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

Nº 43 (2a.època) novembre 2015
URL: www.ambitsaaf.cat
ISSN: 2339-7454

USOS I CRÍTIQUES DELS TEST ESTANDARDITZATS EN ELS SISTEMES EDUCATIUS

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

RESUM
"Usos i crítiques dels test estandarditzats en els sistemes educatius"
Actualment, tant en l'àmbit local com internacional, hi ha un gran desenvolupament i es dona cada cop més importància als sistemes d'avaluació educativa. La inclusió de proves a les iniciatives legislatives educatives, el desenvolupament d'estratègies avaluatives en els diferents nivells de l'administració educativa i les proves internacionals d'avaluació com PISA, són clars indicis d'aquesta expansió.
Encara que en general l'increment de l'avaluació educativa és quelcom de positiu, el seu desenvolupament s'està majoritàriament restringint a una determinada perspectiva metodològica, els sistemes de mesura del rendiment escolar a través d'exàmens externs estandarditzats o test estandarditzats.
Utilitzar un o altre mètode d'avaluació no és indiferent. Cada mètode genera diferents productes avaluatius que poden ser adequats per a determinats usos i, per tant, també poden implicar diferents conseqüències en els objectes avaluats, és a dir, en els centres o en els sistemes educatius.
L'avaluació escolar ha de ser un exercici conscient i coneixedor dels propòsits que es pretenen per poder adequar els mètodes al seu ús. L'objectiu d'aquest article és contextualitzar els tests estandarditzats i analitzar els límits i les crítiques d'aquest tipus de mètode, utilitzant com a estudi de cas el sistema utilitzat a Espanya. El text s'ha estructurat en diferents epígrafs: contextualització històrica, definició i característiques principals dels test, principals crítiques i conclusió.

Paraules Clau: Avaluació. Educació. Currículum

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

CONTEXTUALITZACIÓ HISTÒRICA

L'àmbit educatiu ha estat un dels sectors més prolífics en el desenvolupament metodològic de l'avaluació; gran part dels teòrics**[1]** han elaborat el seu

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

pensament en relació a objectes educatius, pel que gairebé parlar d'història metodològica d'avaluació educativa és parlar de la història de l'avaluació.

Encara que es poden trobar referències avaluatives aïllades en segles anteriors, és a finals del segle XIX quan s'aprecia un volum d'avaluacions suficients que permeten identificar un primer punt de partida de la disciplina. Probablement s'expliqui per les grans transformacions que les revolucions industrials van provocar, el que Madaus i Stufflebeam (2000) han anomenat l'edat de les reformes. Durant els primers anys del segle XX, l'educació com d'altres àmbits de la intervenció política, no es lliuren de la metàfora febril. L'alumnat és vist com "matèria primera" que ha de ser processada a l'escola com "planta de tractament". El deure de la direcció és "fer la seva feina tan efectiva i eficient com sigui possible" (Guba i Lincoln, 1989: 26). Sorgeixen estratègies i tècniques coherents amb aquesta visió amb la intenció d'incrementar la racionalització i l'eficiència dels centres educatius.

El 1904 el Ministeri d'Educació francès encarrega a Alfred Binet un sistema per identificar i poder descartar "els joves mentalment retardats" que puguin dificultar el desenvolupament de la resta del grup (Guba i Lincoln, 1989: 23), conegut posteriorment com a test d'intel·ligència (Monnier, 1992). En el mateix any, l'Associació Nacional d'Educació dels EUA va crear un comitè per estudiar l'ús dels tests en la classificació d'estudiants i en la determinació dels seus progressos. El 1908 es publica el test de raonament en aritmètica i el 1922 Stanford desenvolupa una bateria de test que permeten la valoració simultània dels estudiants en múltiples assignatures (Guba i Lincoln, 1989).

Durant la primera meitat del segle XX es consolida l'ús dels tests que deixen de ser experiències singulars. El 1933 Gertrude Hildreths publica una bibliografia on es trobaven 3.500 test mentals i escales de puntuació; el 1945 va actualitzar el treball recollint 5.200 instruments. El 1958 l'Acta Nacional de defensa de l'Educació als EUA declarava la necessitat d'avaluar els plans educatius. Aquest fet, juntament amb el desenvolupament de les màquines de lectura ràpida dels tests (Madaus, 2004: 77) va generar un moviment conegut com el boom dels tests estandarditzats.

Encara que s'estigués davant un moment expansiu, des d'una perspectiva metodològica, l'escena quedava reduïda als tests estandarditzats i en menor mesura a la proposta de Ralph Tyler d'avaluació per objectius**[2]** (Millbrey i altres, 1991).

El desenvolupament d'un mercat d'avaluació real amb demandes en diferents situacions, amb múltiples actors amb interessos i propòsits diferents, posaven a prova i tensionaven els límits dels mètodes existents. A la fi dels anys 60' apareixen les principals crítiques als mètodes predominants:

- En els test, el focus exclusiu en els resultats de l'alumnat no permet entendre per què passen les coses (Tyler, 1991; Cronbach, 2000).

- Tampoc proveeixen d'una informació fonamentada per prendre decisions sòlides que serveixin per millorar la docència (Cronbach, 2000).

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

- Tant els tests com l'avaluació per objectius aporten només una descripció o una mesura sense proporcionar mecanismes per avaluar suficientment els programes (Guba i Lincoln, 1989).

- No tenen en compte les necessitats i els valors dels diferents actors implicats en una intervenció (Stake, 2004).

Aquestes crítiques van ser l'esperó a finals dels 60 'principis dels 70' pel sorgiment del període metodològicament més creatiu. "L'avaluació es va moure de concepcions monolítiques a més plurals, múltiples mètodes, mesures, criteris, perspectives, audiències i fins i tot interessos" (House, 1990: 24). Van sorgir diferents aproximacions que proposaven altres mirades, sensibilitats, i en tots els casos es van generar noves línies d'estudi en avaluació (Schwandt, 2009).

