzy* los conceptos de pertinencia y no pertinencia de elementos a conjuntos son flexibilizados: La función característica puede asumir un número infinito de valores en el intervalo $[0,1]$, los cuales indican los grados de pertinencias de los elementos a un dato conjunto (TANSCHIEIT, 2004).

Cuando se opta por trabajar con la lógica tradicional en las listas de chequeo, surge una discusión que nos remite a la categoría precisión de los patrones de una verdadera evaluación y la pregunta que surge es: ¿será que se está midiendo lo que real-

⁶Imprecisión intrínseca es aquella asociada con la descripción de las propiedades de un fenómeno y no con la imprecisión asociada a la medición del fenómeno propiamente dicho.

mente se quiere? Esto aun se refuerza más por el hecho de que experiencias en laboratorio indicaron que la contextualización y el razonamiento *fuzzy* pueden ser típicos de la mitad de la población (KOCHEN, 1975).

La misma dificultad que existe en la fase de recopilación de datos también se enfrenta en el tratamiento de la información, por causa de que para realizar una evaluación es imprescindible establecer criterios de excelencia que sirven como parámetros para la elaboración de juicio de valor y que en realidad forman una base de reglas, generalmente fornecidas por especialistas, que son utilizadas para verificar se un resultado atiende o no a un determinado criterio. La lógica *fuzzy* permite tratar las cuestiones relacionadas a la imprecisión intrínseca y al conocimiento subjetivo de los especialistas, forneciendo una herramienta para traducir matemáticamente reglas lingüísticas del tipo SI-ENTONCES. Un problema frecuente, tanto en la práctica de la evaluación como en la de la meta evaluación, es la falta de consenso entre los especialistas, lo que puede generar reglas contradictorias. Estas pueden ser incorporadas a una base de reglas y tratadas de manera adecuada en el ámbito de la lógica *fuzzy*.

Como la creación de modelos tiene un papel importante, no en la reproducción realidad, sino en la contribución para su mejor entendimiento, trata el problema por medio de la lógica *fuzzy* es una opción que se considera bastante adecuada, pues posibilita (i) aceptar respuestas que indiquen el real entendimiento del respondiente con relación a la atención de un determinado patrón, (ii) utilizar reglas lingüísticas fornecidas por especialistas y, cuando sea necesario, incorporar reglas contradictorias en

un mismo modelo y (ii) tratar con la imprecisión intrínseca que generalmente existe en problemas complejos como puede ser el caso de procesos de meta evaluación .

Es importante, sin embargo, enfatizar que no sirve de nada todo el esfuerzo si los resultados de la meta evaluación no se utilizan para el perfeccionamiento de la evaluación albo (PENNA FIRME; LETICHEVSKY, 2002). De esta forma se puede extender la afirmativa de que el valor de una evaluación se refleja en la utilización de sus resultados (PATTON, 1986): El valor de una meta evaluación se refleja en la utilización de sus resultados para la mejora del proceso evaluativo y el fortalecimiento del equipo de evaluación.

Conclusión

La implantación de un proceso de meta evaluación envuelve cuidados especiales para la categoría **precisión**, tanto en lo que se refiere a la precisión de la evaluación a la que se está sometiendo la meta evaluación como en lo que se refiere a la precisión de la propia meta evaluación principalmente con relación a la elección apropiada de métodos y técnicas de recopilación, tratamiento, modelado y análisis de los datos y a su adecuada aplicación.

Cuando se diseña un proceso evaluativo las técnicas de recopilación, de tratamiento y de modelado de datos deben ser escogidas teniendo en consideración: la naturaleza de la evaluación, las cuestiones evaluativas que se pretende responder, el tipo de respondiente, y el perfil de los integrantes del equipo responsable por la evaluación. En este sentido el equipo de meta evaluación precisa dedicar una atención especial en auxiliar a los evaluadores para realizar una elección apropiada y, princi-

palmente, para una correcta aplicación de los métodos y técnicas escogidos. Aun es responsabilidad del meta evaluador celar por la correcta divulgación y apropiada utilización de los resultados. Asegurando que la divulgación sea responsable, ética y precisa y, aun, promoviendo la utilización de los resultados en el sentido de fornecer subsidios para la mejora del foco evaluativo. De forma análoga cuando se diseña una meta evaluación es preciso mantener el mismo cuidado dedicado a la elección y la aplicación de las técnicas y de los

métodos utilizados desde la recopilación de los datos hasta el análisis de los resultados. En este sentido, como la evaluación está siendo empleada en diversas áreas del conocimiento y, en muchos casos, atinge programas cuyos objetivos contienen algún tipo de imprecisión intrínseca, y además, como no siempre es posible obtener consenso entre los meta evaluadores existen situaciones donde es preciso romper con el paradigma de bivalencia de la lógica clásica y valerse de la lógica fuzzy que se basa en el principio de la multi valencia.

