

Aproximación Histórica a la Evaluación Educativa: De la Generación de la Medición a la Generación Ecléctica

Historical Approach to the Educative Evaluation: From Generation of the Measurement to the Generation Eclectic

Noelia Alcaraz Salarirche *

Universidad de Cádiz

El objetivo de este artículo es mostrar el recorrido que el concepto de evaluación ha realizado desde que aparecieron las primeras prácticas evaluadoras en el ámbito educativo hasta la llegada del constructivismo, lo que representa un importante cambio de modelo. Para ello se ha hecho una revisión teórica completa sobre el tema en cuestión que posibilite dar respuestas a cuestiones como: ¿En qué época podemos situar el interés por evaluar? ¿Cuándo aparece la obsesión por la medición? ¿En qué momento es acuñado el término “evaluación educativa”? ¿Quién es considerado el padre de la evaluación en educación? ¿Qué nuevos modelos de evaluación comienzan a surgir en la época de los setenta? Y ¿Qué ocurre cuando el constructivismo aparece en el panorama educativo? Se concluye con una pequeña reflexión sobre si aquellas ideas constructivistas que implicaban cambios muy importantes en los modos de entender la evaluación en los contextos educativos, realmente transformaron las prácticas al uso. La autora propone una nueva generación de la evaluación para los tiempos actuales: la generación ecléctica.

Palabras claves: Evaluación, Medición, Rendición de cuentas, Paradigma, Constructivismo.

The aim of this paper is to show the evolution in educational assessment concept from first assessment practices in education until the constructivism appears what means an important change of model. So that way we make a theoretical study about this issue. In it we can find answers to questions like: what time can we find evaluation interest in? When does measurement obsession appear? What time does educative evaluation concept appear in? Who is considered evaluation father in education? What does new evaluation models appear in 70's? When the constructivist paradigm appears in education, what did it change? This paper shows an historical approach from firsts evaluation practices appear in educational context, to constructivist theories arrive, which carried on very important changes in the ways to understand evaluation in education. For finish, we include a reflexion about if those constructivist ideas what means changes so important in ways to understand assessment in education, really transform real practices. The author proposes a new generation of assessment for new ages: eclectic age.

Keywords: Evaluation, Measurement, Accountability, Paradigm, Constructivism.

*Contacto: noelia.alcaraz@uca.es

1. Introducción¹

A lo largo de todos los tiempos las personas evaluamos y somos evaluadas continuamente. Se puede decir que la evaluación es un proceso natural permanentemente presente en muchos aspectos de nuestras vidas. Evaluamos diariamente con distintos fines. Valoramos las características y posibilidades que nos ofrecen unos productos del mercado con respecto a otros, para comprarlos o no; atribuimos adjetivos a personas, situaciones o cosas, en función de las impresiones que nos dan; elaboramos juicios que nos ponen de parte de alguien o en contra de otro alguien en una disputa; buscamos conocer lo que le gusta o le molesta a nuestra pareja, con la intención de que la relación vaya a buen puerto, entre otros. Evaluamos, valoramos constantemente sin esfuerzo alguno aparente y casi sin darnos cuenta de ello. Igualmente, aunque a diferente nivel, más sistemático, más consciente, ocurre en los ámbitos profesionales: la medicina, la justicia, el periodismo, la construcción, la seguridad y vigilancia... y la educación. Y aunque se podría hablar de los porqués y para qué de la evaluación en otras esferas profesionales, nos ocupamos en este caso, de la evaluación en los contextos educativos.

2. Desde el 2000 a. de C. hasta 1930

El interés por la evaluación se remonta a mucho tiempo atrás. Forrest (citado en Sacristán, 2002) sitúa su primera manifestación histórica en el siglo II (a.C.), tratándose de una práctica china para seleccionar funcionarios. Se trata de las primeras prácticas selectivas de evaluación oral. Lemus (2012) añade que, algo así como un cuestionario de evaluación que utilizaron Sócrates y otros maestros de la época en sus prácticas de enseñanza, alrededor del siglo V (a.C.). Sitúa los sistemas de evaluación educativa (dogmáticos), entre los siglos V y XV (d.C.). No tan lejano, Lemus nombra a dos países precursores de la evaluación, durante el siglo XIX: Estados Unidos, en 1845, comienza a aplicar los test de rendimiento a estudiantes, con objeto de contribuir a la educación de los estudiantes. Y Gran Bretaña, donde existieron comisiones para evaluar los servicios públicos.

Hernández y Guzmán (1991), añaden un dato histórico curioso, dicen que en Gran Bretaña fue creada una comisión para evaluar la educación en Irlanda y que entre sus conclusiones se encontraba que el progreso de los niños (en las escuelas estudiadas) era menor al que debería ser. Para lo cual recomendaban una medida que atajara el problema: “pagar por resultados”, es decir, que parte del salario del profesorado iría en proporción a las calificaciones obtenidas por el alumnado al final del año escolar. Ellos lo resaltan por el valor tan actual que podría tener, y nosotros nos hacemos eco.

Estos mismos autores sitúan en el año 1845, en Boston, el uso de pruebas de rendimiento escolar como fuente básica de información para evaluar a las escuelas y al profesorado. Hernández y Guzmán (1991) también añaden que es entre 1887 y 1898, cuando el pedagogo Joseph Rice, utiliza por primera vez grupos de comparación para medir la eficacia de los programas; “se estudió los conocimientos de ortografía de 33.000 estudiantes de un amplio sector escolar (...). Este estudio está generalmente reconocido

¹ El contenido del presente artículo forma parte de la Tesis Doctoral de la autora, dirigida por Ángel I. Pérez Gómez, en el departamento de Didáctica y Organización Escolar de la Facultad de Ciencias de la Educación, de la Universidad de Málaga.

como la primera evaluación formal sobre un programa educativo realizado en América” según Stufflebeam y Shinkfield (citado en Vélez, 2007, p. 146).