"Scriven (1967), Stufflebeam (1967 i 1971, amb altres) i Stake (1967) van introduir nous models d'avaluació que es diferenciaven radicalment de les aproximacions anteriors. Aquestes conceptualitzacions reconeixien la necessitat d'avaluar els objectius, mirar els recursos, examinar la implementació i la forma d'oferir els serveis, així com mesurar els resultats previstos i no previstos del programa. Ells també emfatitzaven la necessitat d'avaluar el mèrit o el valor de l'objecte que estava sent avaluat "(Madaus i Stufflebeam, 2000: 14).

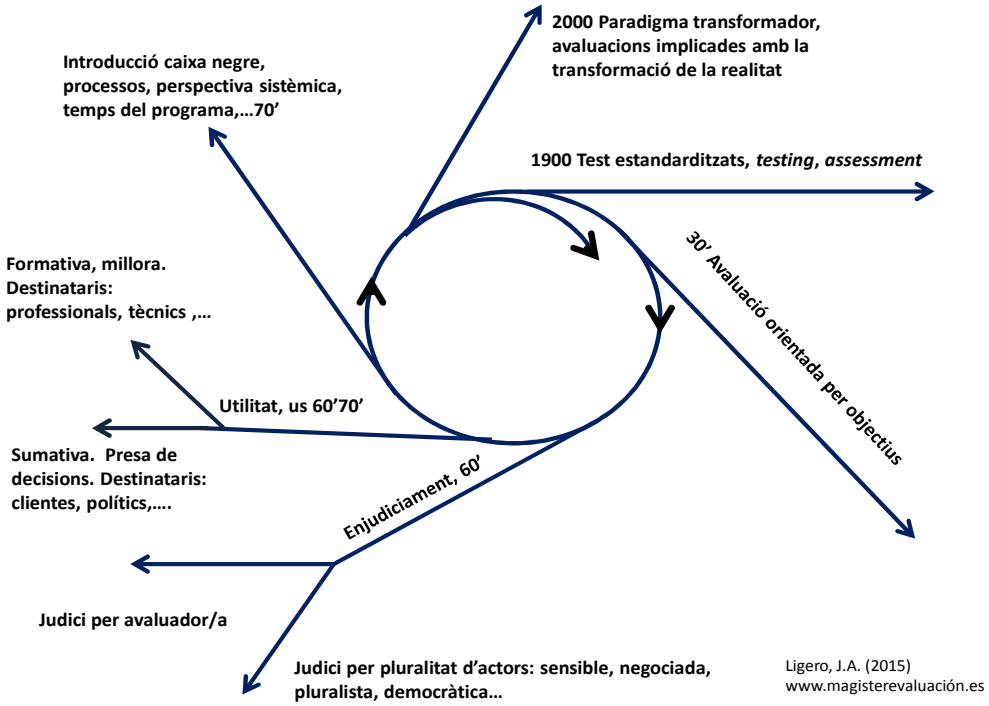
A la mateixa època el paradigma científic tradicional va ser posat en qüestió pel constructivisme que negava l'existència de l'objectivitat tal com s'entenia fins aquell moment. La realitat "són construccions socials mentals, hi pot haver tantes construccions com individus hi hagi (encara que moltes poden ser compartides entre els diferents subjectes)" (Guba i Lincoln, 1989: 43). Per tant, tal com ho descriu Scriven, l'objectivitat, si existeix, es pot definir com un acord intersubjectiu entre els diferents actors implicats (Chen, 1990). "L'avaluació canvia d'un èmfasi prioritari en els mètodes quantitatius, (...) a una actitud en la qual els mètodes qualitatius es converteixen en acceptables" (House, 1990: 25).

Aquest moviment per a alguns autors (House, 1990; Schwandt, 2009) va suposar el naixement real de l'avaluació de programes o l'etapa del seu desenvolupament (1958-1972) (Madaus i Stufflebeam, 2000); el que, en qualsevol cas, ve a assenyalar la diversitat, la innovació i la profusió de mètodes que hi va haver en aquest període.

Als EUA l'arribada de Reagan al govern (1981-1990) va suposar una retallada de l'avaluació professional, almenys des d'un punt de vista extern (House, 1990). En canvi, la nova reforma educativa que es va impulsar durant aquests anys va portar un renovat èmfasi en els tests estandarditzats, no amb propòsits formatius**[3]** sinó més aviat enfocada a oferir un rendiment de comptes públic.

En síntesi, la narració històrica de les metodologies no es pot entendre com un procés evolutiu on els nous mètodes descarten als antics, ja que encaixa més amb la idea d'un espai escènic on l'eix cronològic mostra l'entrada a l'escena de les aproximacions que després hi romanen. En el gràfic següent es representa una síntesi de les principals aproximacions i la seva seqüència d'aparicions.

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



(**Gràfic 1.** Seqüència històrica d'aparició de les diferents aproximacions avaluatives.)

Com a mostra el dibuix, els tests estandarditzats han estat unes primeres propostes metodològiques (faltaria per incloure els sistemes d'inspecció) sobre les quals s'han anat suplementat altres aproximacions, en part per respondre als dèficits i limitacions que mostraven. L'avaluació per objectius emfatitza la singularitat dels centres que no es recollia. Les propostes de finals dels 60` incorporen la importància d'emetre judicis de valor sobre l'objecte i no només mesurar els resultats. Es pren consciència de l'existència de diferents actors amb diferents valors que desborden una mirada única o estandarditzada (Bustelo, 2011). Aquest pluralisme d'actors també fa palès els diferents propòsits de l'avaluació i mostren que un únic mètode no té per què respondre a tots. També en aquesta època, els avaluadors comencen a mirar l'interior de la caixa negra, a descobrir relacions causals i a veure per què es produeixen certs resultats. A més, les perspectives d'indagació s'amplien amb les tècniques qualitatives alhora que se exigeix rigor en la seva aplicació.

Les aproximacions sorgides en els diferents moments poden mantenir desenvolupaments propis i independents del corrent principal, com així ho intenten representar les fletxes tangents, fins i tot fins a generar disciplines diferents com ha estat el cas dels tests i el *assessment***[4]** tal com s'explica més endavant.

