

Consultor i analista de polítiques d'educació

Norberto Bottani

La més bonica del regne: el món de l'educació en alerta davant l'arribada d'un príncep encantador*

Aquest article intenta oferir un panorama crític de la història de les avaluacions internacionals a gran escala, que comença als anys cinquanta del segle passat. L'IEA (Associació Internacional per a l'Avaluació del Rendiment Educatiu) en va ser l'organisme pioner, i les matemàtiques, la primera àrea investigada. El 1958 va aparèixer el primer esborrany d'un projecte d'avaluació comparada del rendiment dels alumnes en diferents sistemes d'ensenyament. L'objectiu principal d'aquest tipus de recerca és arribar a identificar els factors que poden explicar el rendiment dels sistemes educatius per poder contestar a la pregunta crucial de si hi ha un sistema d'ensenyament millor que d'altres i com es pot definir per poder aprendre dels millors. L'experiència en investigació acumulada per l'IEA va demostrar, per primer cop, que es podien mesurar empíricament els resultats de l'educació i aprofitar-ne les conclusions. El principal assoliment d'aquesta associació va ser posar en marxa instruments que feien factible la comparació de dades sobre el rendiment dels alumnes en diferents sistemes d'ensenyament. Fruit d'això, entorn de l'IEA es va crear una comunitat científica internacional especialitzada en aquest camp. El seu monopoli sobre l'avaluació del rendiment escolar va finalitzar amb la intervenció i el desenvolupament, especialment als Estats Units, d'altres associacions d'avaluació que, a mitjà termini, van reduir

* Adaptació d'un text elaborat per a l'Haut Conseil de l'évaluation de l'école. A BOTTANI, N.; VRIGNAUD, P. *La France et les évaluations internationales*. HCcéé, 16 (gener 2005), p. 168. Publicat en castellà a: *Revista de Educación* (2006). Traduït i publicat a VIA amb permís de *Revista de Educación*.

l'autonomia científica de l'IEA. A partir de l'entrada en escena de l'OCDE el 1993, les polítiques internacionals d'avaluació van prendre un nou rumb. Les característiques del projecte PISA –que té una coherència diacrònica, es concentra en tres àmbits i avalua nivells de competència i no de coneixement curricular concret– fan de les seves conclusions una poderosa arma d'influència sobre les polítiques educatives de molts països.

Les avaluacions internacionals a gran escala: una història de cinquanta anys

Fa aproximadament quaranta anys, el 1965, l'IEA (Associació Internacional per a l'Avaluació del Rendiment Educatiu) acabava la seva primera gran enquesta internacional sobre els coneixements en matemàtiques dels alumnes, enquesta que més tard seria anomenada *Enquesta FIMS* (First International Mathematic Study). Aquesta primera prova s'havia preparat en trobades científiques organitzades al si de la UNESCO al llarg del decenni precedent: des del 1952, a l'Institut Internacional d'Educació d'Hamburg es reunien un cop l'any els directors dels instituts d'investigació en educació més importants d'aleshores. Els assumptes tractats en aquestes trobades es referien a diversos aspectes de l'avaluació dels programes d'ensenyament, com ara l'elaboració d'exàmens o la creació de mètodes i instruments per mesurar en quina proporció s'assolien els objectius de l'ensenyament i dels programes escolars. El 1958, els participants a la reunió van ser convidats a debatre la possibilitat d'organitzar avaluacions comparades dels rendiments escolars obtinguts en diferents sistemes d'ensenyament. Posteriorment, Benjamin S. Bloom, de la Universitat de Chicago, va elaborar un primer esborrany de projecte d'avaluació comparada del rendiment escolar en diferents sistemes d'ensenyament.

En aquella època, ningú no tenia experiència en aquest tipus d'investigació i ningú no podia dir amb antelació si aquest tipus d'estudis era viable. El repte científic que aquest projecte suposava no era menor, ja que es tractava d'identificar els factors que podien explicar el rendiment dels sistemes d'ensenyament a fi de poder contestar una pregunta crucial: hi ha un sistema d'ensenyament millor que d'altres? Qui seria aleshores la noia més bonica del regne? Aquesta pregunta un xic iconoclasta no va ser formulada per dirigents polítics sinó per científics; va ser al si de

la comunitat d'investigació on es va adoptar el compromís d'estudiar una bateria de proves amb l'objectiu d'arribar a designar el millor sistema d'ensenyament, aquell que proposés solucions susceptibles d'obtenir les prestacions de conjunt més satisfactòries. Els decisors polítics van entrar en escena molt més tard. Podem preguntar-nos, a propòsit d'això, si els científics reunits entorn d'aquest projecte van actuar d'una manera totalment independent o si van anticipar un desig no formulat dels decisors.

L'avaluació dels sistemes d'ensenyament amb l'ajuda de grans enquestes de masses sobre competències i coneixements dels alumnes ha estat, per tant, concebuda al llarg dels anys cinquanta. Des d'aleshores, durant uns cinquanta anys, s'ha acumulat una àmplia experiència en estadística i metodologia en aquest camp. L'avaluació d'alumnes a gran escala en l'àmbit internacional s'ha convertit en una especialització en si mateixa, un camp científic completament estructurat, amb les seves regles de funcionament, legitimació, control i reconeixement.