Tales estudios tienen lugar en una época en la que muchos-as autores-as recientes, recogiendo las ideas de Stufflebeam y Shinkfield (2005), y Guba y Lincoln (1982, 1989), coinciden en denominar “época pretyleriana”, según los primeros, o “la primera generación de la evaluación”, “la técnica”, según los segundos.

Dicha etapa (la pretyleriana o técnica) está comprendida entre el 2000 a.C. y el año 1930. Reyes (2001), sitúa también hacia 1916 la aplicación de los test de rendimiento y test de inteligencia. Recoge y sintetiza que, en esta generación definida por Guba y Lincoln (1989), el papel del-la evaluador-a se caracterizaba por aplicar los instrumentos de forma técnica. Según Vélez (2007):

Para Stufflebeam y Shinkfield, las evaluaciones realizadas antes de los años 30 estaban centradas en valorar el resultado de los programas educativos en los alumnos a través de la aplicación de test, encuestas, acreditaciones y comparaciones experimentales. (p. 147)

Escudero (2003) reconoce el interés por la medición científica de las conductas humanas, de modo que las ciencias humanas asumen el positivismo de las ciencias físico-naturales. El autor enumera cuatro factores que confluyen en dicho momento y condicionaron la actividad evaluativa. Merece la pena destacar el cuarto: sitúan en el desarrollo de la sociedad industrial, el origen de los mecanismos de acreditación y selección de estudiantes, en función de sus conocimientos. Éste ha permanecido prácticamente inmutable desde entonces hasta hoy. La correcta selección y acreditación del alumnado, cada vez en niveles más bajos del sistema educativo, es una preocupación transversal a todas las sociedades avanzadas desde que las máquinas revolucionaran nuestro mundo.

Para Escudero (2003) medición y evaluación resultaban términos intercambiables. Sin embargo, en la práctica sólo se hablaba de medición. Surgen durante la Primera Guerra Mundial, un especial interés por la aplicación de test de inteligencias y personalidad, puestos al servicio de fines sociales.

El punto más alto del *testing* se sitúa en la década entre 1920 y 1930, donde los tests estandarizados ocupan un espacio privilegiado en el ámbito de la educación, con la intención de medir destrezas escolares, basados en procedimientos de medida de la inteligencia para utilizar con grandes colectivos de estudiantes.

Esta “época pretyleriana” o denominada por Guba y Lincoln (1989) como la “primera generación”, la “generación de la medida”, en la que la persona que evalúa es simplemente proveedora de instrumentos de medición, marca un periodo en el que la evaluación y la medida tenían poca relación con los programas escolares. Los tests informaban algo sobre el alumnado, pero nada de los programas de formación.

Coincidimos con los autores y con Gronlund (1985), en que esta primera generación (referida a la de los *testing*) permanece todavía viva, pues todavía existen textos y publicaciones que utilizan de manera indisoluble evaluación y medida.

3. Desde 1930 hasta 1957

El “periodo tyleriano”, reconocido así por Stufflebeam y Shinkfield (2005), comprendió desde 1930 hasta 1945, recibe este nombre por el que es conocido como el padre de la evaluación educativa, Ralph Tyler (1969). Él fue el primero en acuñar el término.

Hasta el momento, como hemos visto, evaluación y medición eran conceptos intercambiables. Según expone Vivian Durar (2013) en el blog “RED Maestros de Maestros”, el de evaluación era poco frecuente e iba acompañado siempre del de medida. Solía anteponerse el término medición al de evaluación. Es con la aparición de Ralph Tyler (1969), cuando evaluación pasó a un primer plano y la medición a un segundo, aunque eso sí, siempre ligados entre sí. Es así como nace el término evaluación educativa que se remonta a la época de Tyler (1969) en los años 30. Es Tyler (1969), por tanto, quien supera la mera evaluación psicológica (característica del periodo anterior) y sistematiza la evaluación en el ámbito educativo. Dicha propuesta basada en la formulación de objetivos curriculares muy precisos y en la comprobación de la consecución o no de los mismos, se extiende ampliamente por EE.UU.

Este modelo incluye un avance con respecto a la anterior época, en tanto en cuanto (Escudero, 2003):

El objeto del proceso evaluativo es determinar el cambio ocurrido en los alumnos, pero su función es más amplia que el hacer explícito este cambio a los propios alumnos, padres y profesores; es también un medio para informar sobre la eficacia del programa educacional y también de educación continua del profesor. (p.15)

Empieza a abandonarse la evaluación basada en la norma, y surge lo que se conoce como evaluación criterial. Según Stenhouse (1984), la primera nos informa del rendimiento del individuo en comparación con un grupo, mientras que la segunda, indica rendimiento de un individuo en relación con un estándar.

A esta “etapa tyleriana” se le conoce también según Guba y Lincoln (1982, 1989) como la segunda generación de la evaluación, la descriptiva, que comprendería hasta 1957, es decir, esta segunda generación se compuso por dos periodos, el hasta ahora explicado periodo Tyleriano (1930-1945), según Stufflebeam y Shinkfield (2005) y, siguiendo a estos autores, lo que se llamó época de la inocencia o irresponsabilidad social (1945-1957).

Estos doce años se caracterizaron por ser una continuidad de la mentalidad tyleriana (Vélez, 2007):

Las evaluaciones siguen respondiendo a las generaciones de la descripción y de la medición. Se recopilaba información, se describían las actuaciones públicas, y se medían sus resultados, pero no se ofrecían recomendaciones para la mejora de los programas. (p. 149)

Se trata de una época en la que según Escudero (2003:15): “la aportación de la evaluación a la mejora de la enseñanza es escasa debido a la carencia de planes coherentes de acción. Se escribe mucho de evaluación, pero con escasa influencia en el perfeccionamiento de la labor instruccional”.

Stufflebeam y Shinkfield (citado en Vélez, 2007:149) mantienen que “los educadores escribían sobre la evaluación y sobre la recopilación de los datos más importantes, pero al parecer, no traducían estos esfuerzos en intentos de mejorar los servicios educativos”.

