

Capítulo 1

MARCO HISTÓRICO DE LA EVALUACIÓN

Nicolás Ortiz-Ruiz

Antes de avanzar con el contenido específico de este libro, es preciso hacer un breve recorrido histórico del desarrollo de la práctica evaluativa de intervenciones de interés colectivo. En la historia reciente de Occidente, la evaluación marchó a la par de la consolidación de los aparatos estatales, articulada a discusiones teóricas y metodológicas que buscan refinar y mejorar su aplicación y usos. Los registros indican que se empieza a consolidar a comienzos del siglo XVIII en Gran Bretaña y Francia, aunque existen evidencias de hace más de 4000 años, en la civilización china, donde se utilizaba para el reclutamiento de funcionarios públicos (Forrest, 1990). Con mayor precisión, podría afirmarse que, en la modernidad, la evaluación fue avanzando como resultado de la consolidación de los Estados, el capitalismo y la ciencia, toda vez que combina conocimientos y dispositivos metodológicos precisos para medir, valorar y optimizar acciones, con respecto a ciertos intereses y propósitos, dentro de la pretensión moderna de la racionalidad basada en la optimización de ganancias, beneficios y la previsión de posibles efectos deseados y no deseados.

En cuanto a las etapas de desarrollo, existe un amplio consenso acerca de una historia ligada a los principales acontecimientos de los últimos dos siglos, identificando cuatro "generaciones" y seis periodos, los cuales son descritos tomando la revisión realizada por Brousselle *et al.* (2011).

El reformismo (1800-1900)

Durante este período se producen en Europa los diseños más cercanos a lo que hoy se conoce como evaluación. Las coordenadas históricas están marcadas por la racionalidad científica de la Ilustración y su necesidad de la comprobación empírica del conocimiento. Así mismo, con los inicios y primeras etapas de la Revolución Industrial, surgida en Gran Bretaña y luego extendida a otros territorios europeos y de América del Norte, en primera instancia (Hernández y Guzmán, 1991). Durante este periodo, una parte del planeta empieza a vivir acelerados procesos de urbanización, el surgimiento de nuevas clases sociales y nuevas relaciones de producción, que implicaban desafíos en la implementación de intervenciones sociales orientadas a resolver las necesidades de estos sectores sociales emergentes (Brousselle *et al.*, 2011).

Esto implicó un esfuerzo sin precedentes de los gobiernos para diseñar e institucionalizar un conjunto de políticas sociales, en lo que se podría llamar "reformismo". En ese contexto, Europa y América del Norte desarrollaron una infraestructura urbana para combatir los riesgos de enfermedades vinculadas a la insalubridad y fortalecieron los sistemas educativos (Brousselle *et al.*, 2011). Estas iniciativas alcanzaron un punto alto en la década de los ochenta del siglo XIX, cuando varios países de Europa adoptan políticas explícitas de bienestar. En ese entonces, las preocupaciones por la evaluación surgen de la necesidad de saber en qué medida las intervenciones están mejorando las condiciones de vida de las sociedades y de los individuos, para lo cual se impulsó la aplicación de métodos científicos para comprender mejor el fenómeno social (Brousselle *et al.*, 2011).

La idea de tener información para la toma de decisiones se instaura con fuerza en los sistemas educativos, constituyéndose en foco de atención de los gobernantes y científicos sociales. En los Estados Unidos de América, a mediados del siglo XIX se produce información sobre la distribución geográfica de las escuelas, las competencias de los profesores de formación primaria, la disciplina, la supervisión, la situación de la infraestructura y las finanzas de los establecimientos; en suma, datos ligados principalmente a los recursos utilizados en la producción escolar. Se realizaron las primeras pruebas estandarizadas a gran escala para evaluar el rendimiento de los estudiantes, clasificarlos y establecer comparaciones entre escuelas, constituyéndose así, en antecedente para la adopción de pruebas estandarizadas y el uso de los resultados de los estudiantes como criterio para la evaluación escolar (Brousselle *et al.*, 2011).

