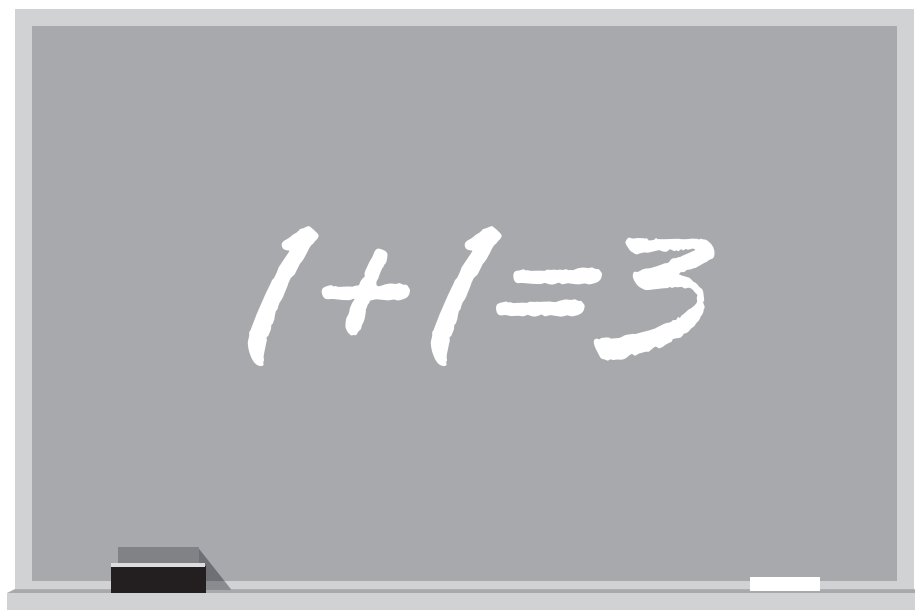


Guia pràctica 11 - Com avaluar l'impacte de les polítiques educatives

Col·lecció Ivàlua de guies pràctiques sobre avaluació de polítiques públiques



ivàlua  Institut Català d'Avaluació de Polítiques Públiques

Institucions membres d'Ivàlua



© 2015, Ivàlua

No es permet la reproducció total o parcial d'aquest document, ni el seu tractament informàtic, ni la seva transmissió en qualsevol forma o per qualsevol mitjà, ja sigui electrònic, mecànic, per fotocòpia, per registre o altres mètodes, sense el permís del titular del Copyright.

Autor:

Miquel Àngel Alegre,

Maquetació i disseny portada: jaumebadosa.es

Primera edició: Abril 2015

Amb la col·laboració de:



Dipòsit legal: B 13257-2015

ÍNDEX

1. PRESENTACIÓ	PÀG. 5
2. QUÈ FUNCIONA EN EDUCACIÓ: COM SABER-HO	PÀG. 7
2.1 NI INTUÏCIÓ, NI EVIDÈNCIA ANECDÒTICA	pàg. 7
2.2 CORRELACIÓ NO ÉS CAUSALITAT: L'EXEMPLE (FICTICI) DEL PAE	pàg. 8
2.3 ELS LÍMITS DE L'“ABANS-DESPRÉS”	pàg. 10
2.4 ALESHORES DE QUÈ ENS HEM DE REFIAR? EVIDÈNCIA EXPERIMENTAL I QUASI-EXPERIMENTAL	pàg. 11
2.4.1 NOTES SOBRE L'AVALUACIÓ EXPERIMENTAL	pàg. 12
2.4.2 NOTES SOBRE ELS DISSENYS QUASI-EXPERIMENTALS	pàg. 15
3. ELS REPTES DE L'EXPERIMENTACIÓ: EXEMPLES	PÀG. 24
3.1 L'AUTONOMIA ESCOLAR: ELS EXPERIMENTS DE LES “CHARTER SCHOOLS”	pàg. 24
3.2 EDUCACIÓ EN LA PETITA INFÀNCIA: PERRY PRESCHOOL I HEAD START	pàg. 26
3.3 EL QUE ES PAGA ALS PROFESSORS: INCENTIUS ECONÒMICS EN ENTORNS COMPLEXOS	pàg. 28
3.4 LES TIC COM A EINA D'ENSENYAMENT-APRENENTATGE: ACCESS TO ALGEBRA I	pàg. 31
3.5 LES TUTORIES INDIVIDUALITZADES: TIME TO READ I SWITCH-ON READING	pàg. 32
3.6 LA IMPLICACIÓ DE LES FAMÍLIES: MALLETE DES PARENTS I READY4K!	pàg. 35
4. ELS REPTES DE L'AVALUACIÓ QUASI-EXPERIMENTAL: EXEMPLES	PÀG. 39
4.1 DISSENYS DE REGRESSIÓ DISCONTÍNUA	pàg. 39
4.1.1 DE QUÈ SERVEIX REPETIR CURS	pàg. 39
4.1.2 BEQUES I AJUTS ALS ESTUDIANTS	pàg. 41
4.2 L'ÚS DE VARIABLES INSTRUMENTALS	pàg. 43
4.2.1 LA IMPORTÀNCIA DE LA RÀTIO D'ALUMNES PER CLASSE	pàg. 43
4.2.2 LA LLIBERTAT D'ESCOLLIR CENTRE EDUCATIU	pàg. 45
4.3 L'ÚS DELS MODELS DE DOBLES DIFERÈNCIES	pàg. 47
4.3.1 LA COMPRENSIVITAT DEL SISTEMA EDUCATIU	pàg. 47
4.3.2 LA DURADA DE LA JORNADA ESCOLAR	pàg. 49
5. I A CASA NOSTRA, COM PODRÍEM AVANÇAR?	PÀG. 53
5.1 ATREVIR-SE A EXPERIMENTAR: APROFITAR LES PROVES PILOT...	pàg. 53
5.2 ... I L'EXCÉS DE DEMANDA	pàg. 55
5.3 AVALUACIÓ QUASI-EXPERIMENTAL	pàg. 56
5.4 ... PREVEIENT L'AVALUACIÓ EN EL DISSENY DEL PROGRAMA	pàg. 57
5.5 SISTEMES D'INFORMACIÓ I ACCÉS A LES DADES	pàg. 61
5.6 QUÈ FUNCIONA I PER QUÈ FUNCIONA: LA IMPORTÀNCIA DE FILAR PRIM	pàg. 62
REFERÈNCIES	PÀG. 65

1. PRESENTACIÓ

A casa nostra és ja un lloc comú la idea que les polítiques públiques s'han d'avaluar; que l'avaluació ha de contribuir a millorar aquestes polítiques, els seus processos, els seus resultats, i al mateix temps ha de servir com a instrument per retre'n comptes davant la ciutadania. En l'àmbit educatiu, aquesta idea és sovint present en el discurs de responsables i gestors públics i en bona part de la producció normativa, alhora que fonamenta la raó de ser d'organitzacions com la Inspecció Educativa o el Consell Superior d'Avaluació del Sistema Educatiu.

I és cert que en l'àmbit educatiu se'n fan moltes d'avaluacions: s'examinen alumnes, es valoren mèrits docents, es diagnostiquen els centres i inspeccionem la seva gestió i processos, es fa seguiment d'indicadors educatius, s'analitza la implementació i els resultats d'alguns programes, etc. De tot això en diem avaluar. És evident, però, que el tipus de coneixement –això és, la mena d'evidència– que unes i altres pràctiques avaluatives aporten és molt diferent, com també ho és l'èmfasi que a casa nostra hem donat a uns i altres tipus d'avaluacions.

A hores d'ara disposem de diversos instruments que ens permeten observar l'evolució dels resultats acadèmics dels centres i els alumnes en diferents etapes educatives (proves de competències, proves diagnòstiques, tests internacionals estandarditzats, registres educatius, etc.). D'altra banda, estem també habituats a inspeccionar i documentar els processos organitzatius, la gestió acadèmica i les activitats formatives que desenvolupen els centres educatius al llarg del curs.

En canvi, hem dedicat pocs esforços a avaluar l'efectivitat dels programes i les intervencions educatives, això és, a tractar de conèixer en quina mesura aconsegueixen tenir impactes sobre la realitat que pretenen modificar. Això fa que a Catalunya, com a d'altres països del nostre entorn, les decisions sobre les polítiques educatives –sobre el seu llançament, el seu manteniment, la seva reforma, la seva supressió– rarament es basin en evidències empíriques sòlides sobre la seva efectivitat.