4. Desde 1957 hasta 1972

Durante el periodo comprendido entre 1957 y 1972, tienen lugar una serie de acontecimientos coincidentes con gran parte de la época de la Guerra Fría. EE.UU se encuentra en plena carrera armamentística con la URSS, de modo que ésta en 1957 lanza el primer satélite artificial de la historia, el Sputnik. Lo que provoca una reacción de desencanto con la escuela pública en los EE.UU.

Se produce un incremento de la presión por la rendición de cuentas que alcanza su máximo esplendor a finales de los 60, con la aparición de un nuevo movimiento que vino a denominarse la era de la *accountability* (Escudero, 2003; MacDonald, 1976; Popham, 1980,1983; Stenhouse, 1984)

Este periodo es reconocido por Stufflebeam y Shinkfield (2005) como “época del realismo” y, por Guba y Lincoln (1982, 1989) la “tercera generación de la evaluación”, la del juicio. Se caracteriza fundamentalmente por reconocer la responsabilidad del personal docente en el logro de los objetivos educativos establecidos (Escudero, 2003).

Este mismo autor sitúa el interés por rendir cuentas en la preocupación creciente por evaluar los proyectos educativos americanos subvencionados con apoyo económico federal; de modo que anualmente debían ser evaluados con objeto de justificar las subvenciones para futuros momentos. Escudero (2003) añade que:

Junto al desencanto de la escuela pública, cabe señalar la recesión económica que caracteriza los finales años sesenta y sobre todo, la década de los setenta. Ello hizo que la población civil, como contribuyentes, y los propios legisladores se preocupasen por la eficacia y el rendimiento del dinero que se empleaba en la mejora del sistema escolar. (p. 17)

Se trata de un periodo en el que hay que empezar a rendir cuentas. Aunque el alumnado seguía siendo sujeto directo de la evaluación, Escudero (2003) apunta que:

Entran en juego, también, todos los demás elementos que confluyen en el proceso educativo (el programa educativo, en sentido amplio, el profesorado, los medios, los contenidos, las experiencias de aprendizaje, la organización, entre otros). Además del propio producto educativo. (p. 17)

Según Stobart (2010) se empieza a juzgar la eficacia de determinadas actividades que pueden ser muy generales, como los servicios médicos o, restringidas a una iniciativa específica como por ejemplo, la reducción del absentismo escolar. Es objeto de atención la organización y pone el caso de los hospitales como ejemplos; ya que al implicar el uso de recursos, quienes lo financian quieren conocer qué se ha conseguido con sus inversiones.

Frente a la ausencia de consecuencias en los programas, propia de la etapa anterior, este otro periodo se caracteriza por todo lo contrario: las cuentas rendidas por las instituciones, en el caso de la escuela, por el profesorado.

Stobart (2010) plantea que uno de los máximos ejemplos de la rendición de cuentas, muy presente en la actualidad con la proliferación de pruebas externas (PISA, PIRLS, entre otros), sea el plan de pago por resultados desarrollado por Robert Lowe en 1862 a través de una ley parlamentaria. Stobart (2010) desarrolla en qué consistió dicho plan, que surge en esta época de preocupación por las inversiones que se hacen en el ámbito educativo. Según él, Lowe (1862) introdujo un sistema de ayudas para la escuelas elementales, con la peculiaridad de que la mayor parte del dinero se distribuía de acuerdo con los resultados de los niños en los exámenes de lectura, escritura y aritmética (Stobart, 2010) (Es lo que se conoció como el plan de enseñanza de las tres erres, las “*three RS*”, las disciplinas básicas: *reading*, *writing* y *arithmetic*). Siendo dirigidos los exámenes por la inspección del Gobierno.

En esta época denominada la generación del juicio y valoración, se crea un contexto en el que surge, a raíz de estas nuevas necesidades de la evaluación, según Escudero (2003), un periodo de reflexión y de ensayos teóricos, unido a la gran expansión de la evaluación de programas que pretende clarificar la multidimensionalidad del proceso evaluativo, enriqueciendo decisivamente el ámbito conceptual y metodológico de la evaluación.

Tal periodo de reflexión está caracterizado especialmente por la aparición de nuevas voces y autores, entre los cuales destacaron Cronbach (1963) y Scriven (1967), cuyas nuevas ideas critican algunos aspectos del pensamiento tyleriano para ir más allá de éste, tratando de superar lo que parecían limitaciones del modelo de evaluación. Nos encontramos ante la generación del juicio, en la que son fundamentalmente las aportaciones de Cronbach (1963) y Scriven (1967) ya que según Mateo (1986 citado en Escudero, 2003, p. 19): “influyeron decisivamente en la comunidad de evaluadores, incidiendo [...] en la evaluación orientada al sujeto, en la línea de evaluación como *assessment*”.

Esta tercera generación se caracteriza según Guba y Lincoln (citado en Escudero, 2003, p.19) por “introducir la valoración, el juicio, como un contenido intrínseco en la evaluación. Ahora el evaluador no sólo analiza y describe la realidad, además, la valora, la juzga con relación a distintos criterios”.

A su vez, para Hernández y Guzmán (1991), Cronbach (1963) y Scriven (1967), son considerados los padres de la evaluación curricular moderna y resaltan el hecho de que Cronbach (1963) estuviera en desacuerdo con los planteamientos de entender a la tarea evaluativa como una mera aplicación de test e instrumentos estandarizados, como una "carrera de caballos" para saber cuál programa triunfaba. Al contrario, Cronbach (1963) reconoce la complejidad de la evaluación y asume la imposibilidad de que ésta pueda ser abordada mediante procedimientos simples o pueda ser entendida únicamente como aplicación de instrumento.

A Cronbach (1967) y Scriven (1963) debemos algunos de los principios que hoy se defienden en lo que respecta a la evaluación educativa. Siendo de los primeros en asociar la evaluación a la toma de decisiones. Hasta entonces los tests habían sido las herramienta por excelencia del proceso evaluador, es Cronbach (1967) quien introduce el uso de cuestionarios, entrevistas, observación sistemática y no sistemática, como técnicas de evaluación. A su vez, a Scriven (1963) le debemos términos tan al uso como: evaluación formativa y evaluación sumativa, o evaluación intrínseca y evaluación extrínseca. En definitiva, se empieza a ampliar el marco de la evaluación educativa, que tras esta generación, sufrirá una apertura mucho mayor, produciéndose lo que se conoce como la “eclosión de los modelos de evaluación”.

