ÀÁF ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING

**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


# USOS Y CRÍTICAS DE LOS TEST ESTANDARIZADOS EN LOS SISTEMAS EDUCATIVOS

**Juan Andrés Ligero Lasa**
Profesor Universidad Carlos III de Madrid
Co-director del Máster en Evaluación de Programas y Políticas de la Universidad Complutense

**RESUMEN**
**"Usos y críticas de los test estandarizados en los sistemas educativos"**
Actualmente tanto en el ámbito local como internacional existe una creciente importancia y desarrollo de los sistemas de evaluación educativa. La inclusión de las pruebas en las iniciativas legislativas educativas, el desarrollo de estrategias evaluativas en diferentes niveles de la administración educativa y las pruebas internacionales de evaluación como PISA son claros indicios de esta expansión.
Aunque en general el incremento de la evaluación educativa es algo positivo, su desarrollo se está restringiendo mayoritariamente a una determinada perspectiva metodológica, los sistemas de medida del rendimiento escolar a través de exámenes externos estandarizados o test estandarizados.
Utilizar un método u otro de evaluación no es indiferente. Cada método genera diferentes productos evaluativos que pueden ser más adecuados para determinados usos y, por tanto, también puede acarrear diferentes consecuencias en los objetos evaluados, es decir, en los centros o en los sistemas educativos.
La evaluación escolar tiene que ser un ejercicio consciente y conocedor de los propósitos que se pretenden para poder adecuar los métodos a su uso. El objetivo de este artículo es contextualizar los test estandarizados y analizar los límites y las críticas de este tipo de método utilizando como estudio de caso el español. El texto se ha estructurado en diferentes epígrafes: contextualización histórica, definición y características principales de los test, principales críticas y conclusión.

**Palabras clave**: Evaluación. Educación. Currículo.

**ABSTRACT**
**"Usages and Reviews of Standardized Tests in Educational Systems"**
Currently, both locally and internationally, there is a growing importance and development of systems of educational evaluation. The inclusion of tests in education legislative initiatives, the development of evaluative strategies at different levels of the educational administration and international evaluation tests such as PISA are clear indicators of this expansion.
Although in general, the increase of this educational evaluation is somewhat positive, its development is being restricted mainly to a certain methodological perspective: systems of measurement of school performance through standard external examinations or standardized tests.
Using one method of evaluation or another is not indifferent. Each method produces different evaluation products that may be more suitable for certain applications and therefore can also lead to different consequences in the evaluated objects, i.e., workplaces or in the educational systems.
School evaluation must be a conscious exercise and aware of the purposes that are intended, to be able to adapt the methods to its use. This article aims to contextualize the standardized tests and analyze the limits and the reviews of this kind of method studying the Spanish case. The text is structured in different sections: historical contextualization, definition and main features of the tests, major reviews and conclusion.

**Key words:** Evaluation. Education. Curriculum.

## CONTEXTUALIZACIÓN HISTÓRICA

El ámbito educativo ha sido uno de los sectores más prolíficos en el desarrollo metodológico de evaluación; gran parte de los teóricos**[1]** han elaborado su

ÀÁF ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


pensamiento sobre objetos educativos, por lo que casi hablar de historia metodológica de evaluación educativa es hablar de la historia de la evaluación.

Aunque se pueden encontrar referencias evaluativas aisladas en siglos anteriores, es a finales del siglo XIX donde se aprecia un volumen de evaluaciones suficientes que permiten identificar un primer punto de partida en la disciplina. Probablemente se explique por las grandes transformaciones que las revoluciones industriales provocaron, lo que Madaus y Stufflebeam (2000) han denominado la edad de las reformas.

Durante los primeros años del siglo XX, la educación como otros ámbitos de la intervención política, no se libran de la metáfora fabril. El alumnado es visto como "materia prima" que debe ser procesada en el colegio a modo de "planta de tratamiento". El deber de la dirección es "hacer su trabajo tan efectivo y eficiente como sea posible" (Guba y Lincoln, 1989:26). Surgen estrategias y técnicas acordes con esta visión con la intención de incrementar la racionalización y la eficiencia de los centros educativos.

En 1904 el Ministerio de Educación francés encarga a Alfred Binet un sistema para identificar y poder descartar a "los jóvenes mentalmente retardados" que puedan dificultar el desarrollo del resto del grupo (Guba y Lincoln, 1989:23), conocido posteriormente como test de inteligencia (Monnier, 1992). En el mismo año la Asociación Nacional de Educación de EE.UU designó un comité para estudiar el uso de los test en la clasificación de estudiantes y en la determinación de sus progresos. En 1908 se publica el test de razonamiento en aritmética y en 1922 Stanford desarrolla la batería de test que permiten la valoración simultánea de los estudiantes en múltiples asignaturas (Guba y Lincoln, 1989).

Durante la primera mitad del siglo XX se consolida el uso de los test dejando de ser experiencias singulares. En 1933 Gertrude Hildreths publicó una bibliografía donde se encontraban 3.500 test mentales y escalas de puntuación; en 1945 actualizó el trabajo recogiendo 5.200 instrumentos. En 1958 el Acta Nacional de defensa de la Educación en EE.UU declaraba la necesidad de evaluar los planes educativos. Este hecho unido al desarrollo de las máquinas de lectura rápida de los test (Madaus, 2004:77) generó un movimiento conocido como el boom de los test estandarizados.

Aunque se estuviera ante un momento expansivo, desde una perspectiva metodológica la escena quedaba reducida a los test estandarizados y en menor medida a la propuesta de Ralph Tyler de evaluación por objetivos**[2]** (Millbrey y otros, 1991).

El desarrollo de un mercado de evaluación real con demandas en diferentes situaciones, con múltiples actores con intereses y propósitos distintos, ponían a prueba y tentaban los límites de los métodos existentes. A finales de los años 60' aparecen las principales críticas a los métodos predominantes:

- En los test, el foco exclusivo en los resultados del alumnado no permiten entender por qué ocurren las cosas (Tyler, 1991; Cronbach, 2000).

- Tampoco provén de una información fundamentada para tomar decisiones sólidas que sirvan para la mejora de la docencia (Cronbach, 2000).

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


- Tanto los test como la evaluación por objetivos aportan solo una descripción o una medida sin proporcionar los mecanismos para enjuiciar suficientemente los programas (Guba y Lincoln, 1989).

- No tienen en cuenta las necesidades y los valores de los diferentes actores implicados en una intervención (Stake, 2004).

Estas críticas fueron el acicate para el surgimiento del periodo metodológicamente más creativo, finales de los 60' principios de los 70'. "La evaluación se movió de concepciones monolíticas a más pluralistas, múltiples métodos, medidas, criterios, perspectivas, audiencias e incluso intereses" (House, 1990:24). Surgieron diferentes aproximaciones que proponían otras miradas, sensibilidades, y en todos los casos generaron nuevas líneas de estudio o pensamiento en evaluación (Schwandt, 2009).