La realització d'enquestes sobre els aprenentatges dels alumnes escolaritzats en els diferents sistemes d'ensenyament ha modificat de manera substancial els estudis comparats tradicionals en obrir la porta a avenços fonamentats sobre metodologies empíriques que mesuren els aprenentatges per jutjar l'eficàcia dels sistemes d'ensenyament. Durant dècades, en el millor dels casos, els estudis de comparació es basaven en la confrontació de dades simples, com ara el percentatge de les taxes de participació en els diferents nivells d'instrucció, les taxes d'escolarització més enllà de l'ensenyament obligatori o la prolongació de la durada de l'escolaritat obligatòria. Així doncs, no es tenien en compte els coneixements que adquirien els alumnes, ja que en general s'estimava que era molt difícil, si no impossible, mesurar els resultats de l'ensenyament amb mesures empíriques. El projecte científic endegat pels pioners de l'IEA va permetre desmentir en part aquesta creença i mostrar que es podia fer alguna cosa per sortir de l'estancament de les autoavaluacions o avaluacions subjectives de naturalesa holística o impressionista.

L'avaluació d'alumnes a gran escala en l'àmbit internacional s'ha convertit en una especialització en si mateixa.

El primer estudi pilot a gran escala de comparació del rendiment dels alumnes

El primer estudi pilot per provar la possibilitat de realitzar avaluacions a gran escala del rendiment dels alumnes en els diferents sistemes d'ensenyament es va dur a terme entre el 1959 i el 1961. Hi van participar dotze països amb una mostra no representativa de 1.000 alumnes que havien complert els tretze anys al principi de l'any escolar, independentment del curs que feien. Per tant, 12.000 alumnes van ser objecte d'aquest test, la qual cosa en aquell temps era una població considerable, tenint en compte

El primer estudi pilot per provar la possibilitat de realitzar avaluacions a gran escala del rendiment dels alumnes en els diferents sistemes d'ensenyament es va dur a terme entre el 1959 i el 1961.

els mitjans tècnics existents per al tractament de dades.¹ D'entrada es van presentar dos problemes: els criteris de definició de la població en qüestió (les edats o el grau escolar) i l'elaboració d'una mostra representativa. En aquest primer exercici, els àmbits que van ser objecte d'estudi van ser la comprensió lectora, les matemàtiques, les ciències i la geografia. D'altra banda, a Anglaterra es va preparar un test no verbal, que va ser igualment realitzat per tots els alumnes. En la realització de la prova no es va imposar cap limitació de temps: els alumnes es podien prendre tot el temps necessari per contestar les preguntes.

El test anava acompanyat d'un breu qüestionari adreçat als estudiants i als directors dels centres. Les dades recollides el 1961 van ser tractades i analitzades per Bob Thorndike, de la Universitat de Colúmbia de Nova York, i els resultats van ser publicats el 1962.

L'objectiu principal d'aquest estudi pilot era provar la possibilitat de construir instruments que es poguessin utilitzar de manera uniforme en diferents sistemes d'ensenyament (com ara el suec o el japonès), a fi de recollir dades que fossin comparables i susceptibles de ser tractades de manera homogènia. En aquest primer estudi pilot es van provar i aplicar molts elements que posteriorment es tornarien a trobar de manera habitual en totes les enquestes internacionals.

Els resultats de l'estudi pilot van ser tan satisfactoris que el grup de promotors va decidir programar posteriorment un estudi real a gran escala. L'estudi pilot havia permès ressaltar i mostrar els punts febles i les llacunes del projecte. Es podia esperar, doncs, que corregint els defectes

de concepció de l'estudi, sobretot pel que fa a la construcció de mostres representatives de la població d'alumnes,² fos possible engegar amb probabilitats d'èxit una operació ambiciosa, mai no intentada anteriorment, que obtingués el beneplàcit de la comunitat científica i, sobretot, el suport dels decisors polítics, sense el qual no hauria estat possible trobar els recursos econòmics per dur a terme aquests projectes.

El naixement de l'IEA i la creació d'un monopoli d'investigacions sobre les avaluacions dels sistemes d'ensenyament

En concloure aquest estudi pilot, havia quedat clar que només podia dur a terme operacions semblants una organització especialitzada en la realització d'aquest tipus de projectes de gran envergadura que aplegués les experteses necessàries per coordinar la participació de molts instituts d'investigació, planificar treballs, assegurar-ne el finançament i la gestió, coordinar el calendari de les proves, supervisar la comparació de les etapes, verificar la uniformitat de les eines, el tractament de dades i càlculs, i assegurar l'anàlisi coherent dels resultats. Així, el 1961 es va constituir una associació per a l'avaluació del rendiment dels alumnes, l'IEA (Associació Internacional per a l'Avaluació del Rendiment Educatiu), la seu de la qual es va establir prop de l'Institut Internacional d'Educació d'Hamburg. El primer president de l'IEA va ser William Douglas Wall, que era també el president de l'organisme britànic NFER (National Foundation for Educational Research). Després de tan sols un any d'activitat, Wall va dimitir i va ser reemplaçat pel suec Torsten Husén, que va ser al càrrec de l'IEA durant més de quinze anys, fins al final del 1978.