5. Las décadas de los setenta y los ochenta

La década de los setenta produce una proliferación de modelos de evaluación que inundan como plantean Escudero (2003) el mercado bibliográfico. Es lo que se conoce, siguiendo la clasificación de Guba y Lincoln (1982, 1989) como “la cuarta generación: la sensible”; y según Stufflebeam y Shinkfield (2005), “la época de la profesionalización”, que la sitúan desde 1973 hasta nuestros días.

Se trata, por tanto, de una época caracterizada por la pluralidad conceptual y metodológica. Guba y Lincoln (1982) nos hablan de más de cuarenta modelos propuestos en estos años, y Mateo (1986) se refiere a la “eclosión de modelos”.

Escudero (2003) apunta una división de periodos durante este época de auge: la primera mitad de la década, en la que las propuestas aún tienen un marcado carácter tyleriano y la segunda mitad de los setenta, con la aparición de los modelos alternativos de evaluación, entre los que destacan: la evaluación responsable de Stake (1975, 1976), con el apoyo de Guba y Lincoln (1982), la evaluación democrática de MacDonald (1976), la

evaluación iluminativa de Parlett y Hamilton (1977) y la evaluación como crítica artística de Eisner (1985).

Esta “eclosión de modelos” según Guba y Lincoln (1982) dan lugar a dos grandes grupos, los cuantitativos y los cualitativos, además de enriquecer considerablemente el vocabulario evaluativo. Se caracteriza por distanciarse ampliamente de las anteriores generaciones, haciendo una nueva apuesta por la evaluación.

Aparecen una serie de propuestas alternativas en lo que respecta a la evaluación de programas que pretenden alejarse del tradicional modelo positivista, tratando de incorporar los principios de un nuevo modelo, el naturalista. Tales principios y tales modelos repercutirán de un modo relevante en el desarrollo de la evaluación de los aprendizajes que, se verá directamente influenciada por los principios de una nueva forma de entender el proceso de E-A como era el constructivismo.

Surge pues, la generación “sensible y constructivista”, cuya esencia queda bien recogida en las siguientes aportaciones:

En las generaciones anteriores, los parámetros han sido contruidos a priori. En la llamada cuarta generación, los límites y parámetros los construyen las personas que participan en la evaluación como elemento importante dentro del mismo proceso de evaluación. (Dobles, citado en Mora, 2004, p. 8)

Según Vélez (2007) y García García, 2003):

En esta generación se toma como punto de partida las preocupaciones, cuestiones y opiniones de los diferentes actores que pueden verse afectados por la evaluación, reconociéndose la pluralidad de valores e intereses coexistentes en la sociedad y su relevancia en la determinación de los propósitos de la evaluación, donde los parámetros de referencia para la emisión de juicios de valor deben ser determinados a través de un procedimiento de diálogo con todos los implicados en un programa o política. (p. 154)

La cuarta generación implica una apuesta decidida por la construcción de los procesos evaluativos desde los contextos, lo que supone una forma particular de enfocar las evaluaciones desde planteamientos interpretativos y sociocríticos. Destacan en esta generación varios aspectos: La aceptación de procesos de negociación en los contextos, la atención a las demandas de los evaluados, la participación de esto en el proceso evaluativo y en la toma de decisiones, donde la evaluación es una herramienta que lo facilita (empoderamiento). (p. 35)

Según Escudero (2006) y Muñoz (2007), la generación sensible y constructivista se apoya en dos elementos fundamentales: La metodología constructivista y la evaluación respondiente de R. Stake (1976). Además son especialmente destacadas las siguientes:

- ✓ La “evaluación democrática” de B. MacDonald (1976), cuya postura, se refleja fundamentalmente, según Gimeno y Pérez Gómez (1989) en los conceptos de “autoevaluación” de Elliot (1993) y del “profesor como investigador” de Stenhouse (1984).
- ✓ La “evaluación iluminativa” de Parlett y Hamilton (1977), quienes plantean el concepto de “ambiente entorno de aprendizaje” como esencial para entender la dependencia entre enseñanza-aprendizaje. Añadiendo que “el desarrollo intelectual de los estudiantes no se puede entender aisladamente, sino sólo dentro de su entorno escolar particular” (Parlett y Hamilton, 1977, citado en Gimeno y Pérez Gómez, 1989, p. 457).
- ✓ La “evaluación como crítica artística” de Eisner (1985), que conceptualiza la enseñanza más como un proceso artístico que tecnológico” (Gimeno y Pérez Gómez, 1989, p. 440);

- ✓ El discurso de House (1994) sobre la enorme importancia social de la evaluación, dándole protagonismos a los términos de moral, ética, justicia y poder; recogido también en el pensamiento de Kushner (2002) que reconoce a la evaluación como una forma de acción política. Dicho pensamiento se vertebra por una preocupación central, la justicia social (Kushner, 2002).

Estos autores protagonizaron el escenario de la evaluación a partir de la década de los setenta. Son, de acuerdo con Rodríguez (2003), los principales precursores, impulsores y defensores de un paradigma alternativo en cuanto a la evaluación en general y por tanto, en lo que respecta a la evaluación escolar. Sus aportaciones dieron lugar a diferentes modelos o enfoques de evaluación, que pueden consultarse en Pérez Gómez (1989) y en House (1994).