"Scriven (1967), Stufflebeam (1967 y 1971, con otros) y Stake (1967) introdujeron nuevos modelos de evaluación que se separaban radicalmente de las aproximaciones anteriores. Estas conceptualizaciones reconocían la necesidad de evaluar los objetivos, mirar los recursos, examinar la implementación y la forma de ofrecer los servicios, así como medir los resultados previstos y no previstos del programa. Ellos también enfatizaban la necesidad de enjuiciar el mérito o el valor del objeto que estaba siendo evaluado" (Madaus y Stufflebeam, 2000:14).

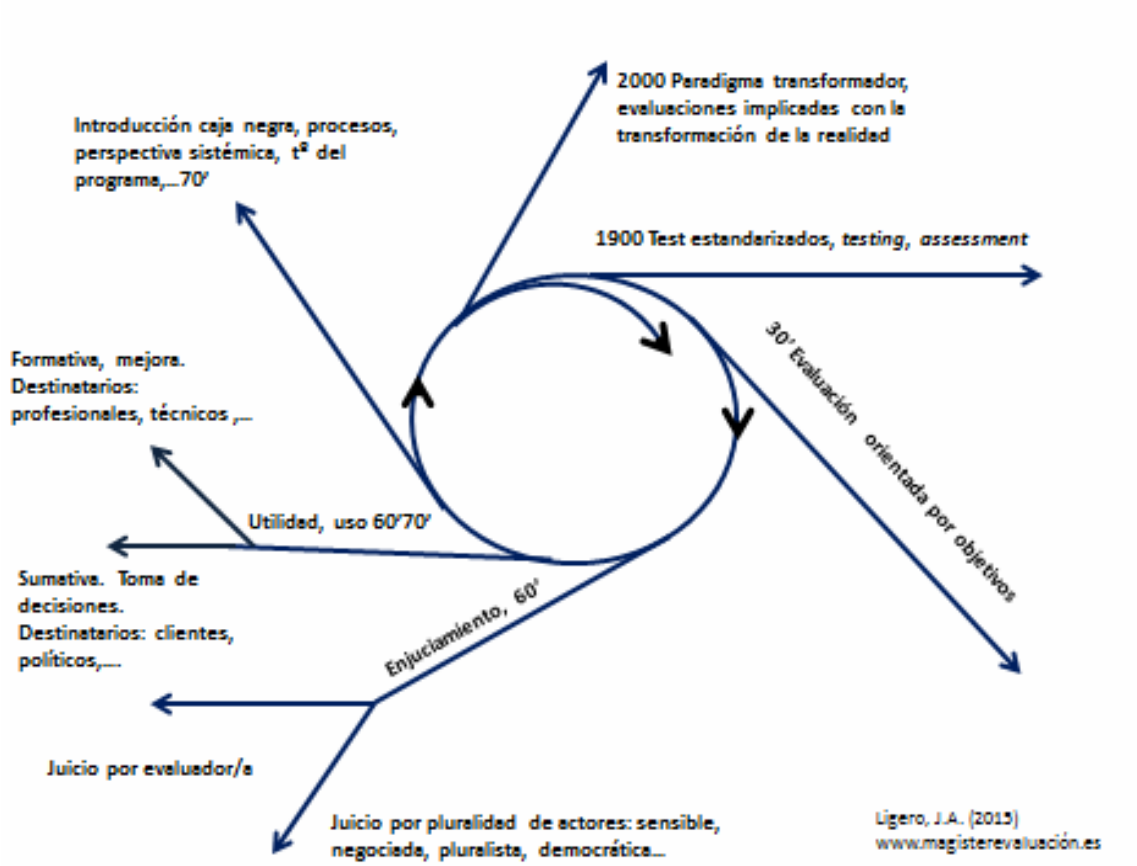
En la misma época el paradigma científico tradicional fue puesto en cuestión por el constructivismo que negaba la existencia de la objetividad tal como se entendía hasta la fecha. La realidad "son construcciones sociales mentales por lo que puede haber tantas construcciones como individuos haya (aunque muchas puede ser compartidas entre los diferentes sujetos)" (Guba y Lincoln,1989: 43). Por consiguiente, tal como lo describe Scriven, la objetividad, si existe, se puede definir como un acuerdo intersubjetivo entre los diferentes actores implicados (Chen, 1990). "La evaluación cambia de un énfasis prioritario en los métodos cuantitativos, (…) a una actitud en la cual los métodos cualitativos se convierten en aceptables" (House, 1990:25).

Este movimiento para algunos autores (House,1990; Schwandt,2009) supuso el nacimiento real de la evaluación de programas o su edad de desarrollo (1958-1972) (Madaus y Stufflebeam, 2000), lo que en cualquier caso viene a señalar la diversidad, la innovación y la profusión de métodos en este periodo.

En EE.UU la llegada de Reagan al gobierno (1981-1990) supuso un recorte de la evaluación profesional, al menos desde un punto de vista externo (House, 1990). En cambio, la nueva reforma educativa que se impulsó durante esos años trajo un renovado énfasis en los test estandarizados, no con propósitos formativos**[3]** sino más bien enfocada a ofrecer un rendimiento de cuentas público.

En síntesis, la narración histórica de las metodologías no se puede entender como un proceso evolutivo donde los nuevos métodos descartan a los antiguos, encaja más con la idea de un espacio escénico dónde el eje cronológico muestra la entrada en la escena de las aproximaciones para luego ya permanecer en ella. En el gráfico siguiente se representa una síntesis de las principales aproximaciones y su secuencia de apariciones.

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


(**Gráfico 1.** Secuencia histórica de aparición de las diferentes aproximaciones evaluativas.)

Como muestra el dibujo, los test estandarizados han sido de las primeras propuestas metodológicas (faltaría por incluir los sistemas de inspección) sobre la que se han ido suplementado otras aproximaciones, en parte para responder a los déficits y limitaciones que mostraban. La evaluación por objetivos enfatiza la singularidad que no se recogía de los centros. Las propuestas de finales de los 60` incorporan la importancia de emitir juicios de valor sobre el objeto y no solamente medir los resultados. Se toma conciencia de la existencia de diferentes actores con diferentes valores que desbordan una mirada única o estandarizada (Bustelo, 2011). Este pluralismo de actores también deja ver los diferentes propósitos a los que la evaluación puede servir y que un único método no tiene por qué responder a todos. También en esta época los evaluadores empiezan a mirar en el interior de la caja negra y a descubrir relaciones causales y por qué se producen los resultados. Además, las perspectivas de indagación se amplían con las técnicas cualitativas a la vez que se les exige a todas rigor en su aplicación.

Las aproximaciones surgidas en los diferentes momentos pueden mantener desarrollos propios e independientes de la corriente principal, como así lo tratan de representar las flechas tangentes, incluso hasta generar disciplinas diferentes como ha sido el caso de los test y el *assessment***[4]** tal como se explica más adelante.