En Europa también se recurre a los métodos experimentales buscando medir las capacidades intelectuales a través de los laboratorios de psicometría (Guba y Lincoln, 1989). Estas primeras experiencias utilizaban pruebas estandarizadas para obtener una imagen objetiva y adecuada de las habilidades de los estudiantes o de otros sujetos, midiendo las diferencias en relación con la norma para clasificar los sujetos observados (Muñoz, 2007). Este tipo de evaluación se inscribe en el paradigma científico y

recurre al método experimental, constituyendo "la primera generación" de la evaluación (Brousselle *et al.*, 2011; Samico *et al.*, 2010).

En salud, durante la Revolución Industrial y las reformas sanitarias, fueron de gran relevancia los estudios estadísticos y epidemiológicos, para el desarrollo de evaluaciones (Rosen, 1986). A comienzos del siglo XIX Gran Bretaña instituyó el registro sistemático de muertes, luego el sistema nacional para registrar las causas de muerte, dando paso a métodos estadísticos sofisticados para monitorear la salud de la población. Algunos estudios famosos como los de Snow, Villermé y Virchow guiaron formas de intervención en salud pública. Para ese entonces, la Sociedad Estadística de Londres, la Sociedad Estadística Americana y la Sociedad Americana de Salud Pública, fundadas respectivamente en 1834, 1858 y 1873, alcanzaron una influencia considerable (Rosen, 1994).

Se crean numerosas comisiones presidenciales o reales con fines de evaluación para estudiar temas específicos basados en evidencias. Edwin Chadwick, por ejemplo, aporta información valiosa sobre las condiciones de vida de las clases trabajadoras emergentes para ajustar la Ley de Pobres que operaba en Gran Bretaña (Checkland y Checkland, 1974). En los Estados Unidos, el Informe sobre la Condición de los Trabajadores de Nueva York, y el Informe de la Comisión Sanitaria de Massachusetts, se apoyaron en datos y estadísticas de salud para documentar diferencias en morbilidad y mortalidad y hacer recomendaciones para mejorar la salud pública. A finales del siglo XIX, las estadísticas en los asuntos de salud pública, ligada a los desarrollos bacteriológicos, ofrecían insumos para la planeación de intervenciones y la evaluación de su utilidad social (Brousselle *et al.*, 2011).

En síntesis, durante este periodo, las actividades de evaluación se concentraron en pruebas estandarizadas y mecanismos de recolección sistemática de datos. La investigación experimental se utilizó en los campos de la educación y la salud para evaluar la utilidad de diversas intervenciones, identificar las causas asociadas con varios problemas y desarrollar nuevas herramientas de intervención. Así mismo, analizar, comparar y clasificar individuos y grupos de la población. Durante este

periodo, la proliferación de comisiones y el uso de herramientas de la estadística, la epidemiología y otras disciplinas, cimentaron las bases para la institucionalización de la práctica evaluativa (Brousselle *et al.*, 2011).

La eficiencia y las pruebas (1900-1930)

En este periodo, los instrumentos de medición son perfeccionados con el propósito de consolidar los esfuerzos de la evaluación de “primera generación”. La figura de la máquina se convierte en la representación del funcionamiento de las organizaciones productivas y políticas, bajo la supremacía de la estandarización, la cuantificación, la uniformidad y la precisión. Taylor en los Estados Unidos, Fayol en Francia y Weber en Alemania, sientan las bases para el estudio de la administración científica del trabajo, la teoría administrativa y las formas de organización burocrática, respectivamente. Sobre estos referentes, las intervenciones económicas, políticas y sociales buscaban mayores utilidades, optimizando su funcionamiento y aplicando pruebas estandarizadas para medir y clasificar individuos respecto a su rendimiento escolar, sus conocimientos y sus capacidades cognitivas (Brousselle *et al.*, 2011).

En el campo de la salud, la epidemiología y la estadística usan bancos de datos para justificar científicamente los programas, siguiendo los parámetros de la gestión clásica en el funcionamiento social. Winslow (1920) resalta la importancia de utilizar toda la “maquinaria social”, refiriéndose a los programas que contribuyan a mejorar la salud de los miembros de la población. En muchos países de Europa y en Estados Unidos, esto se evidenció en la creación de los Servicios Nacionales de Salud Pública, a través de los cuales se implementan proyectos de demostración, al igual que con los programas de control de enfermedades infecciosas.