D'altra banda, actualment es pot trobar qualsevol de les aproximacions avaluatives referides. Per exemple, en un mateix centre escolar es pot trobar un model CIPP (Context*, Input*, Procés i Producte)**[5]** juntament amb test estandarditzats i gaudir tots dos d'acceptació. Hi ha tal profusió de mètodes (22 aproximacions recull Stufflebeam, 2001) que semblaria que escollir entre un i altre és tan sols una

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

qüestió de gustos. És funció de la metodologia ajudar a definir i reconèixer les seves virtuts i limitacions per poder adequar-los al seu ús.

Aquest és l'objectiu del present article. Donat el gran desenvolupament que estan vivint actualment els tests estandarditzats, cal conèixer els seus contorns i limitacions per saber si el mètode s'ajusta als objectius que es pretenen.

DEFINICIÓ I CARACTERÍSTIQUES DELS TEST ESTANDARDITZATS

Els tests o els exàmens externs estandarditzats són maneres diferents d'anomenar el mateix mètode, que consisteix en un qüestionari amb preguntes i respostes tancades o amb un alt grau d'estructuració. Els qüestionaris poden estar adreçats a qualsevol població, però en l'àmbit educatiu és freqüent trobar-los aplicats a estudiants per mesurar el seu acompliment en coneixements, competències o en desenvolupament no cognitiu**[6]**. S'apliquen homogèniament a tots els subjectes seleccionats i el tractament de la informació està estandarditzat, és a dir, és igual en tots els casos.

Als test se'ls atribueix certs avantatges enfront d'altres mètodes:

- Són proves preparades prèviament, normalment per una institució educativa o governamental, pel que la seva aplicació requereix menys elaboració i una reducció de la incertesa sobre els aspectes que han de ser mesurats.
- La seva aplicació procedimentada evita biaixos intencionals.
- L'estandardització de les mesures permet una comparació entre instàncies educatives (aules, escoles i territoris). Es poden elaborar rànquings de més a menys puntuació entre les diferents unitats educatives.
- Aquestes mateixes classificacions poden estimular l'aprenentatge organitzatiu identificant aquelles estratègies pedagògiques que han aconseguit millors resultats.
- Poden oferir informació sobre l'acompliment als diferents actors del sistema educatiu, fomentant la rendició de comptes social del sistema educatiu.
- Contribueixen a la responsabilització dels centres educatius amb els seus resultats, fomentant un model de gestió gerencial (Bhen cit. Ryan i Cousins, 2009) i s'espera que això afavoreixi la generació d'una cultura de qualitat i millora.

En l'actualitat els test tenen un alt grau d'acceptació, s'estan implantant cada cop en un major nombre de països i territoris. Per exemple, l'any 2012 PISA (*Program International Students Assesment*) es va aplicar a 65 països dels cinc continents enquestant 510.000 estudiants que representen una població aproximada de 28 milions de joves de 15 anys (PISA, 2014). A Europa pràcticament tots els països de la Unió Europea s'han sumat a aquest programa.

A més de PISA hi ha altres proves adreçades a estudiants com TIMSS (*Trends in International Mathematics and Science Study*), PIRLS (*Progress in International Reading Literacy Study*) o EECL (Estudi Europeu de Competències Lingüística). També hi ha proves per docents i personal en càrrecs de direcció com l'estudi TALIS

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

Nº 43 (2a.època) novembre 2015
URL: www.ambitsaaf.cat
ISSN: 2339-7454

(*Teaching and Learning Internacional Survey*) i fins i tot es poden trobar proves estandarditzades per la població adulta com PIAAC (Programa per a l'Avaluació Internacional de les Competències dels Adults) (INEE , 2015).

La profusió de test estandarditzats no respon només a un moviment social. En un escenari internacional globalitzat (Rizvi, 2009) existeix una política articulada per a la promoció aquest tipus de sistemes de mesura escolar. Hi ha organismes internacionals que treballen en el foment d'aquesta visió, com ara l'Organització per a la Cooperació i el Desenvolupament Econòmic (OCDE), l'*Association for the Evaluation of Educational Achievement* (IEA) o la iniciativa *World Education Indicator Program* promogut pel Banc Mundial i la UNESCO.

En l'àmbit nacional cada vegada són més els països que implementen sistemes propis de test per observar la qualitat educativa (Ryan i Cousins, 2009). Això també implica un augment d'instituts i oficines gestores d'aquestes proves com la britànica *Office for Standards in Education Children Services and Skill* (Ofsted), la *National Assessment of Educational Progress* (NAEP) als EUA o l'Institut Nacional d'Avaluació Educativa (INEE) d'Espanya.

Prenent Espanya com a estudi de cas, és significatiu veure que una de les principals novetats introduïdes per la nova llei educativa LOMCE (8/2013, 9 de desembre) és la incorporació en tot el sistema educatiu de les proves externes estandarditzades.
"Article 144. Avaluacions individualitzades. (...) En concret, les proves i els procediments de les avaluacions indicades en els articles 29 i 36 bis es dissenyaran pel Ministeri d'Educació, Cultura i Esport, a través de l'Institut Nacional d'Avaluació Educativa. Aquestes proves seran estandarditzades i es dissenyaran de manera que permetin establir valoracions precises i comparacions equitatives, així com el seguiment de l'evolució al llarg del temps dels resultats obtinguts".

D'altra banda, hi ha hagut un gran desenvolupament de les proves a les comunitats autònomes. Per citar alguns exemples, a Catalunya s'apliquen proves externes a 6è de primària i a 4t de l'ESO. A Andalusia, la prova Escala es passa a 2n de primària. Al País Basc i la Comunitat Valenciana l'Avaluació Diagnòstica s'aplica a 4t de primària i 2n d'ESO. En cap dels quatre exemples es fan comparacions entre instàncies educatives, la informació es retorna als centres amb la intenció de generar una inèrcia formativa interna. Un cas diferent és el de Madrid on la prova CDI (Coneixements i Destreses Indispensables) es passa a 6è de primària i permet amb posterioritat comparar col·legis entre si i establir classificacions.
Part de la popularitat d'aquestes proves es deu a que les avaluacions i els seus resultats són usats per diferents actors:

- Es poden fer servir per pares i mares per conèixer l'acompliment escolar del centre i per decidir on matricular els seus fills. Per exemple, a la Comunitat de Madrid els resultats de les proves CDI són àmpliament consultats a través de la pàgina web en els períodes de matriculació escolar.