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


Por otro lado, actualmente se pueden encontrar en funcionamiento cualquiera de las aproximaciones evaluativas referidas. Por ejemplo, en un mismo centro escolar puede hallarse un modelo CIPP (Contexto, *Input*, Proceso y Producto)**[5]** junto con test estandarizados y ambos gozar de aceptación. Tal es la profusión de métodos (22 aproximaciones recoge Stufflebeam, 2001) que pareciera que escoger entre uno y otro es una mera cuestión de gustos. Es función de la metodología definir y reconocer sus virtudes y limitaciones para poder adecuarlos a su uso.

Este es el objetivo del presente artículo, dado el gran desarrollo que están viviendo actualmente los test estandarizados es necesario conocer sus contornos y limitaciones para saber si el método está ajustado a los propósitos pretendidos.

## DEFINICIÓN Y CARACTERÍSTICAS DE LOS TST ESTANDARIZADOS

Los test o los exámenes externos estandarizados son diferentes formas de denominar el mismo método, que consiste en un cuestionario con preguntas y respuestas cerradas o con un alto grado de estructuración. Los cuestionarios pueden estar dirigidos a cualquier población, pero en el ámbito educativo es frecuente encontrarlos aplicados a estudiantes para medir su desempeño en conocimientos, competencias o desarrollos no cognitivos**[6]**. Se aplican homogéneamente a todos los sujetos seleccionados y el tratamiento de la información está estandarizado, es decir, es igual en todos los casos.

A los test se les atribuye ciertas ventajas frente a otros métodos:

- Son pruebas preparadas previamente, normalmente por una institución educativa o gubernamental, por lo que su aplicación requiere de menos elaboración y de una reducción de la incertidumbre sobre los aspectos que deben ser medidos.

- Su aplicación procedimentada evita sesgos intencionales.

- La estandarización de las medidas permite una comparación entre instancias educativas (aulas, colegios y territorios). Se pueden elaborar *rankings* de mayor a menor puntación entre las diferentes unidades educativas.

- Estas mismas clasificaciones puede estimular el aprendizaje organizativo identificando aquellas estrategias pedagógicas que han obtenido mejores resultados.

- Pueden ofrecer a los diferentes actores del sistema educativo información sobre el desempeño, fomentando la rendición de cuentas social sobre el sistema educativo.

- Contribuyen a la responsabilizacion de los centros educativos con sus resultados, fomentando un modelo de gestión gerencial (Bhen cit. Ryan y Cousins, 2009) y se espera que esto favorezca la generación de una cultura de la calidad y la mejora.

En la actualidad los test tienen un alto grado de aceptación, están implantándose cada vez en un mayor número de países y territorios. Por ejemplo, en el año 2012 PISA (*Program International Students Assesment*) se aplicó en 65 países de los 5 continentes encuestando a 510.000 estudiantes que representan a una población

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


de aproximadamente 28 millones de jóvenes de 15 años (PISA, 2014). En Europa prácticamente todos los países de la Unión Europea se han sumado a este programa.

Además de PISA existen otras pruebas dirigidas a estudiantes como TIMSS (*Trends in International Mathematics and Science Study*), PIRLS (*Progress in International Reading Literacy Study*) o EECL (Estudio Europeo de Competencias Lingüística). También hay pruebas para docentes y personal en cargos de dirección como el estudio TALIS (*Teaching and Learning Internacional Survey*) e incluso se pueden encontrar pruebas estandarizadas para la población adulta como PIAAC (Programa para la Evaluación Internacional de las Competencias de los Adultos) (INEE, 2015).

La profusión de test estandarizados no responde solamente a un movimiento meramente social. En una escena internacional globalizada (Rizvi, 2009) existe una política articulada para la promoción este tipo de sistemas de medida escolar. Existen organismos internacionales que trabajan en el fomento de este visión, como por ejemplo la Organización para la Cooperación y el Desarrollo Económica (OCDE), la *Association for the Evaluation of Educational Achievement* (IEA) o la iniciativa *World Education Indicator Program* promovido por el Banco Mundial y la UNESCO.

En el ámbito nacional cada vez son más los países que implementan sistemas propios de test para observar la calidad educativa (Ryan y Cousins, 2009). Esto también implica un creciente papel de institutos y oficinas gestoras de estas pruebas como la británica *Office for Standars in Education Children´s Services* and Skill (Ofsted), la *National Assessment of Educational Progress* (NAEP) en EE.UU. o el Instituto Nacional de Evaluación Educativa (INEE) de España.

Tomando España como estudio de caso, es significativo ver que una de las principales novedades introducidas por la nueva ley educativa LOMCE (8/2013, 9 de diciembre) es la incorporación en todo el sistema educativo de las pruebas externas estandarizadas.

"Artículo 144. Evaluaciones individualizadas. (…) En concreto, las pruebas y los procedimientos de las evaluaciones indicadas en los artículos 29 y 36 bis se diseñarán por el Ministerio de Educación, Cultura y Deporte, a través del Instituto Nacional de Evaluación Educativa. Dichas pruebas serán estandarizadas y se diseñarán de modo que permitan establecer valoraciones precisas y comparaciones equitativas, así como el seguimiento de la evolución a lo largo del tiempo de los resultados obtenidos".

Por otro lado, ha habido un gran desarrollo de los test en las Comunidades Autónomas. Por citar algunos ejemplos, en Cataluña se aplican pruebas externas en 6º de primaria y en 4º de la ESO. En Andalucía, la prueba Escala se pasa a 2º de primaria. En el País Vasco y la Comunidad Valenciana la Evaluación Diagnóstica se aplica a 4º de primaria y 2º de ESO. En ninguno de los cuatro ejemplos se hacen comparaciones entre instancias educativas, la información se devuelve a los centros con la intención de generar una inercia formativa interna. Un caso diferente es el de Madrid donde la prueba CDI (Conocimientos y Destrezas Indispensables) se pasa en 6º de primaria y permite con posterioridad comparar colegios entre sí y establecer clasificaciones.

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


Parte de la popularidad de estas pruebas es debido a que las evaluaciones y sus resultados son usados por distintos actores:

- Se pueden usar por padres y madres para conocer el desempeño escolar del centro y para decidir dónde matricular a sus hijos. Por ejemplo, en la Comunidad de Madrid los resultados de las pruebas CDI son ampliamente consultados a través de la página web en los periodos de matriculación escolar.

- Se usan por la dirección y profesorado para identificar los resultados obtenidos de su centro. A este ejercicio también se le presume la virtud de estimular la mejora de calidad, "porque la supervisión más cercana por parte de padres y administradores proporciona una motivación adicional para maestros y directores para mejorar los resultados escolares de sus estudiantes" (Brindusa et al., 2012).

- Se puede usar por toda la cadena de decisores políticos para tomar decisiones de carácter sumativo del tipo de continuar o no continuar con una determinada política educativa o para actuar sobre algunos centros con puntuaciones fuera de lo esperado.