En qualsevol escenari, més encara en un de crisi econòmica i contenció de la despesa, i encara més quan una bona part dels nostres indicadors educatius estan encara lluny del que seria desitjable, és necessari conèixer si les intervencions educatives funcionen, si produeixen els impactes esperats. I és necessari aprofitar aquest coneixement per millorar el disseny de les mateixes intervencions, per tractar de fer-les més efectives (i més cost-efectives). És cert que l'avaluació d'impacte requereix d'unes condicions de viabilitat i d'uns requeriments metodològics no sempre fàcils d'acomplir. I és també cert que, en el procés de formació i reforma de polítiques educatives (i no només educatives), el recurs a l'evidència i al coneixement forçosament ha de competir amb altres factors (interessos o compromisos polítics o de grups d'interès, inèrcies institucionals, pressions mediàtiques, etc.). Però tot plegat

no justifica el poc pes que, a casa nostra, tenen la producció i l'ús d'evidències d'efectivitat en educació.

El propòsit final d'aquesta guia és contribuir a contrapesar aquesta tendència¹. I amb aquest propòsit, cobreix tres centres d'atenció.

En primer lloc (capítol 2), discutirem el mateix concepte d'evidència en educació. Ens preguntarem què és evidència sòlida i creïble, i què no ho és; de quines proves empíriques ens podem refiar i de quines no. Feta aquesta distinció, pararem breu atenció a les aproximacions i als mètodes susceptibles de generar el tipus d'evidència pel qual apostarem: el mètode experimental (estudis amb assignació aleatòria) i determinades tècniques quasi-experimentals (en particular, aquelles que permeten identificar grups de comparació vàlids).

En segon lloc (capítols 3 i 4), la guia ofereix exemples d'avaluacions realitzades i evidències acumulades a nivell internacional en diferents àmbits de la política i la intervenció educativa. La selecció d'exemples que aquí proposem no té cap ànim d'exhaustivitat ni pretén ser representativa del conjunt d'àmbits d'actuació rellevants en educació. En canvi, la selecció obeeix a la voluntat d'oferir una mostra prou àmplia i equilibrada d'exemples de pràctiques i mètodes d'avaluació (experimentals i quasi-experimentals), que, alhora, connecti amb l'agenda de prioritats de la política educativa a casa nostra.

En tercer lloc (capítol 5), ens fixarem en els espais d'oportunitat que podríem trobar a casa nostra per tal d'afavorir l'avaluació d'efectivitat de les polítiques i intervencions educatives i de cara a promoure decisions en educació més basades en l'evidència.

Notes:

¹ Compartint aquest mateix propòsit, hem de fer referència a la posada en marxa del projecte "Què funciona en educació", liderat per Ivàlua i la Fundació Jaume Bofill. Aquest projecte, que s'engega l'any 2015, inclou una publicació periòdica de revisió d'evidències i tot un conjunt de seminaris i jornades adreçats a promoure l'aprofitament d'aquest coneixement per a la millora de la intervenció i la pràctica educativa a Catalunya. El projecte s'inspira en dues iniciatives d'institucionalització de la perspectiva "què funciona" en educació: el *What Works Clearinghouse (WWC)*, nascut l'any 2002 de la mà de l'*Institute of Education Sciences del Govern Federal dels Estats Units* (<http://ies.ed.gov/ncee/wwc/>); i l'*Educacion Endowment Foundation*, constituïda l'any 2011 al Regne Unit (<http://educationendowmentfoundation.org.uk/>). Part essencial de la tasca d'aquestes dues institucions passa per revisar, qualificar, resumir i difondre el coneixement existent sobre l'efectivitat de diferents línies d'actuació educativa, prenent com a base l'evidència aportada per estudis i revisions sistemàtiques d'especial rellevància.