El principal mèrit científic de l'IEA va ser posar en marxa instruments que feien possible la comparació de dades sobre el rendiment d'alumnes subjectes a diferents sistemes d'ensenyament. Al voltant de l'IEA es va constituir en aquells anys una comunitat científica internacional especialitzada en l'avaluació a gran escala dels coneixements dels alumnes.³ Els estudis de l'IEA van ser objecte de moltes crítiques, que els atribuïen sobretot objectius que no tenien, com ara el d'establir explicacions causals

El 1961 es va constituir una associació per a l'avaluació del rendiment dels alumnes, l'IEA (Associació Internacional per a l'Avaluació del Rendiment Educatiu).

de les diferències entre sistemes d'ensenyament. Efectivament, hi ha una certa ambigüitat en les posicions dels pioners de l'IEA sobre aquest propòsit, i no es pot descartar que un objectiu latent d'aquest tipus hagués inspirat una part dels pioners, i que alguns d'ells cultivessin un disseny polític en un sentit progressista que comptés a recollir proves eloqüents a favor de reformes en l'ensenyament. Tanmateix, l'aposta principal en el moment inicial era una altra, i consistia a desenvolupar proves sobre el rendiment dels alumnes, qüestionaris sobre l'entorn familiar i les seves estratègies d'aprenentatge, i instruments per entendre la proporció de currículums realment impartits, així com el temps real durant el qual els alumnes a classe es trobaven en condicions d'aprenentatge i ensenyament. Des d'aquest punt de vista, l'IEA ha guanyat segurament la partida: els seus estudis han estimulat la investigació de sistemes d'ensenyament i han fet possible la realització d'investigacions a escala mundial que han donat lloc a noves representacions dels sistemes d'ensenyament i n'han modificat la comprensió de l'organització i el funcionament.

La fi del monopoli de l'IEA

L'IEA ha estat durant uns trenta anys l'única organització mundial especialitzada en l'àmbit de la realització d'enquestes internacionals de masses sobre el rendiment dels alumnes. Aquest monopoli es va trencar el 1988 quan l'Administració federal nord-americana, després d'haver finançat l'IEA durant dècades, va decidir encarregar a l'Educational Testing Service de Princeton (ETS) la realització d'una enquesta internacional sobre el rendiment dels alumnes que feia competència a la de l'IEA. Aquesta enquesta, anomenada *IAEP* (International Assessment of Educational Progress), es va centrar en dues assignatures, les matemàtiques i les ciències. Amb aquesta enquesta, repetida el 1991, l'ETS es proposava demostrar que era possible sortir dels camins imposats per l'IEA, evitar els estancaments que alentien considerablement les enquestes de l'IEA i concebre un tipus d'enquesta a escala internacional més ràpida que l'elaborada per l'IEA.

Les dues enquestes IAEP van ser molt criticades per la comunitat científica, que aleshores estava majoritàriament composta per especialistes formats al si de l'IEA. La virulència de la crítica va portar l'Administració

federal nord-americana a renunciar al seu projecte i aturar el finançament del programa IAEP. La polèmica entorn d'aquestes dues enquestes es basava concretament en el fet que els ítems de les proves eren en bona part una transposició a l'àmbit internacional dels ítems de les proves del NAEP (National Assessment of Educational Progress),⁴ és a dir, el test utilitzat als Estats Units i realitzat per l'ETS per avaluar anualment el rendiment d'una mostra d'alumnes nord-americans. Per tant, l'IAEP mesurava les competències i el rendiment dels alumnes d'altres sistemes d'ensenyament prenent com a referència criteris nord-americans, cosa que permetia evitar llargues discussions sobre l'adequació de les proves als currículums nacionals.

La demostració de força que va intentar fer l'Administració federal amb l'IAEP no va tenir èxit. Els Estats Units es van veure obligats a fer marxa enrere i renunciar al seu projecte, però l'operació va aconseguir, tanmateix, afeblir l'IEA i llançar-la als braços dels Estats Units. Es pot dir que la fi del programa IAEP va ser una «victòria pírrica» de l'IEA, la qual, traient partit de la lliçó d'aquest atac, va haver d'accelerar el ritme de les enquestes sense tenir els mitjans per fer-ho. Podia trobar-los als Estats Units, però per rebre'ls s'havia de sotmetre a les exigències de la potència dominant. Durant els anys noranta, l'IEA va programar i realitzar tres grans enquestes: una sobre la comprensió lectora el 1990-1991; una sobre matemàtiques i ciències el 1994-1995, que ha estat des d'aleshores la major enquesta realitzada a escala internacional, amb la participació d'uns quaranta sistemes d'ensenyament diferents, i que va ser repetida el 1997-1998, i, finalment, una enquesta sobre educació cívica el 1999.⁵ Això era insòlit en la història de l'IEA, si es té en compte que en el curs del decenni 1960-1970 l'IEA va realitzar una sola enquesta i que en el període 1980-1992 en va fer dues. Tanmateix, en un sol decenni, el que va del 1990 al 2000, l'IEA, en situació de crisi, va dur a terme quatre enquestes de gran envergadura. Era l'efecte d'una cura de cavall propiciada pels Estats Units. L'Administració federal nord-americana havia perdut la batalla amb l'IAEP, però havia aconseguit tornar a agafar les regnes i perseguir els seus interessos d'una manera molt elegant.