- Además, se usan socialmente y por los medios de comunicación. La información sobre las puntuaciones del sistema educativo son noticia. Baste con observar el movimiento mediático que se genera cuando, por ejemplo, se publican los resultados PISA.

En definitiva, los test estandarizados han conformado un sistema propio de valoración escolar, ampliamente aplicado, respaldado local e internacionalmente, lo que a su vez refuerza su hegemonía metodológica. No obstante, existe un claro conjunto de críticas que identifican con claridad las limitaciones, los problemas y los juicios falsos que pueden ser provocados por los test.

## CRÍTICAS Y LIMITACIONES DE LOS TEST ESTANDARIZADOS

En este artículo he destacado las críticas referidas a los aspectos más instrumentales de los test estandarizados, dejando aparte otras consideraciones de carácter político o social atribuibles a los test y a su función de *accountability***[7]**. Muchas de las críticas ya se han apuntado en la contextualización histórica, entre otras razones porque en gran medida el avance metodológico en evaluación ha venido dado para intentar subsanar algunas de las limitaciones que el *testing* presentaba. De forma sintética las críticas a los test se pueden resumir en cinco puntos:

1. Se evalúan resultados estandarizados que pueden coincidir o no con los objetivos y con el trabajo real del centro educativo.

2. Al tratarse de test a gran escala, suelen tender a medir conocimientos y no objetivos educativos finales o competenciales.

3. Los resultados no son atribuibles a los centros ya que existen otras variables que también influyen en los resultados que no están controladas.

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


4. Tienden a evaluar resultados y no la intervención, es decir, las actividades educativas que provocan dichos resultados.

5. Se trata de exámenes a estudiantes no de evaluaciones de centros o de políticas educativas.

## 1. Se evalúan resultados estandarizados frente a objetivos educativos del centro

Cada colegio está en un entorno que puede tener un determinado tipo de alumnado con su propio bagaje, con sus problemas, con sus fortalezas específicas, con diferentes recursos y apoyos comunitarios. Cada centro o unidad educativa tratará de adaptarse a su realidad, utilizar sus recursos y herramientas pedagógicas para conseguir unos determinados logros.

Esta es la esencia en la que se apoya la lógica de los proyectos educativos de centro. Los objetivos educativos son las metas fijadas en los planes o proyectos para lograr cambios en los estudiantes, adaptándose al entorno para resolver problemas o reforzar potencialidades específicas de la población con la que les toca trabajar. Por lo tanto, los objetivos de un centro en un determinado contexto pueden diferenciarse mucho de los de otro centro que esté en otra situación. Esto no solo no es algo negativo sino que se entiende como un indicio de la obligada adaptación de la instancia educativa a su entorno.

Pero esta diversidad pedagógica, tanto ahora como hace 100 años, despierta el temor a que los diferentes objetivos y los currículos específicos sean una forma desordenada y poco rigurosa de preparar a los jóvenes para su paso a secundaria o a la universidad. La investigación *Eight-Year Study* (Tyler, 1991) concluyó que los estudiantes procedentes de los institutos con currículo abierto**[8]** sacaban incluso mejores resultados que el grupo de comparación procedente de enseñanza reglada y uniforme (Ritchie, 1971).

Indistintamente de los planes educativos de los centros y por tanto de los objetivos educativos que se hayan establecido, los test estandarizados miden siempre los mismos resultados. Trasladado al ámbito del aula, es optar por que el profesorado no desarrolle sus propias pruebas para evaluar el desempeño de los estudiantes en su materia, sino que los exámenes ya vengan dados por una instancia superior.

La evaluación por objetivos nació para responder a la crítica de colegios y profesionales de que los test no miden lo que ellos están tratando de enseñar a los estudiantes y, por tanto, no son un instrumento adecuado para evaluar su trabajo (Madaus, 2004:74). Además, según Tyler, los test se apoyan en una gran desconfianza a los juicios de las docentes en la medida que los exámenes realizados por el profesorado no se consideran rigurosos o confiables para valorar el logro del alumnado.

En definitiva, los test estandarizados pueden dejar sin medir aspectos que realmente trabaja una instancia educativa o sobredimensionar otros, imponiendo implícitamente los objetivos que deben ser trabajados en todas las aulas indistintamente del contexto, diversidad individual o apuesta pedagógica. Tal como lo expresa la LOMCE, "estas pruebas normalizan los estándares de titulación en toda España, indicando de forma clara al conjunto de la comunidad educativa cuáles son los niveles de exigencia requeridos" (…) (Preámbulo, VIII).

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


Bajo esta perspectiva, los test son un instrumento para la homogenización de los objetivos de los centros. Lo que provoca un conflicto entre la lógica de trabajo a través de proyectos y objetivos educativos, frente a resultados estandarizados por los que finalmente se va a evaluar la labor educativa; dejan al profesorado el dilema de decidir por cuál de dos opciones orientar su trabajo.

## 2. Se tiende a medir conocimientos frente a objetivos finales educativos

Existen diferentes clasificaciones de los posibles resultados educativos que pueden darse en los jóvenes, una de las más referenciadas es la taxonomía de Bloom (Madaus, 2004). Diferencia tres grandes grupos objetivos educativos (*Center for Teaching and Learning*, 2015)http://www.xx/:

- Objetivos basados en los conocimientos, entendiendo por conocimientos todo recurso cognitivo utilizado o creado por un sujeto y conservado en la memoria (Pastré, Mayen y Verganud, 2006).

- Objetivos basados en las habilidades, competencias (*skills*), entendiendo por competencias la movilización de dichos contenidos de una forma creativa para resolver problemas significativos y reales.

- Objetivos basados en los aspectos afectivos. También se puede encontrar la denominación habilidades no cognitivas para referirse las actitudes, la motivación y el juicio entre otros aspectos personales frente a las habilidades académicas (Morrison y Schoon, 2013).

En la décadas de los 80' los test fueron criticados, entre otras razones, por centrase demasiado en una valorización de la memoria y los conocimiento en vez de medir el pensamiento activo de los estudiantes (Nevo, 2006). Es más fácil medir un conocimiento de carácter expositivo que una competencia. Por esta razón diversos autores (Madaus, Haney and Kreitzer,2000; Cronbach, 2000) sostienen que los test no suelen ofrecer información sobre las capacidades de pensamiento, por lo que finalmente no son útiles para diagnosticar y valorar el desarrollo educativo de los jóvenes.

Si la evaluación de la enseñanza se centra en los conocimientos, puede darse la tendencia a relegar a un segundo lugar los aspectos competenciales o no cognitivos, que contradictoriamente son los objetivos finales de un sistema educativo. Cronbach (2000) advierte que la educación que se centra solamente la adquisición de un conocimiento expositivo puede no promover, e incluso interferir, con los resultados educativos más importantes como son los procesos de pensamiento.