A partir de 1920 empiezan a surgir señales de cambio de la evaluación. Tanto en educación como en salud pública es usada para cuestionar las políticas y los programas. En los Estados Unidos, la emergencia de múltiples agencias nacionales de salud y los problemas de coordinación que supuso la proliferación de intervenciones públi-

cas, hizo que se cuestionara la pertinencia de las acciones y la utilización eficiente de los recursos. Sin embargo, este periodo estuvo marcado por la lógica del reformismo, fundamentalmente normativa, limitándose a actividades de medida, recolección sistemática de datos y análisis cuantitativos, inclusive, usando como sinónimos la medición y la evaluación (Brousselle *et al.*, 2011).

La edad de la inocencia (1930-1957)

Esta etapa denominada “segunda generación” está marcada por la crisis económica de finales de la década de los veinte, la postguerra y por el auge de las políticas keynesianas en Estados Unidos (New Deal) y Europa (Estado de bienestar/providencia) (Guba y Lincoln, 1989). En este periodo el Estado fue protagonista de la intervención social y motor del progreso, por causa de su acción directa en la vida económica y social, capaz de establecer los equilibrios fundamentales y de favorecer el crecimiento. Se denomina la edad de la inocencia porque se confía en los méritos de los programas sociales cuya finalidad solo puede ser positiva en la senda del progreso y los conocimientos científicos pueden mejorar dichos programas (Rossi *et al.*, 2018).

Durante la década de los treinta, en los Estados Unidos se vive una crítica de los programas educativos, lo que llevó a reformas que marcan la segunda postguerra. Se implementan programas bajo una pedagogía progresista orientada por objetivos que tuvieran en cuenta las necesidades físicas, emocionales e intelectuales de los estudiantes. Se propone, en consecuencia, un tipo de evaluación que en lugar de clasificar los individuos en la búsqueda de desviaciones en relación con lo “normal”, determine en qué medida los objetivos pedagógicos cognitivos, afectivos o psicomotores eran alcanzados (Tyler, 1949).

Así, la evaluación se torna en una herramienta para mejorar los programas y las intervenciones, dando paso a la evaluación formativa. Este tipo de evaluación se funda en la idea de que la eficacia de un programa no depende únicamente de las características de los participantes o beneficiarios, sino también del conjunto de elementos ligados a la

estructura del programa, sus estrategias y recursos. Esto es, de los elementos que constituyen las intervenciones para responder a los problemas y necesidades para las cuales son implementadas. En esta nueva etapa el evaluador describe y mide (Samico *et al.*, 2010; Muñoz, 2007).

La Segunda Guerra Mundial generó en el campo militar necesidades que ampliaron el papel de la evaluación, impulsándola. En el área de la salud, se creó, en las zonas de guerra, el Centro para el Control de la Malaria¹, cuya función era vigilar la prevalencia de la enfermedad y apoyar a los Estados miembros en el diseño de sus programas de salud pública (Gómez y Khoshnood, 1991). Terminada la guerra, este impulso extraordinario no cesó, pues las inversiones posteriores de los Estados en múltiples campos como la salud, la educación, la vivienda, entre otros, debían ser documentadas (Rossi *et al.*, 2018). En esa misma época se ve reforzada la necesidad de la evaluación debido a la creación de las agencias de desarrollo económico y social de las Naciones Unidas, el Banco Mundial, la Organización Mundial de la Salud (OMS) y el Programa de las Naciones Unidas para el Desarrollo (PNUD) (Brousselle *et al.*, 2011).

Algunos avances en las ciencias sociales trajeron aportes a la disciplina de la evaluación. Son destacables los aportes que desde la psicología realiza Elton Mayo entre 1927 y 1932 (fábrica de Western Electric Hawthorne, en Chicago) al llamar la atención sobre los aspectos cualitativos del sistema de producción que hasta ahora no habían sido considerados porque el foco estaba puesto en los aspectos cuantitativos (Mayo, 1946). Lo mismo ocurre en la década de los cuarenta con los trabajos de Lewin y de Lippitt en los que muestran que la evaluación puede ser parte constitutiva de las intervenciones y no acciones realizadas por expertos, facilitando con ello procesos de cambio (Lippitt, 1965; Lewin, 1951).