Rodríguez (2003) reconoce ciertas distancias entre esas aportaciones pero asume que, todas tienen en común la adopción de fundamentos epistemológicos y metodológicos naturalistas para conocer, analizar y explicar los fenómenos y los procesos conforme a su naturaleza social, política y ética, cambiando así, el sentido de la evaluación, que ahora se entiende como (Rodríguez, 2003):

La valoración informada de los procesos pedagógicos, principal referente de la calidad de la educación [...] Como una estrategia para la comprensión de la enseñanza y el aprendizaje [...] como una manera de conocer los procesos, para entender mejor los intercambios pedagógicos, con la intención de revisarlos, cambiarlos y, en definitiva, mejorarlos. (pp. 392-393)

6. El nacimiento de un nuevo paradigma

El paradigma naturalista, denominado así en un primer momento por Guba y Lincoln (1989) según Vélez (2007), es para estos autores más que una reacción al paradigma positivista, “no es una postura reactiva sino preactiva, que parte de sus propias asunciones y no sólo, de carácter metodológico, sino lo que es más importante, asunciones diferentes epistemológicas y ontológicas” (Guba y Lincoln citado por Vélez, 2007, p. 155).

Un cambio ontológico, esto es, una forma diferente de entender la realidad. Se enfrenta a la postura tradicional, que considera que existe una única realidad, objetiva, que opera de acuerdo a leyes predeterminadas de causa-efecto. Al contrario, el paradigma naturalista entiende que existen múltiples realidades no gobernadas por leyes causales predeterminadas, con lo que para aprehenderla es necesario conocer las diferentes visiones que de ella se tiene. El relativismo según Guba y Lincoln (2002, p. 129) “supone relaciones sociales múltiples, comprensibles y en ocasiones opuestas, que son producto del intelecto humano, pero que pueden cambiar al volverse sus constructores más informados y sofisticados”.

Un cambio epistemológico, esto es según Guba y Lincoln (1989, citado en Vélez, 2007) una forma diferente de relación entre el conocedor y el conocimiento, entre el evaluador y la realidad. Carr y Kemmis (1998, p. 77) admiten el término positivismo para designar “un estilo de pensamiento informado por determinados supuestos acerca de la naturaleza del conocimiento”. Los autores añaden que:

El más importante de ello es ‘la regla del fenomenalismo’; el postulado de que el conocimiento válido solo puede establecerse por referencia a lo que se ha manifestado a través de la experiencia. Es decir, que se postula que la etiqueta de ‘conocimiento’ solo puede adscribirse a lo fundamentado en la ‘realidad’ tal como la aprehenden nuestros sentidos.

Una de las consecuencias fundamentales del fenomenalismo es la creencia de que los juicios de valor, como no pueden basarse en el conocimiento empírico, no pueden acceder a la categoría de conocimiento válido. (p. 77)

La relación que se establece desde esta óptica entre el agente evaluador y la realidad, se fundamenta en que las interferencias entre ambos han de ser mínimas. De ahí, que no es hasta la tercera generación de la evaluación cuando se introduce como elemento de ésta, el juicio del evaluador.

Y un cambio metodológico, frente a la metodología experimental de la que ya hemos hablado, se apuesta por el uso de técnicas hermenéuticas y el intercambio dialéctico cuyo objetivo es la reconstrucción de las construcciones previas (Guba y Lincoln, 2002).

Sintetizando, se puede decir que el paradigma naturalista se basa en la indagación fenomenológica, usando aproximaciones cualitativas para entender de manera inductiva y holística el comportamiento humano (la realidad) en un contexto específico, mientras que el positivista usa enfoques cuantitativos y experimentales para contrastar generalizaciones utilizando para ello el método hipotético-deductivo (Patton citado por Vélez, 2007, p. 52).

De acuerdo con Wrigley (2013, p. 74) “la evaluación escolar se ha enfocado abrumadoramente en exámenes y resultados de pruebas, los cuales se pueden representar numéricamente, mientras otros aspectos de la educación han sido marginados o simplemente invisibilizados”.

El nuevo modelo, pues, aparece como una brisa de aire fresco que abre nuevas posibilidades evaluativas, ya que en el desarrollo de la evolución de la evaluación que venimos haciendo, la llegada del nuevo paradigma considera relevante la necesidad de fomentar el intercambio de opiniones, valores y experiencias de las personas implicadas en la evaluación, a través de la utilización de métodos participativos (Vélez, 2007). Como ya se expuso, ésta sería la principal preocupación de la cuarta generación de la evaluación, según la clasificación de Guba y Lincoln (2002). Recuérdese también el nexo común de esta generación, según Rodríguez (2003) la adopción de fundamentos epistemológicos y metodológicos naturalistas para conocer, analizar y explicar los fenómenos y los procesos conforme a su naturaleza social, política y ética.

Se trata de una excelente ocasión para tratar de recuperar el sentido de la evaluación (Alcaraz, Fernández y Sola, 2012):

Acercándose más a la comprensión del proceso educativo que tiene lugar en las aulas y alejándose por completo de la intención medidora o sancionadora que a menudo se asocia a los procesos de evaluación, confundiéndose así el acto de calificar con el de evaluar. (p. 27)

Para concluir con las generaciones de la evaluación se muestra en la siguiente tabla una síntesis de las principales características de la evaluación en sus diferentes etapas. Para ello vamos a continuar con la clasificación por generaciones de Guba y Lincoln (1989).