L'entrada en escena de l'OCDE

El cessament de les enquestes IAEP no va ser, per tant, una victòria per a l'IEA, que es va veure obligada a modificar les seves pràctiques i adaptar-se a les exigències del seu «arrendatari» principal, tal com acabem de veure. L'estratègia dels Estats Units va aconseguir afeblir l'IEA, la qual, en conseqüència, va perdre la seva autonomia científica i, sobretot, va quedar englobada *coram populi* en l'esfera d'influència dels Estats Units. De fet, es va obrir la porta a un tercer actor, l'OCDE, que va entrar en acció el 1993, quan l'organització es va adonar que mai no obtindria la informació que necessitava sobre els aprenentatges dels alumnes per completar el seu conjunt d'indicadors. L'antiga IEA no ho havia aconseguit i la nova IEA estava massa sotmesa al control dels Estats Units per obtenir la confiança de tots els països. El 1997, els ministeris d'Educació dels països de l'OCDE van decidir posar en marxa un cicle d'enquestes autònom i d'un altre tipus sobre els coneixements dels alumnes. En la primavera del 2000, una trentena de països van participar en el primer estudi del Programa internacional d'avaluació d'estudiants (PISA, en les seves sigles en anglès).

L'OCDE va obtenir, per tant, el que l'IEA mai no havia aconseguit en quaranta anys, és a dir, no solament atraure l'atenció dels decisors polítics sobre aquestes enquestes, sinó també reorientar les polítiques d'ensenyament en molts països, ja que PISA va esdevenir un referent per justificar tot tipus de decisions o reformes (OCDE, 2004).

Les novetats de PISA

Seria fals afirmar que no hi ha solució de continuïtat entre PISA i els estudis anteriors de l'IEA. PISA, tot i que reprèn molts coneixements metodològics de l'IEA, representa una ruptura amb les enquestes anteriors. Les principals novetats de PISA, que diferencien els seus estudis dels de l'IEA, són les següents.

La periodicitat triennial de les proves

La primera novetat de PISA va consistir a programar un cicle d'enquestes amb una periodicitat fixada amb l'objectiu de construir sèries històriques de dades comparables sobre les competències dels alumnes. PISA va adoptar una periodicitat triennial. Les raons d'aquesta elecció no són clares,

però es pot suposar que l'objectiu de l'OCDE era demostrar, no només que era possible organitzar enquestes a gran escala a intervals propers, com ja ho havia intentat fer abans l'ETS, sinó que tenia també la capacitat de mantenir aquest ritme durant un llarg període, com ho és un període decennal.⁶ Una periodicitat de tres anys posa sota pressió els sistemes d'ensenyament, i fa que les administracions, els directors dels centres i els instituts de recerca encarregats de fer les enquestes no se'n desentenguin. Tothom ha d'adaptar-s'hi, tot i que la quantitat de dades produïdes cada tres anys no sigui explotada; això tant se val mentre hi hagi els resultats que els decisors polítics desitgen. Sense l'acord, el suport o la neutralitat benèvola dels decisors polítics, que d'altra banda paguen els costos dels estudis, l'OCDE no hauria pogut fer el seu programa. Un altre factor important per a l'èxit és el suport d'una part de la comunitat científica, la qual, lluny de ser reticent a un ritme desenfrenat de recollida i producció de dades, accepta la situació i en treu un benefici, no només científic sinó també econòmic, participant en aquesta operació. Si hi ha reserves o reticències, aquestes no són visibles o no apareixen en els informes.⁷

La primera novetat de PISA va consistir a programar un cicle d'enquestes amb una periodicitat fixada amb l'objectiu de construir sèries històriques de dades comparables sobre les competències dels alumnes.

La coherència diacrònica de les proves

El principi de la repetició d'un estudi en el mateix camp disciplinari no és en si mateix una novetat. L'IEA havia dut a terme un programa d'estudis successius en les àrees de les matemàtiques i la lectura, ja que l'interès de fer comparacions diacròniques dels resultats obtinguts per un mateix grup d'edat dins un mateix sistema d'ensenyament (però en diferents dates) és evident per a tothom. Tanmateix, amb PISA és la primera vegada que es fa un programa que fixa d'entrada la periodicitat de les proves. Per tant, els beneficis de la participació en PISA només podran sumar-se si un sistema d'ensenyament s'hi compromet al llarg de dos cicles d'enquestes, és a dir, durant un període de dotze anys. Per exemple, segons el pla PISA, només després de nou anys s'avaluaran de nou, amb la mateixa profunditat analítica, les competències en comprensió lectora, àmbit avaluat en la primavera del 2000 i que no serà matèria principal fins al test del 2009.

Aquesta estratègia permet l'OCDE assegurar-se el compromís dels països (els clients!) durant un llarg període de temps, obligant-los a participar en l'enquesta i a pagar aquesta participació durant dotze anys per treure'n, en contrapartida, el ple benefici de la participació en el programa. Per descomptat, és possible retirar-se del programa, però en aquest cas es perd l'avantatge que ofereix la possibilitat de realitzar comparacions a llarg termini.

Concentració en tres àmbits

Els tres àmbits que té en compte l'OCDE són la comprensió lectora, les matemàtiques i les ciències.

La tercera novetat del programa PISA és l'elecció de tres àmbits que se sotmeten a examen de manera simultània, tot i que amb un ordre d'importància diferent. Els tres àmbits que té en compte l'OCDE són la comprensió lectora, les matemàtiques i les ciències. Aquests tres àmbits són objecte de les proves cada vegada, però el lloc reservat en el test per a cadascun d'ells és diferent –en el programa PISA es parla d'*àmbits principals* i *àmbits secundaris*. En cada enquesta s'avalua en detall una d'aquestes matèries, que representa més o menys dos terços del temps de la durada total de les proves. D'altra banda, com ja es feia també a les enquestes de l'IEA, els alumnes disposen de vint minuts per respondre un qüestionari contextual sobre la seva motivació, el seu entorn familiar i cultural, les seves companyies i els seus interessos. El creuament de les respostes del qüestionari amb els resultats obtinguts en el test permet examinar els efectes de les diferents estratègies d'aprenentatge dels alumnes segons el seu perfil social, cultural i econòmic.