2. QUÈ FUNCIONA EN EDUCACIÓ: COM SABER-HO

Ens preguntem, per tant, per l'efectivitat de les intervencions educatives, pels seus impactes, per si funcionen o no funcionen a l'hora d'incidir en la realitat a la qual s'adrecen. I partim de la base que aquesta pregunta requereix respostes basades en evidències empíriques sòlides, evidències que estableixin de forma robusta les relacions de causalitat entre la política i eventuais canvis en la problemàtica en qüestió. Arribats a aquest punt, caldria que ens preguntéssim: què hauríem de considerar evidència empírica creïble i sòlida en el camp educatiu? Quin tipus d'evidència necessitem obtenir o acumular per determinar amb garanties que una política o programa educatiu funciona o no funciona, que és més o menys efectiu? Preguntem-nos, primer, de quina "evidència" no ens hauríem de refiar.

2.1 NI INTUÏCIÓ, NI EVIDÈNCIA ANECDÒTICA

Quan no s'acompanya d'un contrast empíric fiable, la intuïció pot arribar a ser un mal company de viatge, una mala consellera en el procés de producció i canvi de les polítiques públiques. Per exemple, un programa de digitalització de les aules que proveeixi a cada alumne d'un ordinador personal, fer que els alumnes incrementin el nombre d'hores que passen diàriament a l'escola, fer repetir curs a partir d'un determinat llindar d'assignatures suspeses, treballar amb grups-classe reduïts de nivell homogeni, incentivar econòmicament el professorat, permetre que les famílies tinguin major capacitat per triar l'escola que volen per als seus fills, o que les escoles tinguin major llibertat per definir els continguts del currículum, etc.; totes elles poden semblar apostes raonables, de les quals podríem esperar efectes positius per al progrés educatiu dels alumnes. Molt probablement sigui això el que ens dicti la intuïció. Doncs bé, tindrem ocasió de constatar com l'evidència internacional no és sempre concloent quant a la capacitat d'impacte d'aquestes mesures; el que equival a dir que l'evidència empírica és, o pot ser sovint, contra-intuïtiva.

En segon lloc, hauríem també de desconfiar d'allò que podríem denominar evidència "anecdòtica" o "circumstancial". Amb aquest terme ens referim a aquelles dades empíriques obtingudes sense cap intenció ni criteri de validesa o representativitat, en general basades en el coneixement personal directe (experiència pròpia) o indirecte (coneixement típic, o del boca-orella). En efecte, podem conèixer casos d'alumnes repetidors que, gràcies al fet de repetir, han pogut prendre consciència de la seva situació acadèmica i aconseguir reconduir la seva trajectòria educativa; o casos d'alumnes que relaxen el seu nivell d'esforç en rebre ajuts econòmics a la continuïtat escolar; o casos d'alumnes vulnerables (acadèmicament i socialment) que, en compartir aula amb companys d'altres nivells acadèmics i entorns social, experimenten processos de desplaçament i acaben abandonant els estudis; o casos d'alumnes també vulnerables que rendeixen poc a la primària malgrat haver passat per una escola bressol. Podem haver conegut aquestes situacions en primera persona o com a familiars implicats, o haver-ne sentit a parlar. Tanmateix elevar la dada anecdòtica (sovint qüestionable

en ella mateixa, i necessàriament tendenciosa) a categoria d'evidència és del tot incorrecte.

2.2 CORRELACIÓ NO ÉS CAUSALITAT: L'EXEMPLE (FICTICI) DEL PAE

Que les mesures de dues variables es trobin correlacionades no implica que entre l'una i l'altra existeixi una relació de causalitat. Aquesta màxima, que en el pla teòric esdevé incontestable, a la pràctica no sempre resulta tan òbvia. Vegem-ho a través d'un exemple.

Imaginem-nos que volem conèixer l'efectivitat d'un nou programa de millora dels aprenentatges instrumentals (català, castellà i matemàtiques) adreçat a l'alumnat que comença l'ESO amb dèficits significatius en aquestes àrees; en direm Programa d'Acceleració Educativa o PAE. A través d'un treball en grups reduïts, que es perllonga durant una part significativa de l'horari lectiu, el programa proveeix els alumnes que hi participen d'un nombre d'hores de classe en matèries instrumentals superior al de la resta d'alumnes. Tot plegat amb l'objectiu de reforçar i millorar a curt i mig termini el progrés educatiu d'aquest alumnat acadèmicament vulnerable. La participació en el programa, tant per part dels centres educatius com per part dels alumnes, és de caràcter voluntari. En el cas dels centres, la decisió de sol·licitar el programa ha de ser aprovada per una àmplia majoria del consell escolar. En el cas dels alumnes, la participació en el programa ha de ser sol·licitada per part de la família².