## 3. Resultados no atribuibles a los centros o al sistema educativo

Uno de los propósitos de los test es utilizar las medidas estandarizadas para poder comparar resultados entre sí y sacar conclusiones sobre la calidad y la eficacia de las diferentes instancias educativas (aulas, escuelas, territorios, países,…). Para poder emitir dichos juicios de valor se parte de una premisa muy clara, los logros son debidos a la intervención educativa.

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


Pero en el contexto en el que se desarrolla la intervención educativa existen otras variables (personales, familiares, del entorno,…) que pueden estar influyendo, contribuyendo, aminorando o anulando los efectos de la acción escolar (Alvira, 1991). De hecho, el programa se puede considerar como un factor más entre otros. Una de las variables con mayor influencia en los logros académicos es el estatus económico, social y cultural de las familias a las que pertenecen los estudiantes.

"Entre las variables que más determinan este índice se encuentran el nivel de estudios de los padres, las expectativas que tienen sobre los estudios de sus hijos o el número de libros que hay en el hogar. Esta relación entre resultados y estatus social, económico y cultural de las familias es incuestionable (como ya señaló Coleman hace medio siglo)" (…). (Ministerio de Educación, 2007).

Esto quiere decir que aparentemente un centro puede obtener buenos resultados pero que no sean debidos a su desempeño sino al estatus socioeconómico de sus estudiantes y a la inversa, puede ser que un colegio haga una labor excepcional y obtenga bajas puntuaciones debido a otros factores externos.

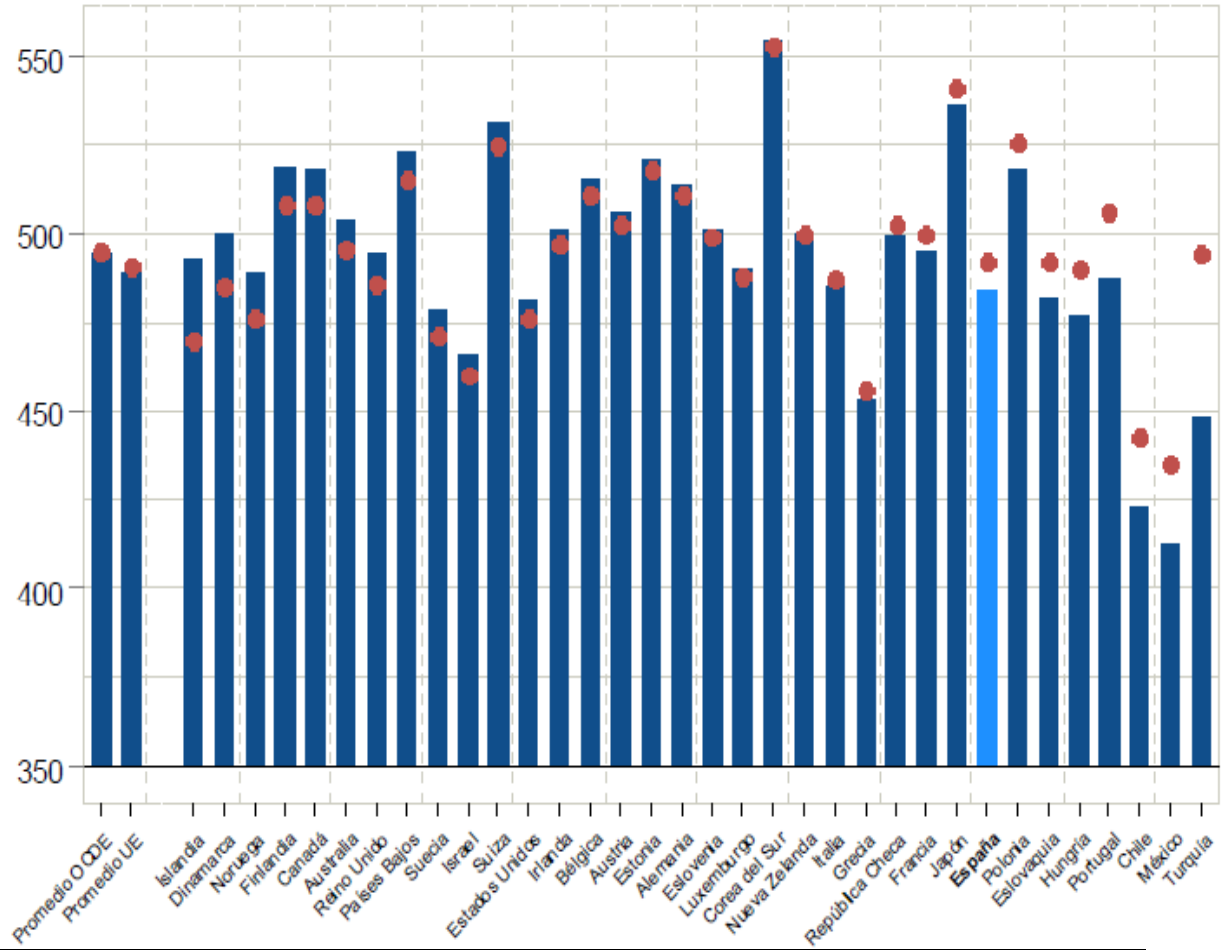
Por lo tanto, una de las cuestiones clave en evaluación es analizar si los cambios que se observan son atribuibles a la intervención o en qué medida la intervención ha contribuido a dichas modificaciones. Bajo el paradigma positivista o postpostivista**[9]** se han elaborado diferentes estrategias para controlar el efecto de otras posibles variables externas, lo que normalmente "requiere de una combinación de diseño experimental, estadística inferencial, observación empírica y una teoría substantiva" (Cook, 2004:88). La tendencia actual es resumir las diferentes estrategias en dos grandes grupos, estrategias de comparación apoyadas en el azar (diseños experimentales y cuasiexperimentales) y modelización estadística.

En el caso del estudio PISA para poder tener en cuenta este factor ha elaborado un Índice Social, Económico y Cultural (ESCS en sus siglas en inglés) que refleja la ocupación profesional y el nivel educativo de los padres y madres, así como los recursos disponibles en el hogar a través, por ejemplo, del número de libros en casa. Del análisis de este índice se pueden obtener varias conclusiones:

- Los países o territorios tienen diferentes estatus socioeconómicos.
- El estatus tiene altas correlaciones con los resultados académicos.

Sí se controla estadísticamente ESCS resulta interesante observar *las puntuaciones que tendrían los participantes en PISA si el valor de sus ESCS equivaliese al nivel promedio de la OCDE, es decir, a cero* (PISA, 2014: 99), que es una de los mecanismos para detraer el efecto de variables intervinientes.

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


| | | | |
|---|---|---|---|
| | Puntuación Media en Matemáticas sin ESCS | | Puntuación Media en matemáticas |

(Fuente: PISA, 2014: 100)
(**Gráfico 2:** Puntuaciones medias en matemáticas de los países de la OCDE, descontando ESCS)

Cuando las barras azueles están en una posición más baja que los puntos rojos indica *que los buenos resultados obtenidos por algunos países se deben en parte a su alto índice de bienestar sociocultural. Por el contrario, en países como Turquía, Chile, Japón, Polonia o España, al tener en cuenta el contexto social, económico y cultural se produce una mejora de los resultados de los alumnos de 15 años* (PISA, 2014: 99).