Tabla 1. Principales características de la evaluación a lo largo del siglo XIX

1ª GENERACIÓN DE LA EVALUACIÓN (HASTA 1930)	2ª GENERACIÓN DE LA EVALUACIÓN (1930-1957)	3ª GENERACIÓN DE LA EVALUACIÓN (1957-1972)	4ª GENERACIÓN DE LA EVALUACIÓN (DESDE 1973)
Se asume el positivismo de las ciencias físico-naturales	Tyler acuña el término de “evaluación educativa”	La evaluación se asocia a la toma de decisiones	Se produce una “eclosión” de modelos de evaluación
Interés por la medición científica de las conductas humanas	Se va más allá de la evaluación psicológica	Interés por rendir cuentas (<i>accountability</i>). El alumnado seguía sujeto directo de la evaluación, pero además, también el profesorado, los medios, los contenidos, las experiencias de aprendizaje, la organización...	Entre esos modelos aparecen los cualitativos, impregnados por la lógica del constructivismo. El interés se basa en la comprensión y reconstrucción de los acontecimientos educativos
A través de la aplicación de instrumentos de forma técnica (tests de inteligencia y personalidad)	A partir de la propuesta de B. Bloom se define el currículum en términos de conductas (a través de objetivos muy precisos). Predomina el uso de tests.	Cronbach incluye los cuestionarios, las entrevistas, y la observación como técnicas de evaluación	Es necesario fomentar el intercambio de opiniones, valores y experiencias entre los participantes de un programa a través de la utilización de métodos participativos
Valoración de resultado de los programas educativos en estudiantes a través de la aplicación de tests, encuestas, acreditaciones y comparaciones experimentales	La evaluación se encarga de verificar la consecución de los objetivos propuestos, con objeto de valorar la eficacia de los programas. Aunque, la aportación de la evaluación a la mejora de la enseñanza es escasa	La evaluación sirve para rendir cuentas. Y éstas deben tener incidencia directa en los programas, los proyectos o las instituciones. Las decisiones deben mejorar la enseñanza	La evaluación es una herramienta que facilita el empoderamiento, la emancipación de los individuos
Medición es sinónimo de evaluación, anteponiéndose el término medición	El término evaluación se antepone al de medición. Se habla de evaluación y medición	Se incluye el término “juicio”. Se pretende valorar el mérito o valor de los programas. El evaluador es un juez, que emite juicios.	Deja de hablarse de “medición”, para avanzar en el concepto de evaluación educativa
“Evaluación basada en la norma”. Informa sobre el rendimiento del individuo en comparación con un grupo.	“Evaluación basada en criterios”. Indica el rendimiento de un individuo en relación con un estándar.	Predomina la “evaluación criterial”. Scriven acuña los términos “evaluación formativa” y “sumativa”, o “evaluación intrínseca” y “extrínseca”.	Aparecen modelos alternativos (“Evaluación respondiente” de Stake, “democrática” de McDonald, “iluminativa” de Parlett y Hamilton, la “evaluación como crítica artística” de Eisner, entre otros).

Fuente: Elaboración propia.

Ocurre que la vinculación de la evaluación al examen aún está muy presente. Como plantean Román y Murillo (2009, citado en Alcaraz, Fernández y Sola, 2009, p. 27): “más

que evaluar para aprender y comprender cómo y por qué aprenden lo que aprenden niños, jóvenes y adultos en su proceso educativo, la evaluación de los aprendizajes parece haberse conformado con medir cuanto rinden o logran los estudiantes”. De acuerdo con Álvarez Méndez (2001), se mezclan las funciones que se les asignan a la evaluación y se produce una gran confusión en torno al concepto, ya que muchas de esas funciones llegan a contradecirse:

Funciones relacionadas con la formación, selección, certificación, ejercicio de autoridad, mejora de la práctica docente, de motivación, orientación; funciones administrativas, académicas de promoción o recuperación, de información, de retroalimentación, de control. Tantas que, al tiempo que aumentan en proporción, pierden en transparencia. (p. 24)

Por tanto, a pesar de los intentos de diferenciación y separación de funciones realizados por Scriven (1967) al acuñar los términos de “evaluación formativa” y “evaluación sumativa” sigue existiendo una importante confusión. Stake (2006), insiste en clarificar ambas funciones y plantea con suma claridad la dicotomía entre ambas diciendo que cuando un cocinero prueba la sopa, está haciendo ‘evaluación formativa’, y cuando es el invitado quien la prueba, esa es una “evaluación sumativa”. Se trata de entender que en el primer caso, la evaluación se concibe como parte de un proceso de cambio que aporta información que contribuye a cambiar al evaluando, todavía en proceso de desarrollo. Mientras que en el segundo, ésta centra sus esfuerzos en medir los efectos del evaluando. Stake (2006) añade que aunque ambas podrían estar teniendo lugar al mismo tiempo, conviene mantener separadas las funciones de mirar hacia delante con actitud formativa y mirar hacia atrás con actitud sumativa. Hablamos de evaluación frente a calificación. La diferencia queda planteada haciéndose patente que una y otra pretenden cosas distintas; nos encontramos ante dos conceptualizaciones completamente opuestas en lo que respecta a la finalidad de ambas.

Sin embargo, aunque el esfuerzo por clarificar y diferenciar la evaluación de la calificación se hace ya viejo, aún hoy sigue existiendo confusión sobre los términos debido quizá a lo prolífero del campo y a la multitud de conceptos que se vienen utilizando para designar a ambas, perdiéndose en muchos casos el referente al que acuden estos; y siendo no poco frecuente, encontrarnos con la palabra evaluación seguida de los más variados apelativos (evaluación para la mejora, evaluación continua, evaluación para el aprendizaje...), que aunque parece que tratan de separar la acción de evaluar del acto calificador, son en nuestra opinión redundancias del término que en el peor de los casos acaban sumergiéndonos en un marasmo de sustantivos y adjetivos inútiles que añaden complejidad al asunto. De acuerdo con Rodríguez: “añadir apellidos al mismo concepto para distinguirlo de lo que verdaderamente ya de por sí no representa, resulta un pleonasma poco clarificador que no elimina la confusión” (2003, p. 405).

La consecuencia siguiente es la existencia de una bibliografía sobre el tema amplia y variada, pero confusa, pues hay que hacer un esfuerzo extra por tratar de diferenciar constantemente cuándo una nueva expresión se refiere a evaluación o a calificación.

Ocurre que mientras sobre el papel el concepto de evaluación ha sufrido numerosas transformaciones de apellidos, en la práctica aún no hemos sido capaces de diferenciar la ‘evaluación formativa’ de la “sumativa”, o lo que es lo mismo, la evaluación de la calificación. Así pues, el último informe Talis publicado por MECD (2014) expone que el porcentaje de docentes en los países de la OCDE de acuerdo con el uso de prácticas constructivistas de enseñanza-aprendizaje es superior al 74%, ascendiendo la cifra a un 87% en España. Dato que contrasta con el hecho de que, según este mismo informe el

porcentaje de profesorado que declara utilizar con frecuencia exámenes estandarizados como método de evaluación del alumnado es de 66% para la OCDE y 76% para España. Lo cual puede ser indicador de cómo pese a encontrarnos bajo el paraguas de un nuevo paradigma que hace hincapié en la función formativa de la evaluación, aún existe un predominio de prácticas evaluadoras en las aulas que ponen el acento en la función sumativa de la misma. En este mismo sentido se pronuncian Álvarez Méndez (2001) y Sverdlick (2012).