Per tant, volem saber si el PAE ("tractament" en qüestió) pot tenir impactes significatius sobre el rendiment dels alumnes en les àrees curriculars en qüestió i sobre els seus nivells d'abandonament prematur. Imaginem-nos que, amb aquest objectiu, comencem a cercar la presència de possibles correlacions significatives entre el fet de participar en el programa i determinats resultats (o outcomes) –per exemple, rendiment acadèmic i nivells de graduació. I posem que, efectivament, identifiquem l'existència d'associacions significatives a tres nivells.

En el pla de la comparativa internacional, observem que els alumnes que inicien la secundària amb més dèficits formatius tendeixen a obtenir millors resultats (en rendiment i graduació) en aquells sistemes educatius on s'aposta de manera decidida per programes comparables al PAE que en aquells altres sistemes on aquest tipus d'intervencions són residuals. Per exemple, veuríem que a mida que els països augmenten el percentatge d'alumnes inscrits en escoles amb un PAE (o similar), augmenten també el rendiment mitjà i nivells d'acreditació del seu alumnat més vulnerable. I podria perfectament ser que aquesta associació es mantingués significativa a igualtat d'altres variables de context que poguéssim considerar rellevants (PIB per càpita dels països, nivells de despesa en educació, volum total d'alumnes amb dèficits formatius a primària i secundària, etc.).

A un segon nivell, situats ja a Catalunya, podríem detectar que és justament en els centres amb PAE allà on l'alumnat vulnerable obté, com a mitjana, millors resultats. I podria també ser que aquesta associació entre PAE i progrés acadèmic es mantingués significativa a igualtat d'altres

variables clau de centre (percentatge d'alumnat vulnerable a l'escola, recursos materials i humans, ràtio mitjana d'alumnes-classe en els grups ordinaris, titularitat, etc.).

Finalment, a un tercer nivell, podríem observar que, dins del mateix centre, els joves vulnerables que reben el PAE tendeixen a presentar uns nivells de rendiment i graduació superiors als joves vulnerables que no van sol·licitar participar-hi; una tendència que podria romandre significativa fins i tot quan tenim en compte altres variables individuals previsiblement relacionades amb els outcomes considerats (variables com el sexe, la procedència, l'estatus socioeconòmic de pares i mares, etc.).

Identificades aquestes tres associacions, podríem aleshores concloure que el PAE té impactes positius sobre les oportunitats educatives de l'alumnat més vulnerable? Tot apuntaria cap a una resposta afirmativa. En termes de resultats acadèmics, els països que aposten pel PAE van millor, les escoles que el tenen van millor, els alumnes que el reben van millor. I en canvi la resposta a l'interrogant plantejat seria "no necessàriament". Correlació no és causalitat, i les troballes fins ara exposades són eminentment correlacionals.

En el nivell de la comparativa entre països, són molts els factors que poden estar mediant en l'associació entre nivell d'extensió del PAE i resultats educatius, factors no sempre fàcils de controlar i que, tanmateix poden estar relacionats tant amb la variable de tractament com amb els outcomes d'interès –tindríem aquí el que s'anomena un problema d'"endogeneïtat"³. Per exemple, pot ser que programes com el PAE tinguin tendència a consolidar-se en sistemes educatius poc segregats, més comprensius i/o més madurs en l'aplicació de mesures efectives d'atenció a la diversitat, o fins i tot en societats particularment homogènies des d'un punt de vista socioeconòmic i cultural. I podria ser que fossin aquestes variables les veritables facilitadores de la recuperació dels assoliments formatius dels alumnes més vulnerables.