Como se puede leer en el gráfico 2 una vez que se controlan los datos los resultados obtenidos pueden distar mucho de la información aparente. Por ejemplo, en el caso de España el rendimiento de los jóvenes en matemáticas es mejor que en Islandia, Dinamarca, Noruega, Reino Unido, Suecia o Estados Unidos entre otros países, una imagen bastante más positiva de lo que el imaginario colectivo nos ofrece.

Teniendo en cuenta que en muchas ocasiones los test sirven de base para tomar decisiones sobre los grupos, centros o sistemas educativos e incluso se llegan a utilizar como referencia para determinar el salario del profesorado, las estimaciones tienen que ser rigurosas y válidas.

ÀÁF ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


Si no se controla la influencia de variables externas se van a emitir juicios injustos sobre los sistemas educativos. Además de ser una mala praxis evaluadora, el hecho de distribuir informes de evaluación sesgados, tal como dice Tyler (1991:16), "es realmente un crimen de cuello blanco".

## 4. Se evalúan solo los resultados y no la intervención educativa

Los test estandarizados miden los resultados finales en los estudiantes. Con esa información, tal como cité anteriormente, se enjuicia la totalidad del centro o de una determinada política educativa. Se asume el supuesto de que si ha habido buenos resultados es porque la formación que se ha dado ha sido buena o si los resultados han sido malos es a causa de que la propuesta docente no ha sido adecuada

Pero quizás detrás de unos malos resultados puede haber buenas docentes, con una buena práctica, ante un grupo con dificultades de aprendizaje, con una ratio elevada o en contextos sociales poco motivadores. O al contrario, detrás de unos buenos resultados puede haber una pésima docencia, con profesoras poco competentes pero con familias que suplen los déficits con profesorado de apoyo extraescolar. También ya he expuesto la asociación entre estatus socioeconómico y logros académicos, otro ejemplo en el que los resultados no tienen por qué dar información sobre la bondad de la docencia o el currículo. En definitiva, son demasiadas largas las cadenas causales para evaluar la calidad de un centro solamente con el dato final sumativo.

Debido a estas limitaciones Stufflebeam (2001) denomina a la evaluación de resultados como cuasi-evaluaciones, ya que emiten una valoración de un todo (grupo, centro escolar, política educativa,….) teniendo en cuenta sólo una parte: los resultados. También se les ha denominado evaluaciones de caja negra (Chen, 1990; Weiss, 1998) por no tener en cuenta los procesos de trabajo (docencia, coordinación, programación, tutorías, resolución de conflicto, actualización del currículo,…) que son las cosas que realmente producen los cambios en el alumnado. Lo que se quiere resaltar con la metáfora de caja negra es que no se sabe nada de cómo se hacen las cosas y, por lo tanto, no se pueden establecer relaciones lógicas entre un determinado tipo de docencia y los logros obtenidos.

Las propuestas alternativas buscan abrir, desentrañar y evaluar lo que hay dentro del programa. Ampliar la perspectiva de resultados con el análisis del contexto, de los métodos de trabajo y los medios con los que se cuenta (Tyler, 1991). La vinculación entre procesos y resultados es probablemente una de las contribuciones más importantes a la teoría de la evaluación (Greene,1999; Shadish, Cook, and Leviton, 1991) ya que busca la articulación causal entre lo que se hace, por ejemplo las clases y lo que se consigue, los resultados en los estudiantes.

Por más que se quiera enfatizar la orientación de los servicios públicos a los resultados finales, no se pueden invisibilizar los mecanismos que los producen, entre otras razones porque perdemos el conocimiento de cómo se hacen las cosas, de cómo se logran los éxitos en el campo educativo.

Cronbach (2000) entre otros muchos defiende que en la medida de lo posible la evaluación debe ser usada para entender como el curso produce sus efectos y qué variables influyen en la eficacia (los resultados). De este modo la evaluación no es

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


solamente enjuiciamiento sino también comprensión de los mecanismos causales. Entender cómo se ha generado el aprendizaje educativo es el primer paso para mejorarlo.

En resumen, los test juzgan a las escuelas pero no les dotan de información para comprender los fallos o los éxitos. No son sistemas útiles para la mejora de los centros, sencillamente porque no ofrecen información sobre lo que hay que mejorar. Realmente los test están orientados hacia un propósito sumativo. Tal como lo explica Weiss (1998:32) la evaluación formativa ayuda a desarrollar el programa y la sumativa a rendir juicios sobre él.

Normalmente los políticos, la dirección, las profesionales y los participantes tiene diferentes propósitos hacia la evaluación (Greene, 2007; Patton, 2008). Las directivas, profesoras y, frecuentemente, los participantes demandan información para entender por qué pasan las cosas e información para buscar soluciones o mejorar. Por su parte, los políticos u otros actores con capacidad de decisión suelen requerir a la evaluación una información para emitir juicios finales sobre el programa, que les valga para rendir cuentas públicas y para tomar grandes decisiones.

Por consiguiente, los test estandarizados responden a los intereses de los decisores políticos más que a los de las profesionales, dirección, alumnado y familias. Por mucho que se invoque "la evaluación para todos los propósitos es un mito" (Weiss, 1998:3).

## 5. Valoración de estudiantes no evaluaciones de centros o de sistemas educativos

En el ámbito educativo algunos autores han establecido una diferenciación terminológica más o menos aceptada en función del tipo de objeto que se analiza (estudiantes, profesionales y programas) (Schwandt, 2009: 19). El ejercicio de valoración de los logros de los estudiantes es definido como *assessment***[10]**, la valoración del desempeño docente como *appraisal***[11]** y cuando el foco está puesto en los programas, centros o servicios se denomina evaluación.

Nevo (2006:447) describe como el *testing* se transformó de alguna manera en una palabra sucia y en el contexto de evaluación de estudiantes se empezó a usar el término alternativo *assessment*. Tal como se recoge en la Enciclopedia de la Evaluación, *assessment* es la opción para describir la valoración de la calidad del trabajo de los estudiantes con el fin de determinar el nivel de logro que han alcanzado (Mabry, 2005).

Inicialmente los test estandarizados estaban orientados a la emisión de juicios sobre las personas, con la intención de seleccionar estudiantes para una formación avanzada, clasificar o diagnosticar competencias (Cronbach, 2000). Aunque con el tiempo, los test "gradualmente cambiaron la medición de resultados por otros objetos como programas, colegios, profesorado y sistemas educativos" (Nevo, 2009: 292). Esta translación del objeto la han definido Guba y Lincoln (1989) como una deficiencia seria engendrada en las primeras generaciones de métodos evaluativos**[12]**.