- Son utilitzats per la direcció i el professorat per identificar els resultats obtinguts en el seu centre. A aquesta pràctica també se li atorga la virtut d'estimular la millora de la qualitat, "perquè la supervisió més propera per part de pares i administradors proporciona una motivació addicional per a mestres i directors per millorar els resultats escolars dels seus estudiants" (Brindusa et al., 2012).

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

- Es pot usar per tota la cadena de decisors polítics per prendre decisions de caràcter sumatiu del tipus de continuar o no continuar amb una determinada política educativa o per actuar sobre alguns centres amb puntuacions fora del que s'esperava.

- A més, son utilitzats socialment i pels mitjans de comunicació. La informació sobre les puntuacions del sistema educatiu són notícia. N'hi ha prou amb observar el moviment mediàtic que es genera quan, per exemple, es publiquen els resultats PISA.

En definitiva, els tests estandarditzats han conformat un sistema propi de valoració escolar, àmpliament aplicat, amb el suport local i internacionalment, el que al seu torn reforça la seva hegemonia metodològica. Tanmateix, hi ha un clar conjunt de crítiques que identifiquen amb claredat les limitacions, els problemes i els falsos judicis que poden provocar.

CRÍTIQUES I LIMITACIONS DELS TESTS ESTANDARDITZATS

En aquest article he destacat les crítiques referides als aspectes més instrumentals dels tests estandarditzats, deixant de banda altres consideracions de caràcter polític o social atribuïbles als tests i a la seva funció d'*accountability***[7]**. Moltes de les crítiques ja s'han apuntat a la contextualització històrica, entre d'altres raons perquè en gran mesura l'avanç metodològic en avaluació s'ha creat per intentar solucionar algunes de les limitacions que el *testing* presentava. De forma sintètica les crítiques als tests es poden resumir en cinc punts:

1. S'avaluen resultats estandarditzats que poden coincidir o no amb els objectius i amb el treball real del centre educatiu.

2. Degut al fet de ser test a gran escala, solen tendir a mesurar coneixements i no objectius educatius finals o competencials.

3. Els resultats no són atribuïbles als centres, ja que hi ha d'altres variables que no es controlen que també influeixen en els resultats.

4. Tendeixen a avaluar resultats i no la intervenció, és a dir, les activitats educatives que provoquen aquests resultats.

5. Es tracta d'exàmens a estudiants, no d'avaluacions de centres o de polítiques educatives.

1. S'avaluen resultats estandarditzats davant objectius educatius del centre

Cada col·legi es troba en un entorn que pot tenir un determinat tipus d'alumnat amb el seu propi bagatge, amb els seus problemes, amb les seves fortaleses específiques, amb diferents recursos i suports comunitaris. Cada centre o unitat educativa cercarà d'adaptar-se a la seva realitat, d'utilitzar els seus recursos i eines pedagògiques per aconseguir uns determinats objectius.

Aquesta és l'essència en la qual es recolza la lògica dels projectes educatius de centre. Els objectius educatius són les fites fixades en els plans o projectes per

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

aconseguir canvis en els estudiants, adaptant-se a l'entorn per a resoldre problemes o reforçar potencialitats específiques de la població amb la que els toca treballar. Per tant, els objectius d'un centre en un determinat context poden ser molt diferents d'un altre centre que estigui en una altra situació. Això, no només no és negatiu, sinó que s'entén com una senyal de l'obligada adaptació de la institució educativa al seu entorn.

Però aquesta diversitat pedagògica, tant ara com fa 100 anys, desperta el temor que els diferents objectius i els currículums específics siguin una forma desordenada i poc rigorosa de preparar els joves per al seu pas a secundària o a la universitat. La investigació *Eight-Year Study* (Tyler, 1991) va concloure que els estudiants procedents dels instituts amb currículum obert**[8]** treien fins i tot millors resultats que el grup de comparació procedent d'ensenyament reglat i uniforme (Ritchie, 1971).

Indistintament dels plans educatius dels centres i per tant dels objectius educatius que s'hagin establert, els tests estandarditzats mesuren sempre els mateixos resultats. Traslladat a l'àmbit de l'aula, és optar per que el professorat no desenvolupi les seves pròpies proves per avaluar l'acompliment dels estudiants en la seva matèria, sinó que els exàmens ja vinguin donats per una instància superior. L'avaluació per objectius va néixer per respondre a la crítica de col·legis i professionals que els tests no mesuren el que ells estan tractant d'ensenyar als estudiants i, per tant, no són un instrument adequat per a avaluar el seu treball (Madaus, 2004: 74). A més, segons Tyler, els tests es recolzen en una gran desconfiança en els judicis dels docents, en la mesura que els exàmens realitzats pel professorat no es consideren rigorosos o fiables per valorar l'assoliment de l'alumnat.

En definitiva, els tests estandarditzats poden deixar sense mesurar aspectes que realment treballa una instància educativa o sobredimensionar-ne d'altres, imposant implícitament els objectius que s'han de treballar a totes les aules indistintament del context, diversitat individual o aposta pedagògica. Tal com ho expresa la LOMCE, "aquestes proves normalitzen els estàndards de titulació a tot Espanya, indicant de forma clara al conjunt de la comunitat educativa quins són els nivells d'exigència requerits" (...) (Preàmbul, VIII).