REFERÊNCIAS

BICHELMeyer, B. A. *Usability evaluation report*. Kalamazoo, MI: Western Michigan University, 2002.

BOLLEN, K. A. *Structural equations with latent variable*. New York: John Wiley & Sons, 1998.

BRYK, A.; RAUDENBUSH, S. *Hierarchical linear models*. Newbury Park, CA: Sage Publications, 1992.

FLETCHER, P. *À procura do ensino eficaz: relatório de pesquisa*. Brasília, DF: PNUD/MEC/SAEB, 1997.

FRANSES, P. H.; GELUK, I.; HOMELEN, V. P. Modeling item nonresponse in questionnaires. *Quality & Quantity*, Dordrecht, v. 33, p. 203-213, 1999.

GOLDSTEIN, H. *Multilevel statistical models*. London: Edward Arnold, 1995.

GOLDSTEIN, H. et al. *A user's guide to MlwiN: multilevel models project*. London: Institute of Education, University of London, 1999.

JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION. *Standards for evaluations of educational programs, projects and materials*. New York: Sage Publications, 1981.

_____. *The program evaluation standards: how to assess evaluations of educational programs*. 2nd ed. Thousand Oaks, CA: Sage Publications, 1994.

MORTIMORE, P. The nature and findings of school effectiveness research in primary sector. In: RIDDELL, S.; BROWN, S. (Ed.). *School effectiveness research: its messages for school improvement*. London: HMSO, 1991.

PATTON, M. Q. *Utilization-focused evaluation*. 3rd ed. Thousand Oaks, CA: Sage, 1997.

PENNA FIRME, T.; BLACKBURN, R.; PUTTEN, J. V. Avaliação de docentes e do ensino. In: SOUZA, E. C. B. M. (Org.). *Curso de especialização em avaliação à distância*. Brasília, DF: Universidade de Brasília, Unesco, 1998. v. 5.

PENNA FIRME, T.; LETICHEVSKY, A. C. O desenvolvimento da capacidade de avaliação no século XXI: enfrentando o desafio através da meta-avaliação. *Ensaio: avaliação e políticas públicas em educação: revista da Fundação Cesgranrio*, Rio de Janeiro, v. 10, n. 36, jul./set. 2002.

PFEFFERMANN, D. et al. Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society*, London, n. 1, p. 23-40, 1998. Series B. Statistical methodology.

RASBASH, J. et al. *MlwiN Beta version: multilevel models project*. London: Institute of Education, University of London, 1999.

RAUDENBUSH, S.; BRYK, A. *Hierarchical linear models: applications and data analysis*. 2nd ed. Newbury Park, CA: Sage, 2002.

RAUDENBUSH, S.; SAMPSON, R. Ecometrics: toward a science of assessing ecological settings, with application to the systematic observation of neighborhoods. *Sociological Methodology*, San Francisco, n. 29, p. 1-41, 1999.

SCRIVEN, M. The methodology of evaluation. In: TYLER, R. W.; GAGNÉ, R. M.; SCRIVEN, M. *Perspectives of curriculum evaluation*. Chicago, IL: Rand McNally, 1967. (American Educational Research Association monograph series on curriculum evaluation; n. 1).

SHADISH, W. R.; COORK, T. D.; LENITON, L. C. *Foundations of program evolution: theories of practice*. Newbury Park, CA: Sage, 1991.

SHADISH, W. R. et al. *Guiding principles for evaluators*. San Francisco: Jossey-Bass, 1995. (New Directions for Program Evaluation; n. 66).

STAKE, R. E. *Standards-based & responsive evaluation*. Thousand Oaks, CA: Sage, 2004.

STUFFLEBEAM, D. Empowerment evaluation, objectivist evaluation, and evaluation standards: where the future of evaluation should not go and where it needs to go. *Evaluation Practice*, Beverly Hills, CA, v. 15, n. 3, p. 321-338, 1994.

TANSCHIEIT, R. *Sistemas fuzzy*. Rio de Janeiro: Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, 2004.

WINGATE, L. A. The evaluation checklist project: the inside scoop on content, process, policies, impact, and challenges. In: AMERICAN EVALUATION ASSOCIATION CONFERENCE, 2002, Washington, DC. *Paper presented ...* Washington, DC, 2002.

YANG, M. et al. *MlwiN macros for advanced multilevel modelling, version 2.0: multilevel models project*. London: Institute of Education, University of London, 1999a.

_____. The use of assessment data for school improvement purposes. *Oxford Review of Education*, Dorchester-on-Thames, Oxford, n. 25, p. 469-483, 1999b.