Una població estable i fàcilment identificable: els alumnes de quinze anys

Una altra originalitat del programa PISA consisteix en la composició de la població avaluada, els joves de quinze anys, independentment del curs que facin. Això significa que en la mostra representativa de la població d'alumnes de quinze anys hi ha alumnes de cursos escolars diferents. PISA, doncs, no verifica el que s'ha après en un curs escolar, sinó el nivell cultural dels joves de quinze anys al país que és objecte de l'enquesta. Als països de l'OCDE la majoria dels joves de quinze anys van a escola i per

això és relativament fàcil trobar-los quan s'elabora una mostra representativa de la població escolar. Una enquesta sobre els joves de quinze anys és menys cara: se sap on són els joves i es pot fer el test a l'escola. Tanmateix, aquesta opció presenta un inconvenient: en els sistemes d'ensenyament on s'aplica la repetició de curs a gran escala, en la mostra de l'enquesta es trobarà una proporció més o menys important d'alumnes de quinze anys amb un retard d'un any o potser dos anys respecte a la classe teòrica on es troben la majoria dels alumnes de quinze anys que tenen una escolaritat regular. Per tant, aquests alumnes fan el test sense haver completat tota l'escolaritat obligatòria, cosa que pot presentar un desavantatge respecte a una mostra de joves de quinze anys que hagin estat escolaritzats en un sistema d'ensenyament en el qual no s'aplica gairebé mai la repetició de curs.

L'opció de donar preferència al grup d'edat com a criteri de composició de la mostra avaluada en comptes del curs presenta una altra particularitat, que és contradictòria amb un dels objectius que pretén PISA. L'elecció del grup d'edat de quinze anys es justifica pel fet d'avaluar les competències al final de l'escolaritat obligatòria, atès que s'estima que la majoria dels joves dels països de l'OCDE completen la seva escolaritat obligatòria en aquesta edat. Però aquesta suposició no és del tot correcta. Per verificar realment el nivell de competències adquirit dins un sistema d'ensenyament determinat pel conjunt dels alumnes que acaben l'escolaritat obligatòria, no s'hauria d'haver escollit el grup d'edat com a criteri per constituir la població avaluada, sinó l'últim curs de l'ensenyament obligatori, cosa que hauria implicat la constitució d'una mostra d'alumnes composta d'individus amb edats diferents. Però l'OCDE no ha escollit aquesta opció, més complexa de gestionar.

Als països de l'OCDE la majoria dels joves de quinze anys van a escola i per això és relativament fàcil trobar-los quan s'elabora una mostra representativa de la població escolar.

Proves centrades en les competències i no en el currículum escolar

PISA no proposa avaluar el que s'aprèn a l'escola d'acord amb els programes que els docents han d'impartir. Aquesta operació s'havia revelat molt àrdua en el transcurs de les enquestes de l'IEA. Amb l'enquesta TIMSS (Third International Mathematics and Science Study), l'IEA va encarregar

una enquesta exhaustiva de programes d'ensenyament de matemàtiques amb vista a interpretar els resultats de les proves, però no va aconseguir establir la conjunció entre programes, proves i resultats.⁸ L'estructura i l'organització dels programes d'ensenyament és un trencaclosques per als «comparativistes». PISA resol aquestes dificultats deixant-les de banda: les proves de l'OCDE no se centren en els programes, sinó en les competències o aptituds que es considera que són indispensables per portar una existència autònoma i independent en les societats democràtiques amb una economia de mercat.⁹

Per aquests motius, les empreses que competeixen amb l'IEA en el camp de les avaluacions internacionals de competències han seguit un altre camí. Les dues enquestes de l'ETS que es van realitzar el 1988 i el 1991, conegudes amb les sigles IAEP 1 i 2 (International Assessment of Educational Progress), van ser concebudes per avaluar els coneixements en dues disciplines, les matemàtiques i les ciències. Tanmateix, per neutralitzar la relació entre proves i programes nacionals d'ensenyament, l'ETS va adoptar i adaptar parcialment els ítems del test del programa d'avaluació federal nord-americà NAEP (National Assessment of Educational Progress). Tot i així, el problema de la concordança entre test internacional i programes d'ensenyament es va resoldre tenint com a únic referent el programa d'ensenyament en vigor en només un dels sistemes d'ensenyament. Les especificacions dels programes d'ensenyament en vigor en la resta de sistemes d'ensenyament només es van tenir en compte en el moment de la interpretació dels resultats, que es van ponderar en funció dels programes.