Evaluación y *assessment* son diferentes términos que implican diferentes miradas sobre diferentes objetos. Recapitulando los elementos que han salido en la

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


exposición de las críticas, las dos aproximaciones se diferencian en los siguientes puntos:

| Criterios | Test, *assessment* | Evaluación de programas |
|---|---|---|
| **Objeto de análisis** | Su foco es el alumnado. | Su foco es la instancia educativa: programa, aula, curso, centro, sistema, política… |
| **Dimensiones a tener en cuenta del objeto analizado** | Resultados. | Resultados, procesos (actividades, implementación, docencia) y elementos estructurales. |
| **Tipo de resultados a evaluar** | Tiende a mirar conocimientos de forma estandarizada. | Mira todos los posibles cambios en las personas provocados por acción educativa (con diferentes niveles de abstracción). El énfasis en que sean debidos a la intervención resalta la importancia otorgada a los mecanismos metodológicos para poder hablar de atribución o contribución del programa en los resultados. |
| **Usuarios principales** | Principalmente responde a las necesidades de políticos y otros decisores políticos. | Puede tener en cuenta las necesidades de una amplia variedad de actores, incluidos dirección, profesorado, alumnado y familias. |
| **Propósitos** | Rendimiento de cuentas público (*accountability*) y apoya la gran toma de decisiones del tipo continuar o no, expandirse, modificar,… | Además del rendimiento de cuentas, puede contribuir a la comprensión de la unidad evaluada, metiéndose en la caja negra del programa, y por lo tanto se orienta sustentadamente hacia la mejora. |

El problema no está en la confusión de términos, sino en la confusión de métodos. Se aplican las lógicas del examen a estudiantes para extraer conclusiones sobre todo el sistema educativo, con todas las consecuencias negativas que esto conlleva: homogenización e imposición de objetivos, mirada limitada, juicio injustos sobre los centros, descapitalización de conocimientos sobre la práctica docente y desorientación de todo el sistema con relación al propósito de mejora.

## CONCLUSIÓN

Los test estandarizados son una herramienta muy versátil para examinar determinados conocimientos. Su facilidad de aplicación, el reconocimiento institucional que tienen y el poder de la comparación entre escuelas o sistemas, son aspectos que resultan muy atractivos.

La emergencia que están teniendo los test puede hacer pensar que se está ante un descubrimiento técnico reciente, que nos muestra una forma más eficiente y objetiva de evaluar la acción pedagógica. Pero los test, tal cual los estamos viendo, son una de las aproximaciones más antiguas para la valoración de estudiantes.

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING

**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


Por su parte, las críticas a los test han supuesto retos metodológicos que han acabado generando alternativas y han contribuido de forma sustancial al levantamiento del cuerpo teórico de la evaluación de programas educativos.

No obstante, es frecuente encontrar una confusión de términos y métodos entre *assessment* y evaluación. Lo que puede hacer creer que los test son válidos para todos los propósitos evaluativos, pero esto no es así y, tal como se ha visto en las críticas anteriores, pueden traer graves consecuencias negativas al sistema escolar. Utilizando como referencia el caso español, es aún más llamativo el gran esfuerzo político y económico que se está haciendo por extender los test a todos los centros y aulas en contraste con los recortes presupuestarios que se están produciendo en el mismo momento en el ámbito educativo.

Es cierto, que el sistema educativo debe ser evaluado y revisado, pero lo mismo hay que hacer con los modelos de evaluación ya que también son recursos públicos. La propuesta de test estandarizados debe ser analizada para saber si genera el valor social que se le presupone. Bajo mi punto de vista basta con hacerse dos preguntas ¿qué es lo que se quiere conseguir con los test? Y ¿para qué están sirviendo real y concretamente?

Afortunadamente se pueden encontrar muchos métodos y aproximaciones de evaluación que se adaptan a diferentes propósitos y usos, los metodólogos y los teóricos han hecho su trabajo. Ahora, si nosotros utilizamos un tenedor para tomar la sopa ya solo es responsabilidad nuestra.

**Notas:**

**[1]** He procurado utilizar un lenguaje inclusivo desde una perspectiva de género. Cuando he necesitado usar las terceras personas del singular o del plural he optado por la convención de referirme a todo el profesorado y dirección como profesoras y directoras (femenino) y al resto de los actores (estudiantes, autores,…) en masculino, indistintamente de que haya hombres y mujeres en todos los grupos mencionados.

**[2]** La evaluación por objetivos es el "el proceso de determinar en qué medida los objetivos educativos son efectivamente cumplidos" (Nevo, 2006:442).

**[3]** Evaluación formativa: Se trata de evaluaciones diseñadas, realizadas y destinadas a apoyar los procesos de mejora, normalmente encargados o realizados por alguien y entregados a alguien que pueda llevar a cabo las mejoras (Scriven, 1991:19).

**[4]** Dada la dificultad de encontrar un término adecuado en castellano, he optado por mantener la voz inglesa de *assessment*. Usualmente se encuentra traducido como evaluación aunque probablemente la expresión más cercana a su significado técnico sea "examen de estudiantes".

**[5]** La característica central del modelo está definida por el acrónimo CIPP que representa la evaluación de Productos, Procesos, Inputs y Contexto de cualquier entidad (Stufflebeam, 2005).

**[6]** Desarrollo no cognitivo: Concepto introducido por Bowles & Gintis que hace referencias a cuestiones como las actitudes, la motivación y el juicio entre otros aspectos personales (Morrison y Schoon, 2013).

**[7]** En evaluación y calidad es usual encontrar el término *accountability* en inglés por su difícil traducción al castellano. Se puede entender como rendimiento de cuentas público (aunque no hace mención exclusivamente a la contabilidad) o responsabilización (Echebarria, 2005).

**[8]** En EE.UU en aquella época había muchos colegios e institutos inspirados en el movimiento de renovación pedagógica conocido como educación progresiva que trabajaban respetando la diversidad del alumnado y generando diferentes procesos pedagógicos.

**[9]** Existen otras aproximaciones de carácter constructivista u otras bajo el denominado paradigma transformador.

**[10]** Como comentaba en una cita anterior, mantengo la voz inglesa de *assessment* ya que no se encuentra un paralelismo similar en castellano a los términos "*evaluation*" – "*assessment*".

**[11]** Igual que con *assessment* he optado por mantener el término de *appraisal* en inglés a la espera de un consenso sobre una traducción técnica al castellano.

**[12]** Para evaluar un programa educativo es cierto que es necesario evaluar sus resultados pero articulando esas piezas de información dentro de un esquema interpretativo más amplio que nos permita comprender y enjuiciar de forma completa la intervención.

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING

**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


**Referencias Bibliográficas**

Alvira, F. (1991) *Metodología de la evaluación de programas.* Madrid: CIS.