Sota aquesta perspectiva, els tests són un instrument per a l'homogeneïtzació dels objectius dels centres. El que provoca un conflicte entre la lògica de treball a través de projectes i objectius educatius, en front resultats estandarditzats pels que finalment s'avaluarà la tasca educativa; deixen al professorat el dilema de decidir per quina de les dues opcions orienten el seu treball.

2. Es tendeix a mesurar coneixements davant objectius finals educatius

Existeixen diferents classificacions sobre els possibles resultats educatius que poden mostrar-se en els joves, una de les més referenciades és la taxonomia de Bloom (Madaus, 2004). Distingeix tres grans grups d'objectius educatius (Center for Teaching and Learning, 2015):

- Objectius basats en els coneixements, entenent per coneixements tot recurs cognitiu utilitzat o creat per un subjecte i conservat a la memòria (Pastré, Mayen i Verganud, 2006).

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

- Objectius basats en les habilitats, competències (skills), entenent per competències la mobilització d'aquests continguts d'una forma creativa per resoldre problemes significatius i reals.

- Objectius basats en els aspectes afectius. També es pot trobar la denominació habilitats no cognitives per referir-se a les actituds, la motivació i el judici, entre d'altres aspectes personals davant les habilitats acadèmiques (Morrison i Schoon, 2013).

A la dècades dels 80' els test van ser criticats, entre d'altres raons, per centrar-se massa en valorar la memòria i els coneixement en lloc de mesurar el pensament actiu dels estudiants (Nevo, 2006). És més fàcil mesurar un coneixement de tipus expositiu que una competència. Per aquesta raó, diversos autors (Madaus, Haney and Kreitzer, 2000; Cronbach, 2000) sostenen que els test no solen oferir informació sobre les capacitats de pensament, de manera que finalment no són útils per diagnosticar i valorar el desenvolupament educatiu dels joves.

Si l'avaluació de l'ensenyament es centra en els coneixements, pot relegar-se a un segon lloc els aspectes competencials o no cognitius, que contradictòriament són els objectius finals d'un sistema educatiu. Cronbach (2000) adverteix que l'educació que es centra només en l'adquisició d'un coneixement expositiu pot no promoure, i fins i tot interferir, en els resultats educatius més importants que són els processos de pensament.

3. Resultats no atribuïbles als centres o al sistema educatiu

Un dels propòsits dels tests és utilitzar les mesures estandarditzades per poder comparar resultats entre si i treure conclusions sobre la qualitat i l'eficàcia de les diferents instàncies educatives (aules, escoles, territoris, països, ...). Per poder emetre aquests judicis de valor, es parteix d'una premissa molt clara, els èxits són deguts a la intervenció educativa.

Però en el context en què es desenvolupa la intervenció educativa hi ha altres variables (personals, familiars, de l'entorn, ...) que poden estar influint, contribuint, minorant o anul·lant els efectes de l'acció escolar (Alvira, 1991). De fet, el programa es pot considerar com un factor més entre d'altres. Una de les variables amb major influència en els èxits acadèmics és l'estatus econòmic, social i cultural de les famílies a les que pertanyen els estudiants.

"Entre les variables que més determinen aquest índex es troben el nivell d'estudis dels pares, les expectatives que tenen sobre els estudis dels seus fills o el nombre de llibres que hi ha a la llar. Aquesta relació entre resultats i estatus social, econòmic i cultural de les famílies és inqüestionable (com ja va assenyalar Coleman fa mig segle) "(...). (Ministeri d'Educació, 2007).

Això vol dir que aparentment un centre pot obtenir bons resultats però que no siguin deguts al seu acompliment sinó a l'estatus socioeconòmic dels seus estudiants i al revés, pot ser que un col·legi faci una tasca excepcional i obtingui baixes puntuacions causa d'altres factors externs.

Per tant, una de les qüestions clau en avaluació és analitzar si els canvis que s'observen són atribuïbles a la intervenció o, en quina mesura la intervenció ha contribuït a aquestes modificacions. Sota el paradigma positivista o

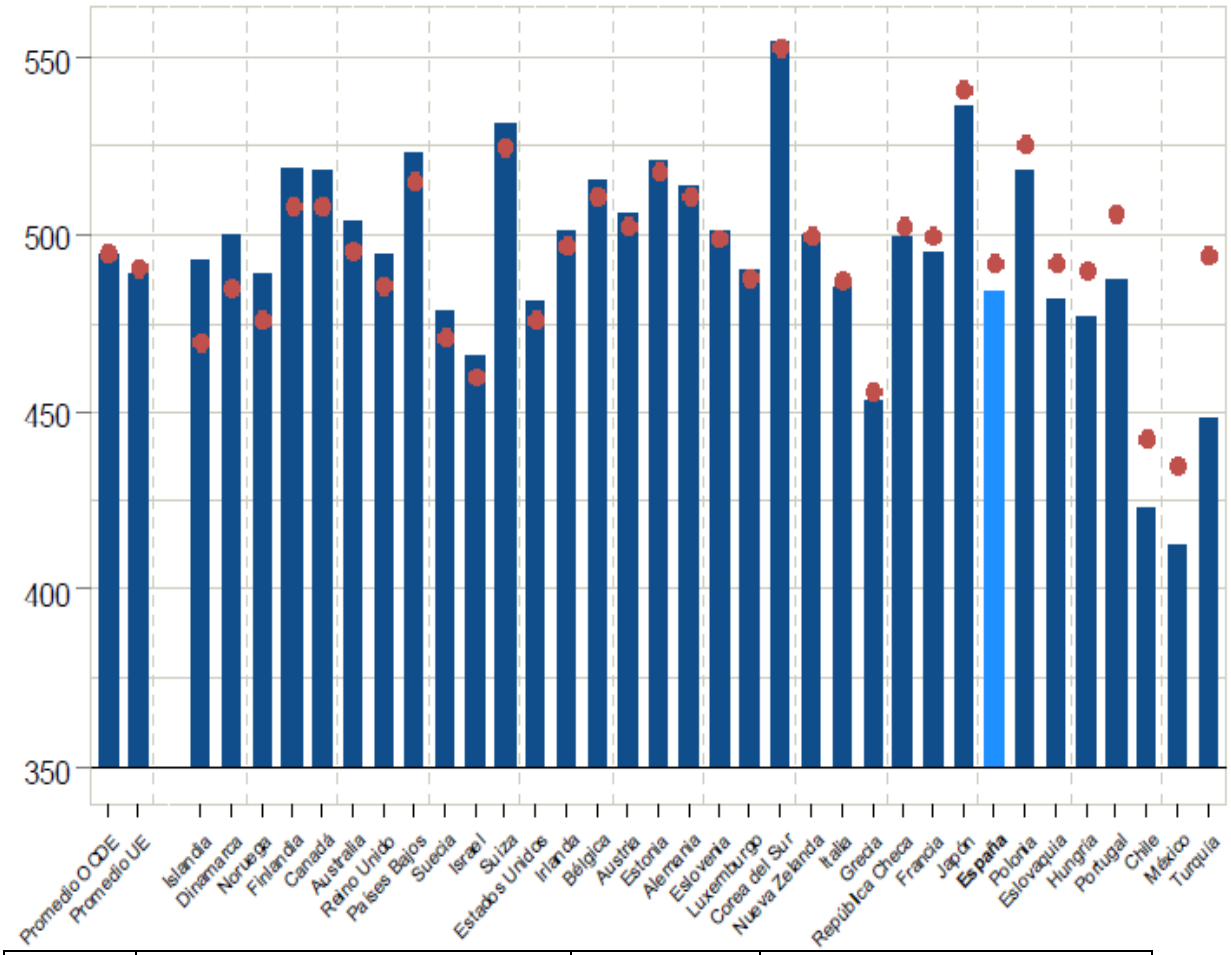
ÀMBITS DE PSICOPEDAGOGIA I ORIENTACIÓ
ÁMBITOS DE PSICOPEDAGOGIA Y ORIENTACION
FIELDS OF EDUCATIONAL PSYCHOLOGY AND COUNSELLING
Nº 43 (2a.època) novembre 2015
URL: www.ambitsaaf.cat
ISSN: 2339-7454