Quan l'OCDE va prendre la decisió de realitzar avaluacions internacionals a gran escala amb un ritme sostingut, també es va enfrontar amb el problema de les diferències existents entre els programes d'ensenyament. Per procedir de manera ràpida a la realització i l'aplicació de l'enquesta, calia resoldre prèviament aquesta qüestió. Atès que el camí que havia seguit l'ETS no era viable perquè no era políticament correcte, només quedava una solució: alliberar-se totalment dels programes d'ensenyament en l'elaboració de les proves. Aquest és el camí que ha seguit l'OCDE amb PISA: s'ha renunciat a mesurar el que els alumnes aprenen a l'escola (almenys en teoria), i es mesura, per contra, el nivell de compe-

tències que es considera que és indispensable als quinze anys per viure en societats democràtiques i amb una economia de mercat. De fet, les proves PISA han estat concebudes com a proves curriculum free, és a dir, proves neutres en relació amb els programes d'ensenyament. El test PISA té com a objectiu avaluar la presència o l'absència d'aptituds que es consideren «essencials» per a la vida adulta.¹⁰ En paraules de la mateixa OCDE, «es tracta de l'aportació més important i ambiciosa del programa OCDE/PISA».¹¹

El programa PISA s'allibera dels límits imposats per la necessitat de trobar un denominador comú constituït pel contingut específic de l'ensenyament impartit a les escoles dels països participants. L'OCDE justifica aquest enfocament de l'avaluació per l'exigència de valorar el capital humà, definit com «els coneixements, les qualificacions, les competències i altres qualitats que posseeix un individu i que interessen al benestar personal, social i econòmic». L'ambició de l'OCDE és arribar a mesurar de manera apropiada l'estoc de capital humà d'un país. PISA marca, així, un gir respecte als estudis anteriors a gran escala i consagra el fet que l'avaluació de coneixements i continguts escolars no és suficient per si mateixa. El programa PISA transmet un nou missatge: l'eficàcia dels sistemes d'ensenyament no es mesura en relació amb l'ensenyament i l'adquisició de coneixements, sinó a través d'avaluacions centrades en el desenvolupament del capital humà. Els paràmetres per avaluar un sistema ja no són interns al sistema, sinó externs.

L'ambició de l'OCDE és arribar a mesurar de manera apropiada l'estoc de capital humà d'un país.

La politització del programa d'avaluació

Amb PISA es produeix un relleu entre la comunitat científica que havia portat de manera més o menys autònoma els estudis a gran escala durant uns trenta anys i la comunitat dels polítics que decideixen, gestionen els afers i dicten o imposen les seves prioritats. La transició és clara: PISA té recursos i ha pogut imposar-se perquè està al servei del que esperen els que decideixen i tenen la responsabilitat de dirigir i regular els sistemes d'ensenyament i dels aparells burocràtics que els governen.

Quan a la tardor del 1997 els governs de l'OCDE van decidir crear PISA i repartir-se entre ells els costos del seu finançament, va començar una nova època en la història de les avaluacions internacionals.

Quan a la tardor del 1997 els governs de l'OCDE van decidir crear PISA i repartir-se entre ells els costos del seu finançament, va començar una nova època en la història de les avaluacions internacionals a gran escala. En efecte, per primer cop els governs van assumir la responsabilitat de programar avaluacions a gran escala amb l'objectiu d'obtenir informacions que s'estimava que eren necessàries per regular els sistemes d'ensenyament i no només per a objectius relatius als coneixements proposats per la comunitat científica. La decisió es va prendre sense prestar gaire atenció a les repercussions d'aquest procés a mitjà i llarg termini. Tanmateix, des de fa molt de temps sabem que la realització de l'avaluació a les classes i sobre els sistemes d'ensenyament provoca comportaments d'adaptació que no més coneixem en part. Les avaluacions posen sota pressió el sistema i els seus actors. Un gran nombre de proves, senyals i declaracions indiquen que PISA tampoc no ha pogut evitar aquest risc. S'entén que, a mesura que passin els anys, la repetició cíclica de les avaluacions previstes en el programa empitjorarà les reaccions d'adaptació i preparació davant l'avaluació, cosa que distorsionarà els resultats de les avaluacions i les comparacions en un grau que encara no es pot mesurar.

En aquestes qüestions hi ha subjacent el conflicte clàssic entre exigències i prioritats científiques, d'una banda, i imperatius polítics, de l'altra, entre coneixement i canvi, entre objectius teòrics i control del servei. En finançar PISA, les autoritats educatives responsables dels sistemes d'ensenyament esperaven no només obtenir de manera ràpida dades sobre els efectes de les polítiques educatives i les reformes escolars, sinó, sobretot, controlar l'eficàcia dels sistemes i obtenir informació sobre diverses formes de regulació de l'ensenyament. Aquesta aposta s'ha guanyat d'una manera aclaparadora si ens cenyim a les repercussions de l'enquesta PISA en les polítiques nacionals d'educació de molts països.¹²

La flexibilitat del programa

Una última particularitat de PISA és la seva flexibilitat: els sistemes d'ensenyament que participen en el programa poden afegir al test principal