Sin embargo, no ocurre igual con la terminología empleada para aludir al concepto de evaluación. De modo que desde su nacimiento hasta hoy, la aparición de tantas nuevas voces y terminologías sobre evaluación han añadido dificultad al hecho de diferenciar lo esencial de dos conceptos absolutamente opuestos como son los términos sencillos de evaluación y calificación. De acuerdo con Popham (1980) existe una jerga bastante confusa ya que, son tantas las terminologías que surgen que se está perdiendo de vista lo esencial del término evaluación, que igual se acaba usando para “un roto que para un descosido”.

Al tiempo que se sofistican los conceptos, se mantienen las prácticas. Se produce una especie de enmascaramiento de prácticas “evaluadoras” tradicionales, ocultas tras novedosos e imaginativos apellidos de la evaluación. Encontrándonos ante la circunstancia de que, al tiempo que aumenta la terminología referida a la evaluación y se modernizan términos, las prácticas más propias de la calificación que de ésta, se mantienen. Es lo que podríamos llamar el nacimiento de “la quinta generación de la evaluación”, la “generación perdida o ecléctica”.

Tras las generaciones de la medición, la descripción, la del juicio y la sensible, recogidas por Guba y Lincoln (1982, 1989), cada una de las cuales definía una época y unas características concretas en torno a la manera de entender la “evaluación”, llegamos a la actualidad, donde hay tal variedad y mezcla de prácticas que, resulta difícil encontrar un elemento común que pueda caracterizar a la evaluación de nuestros tiempos, más allá del eclecticismo y la confusión. Se produce una mezcla de funciones y los esfuerzos se dirigen a hacerlas combinar, queremos que la evaluación suponga una comprensión y una mejora, pero a la vez queremos también que sirva para seleccionar y acreditar. Las primeras son funciones de la evaluación, las segunda, de la calificación. En muchos discursos –teóricos– no estamos dispuestos a renunciar a una parte de esas funciones, y si lo hacemos, defendemos las que tienen que ver con la evaluación; mientras que en la práctica no se renuncia bajo ningún concepto a las relativas a la calificación. Como mucho, los esfuerzos se dirigen a matizar la calificación vinculándola a la evaluación. Es lo que popularmente se conoce como “estar en el plato y en la tajada”.

8. Conclusión

Históricamente el interés por la evaluación se remonta a tiempos muy antiguos. Algunos autores lo sitúan en el 2000 a.C. Desde entonces hasta la actualidad, el concepto de evaluación ha ido modificándose a través de diferentes conceptualizaciones y características que han ido dando lugar a una serie de etapas. En el presente artículo se ha seguido la clasificación de Guba y Lincoln (1982, 1989) referida a las cuatro “generaciones de la evaluación”, las cuales han comprendido cuatro periodos de tiempos claves en la manera de conceptualizar la evaluación. Así hasta 1930 el término medición se antepuso al de evaluación que cobra un mayor protagonismo con la llegada de Tyler (1969), quien es reconocido como el padre de la evaluación educativa. En la etapa

tyleriana se invierte el orden y la evaluación es antepuesta a la medición. Sin embargo, no es hasta 1957 cuando la evaluación se empieza a utilizar con intenciones de mejora, asociándola a la toma de decisiones que favorezca la mejora de los programas y de la enseñanza, pero no directamente a los individuos, sujetos de aprendizaje. Siendo a partir de 1973 cuando aparecen propuestas alternativas al modelo tradicional basado en el positivismo y se empieza a entender la evaluación como herramienta de empoderamiento de los sujetos. Se incorporan nuevos principios que acaban repercutiendo en las formas de entender la evaluación de los aprendizajes en el ámbito escolar que ya ha comenzado a impregnarse de ideas constructivistas que alteran la antigua interpretación del modelo de enseñanza-aprendizaje. Deja de hablarse de medición y se avanza en la evaluación educativa. Desde entonces hasta la actualidad, la evaluación educativa se encuentra presente en cualquier debate sobre educación. Sus funciones, sus intenciones, incluso sus nombres proliferan y aumentan su complejidad hasta tal nivel que acabamos encontrándonos en un momento histórico donde a pesar de que el conocimiento sobre la evaluación es mayor que nunca, la confusión es grande. De modo que, muchos de los intentos que se hicieron en las anteriores décadas por distinguir medición, evaluación y calificación no obtienen recompensa. La literatura sobre las nomenclaturas que acompañan al término evaluación es tan extensa que provoca confusión en cuanto al verdadero significado de dicho término. Ocurre entonces que muchas de las prácticas que dicen llamarse evaluación, pretenden ser a su vez prácticas de calificación o de medición. De ahí que hayamos denominado a la época actual como la quinta generación de la evaluación: la evaluación ecléctica.