postpostivista**[9]** s'han elaborat diferents estratègies per controlar l'efecte d'altres possibles variables externes, el que normalment "requereix d'una combinació de disseny experimental, estadística inferencial, observació empírica i una teoria substantiva" (Cook, 2004: 88). La tendència actual és resumir les diferents estratègies en dos grans grups, estratègies de comparació recolzades en l'atzar (dissenys experimentals i quasiexperimentals) i modelització estadística.

En el cas de l'estudi PISA per poder tenir en compte aquest factor s'ha elaborat un Índex Social, Econòmic i Cultural (ESCS en les seves sigles en anglès) que reflecteix l'ocupació professional i el nivell educatiu dels pares i mares, així com els recursos disponibles a la llar a través, per exemple, del nombre de llibres a casa. De l'anàlisi d'aquest índex es poden obtenir diverses conclusions:

- Els països o territoris tenen diferents estatus socioeconòmics.
- L'estatus té altes correlacions amb els resultats acadèmics

Si es controla estadísticament l'ESCS resulta interessant observar les puntuacions que tindrien els participants en PISA si el valor dels seus ESCS equivalgués al nivell mitjà de l'OCDE, és a dir, a zero (PISA, 2014: 99), que és una dels mecanismes per detreure l'efecte de les variables que hi intervenen.



	Puntuació Mitjana en Matemàtiques sense ESCS		Puntuación Mitjana en matemàtiques

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

(Font: PISA, 2014: 100)
(**Gràfic 2.** Puntuacions mitjanes en matemàtiques dels països de l'OCDE, descomptant ESCS.)

Quan les barres blaus son en una posició més baixa que els punts vermells indica que els bons resultats obtinguts per alguns països es deuen en part al seu alt índex de benestar sociocultural. Per contra, en països com Turquia, Xile, Japó, Polònia o Espanya, si es té en compte el context social, econòmic i cultural s'observa una millora dels resultats dels alumnes de 15 anys (PISA, 2014: 99).

Com es pot veure en el gràfic 2, si controlem les dades dels resultats obtinguts poden ser molt distants de la informació aparent. Per exemple, en el cas d'Espanya el rendiment dels joves en matemàtiques és millor que a Islàndia, Dinamarca, Noruega, Regne Unit, Suècia o els Estats Units entre d'altres països, una imatge bastant més positiva del que l'imaginari col·lectiu ens ofereix.

Tenint en compte que en moltes ocasions els tests serveixen de base per prendre decisions sobre els grups, centres o sistemes educatius i fins i tot s'arriben a utilitzar com a referència per determinar el salari del professorat, les estimacions han de ser rigoroses i vàlides.

Si no es controla la influència de les variables externes s'emeten judicis injustos sobre els sistemes educatius. A més de ser una mala praxi avaluadora, el fet de distribuir informes d'avaluació esbiaixats, tal com diu Tyler (1991: 16), "és realment un crim de coll blanc".

4. S'avaluen només els resultats i no la intervenció educativa

Els tests estandarditzats mesuren els resultats finals en els estudiants. Amb aquesta informació, tal com he citat, es jutja la totalitat del centre o d'una determinada política educativa. S'assumeix el supòsit que si hi ha hagut bons resultats, és perquè la formació que s'ha donat ha estat bona o si els resultats han estat dolents la causa es que la proposta docent no ha estat adequada.

Però potser darrere d'uns mals resultats pot haver-hi bons docents, amb una bona pràctica, davant d'un grup amb dificultats d'aprenentatge, amb una ràtio elevada o en contextos socials poc motivadors. O al contrari, darrere d'uns bons resultats pot haver-hi una pèssima docència, amb professors poc competents però amb famílies que supleixen els dèficits amb professorat de suport extraescolar. També ja he exposat l'associació entre estatus socioeconòmic i èxits acadèmics, un altre exemple de com els resultats no donen necessàriament informació sobre la bondat de la docència o el currículum. En definitiva, són massa llargues les cadenes causals per avaluar la qualitat d'un centre només amb els resultats sumatius finals.