dues parts optatives (qüestionari o test). En el primer estudi de PISA, l'any 2000, es va proposar als països dos qüestionaris optatius, l'un sobre competències transversals i l'altre sobre competències informàtiques. El primer d'aquests dos qüestionaris el va elaborar, en el marc del projecte INES de l'OCDE-CERI, una xarxa de científics que es preguntaven sobre els productes de l'educació escolar. Després de prop de set anys de treball, aquest grup va aconseguir incloure un qüestionari en el paquet d'instruments proposats pel programa PISA. Un procés semblant va portar a l'elaboració del qüestionari sobre les competències informàtiques. Un total de 26 països, dels 32 que conformen la mostra, han utilitzat els qüestionaris sobre les competències transversals, i 20 de 32 el de l'ús dels ordinadors i la disponibilitat dels ordinadors a les escoles i les llars, així com l'apreciació que donaven els estudiants sobre les seves competències informàtiques. Aquestes parts han estat negociades internacionalment; d'altres s'han creat al si d'un sistema d'ensenyament o en col·laboració amb certs sistemes. Per exemple, els sistemes d'ensenyament germanoparlants van estudiar junts un qüestionari sobre els docents que no es va dur a terme. Qüestionaris i mòduls complementaris, com ara el qüestionari sobre estratègies d'aprenentatge dels alumnes, han estat molt explotats en la interpretació dels resultats del test principal. Tot i que és difícil incloure en un programa d'avaluació ja de per si molt carregat altres parts que allarguen el temps de durada de les proves amb conseqüències no previsible sobre la motivació i la concentració dels joves, i per tant sobre les possibles distorsions en els resultats, aquesta eventualitat obre camins per recollir altres informacions sobre els centres. Res no impedeix que el programa PISA es vagi modulant, tret de la seva rigidesa i la seva durada, que el converteixen en un programa omnívor que no deixa fàcilment espais per a explotacions complementàries.

Conclusió: la fiabilitat de les enquestes internacionals a gran escala

Les enquestes internacionals, i en concret PISA, continuen sent objecte de moltes crítiques. L'explotació realitzada des del punt de vista polític i per mitjà dels resultats de PISA no suscita confiança en els sectors educatius. També hi ha una sèrie de problemes metodològics que cal resoldre i aclarir, com ara el model de mesura emprat, les condicions de validesa dels

Aquests estudis són una mina d'informació sobre els sistemes d'ensenyament.

Ítems del test, la reducció de distorsions culturals o les qüestions relatives a la traducció. Al llarg dels darrers quaranta anys s'han fet avenços considerables pel que fa a la metodologia i l'organització. La qualitat de les últimes grans enquestes internacionals, com PISA, PIRLS i TIMSS (Trends in Mathematics and Science), és admirable. Aquests estudis són una mina d'informació sobre els sistemes d'ensenyament; d'altra banda, l'exigència de la comparació revela els seus punts febles. Tanmateix, són insubstituïbles, ja que no hi ha altres mitjans per documentar-se sobre les prestacions dels serveis d'ensenyament a fi de fer comparacions que diguin alguna cosa vàlida sobre l'estat de salut d'un sistema i produeixin dades empíriques que puguin ser discutides, posades a prova o rebutjades sense més ni més.

El revés de la moneda és la temptació de la classificació per designar «la noia més bonica del regne». Tanmateix, aquest mirall que ens mostren els estudis internacionals no és perfecte. És un mirall que deforma. Només un esforç considerable i crític des del punt de vista científic podrà ajudar a corregir-ne els defectes. Més que mai, cal fer un debat democràticament obert i una discussió transparent sobre la naturalesa i els límits d'aquestes enquestes. La seva qualitat és indiscutible, la qualitat de les avaluacions és evident, les tècniques de mostreig s'han perfeccionat i el tractament estadístic ha avançat, sobretot pel que fa a la metodologia de les comparacions internacionals. Coneixem millor els límits i les traves d'aquests estudis i comencem a dominar-los millor. Per tant, s'imposa una conclusió: si volem controlar els efectes d'aquestes estratègies i dominar aquestes operacions, poder pronunciar-nos-hi i influir-hi, cal participar-hi, ja que només així farem progressar la ciència de l'educació, però també podrem adquirir un coneixement íntim dels mecanismes, les lluites de poder i els grups de pressió que forgen aquests estudis entre bastidors. Aquest coneixement és indispensable per actuar amb eficàcia, sense cometre errors greus que puguin restar credibilitat a les queixes i les preses de posició que caldrà plantejar davant la comunitat científica.

Referències bibliogràfiques

- OECD. *What Makes School Systems Perform? Seeing School Systems Through the Prism of PISA*. París, 2004.
- SALGANIK, L. H.; RYCHEN, D. S. (ed.) (DeSeCo). *Defining and Selecting Key Competencies*. Seattle: Hogrefe & Huber Publ., 2001.
- SALGANIK, L. H.; RYCHEN, D. S. (ed.). *Key Competencies for a Successful Life and Well-functioning Society*. 2003.
- HAUT CONSEIL DE L'ÉVALUATION DE L'ÉCOLE:
- CEARD, M. T.; REMOND, M.; VARIER, M. «L'application des compétences des élèves et des jeunes en lecture et en écriture et l'évolution de ces compétences dans le temps» [informe] (desembre del 2003).
- SALINES, M.; VRIGNAUD, P. «Apprécier et certifier les acquis des élèves en fin de collège: diplôme et évaluations-bilan» [informe] (juny del 2001).