Referencias

- AA.VV. (2009). *Guía para la evaluación de competencias en el prácticum de los estudios de maestro/a*. Barcelona: Agència per a la Qualitat del Sistema Universitari de Catalunya.
- Alcaraz, N., Fernández, M. y Sola, M. (2012). La voz del alumnado en los procesos de evaluación docente universitaria. *Revista Iberoamericana de Evaluación Educativa*, 5(2), 27-39.
- Alvárez Méndez, J.M. (2001). *Evaluar para conocer, examinar para excluir*. Madrid: Morata.
- Carr, W. y Kemmis, S. (1988). *Teoría crítica de la enseñanza. La investigación acción en la formación del profesorado*. Barcelona: Martínez Roca.
- Cronbach, L.J. (1963). Course improvement through evaluation. *Teachers College Record*, 64, 672-683.
- Dobles, M.C. (1996). *Acreditación en educación superior: orígenes, prácticas actuales, perspectivas*. Comunicación personal.
- Durar, V. (2013). *Red Maestros de Maestros*. Recuperado de http://www.rmm.cl/index_sub.php
- Escudero, T. (2003). Desde los tests hasta la investigación evaluativa actual. Un siglo, el XX, de intenso desarrollo de la evaluación en educación. *RELIEVE*, 9(1), 11-43.
- Eisner, W.E. (1985). *The art of educational evaluation*. Londres: The Falmer Press.
- Elliot, J. (1993). *El cambio educativo desde la investigación-acción*. Madrid: Morata.
- Forrest, G. (1990). Oral examinations. En: H. Walberg y G. Haertel, (Eds.), *The international encyclopedia of educational evaluation* (pp. 42-86). Oxford: Pergamon Press.
- García García, M. (2003). *Investigación Evaluativa del Programa de Prácticas en Empresas de la Universidad de Cádiz*. Tesis doctoral no publicada, Universidad de Cádiz, España.

- Gimeno Sacristán, J. y Pérez Gómez, A.I. (1989). *Comprender y transformar la enseñanza*. Madrid: Morata.
- Gronlund, N.E. (1985). *Measurement and evaluation in teaching*. Nueva York: Macmillan.
- Guba, G.E. y Lincoln, Y.S. (1982). *Effective evaluation*. San Francisco, CA: Jossey Bass Publishers.
- Guba, G.E. y Lincoln, Y.S. (1989). *Fourth Generation Evaluation*. Nueva York: Sage Publications.
- Guba, G.E. y Lincoln, Y.S. (2002). Paradigmas en competencias en la investigación cualitativa” En C. Deman y J.A. Haro (Comps.), *Por los rincones. Antología de métodos cualitativos en la investigación social* (pp.113-145). El Colegio de Sonora: Hermosillo Sonora.
- Hernández Rojas, G. y Guzmán. J. (1991). *Evaluación curricular*. Recuperado de <http://es.scribd.com/doc/6571177/Historia-de-La-Evaluacion-Curricular>
- House, E.R. (1994). *Evaluación, ética y poder*. Madrid: Morata.
- Kushner, S. (2002). *Personalizar la evaluación*. Madrid: Morata.
- Lemus Alvarado, M. (2012). *La evaluación educativa tiene historia*. Recuperado de <http://es.scribd.com/doc/80501271/La-evaluacion-educativa-tiene-historia>
- Macdonald, B. (1976). Evaluation and the control of education. En D. Tawney (Ed.), *Curriculum evaluation today: trends and implications* (pp. 125-136). Londres: McMillan.
- Mateo, J. (1986). *Proyecto docente e investigador*. Barcelona: Universidad de Barcelona.
- MECD (2014). *Informe español Talis 2013: Estudio internacional de la enseñanza y el aprendizaje*. Madrid: INEE.
- Mora Vargas, A.I. (2004). La evaluación educativa: conceptos, periodos y modelos. *Revista electrónica actualidades investigativas en Educación*, 2, 1-28.
- Muñoz Cuenca, G.A. (2007). Un nuevo paradigma: la quinta generación de evaluación. *Laurus*, 23, 158-198.
- Nunnally, J.C. (1978). *Psychometric theory*. Nueva York: McGraw Hill.
- Parlett, M. y Hamilton, D. (1977). Evaluation as illumination: A new approach to the study of innovative programmes. En D. Hamilton et al. (Eds.), *Beyond the numbers game* (pp. 85-99). Londres: MacMillan.
- Planchard, E. (1960). *La investigación pedagógica*. Madrid: Ediciones Fas.
- Popham, W.J. (1980). *Problemas y técnicas de la evaluación educativa*. Madrid: Anaya.
- Popham, W.J. (1983). *Evaluación basada en criterios*. Madrid: Magisterio Español, S. A.
- Reyes Herrera, L. (2001). Evaluación de los aprendizajes en la educación superior. *Tecné, episteme y didaxis: revista de la facultad de Ciencia y Tecnología*, 9, 68-79.
- Rodríguez Sánchez, A. (2003). *El conocimiento escolar. La experiencia educativa del conocimiento en un aula a través de un proceso de investigación en la acción*. Tesis doctoral no publicada. Universidad de Málaga, España.
- Scriven, M. (1967). The methodology of evaluation. En M. Scriven (Ed.), *Perspectives of Curriculum Evaluation*, (pp. 39-83). Chicago: Rand McNally and Company.
- Stake, R.E. (1975). *Evaluating the arts in education: A responsive approach*. Columbus, OH: Me-riril.
- Stake, R.E. (1976). A theoretical stament of responsive evaluation. *Studies in Educational Evaluation*, 2, 19-22.
- Stake, R.E. (2006). *Evaluación comprensiva y evaluación basada en estándares*. Barcelona: Graó.

- Stenhouse, L. (1984). *Investigación y desarrollo del curriculum*. Madrid: Morata.
- Stobart, G. (2010). *Tiempos de pruebas: Los usos y abusos de la evaluación*. Madrid: Morata.
- Stufflebeam, D.L. y Shinkfield, A.J. (2005). *Evaluación sistemática. Guía teórica y práctica*. Barcelona: Paidós/MEC.
- Sverdlick, I. (2012). *¿Qué hay de nuevo en evaluación educativa?* Buenos Aires: Noveduc.
- Tyler, R.W. (Ed.). (1969). *Educational evaluation: New roles, new means*. Chicago, IL: University of Chicago Press.
- Vélez Méndez, C. (2007). El cambio de paradigma en evaluación de políticas públicas: el caso de la cooperación al desarrollo. *Nuevas Políticas Públicas: anuario multidisciplinar para la modernización de las Administraciones Públicas*, 3, 145-170.
- Wrigley, T. (2013). Repensando el cambio escolar y el papel de la evaluación: La experiencia de dos países anglófonos. *Revista Iberoamericana de Evaluación Educativa*, 6(2), 73-90.