A causa d'aquestes limitacions Stufflebeam (2001) defineix les avaluacions de resultats com quasi-avaluacions, ja que emeten una valoració d'un tot (grup, centre escolar, política educativa, ...), tenint en compte només una part: els resultats. També se'ls ha denominat avaluacions de caixa negra (Chen, 1990; Weiss, 1998) per no tenir en compte els processos de treball (docència, coordinació, programació, tutories, resolució de conflicte, actualització del currículum, ...) que són els factors que realment produeixen canvis en l'alumnat. El que es vol ressaltar amb la metàfora de caixa negra és que no es sap res sobre com es fan les coses i,

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

per tant, no es poden establir relacions lògiques entre un determinat tipus de docència i els èxits obtinguts.

Les propostes alternatives busquen obrir, analitzar i avaluar el que hi ha dins el programa. Ampliar la perspectiva de resultats amb l'anàlisi del context, dels mètodes de treball i els mitjans amb què es compta (Tyler, 1991). La vinculació entre processos i resultats és probablement una de les contribucions més importants a la teoria de l'avaluació (Greene, 1999; Shadish, Cook, and Leviton, 1991) ja que busca l'articulació causal entre el que es fa, per exemple les classes i el que s'aconsegueix, els resultats que obtenen els estudiants.

Per més que es vulgui emfatitzar l'orientació dels serveis públics als resultats finals, no es poden invisibilitzar els mecanismes que els produeixen, entre d'altres raons perquè perdem el coneixement de com es fan les coses, de com s'aconsegueixen els èxits en el camp educatiu.

Cronbach (2000), entre molts d'altres, defensa que en la mesura del possible l'avaluació s'hauria d'utilitzar per comprendre com el curs produeix els seus efectes i quines variables influeixen en l'eficàcia (els resultats). D'aquesta manera l'avaluació no és només emetre judicis sinó també comprendre els mecanismes causals. Comprendre com s'ha generat l'aprenentatge educatiu és el primer pas per millorar-lo.

En resum, els tests jutgen les escoles però no les doten d'informació per comprendre els fracassos ni els èxits. No són sistemes útils per a la millora dels centres, senzillament perquè no ofereixen informació sobre el que cal millorar. Realment els tests estan orientats amb una finalitat sumativa. Tal com explica Weiss (1998: 32) l'avaluació formativa ajuda a desenvolupar el programa i la sumativa a emetre judicis sobre ell.

Normalment els polítics, la direcció, les professionals i els participants tenen diferents propòsits envers l'avaluació (Greene, 2007; Patton, 2008). Les directives, professores i, freqüentment, els participants demanen informació per entendre perquè passen les coses i informació per buscar solucions o millorar. Per la seva banda, els polítics o altres actors amb capacitat de decisió solen requerir a l'avaluació una informació per emetre judicis finals sobre el programa, que els serveixin per retre comptes públics i per prendre grans decisions.

Per tant, els tests estandarditzats responen als interessos dels decisors polítics més que als de les professionals, direcció, alumnat i famílies. Per molt que es demani, "l'avaluació per a tots els propòsits és un mite" (Weiss, 1998: 3).

5. Valoració d'estudiants no avaluacions de centres o de sistemes educatius

En l'àmbit educatiu alguns autors han establert una diferenciació terminològica més o menys acceptada en funció del tipus d'objecte que s'analitza (estudiants, professionals i programes) (Schwandt, 2009: 19). L'exercici de valoració dels èxits dels estudiants és definit com *assessment***[10]**, la valoració del desenvolupament docent com *appraisal***[11]** i quan el focus està posat en els programes, centres o serveis s'anomena avaluació.

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

Nevo (2006: 447) descriu com el *testing* es va transformar d'alguna manera en una paraula bruta i en el context d'avaluació d'estudiants es va començar a usar el terme alternatiu *assessment*. Tal com es recull en l'Enciclopèdia de l'Avaluació, *assessment* és l'opció per descriure la valoració de la qualitat del treball dels estudiants per tal de determinar el nivell d'assoliment que han assolit (Mabry, 2005).

Inicialment els tests estandarditzats estaven orientats a l'emissió de judicis sobre les persones, amb la intenció de seleccionar estudiants per una formació avançada, classificar o diagnosticar competències (Cronbach, 2000), encara que amb el temps, els tests "gradualment van canviar la mesura dels resultats per altres objectes com programes, col·legis, professorat i sistemes educatius" (Nevo, 2009: 292). Aquesta translació de l'objecte l'han definit Guba i Lincoln (1989) com una deficiència seriosa creada en les primeres generacions de mètodes avaluatius**[12]**.

Avaluació i assessment són diferents termes que impliquen diferents mirades sobre diferents objectes. Recapitulant els elements que han sortit en l'exposició de les crítiques, les dues aproximacions es diferencien en els següents punts:

Criteris	Test, *assessment*	Avaluació de programes
Objecte d'anàlisi	El seu focus es l'alumne.	El focus és la instància educativa: programa, aula, curs centre, sistema, política ...
Dimensions a tenir en compte sobre l'objecte analitzat	Resultats	Resultats, processos (activitats, implementació, docència) i elements estructurals.
Tipus de resultats a avaluar	Tendeix a mirar coneixements de forma estandarditzada.	Mira tots els possibles canvis en les persones provocats per acció educativa (amb diferents nivells d'abstracció). L'èmfasi en el fet que siguin deguts a la intervenció ressalta la importància atorgada als mecanismes metodològics per a poder parlar d'atribució o contribució del programa en els resultats.
Usuaris principals	Principalment respon a les necessitats de polítics i d'altres decisors polítics.	Pot tenir en compte les necessitats d'una àmplia varietat d'actors, inclosos direcció, professorat, alumnat i famílies.
Propòsits	Rendiment de comptes públic (*accountability)* i ajut en la gran presa de decisions del tipus continuar o no, expandir-se, modificar, ...	A més del rendiment de comptes, pot contribuir a la comprensió de la unitat avaluada, ficant-se en la caixa negra del programa, i per tant s'orienta cap a la millora.