NOTES

- 1 Les tècniques d'avaluació comparades es van desenvolupar als Estats Units al llarg dels anys trenta, com ara en l'enquesta «Eight-year Study» (Vuit anys d'estudi), que es va dur a terme entre el 1936 i el 1941 sota la direcció de Ralph Tyler (vegeu *International Encyclopedia of Educational Evaluation*, Pergamon Press, 1990, p. 170-171).
- 2 Per poder obtenir mostres probabilistes vàlides, l'anglès Gilbert Peaker va elaborar un manual de mostreig, que es va convertir des d'aleshores en un text de referència per a tots els estudis successius de l'IEA.
- 3 L'IEA ha tingut un paper clau en la constitució i el desenvolupament de la disciplina a escala mundial, però el desenvolupament de l'àmbit de l'avaluació a gran escala del rendiment dels alumnes no només s'ha d'atribuir a l'èxit de l'IEA. Per exemple, hi han tingut un paper principal els científics que als Estats Units han endegat i realitzat el NAEP (National Assessment of Educational Progress) des del 1964.
- 4 El NAEP és el programa d'avaluació que es va endegar el 1964 als Estats Units per observar i mesurar les millors dels alumnes nord-americans de quatre grups d'edat. La primera enquesta sobre una mostra nacional es va fer el 1969. El pressupost inicial del NAEP va ser de prop de 2 milions de dòlars americans. Després del 2001, amb la reforma escolar proposada pel president Bush, el NAEP es va convertir en una peça central del programa de millora de l'ensenyament, amb un pressupost de més de 100 milions de dòlars. Sens dubte, el NAEP ha estat un laboratori formidable de renovació i experimentació de nous enfocaments en matèria d'avaluació. Vegeu: JONES, L. V.; OLKIN, I. *The Nation's Report Card. Evolution and Perspectives*. Bloomington: Phi Delta Kappa Educational Foundation, 2004.
- 5 A aquestes enquestes convé afegir l'enquesta sobre les noves tecnologies a l'escola (IEA-SITES), realitzada el 1998-1999.

6 Convé assenyalar que l'IEA també va adoptar als anys noranta –per tant, després de la crisi desencadenada per l'IAEP– una modalitat periòdica en la programació d'enquestes. Així, tot i conservar les sigles TIMSS, que signifiquen «tercera enquesta internacional sobre les matemàtiques i les ciències», quan el 1994-1995 es va fer l'enquesta es tractava de la tercera que l'IEA feia sobre aquest tema. Amb la rèplica de l'enquesta en els anys següents es va haver de lletrejar les sigles d'una altra manera: el significat de la *T* va canviar de «tercera» a «tendències» (*trends*, en anglès).

7 És interessant assenyalar que en certs països, en el moment de presentar els resultats de PISA 2000 i PISA 2003, es van verificar quines reformes d'ensenyament havien viscut els alumnes que havien estat sotmesos a les proves. Per exemple, en el moment de la presentació dels resultats de PISA 2003, el president de la Conferència dels Caps de Departaments d'Educació de les regions helvètiques, M. Stöckling, va declarar que la posició relativament bona dels alumnes suïssos demostrava l'eficàcia de les reformes de l'ensenyament realitzades als anys noranta.

8 L'anàlisi dels programes d'ensenyament va ser objecte d'una atenció particular en el cas de l'enquesta TIMSS, que es va desenvolupar entre 1994-1995. En aquella ocasió, un grup específic de treball dirigit per William H. Schmidt, de la Universitat de Michigan, va analitzar els programes d'ensenyament oficials i els manuals de matemàtiques que es feien servir en els sistemes d'ensenyament que van participar en aquest estudi. El projecte SMSO (Survey of Mathematics and Science Opportunities) va permetre constituir un banc de dades sobre els programes de matemàtiques i els manuals escolars utilitzats a classe per a l'ensenyament de les matemàtiques. Vegeu l'obra *Characterizing Pedagogical Flow: An Investigation of Mathematics and Science Teaching in Six Countries*, publicada per Kluwer Academic Publishers (Dordrecht) el 1996. Malauradament, aquesta anàlisi dels programes versava només sobre un petit nombre de sistemes d'ensenyament i, sobretot, es desenvolupava en paral·lel a l'enquesta TIMSS i no la precedia, com hauria estat lògic per construir una avaluació orientada segons el currículum. El problema de l'adequació entre sistemes d'ensenyament i avaluacions internacionals ha constituït un obstacle permanent per a l'IEA.

9 Per exemple, els resultats insuficients obtinguts pels alumnes d'un sistema d'ensenyament en càlcul algebraic o geometria espacial poden explicar-se pel fet que aquestes àrees de l'ensenyament de les matemàtiques no s'han tractat a l'escola o ni tan sols estan incloses dins del programa d'ensenyament.

10 OCDE. *Mesurer les connaissances et les compétences des élèves. Un nouveau cadre d'évaluation*, París, 1999. En els documents de l'OCDE s'esmenta un nou referent: les competències necessàries per tenir èxit a la vida, que és menys precís que els coneixements o les competències que es fixen en els programes d'ensenyament.

11 Posteriorment, l'OCDE ha patrocinat un projecte finançat pels Estats Units i la Confederació Helvètica que té com a objectiu definir components que permetin crear dins els comptes nacionals una cistella de béns educatius per tal d'elaborar un índex PPA (poder i paritat de compra) per a l'educació. La necessitat d'aquest índex s'ha fet notar en produir els indicadors financers de «Regards sur l'éducation», que són comparables només en ponderar els valors nacionals en funció del cost de vida existent als diferents països. Tanmateix, l'índex PPA produït i utilitzat per l'OCDE s'ha establert sobre la base del càlcul del cost de la cistella de béns essencials, que no inclou béns relatius a l'educació. És per identificar aquests béns educatius que s'ha creat el projecte DeSeCo (Definició i Selecció de Competències Clau). Posteriorment, el projecte ha estat l'aposta per a l'elaboració d'avaluacions d'enquestes sobre les competències dels adults (Project ALL) i d'alumnes de quinze anys (projecte PISA).