Brindusa, A., Cabrales, A., Sainz, J. y Sanz, I. (2012) Publicación de los resultados de las pruebas estandarizadas externas: ¿tiene ello un efecto sobre los resultados escolares. En A. Cabrales y A. Ciccone, *La educación en España una visión académica*. Fedea Monografías. Recuperado en: http://www.fedea.net/educacion/monografia-2013/web-monografia-educacion-2013.pdf.

Bustelo, M. (2011). *Last but not least: gender sensitive evaluations as a forgotten piece of the policymaking process*. Paper presented at ECPR General Conference. Reykjavik, August 25-27th 2011.

Center for Teaching and Learning (2015) *Bloom´s Taxonomy Educational Objectives*. Recuperado en: http://teaching.uncc.edu/learning-resources/articles-books/best-practice/goals-objectives/blooms-educational-objectives#sthash.3wEM2rqA.dpuf.

Chen, H.T. (1990) *Theory-Driven Evaluations*. CA: Sage.

Cook, T.D (2004) Causal Generalization: How Campbell and Cronbach Influenced My Theoretical Thinking on This Topic, Incluiding in Shadish, Cook and Campbell. In M. Alkin (Ed.) *Evaluation Roots.* California: Sage.

Cronbach, L.J. (2000). Course Improvement Through Evaluation. In D. L. Stufflebeam, G.Gl Madaus and T. Kellagha (Eds.) *Evaluation Models Viewpoints On Educational and Human Services Evaluation* (pp. 235-248). London: Kluwer Academic Publishers.

Echebarria, k (2005) Responsabilización y responsabilización gerencial: instituciones antes que instrumentos. En CLAD *Responsabilización y evalución de la gestión pública.* Venezuela: CLAD.

España. Ley Orgánica 8/2013, de 9 de diciembre, de Mejora de la Calidad Educativa. *Boletín Oficial del Estado*, 10 de diciembre de 2013, núm. 295, p. 97858 -97921.

Greene, J. (2007). *Mixed Methods in Social Inquiry*. John Wiley & Sons.

Guba, E.G. and Lincoln, Y.S. (1989). *Fourth Generation Evaluation*. CA: Sage.

House, E. (1990). Trends in Evaluation in *Educational Researcher* Vol. 19, No. 3 (Apr., 1990), pp. 24-28.

INEE (2015) *Comunidades Autónomas*. Recuperado en: http://www.mecd.gob.es/inee/portada.html.

Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation2nd edition. Thousands Oaks, CA: Sage.*

Mabry, l. (2005) Assessment. In S. Mathison (Ed.) *Encyclopedia of Evaluation*. Thousand Oaks, CA: Sage.

Madaus, G.F. (2004) Ralph W. Tyler`s Contribution to Program Evaluation. In M. Alkin (Ed.) *Evaluation Roots.* California: Sage.

Madaus, G.F., Haney, W. and Kreitzer, M. (2000). The Role of Testing in Evaluation. In D. L. Stufflebeam, G.Gl Madaus and T. Kellaghan (Eds.) *Evaluation Models Viewpoints On Educational and Human Services Evaluation* (pp. 113-126). London: Kluwer Academic Publishers.

Madaus, G. and Stufflebeam, D. (2000) Program Evaluation: A historical overview. In D. L. Stufflebeam, G.Gl Madaus and T. Kellaghan (Eds.), *Evaluation Models Viewpoints On Educational and Human Services Evaluation* (pp. 3-18). London: Kluwer Academic Publishers.

Milbrey, W. Mc Lauglhlin and Phillips, D.C (1991) *Evaluation and Education: At Quarter Century*. Chicago, Illinois: The National Society for the Study of Education.

Ministerio de Educación (2007) *Educación Primaria. Evaluación general del sistema educativo.* Madrid: Ministerio de Educación.

Monnier, E. (1992). *Evaluación de la acción de los poderes públicos.* Madrid: Instituto de Estudios Fiscales.

Morrison, L. & Schoon, I. (2013) *The Impact of Non-Cognitive Skills on Outcomes for Young People*. Literature review. London: Institute Of Education. Recuperado en: www.ioe.ac.uk.

Nevo, D. (2009*).* Accountability and Capacity Building: Can they live together? In K.E. Ryan and J. B. Cousins (Eds.) *The Sage International Handbook of Educational Evaluation*(pp.291-304). CA: Sage.

(2006) Evaluation in Education. In I.F. Shaw, J.C. Greene and M.M. Mark (Eds.), *The Sage Handbook of Evaluation* (pp. 441-460). London: Sage.

Pastré, P., Mayen, P. et Vergnaud, G.(2006) La didactique professionnelle, *Revue Française de Pédagogie* (janvier-mars)

Patton, M.Q. (2008).*Utilization-Focused Evaluation.* California: Sage.

PISA (2014) *PISA 2012 Programa para la evaluación internacional de los alumnos*. Madrid: OCDE y Ministerio de Educación, cultura y Deporte.

Ritchie, C. (1971). Can We Afford to Ignore It? *Educational Leadership,* 484-485.

Rizvi, F. (2009) Globalization and Policy Research in Education. In K.E. Ryan and J. B. Cousins (Eds.) *The Sage International Handbook of Educational Evaluation* (pp.3-18). CA: Sage.

Ryan, K.E. and Cousins J.B. (2009) Introduction. In K.E. Ryan and J. B. Cousins (Eds.) *The Sage International Handbook of Educational Evaluation* (pp.ix-xvii). CA: Sage.

ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING

**Nº 43 (2a. época) noviembre 2015**
**URL**: www.ambitsaaf.cat
**ISSN**: 2339-7454


Schwandt, T.A (2009) Globalizing Influences on the Western Evaluation Imaginary. In K.E. Ryan and J. B. Cousins (Eds.) *The Sage International Handbook of Educational Evaluation* (pp.19-36). CA: Sage.

Scriven, Michael. (1991). Beyond Formative and Summative Evaluation. In W. . Milbrey,W. Mclaughlin y D.C Phillips (Eds.) *Evaluation and Education: At Quarter Century*. Chicago, Illinois: The National Society for the Study of Education.

Shadish, W., Cook, T.& Leviton, L. (1991). *Foundations of Program Evaluation*. Ca: Sage.

Stake (2006). *Evaluación comprensiva y evaluación basada en estándares*. Barcelona: Grao.

Stufflebeam, D. (2005) CIPP Model. In S. Mathison (Ed.) *Encyclopedia of Evaluation*. Thousand Oaks, CA: Sage.

Stufflebeam, D. (2001). *Evaluation Models*. San Francisco: New Directions in Program Evaluation.

Tyler, R. (1991). General Statement on Program Evaluation. In M.W. McLaughlin and D.C. Phillips (Eds.) *Evaluation and Education: At Quarter Century*. Chicago, Illinois: The National Society for the Study of Education.

Weiss, C. (1998). *Evaluation.* New Jersey: Prentice – Hall.

**Correspondencia con el autor:** Juan Andrés Ligero Lasa. (www.magisterevaluacion.es). E-mail: jligero@uc3m.es.