El problema no es la confusió de termes, sinó la confusió de mètodes. S'apliquen les lògiques de l'examen a estudiants per extreure conclusions sobre tot el sistema educatiu, amb totes les conseqüències negatives que això comporta: homogeneïtzació i imposició d'objectius, mirada limitada, judicis injustos sobre els centres, descapitalització de coneixements sobre la pràctica docent i desorientació de tot el sistema amb relació al propòsit de millora.

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

Nº 43 (2a.època) novembre 2015
URL: www.ambitsaaf.cat
ISSN: 2339-7454

CONCLUSIÓ

Els tests estandarditzats són una eina molt versàtil per a examinar determinats coneixements. La seva facilitat d'aplicació, el reconeixement institucional que tenen i el poder de la comparació entre escoles o sistemes, són aspectes que resulten molt atractius.

L'emergència que estan tenint els test pot fer pensar que s'està davant d'un descobriment tècnic recent, que ens mostra una manera més eficient i objectiva d'avaluar l'acció pedagògica. Però els test, tal qual els estem veient, són una de les aproximacions més antigues en la valoració d'estudiants.

Per la seva banda, les crítiques als tests han plantejat reptes metodològics que han acabat generant alternatives i han contribuït de manera substancial a augmentar el cos teòric de l'avaluació dels programes educatius.

No obstant això, és freqüent trobar una confusió de termes i mètodes entre *assessment* i avaluació. Això pot fer pensar que els tests són vàlids per a qualsevol propòsit avaluatiu, però això no és així i, tal com s'ha vist a les crítiques anteriors, pot ocasionar greus conseqüències negatives al sistema escolar. Utilitzant com a referència el cas espanyol, és encara més escandalós el gran esforç polític i econòmic que s'està fent per estendre els tests a tots els centres i aules, en contrast amb les retallades pressupostàries que s'estan produint en el mateix moment en l'àmbit educatiu.

És cert, que el sistema educatiu ha de ser avaluat i revisat, però el mateix cal fer amb els models d'avaluació ja que també són recursos públics. La proposta de test estandarditzats ha de ser analitzada per saber si genera el valor social que se li pressuposa. Sota el meu punt de vista n'hi ha prou amb fer-nos dues preguntes clau: Què es vol què aconseguir amb els tests? I per a què estan servint en realitat i en concret?

Afortunadament podem trobar molts mètodes i aproximacions d'avaluació que s'adapten als diferents propòsits i usos, els metodòlegs i els teòrics han fet la seva feina. Ara, si nosaltres fem servir una forquilla per prendre la sopa ja només és responsabilitat nostra.

Notes:

[1] He procurat utilitzar un llenguatge inclusiu des d'una perspectiva de gènere. Quan utilitzat les terceres persones del singular o del plural he optat per la convenció de referir-me a tot el professorat i direcció com professores i directores (femení) i la resta dels actors (estudiants, autors, ...) en masculí, encara que hi hagi homes i dones en tots els grups esmentats.

[2] L'avaluació per objectius es el "el procés de determinar en quina mesura els objectius educatius efectivament s'acompleixen" (Nevo, 2006:442).

[3] Avaluació formativa: Es tracta d'avaluacions dissenyades, fetes i destinades a donar suport als processos de millora, normalment encarregats o realitzats per algú i lliurats a algú que pugui dur a terme les millores (Scriven, 1991: 19).

[4] Donada la dificultat de trobar un terme adequat en castellà, l'autor opta per mantenir la veu anglesa d'*assessment.* Usualment es troba traduït com a avaluació encara que probablement l'expressió més propera al seu significat tècnic sigui "examen d'estudiants".

[5] La característica central del model està definida per l'acrònim CIPP que representa l'avaluació de Productes, Processos, Inputs i Context de qualsevol entitat (Stufflebeam, 2005).

[6] Desenvolupament no cognitiu: Concepte introduït per Bowles & Gintis que fa referències a qüestions com les actituds, la motivació i el judici entre d'altres aspectes personals (Morrison i Schoon, 2013).

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

Nº 43 (2a.època) novembre 2015
URL: www.ambitsaaf.cat
ISSN: 2339-7454

[7] En avaluació i qualitat és usual trobar el terme *accountability* en anglès per la seva difícil traducció al castellà. Es pot entendre com a rendiment de comptes públic (encara que no fa esment exclusivament a la comptabilitat) o responsabilització (Echebarria, 2005).
[8] Als EUA en aquella època hi havia moltes escoles i instituts inspirats en el moviment de renovació pedagògica conegut com educació progressiva que treballaven respectant la diversitat de l'alumnat i generant diferents processos pedagògics.
[9] Existeixen altres aproximacions de caràcter constructivista o d'altres sota el denominat paradigma transformador.
[10] Com he comentat en una cita anterior, mantinc la veu anglesa de *assessment* ja que no es troba un paral·lelisme similar en castellà als termes *"evaluation" - "assessment".*
[11] Igual que amb *assessment* he optat per mantenir el terme de *appraisal* en anglès a l'espera d'un consens sobre una traducció tècnica al català.
[12] Per avaluar un programa educatiu és cert que cal avaluar-ne els resultats però articulant aquestes peces d'informació dins d'un esquema interpretatiu més ampli que ens permeti comprendre i valorar d'una manera global la intervenció

Referències Bibliogràfiques:

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. A 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*. Recuperat a : 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*. Recuperat a : 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.

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

Nº 43 (2a.època) novembre 2015
URL: www.ambitsaaf.cat
ISSN: 2339-7454

Morrison, L. & Schoon, I. (2013) *The Impact of Non-Cognitive Skills on Outcomes for Young People*. Literature review. London: Institute Of Education. Recuperat a: 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.

Nevo, D. (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.

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.

Correspondència amb l'autor: *Juan Andrés Ligero Lasa*. (www.magisterevaluacion.es). E-mail: jligero@uc3m.es.