En síntesis, se experimentaron varios cambios en la evaluación: el evaluador amplió su función de recolectar datos de forma sistemática y realizar mediciones, para también describir los programas, comprender su estructura y contenido, identificar

sus fortalezas y debilidades a fin de determinar las razones del cumplimiento o no de los objetivos y, en esa medida, hacer mejores recomendaciones. Por otra parte, la atención deja de estar solo en las características de los individuos e incorpora la intervención como aspecto clave. La “medición” y la “evaluación”, por su parte, se convierten en dos conceptos distintos, considerando que la medición está al servicio de la evaluación. Adicionalmente, los instrumentos existentes y las herramientas apuntan a identificar mejor los resultados de las intervenciones y a aplicar el procedimiento de medición a varios aspectos cualitativos de estas (Brousselle *et al.*, 2011).

La expansión (1957-1972)

La feroz competencia entre Estados Unidos y la Unión Soviética durante la Guerra Fría desencadenó cuestionamientos en el sistema educativo norteamericano, lo que generó fuertes inversiones para su mejoramiento, decisión que fue acompañada por el imperativo de evaluar los nuevos programas educativos. No obstante, se perciben dos vacíos en la evaluación, a saber: por un lado, aunque la evaluación se basara en el cumplimiento de los objetivos, no permitía juzgar el valor de esos objetivos, ni probar su pertinencia en relación con las necesidades. Por otro lado, el hecho de que la medición de los objetivos sea *a posteriori* hace que solo puedan detectarse las deficiencias *ex post* (Cronbach, 2000). Así entonces, durante este periodo se desarrollan múltiples modelos que ponen en cuestión los objetivos y contribuyen a identificar los factores para alcanzarlos. La evaluación es concebida como una herramienta para aumentar la eficacia de los programas (Brousselle *et al.*, 2011).

Ligado a esto, el auge de los Estados de bienestar trajo una fecunda implementación y experimentación de programas sociales. En ese contexto, la evaluación es considerada como una herramienta necesaria para diseñar los programas, agilizar la planificación y asignación de recursos públicos y garantizar la rendición de cuentas. Se pregona la creación de una sociedad de experimentación en la que las políticas y los programas estarían sujetos a

¹ Antecedió a los Centros de Control y Prevención de Enfermedades (CDC) de los Estados Unidos de América.

evaluación científica antes de ser implementados (Campbell y Stanley, 1966). Bajo esta perspectiva, las ciencias sociales ofrecen herramientas para fortalecer la intervención de los asuntos sociales, lo cual crea expectativas frente a la evaluación y le da un lugar destacado en la gestión del Estado como disciplina basada en la ciencia.

En resumen, este cuarto período se caracteriza por una ampliación del marco conceptual y las funciones de evaluación. La evaluación debe permitir la emisión de un juicio sobre una intervención; por ende, al evaluador se le asigna el papel de juez, valiéndose para ello de la descripción y de la medición. La evaluación experimenta una expansión y adquiere mayor legitimidad, soportada en importantes inversiones para su desarrollo y una amplia participación de expertos (Brousselle *et al.*, 2011).

La profesionalización y la institucionalización de la disciplina (1973-1989)

En este periodo, los postulados de la llamada "tercera generación" todavía constituyen el marco principal de referencia. El énfasis permanece en el papel de juez del evaluador y las funciones de medición y descripción son reforzadas por la revolución informática. Pese a los vínculos entre la evaluación como práctica científica y la política, hubo algunos desencuentros, pues mucha de la evidencia que era significativa para los científicos no lo era para los decisores y viceversa, o información científicamente válida entraba abiertamente en conflicto con poderes, intereses e ideologías que tenían mayor capacidad de decisión. En el mismo sentido, los tiempos de la evaluación no coincidían necesariamente con los tiempos políticos, lo que dificultaba su sincronización con el uso de la información (Brousselle *et al.*, 2011).

En este contexto, varios teóricos influyentes, como Carol Weiss y Robert Stake, proponen un enfoque pragmático de la evaluación en el que se favorezca al mismo tiempo la participación de usuarios y clientes en el proceso de producción de evaluación y el mayor uso de sus resultados (Stake, 1983; Weiss, 1990). En la década de los setenta hubo

algunos cambios, entre ellos: nuevos mecanismos para la difusión del conocimiento producido, a través de libros de referencia, monografías, colecciones, enciclopedias y revistas, igualmente, se crearon agencias estatales, programas de formación en la materia y asociaciones comprometidas con la promoción de la evaluación, apoyo a evaluadores y avance del conocimiento (Brousselle *et al.*, 2011).

Recapitulando, este período se caracteriza por la evolución hacia una evaluación profesionalizada, aunque esto aún no es un proceso terminado. Los métodos utilizados corresponden al enfoque de evaluación de la "tercera generación", enfatizando el papel de juez del evaluador; sin embargo, los criterios de juicio que cobran mayor relevancia son los aspectos económicos de programas. Por otra parte, la revolución informática favoreció el refinamiento de estándares de medición existentes y mejoró las capacidades de manejo de datos (Brousselle *et al.*, 2011).

Las dudas (desde 1990)

Después de este breve recuento histórico, son claros los aportes de la evaluación y su utilidad para medir, controlar, administrar, prever, definir prioridades, escoger entre diferentes programas y clasificar individuos (Vieira-da-Silva, 2014; Patton, 2002; Brousselle *et al.*, 2011). Sin embargo, persisten las críticas sobre las certezas y el ideal de verdad que subyace al método científico positivista, dando lugar a las dudas. Con ese espíritu, Guba y Lincoln publican el libro *Fourth Generation Evaluation*, en el cual presentan un nuevo abordaje posible de la evaluación (Guba y Lincoln, 1989).

Estos autores llaman la atención sobre los principios que fundamentan la evaluación y sobre la naturaleza contextual y negociada tanto de los alcances y el foco como de sus resultados (Figueira *et al.*, 2010). En este sentido, afirman que, pese a la pretensión de objetividad, la práctica de la evaluación siempre va a estar condicionada por los valores y decisiones de los evaluadores, así como su instrumentalización por parte de agentes y estrategias políticas. Cuestionan cómo la preponderancia del método experimental y de mediciones

cuantitativas, bajo relaciones directas de causalidad, ocultan aspectos contextuales, políticos y culturales que no pueden ser captados por las mediciones (Guba y Lincoln, 1989).

Por consiguiente, la evaluación de “cuarta generación” toma en cuenta los desarrollos anteriores, buscando llenar los vacíos y alcanzar mayor complejidad en la captación de la realidad. Para ello, se convierte en un instrumento de negociación y empoderamiento implementando prácticas participativas. Este tipo de evaluación, llamada “constructivista”, señala que mientras en las generaciones anteriores los parámetros han sido construidos *a priori*, en esta, los límites y parámetros son construidos por quienes participan en la intervención/evaluación, siendo esto un elemento importante dentro del mismo proceso de evaluación (Dobles, 1996).

Se enmarca, por tanto, dentro del paradigma interpretativo o cualitativo, reconociendo que no existe realidad objetiva, sino varias realidades, en tanto construcción social (dimensión ontológica). En coherencia, niega la existencia del dualismo sujeto que estudia y objeto estudiado, proponiendo la interacción entre ellos, a través de la cual se produce o emerge el conocimiento de la evaluación (dimensión epistemológica) (Alcaraz, 2016).

La evaluación desde esta perspectiva deja de ser un campo reservado para especialistas, tornándose en una práctica abierta a todos los involucrados en su desarrollo. En consecuencia, al interés técnico de las evaluaciones, se suma uno práctico, potencialmente político y liberador que favorece el diálogo, la reflexión crítica y la comprensión mutua de los diferentes actores quienes participan también en el mejoramiento de las intervenciones (Guba y Lincoln, 1989). En esta tarea, es fundamental el papel del evaluador como moderador, adicionalmente de sus funciones de descripción, valoración y juzgamiento. Por eso, el evaluador debe, a toda costa, favorecer las condiciones necesarias para que todos los actores involucrados puedan participar en cada una de las etapas y actividades de la evaluación (Brousselle *et al.*, 2011; House y Howe, 1999; Mertens, 1999).

Las ramas de la evaluación

Para cerrar este capítulo, es conveniente señalar los aportes de Alkin (2004) en su intento por sistematizar y organizar las teorías de la evaluación usando la figura de un árbol. En una de las raíces sitúa el control y la rendición de cuentas a la sociedad (*accountability*) y, en la otra, la investigación social; esta última surgida de la preocupación por disponer de métodos sistemáticos que ofrezcan información para la rendición de cuentas. Adicionalmente, el árbol tiene tres ramas: Una central, que corresponde a los métodos, acompañada por dos ramas laterales, una que recoge a los autores que enfatizan en la importancia del uso de la evaluación y la otra en la que desataca como finalidad principal de la evaluación la emisión de juicios y valoraciones. Aunque, sin lugar a dudas, la importancia de cada una de las ramas es susceptible de discusión, para Alkin la rama central que comprende los métodos es la más relevante, toda vez que la evaluación es un ejercicio de investigación social aplicada y, por tanto, es primordial la total claridad en la metodología y la selección y diseño de los modelos más adecuados para cada caso de estudio. Ubica como un referente fundacional a Campbell en su empeño por desarrollar enfoques experimentales y cuasiexperimentales de la evaluación. Igualmente, menciona los aportes de Mark, Henry y Julnes², Boruch, Chen, Weiss y Cronbach (Christie y Alkin, 2020). Estos dos últimos autores, con aportes relevantes que vinculan la evaluación con la política, considerando la evaluación como parte integral de una política de investigación enfocada en la elaboración y ejecución de políticas sociales (Alkin, 2004).

La rama de la valoración o el juicio incluye autores como Scriven o Eisner, que consideran que el

² Estos autores desarrollan la Evaluación Realista Emergente como un nuevo modelo de evaluación integral que ofrece nociones reconceptualizadas de utilización, métodos y valoración. La metodología es el pilar fundamental: 1) una metodología de evaluación que da prioridad al estudio de los mecanismos generativos; 2) atenta a múltiples niveles de análisis; y 3) combina métodos apropiadamente. Los mecanismos se definen como «las causas subyacentes de los cambios que son observados» (Mark *et al.*, 2000).

principal atributo de una evaluación es el juicio que se puede hacer sobre una intervención. Afirman que los evaluadores deben atender todas las etapas del proceso, poniendo atención especial al momento final de establecer juicios sobre la intervención evaluada. Para autores como Eisner, por ejemplo, esto se logra con la experticia del evaluador en el tema evaluado (Christie y Alkin, 2020). Autores como Stake, Guba y Lincoln, House y Howe y Greene se adscriben a los enfoques constructivistas e indican que no existe un valor verdadero para todo; es decir, reconocen la relatividad en el comportamiento de las intervenciones sociales, la importancia de la deliberación y la negociación, así mismo, la necesidad de considerar el contexto al igual que la perspectiva de los contratantes, ejecutores o participantes (Christie y Alkin, 2020).

En la tercera rama se ubican autores cuya preocupación responde a la pregunta ¿para qué la evaluación? Dicha pregunta es primordial para definir el diseño de cualquier evaluación. Autores como Cousins, Preskill y King destacan el carácter participativo de las evaluaciones y con ello el potencial para generar aprendizajes y transformación organizativa por parte de quienes participan en la ejecución de las intervenciones. Michael Patton, bajo un punto de vista realista, afirma que no es suficiente pensar en la evaluación como información para la toma de decisiones, sino que es necesario que los evaluadores piensen y apoyen a los contratantes/ejecutores en los tipos de decisiones que son posibles de tomar con la información disponible (Alkin, 2004).

Referencias

- Alcaraz, N. (2016). Aproximación histórica a la evaluación educativa: de la generación de la medición a la generación ecléctica. *Revista Iberoamericana de Evaluación Educativa*, 8(1), 11-25. <https://revistas.uam.es/riee/article/view/2973>
- Alkin, M. C. (ed.) (2004). *Evaluation toots: tracing theorist' views and influences*. Sage Publications.
- Brouselle, A., Champagne, F., Contandriopoulos, A. P. y Hartz, Z. (2011). Avaliação: conceitos e métodos. *Cadernos de Saúde Pública*, 28(7), 1414-1415. <https://doi.org/10.1590/s0102-311x2012000700019>
- Campbell, D. T. y Stanley, J. C. (1966). *Experimental and Quasi-Experimental Designs for Research* (1.a ed.). Wadsworth Publishing.
- Checkland, S. G. y Checkland, O. (1974). *The Poor Law Report of 1834* (1.a ed., Vol. 1). Penguin Books.
- Christie, C. A. y Alkin, M. C. (2020). Reexaminando el árbol de la teoría de la evaluación. *Cuadernos del CLAEH*, Segunda serie, 39(112), 135-147.
- Cronbach, L. J. (2000). Course Improvement Through Evaluation. En D. L. Stufflebeam, G. F. Madaus y T. Kellaghan (eds.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (pp. 235-247). Springer. https://doi.org/10.1007/0-306-47559-6_14
- Dobles, M. C. (1996). Acreditación en educación superior: orígenes, prácticas actuales, perspectivas. *Comunicación personal*. En Memoria Foro Taller Internacional sobre Tendencias Actuales en la Medición y Evaluación Educativa. Compiladora Leda Badilla. Universidad de Costa Rica.
- Figueira, A. C., de Frias, P. G. y Monteiro, L. (2010). *Conceitos básicos para a prática nas instituições*. Bases conceituais e operacionais. MedBook.
- Forrest, G. (1990). Oral examinations. En H. Walberg y G. Haertel (eds.), *The international encyclopedia of educational evaluation* (pp. 42-86). Pergamon Press.
- Gómez, O. y Khoshnood, B. (1991). *La evolución de la salud internacional en el siglo XX*. Salud Pública de México. <https://www.saludpublica.mx/index.php/spm/article/view/5409>
- Guba, E. y Lincoln, Y. S. (1989). *Fourth Generation Evaluation* (Illustrated ed.). Sage.
- Hernández, G. y Guzmán, C. (1991). Evaluación curricular. En *Módulo Bases Psicopedagógicas, Maestría en Tecnología Educativa* (pp. 59-88). Unidad 5, Diseño Curricular I, Tema IV. ILCE-OEA, México.
- House, E. y Howe, K. (1999). *Values in evaluation and social research*. Sage.
- Lewin, K. (1951). Field Theory in Social Science: Selected Theoretical Papers. *Social Service Review*, 25(3), 409-410. <https://doi.org/10.1086/638467>
- Lippitt, R. (1965). The use of social research to improve social practice. *American Journal of Orthopsychiatry*, 35(4), 663-669. <https://doi.org/10.1111/j.1939-0025.1965.tb00435.x>

- Mark, M., Henry, G. y Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. Jossey-Bass.
- Mayo, E. (1946). *The Human Problems of an Industrial Civilization* (2.ª ed.). Routledge.
- Mertens, D. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation*, 20(1), 1-14.
- Muñoz, G. A. (2007). Un nuevo paradigma: "la quinta generación de evaluación". *Laurus*, 13(23), 158-198. <https://www.redalyc.org/articulo.oa?id=76102309>
- Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods* (3.a ed.). Sage.
- Rosen, G. (1986). *De la policía médica a la medicina social: Ensayos sobre historia de la atención a la salud*. Siglo XXI.
- Rosen, G. (1994). *Uma história da saúde pública* (1.a ed.). UNESP.
- Rossi, P. H., Lipsey, M. W. y Henry, G. T. (2018). *Evaluation: A Systematic Approach* (8.a ed.). Sage.
- Samico, I., Felisberto, E., Figueiró, A. C. y Frias, P. G. (2010). Avaliação em saúde: bases conceituais e operacionais. *Cadernos de Saúde Pública*, 27(1), 195. <https://doi.org/10.1590/s0102-311x2011000100021>
- Stake, R. E. (1983). Program Evaluation, Particularly Responsive Evaluation. *Evaluation Models*, 6, 287-310. https://doi.org/10.1007/978-94-009-6669-7_17
- Tyler, R. (1949). *Basic Principles of Curriculum and Instruction* (1.a ed.). University of Chicago.
- Vieira-da-Silva, L. M. (2014). *Avaliação de políticas e programas de saúde* (1.a ed.). SciELO - Editora Fiocruz. <https://doi.org/10.1590/0102-311xre030315>
- Weiss, C. H. (1990). *Investigación evaluativa: métodos para determinar la eficiencia de los programas de acción* (1.a ed.). Trillas.
- Winslow, C. A. (1920). *The Untilled Fields of Public Health*. Science. <https://science.sciencemag.org/content/51/1306/23>