En el nivell de la comparativa entre escoles i entre alumnes dins les escoles, cal comptar amb els problemes d'endogeneïtat procedents de possibles "biaixos de selecció"⁴; biaixos en l'accés al programa que acaben induint a comparacions entre escoles (tractades i no tractades) i entre alumnes (tractats i no tractats) força dissímils en determinades característiques. Això és particularment evident en el cas de programes on la participació és voluntària, com el PAE. Pel que fa a les escoles, fins i tot quan comparem centres similars en allò observable (característiques tangibles i mesurables), hauríem de comptar que un centre amb PAE segurament serà diferent a un altre sense PAE en aspectes que van més enllà del fet de disposar o no del programa en qüestió; per exemple, en cultures institucionals i dinàmiques organitzatives estretament relacionades amb la voluntat d'accedir al programa i amb la capacitat de treure-hi partit. Igualment, les famílies i alumnes que aposten pel PAE probablement es diferenciïn d'aquelles altres que, podent-ho fer, no ho fan, en aspectes estratègics i motivacionals clarament associats al seu rendiment posterior –aspectes no observables, intangibles, que poden escapar al control estadístic.

2.3 ELS LÍMITS DE L'“ABANS-DESPRÉS”

Tampoc no ens podem refiar dels resultats de la simple comparativa entre com es presentava la realitat en qüestió abans de la implementació del programa i com es presenta després. Per una qüestió molt clara: la dimensió sobre la qual una determinada intervenció pretén generar un impacte està sotmesa a la influència d'altres factors externs a més de la mateixa intervenció objecte d'avaluació. Així, els canvis que s'observen en els outcomes d'interès un cop transcorregut el programa o intervenció no tenen perquè ser necessàriament efecte de la seva implementació. Tornem a l'exemple anterior.

Imaginem que identifiquem un escenari pre-programa (centres abans del PAE) i un escenari post-programa (centres després del PAE), i que procedim a comparar els resultats escolars dels alumnes vulnerables en l'un i en l'altre escenaris. Imaginem que aquest esquema, que denominem “abans-després”, ens permet detectar un increment significatiu d'aquests resultats en l'escenari post-programa; això és, es produeix una millora en el rendiment i nivells de graduació de l'alumnat vulnerable un cop els centres comencen a desenvolupar el PAE. És aquest canvi conseqüència del programa? La resposta tornaria a ser, no necessàriament.

La comparació entre els resultats pre- i post-programa, especialment quan l'únic que s'observa són les escoles afectades pel programa, torna a trobar-se subjecta a les limitacions pròpies de la possible omisió de variables rellevants. Entre l'abans i el després del programa, coincidint justament amb el seu llançament i implementació, poden produir-se determinades circumstàncies o canvis en el context que, amb relació o no amb el desenvolupament del programa, siguin a la pràctica els veritables causants de les millores en els resultats observats post-programa. Imaginem que el PAE va començar a implantar-se a casa nostra l'any 2005, i que pocs anys després (i aquí no ens hem d'imaginar res), a partir de 2008, comença a evidenciar-se un descens significatiu del nombre d'alumnes que abandonen prematurament els estudis. La temptació a atribuir el mèrit (o part del mèrit) d'aquest descens al programa és gran. I tanmateix, l'associació abans-després, per si sola, no ens permet determinar si el programa hi ha tingut molt, poc, o gens a veure. De fet, semblaria raonable situar el principal responsable d'aquest descens en la contracció que la crisi econòmica ha generat en l'oferta d'ocupació, i que ha reduït dràsticament el cost d'oportunitat d'estudiar: allò que en el passat representava un incentiu a abandonar els estudis (facilitat per trobar feina remunerada) es transforma ara en un incentiu a romandre o retornar al sistema escolar (dificultats per trobar feina, particularment per als joves menys formats).

L'exercici de comparar els resultats en qüestió abans i després d'una determinada actuació educativa pot revestir un gran interès en el marc d'una avaluació d'impacte. No obstant això, per si sol, aquests exercici no és suficient per pronunciar-se sobre l'efectivitat de l'actuació.

2.4 ALESHORES DE QUÈ ENS HEM DE REFIAR? EVIDÈNCIA EXPERIMENTAL I QUASI-EXPERIMENTAL

Per saber si un programa com el PAE funciona, això és, si és efectiu, si produeix impactes positius rellevants sobre la realitat a la qual s'adreça, el que idealment hauríem de poder fer és comparar allò que els passa als alumnes o escoles que hi participen (escenari factual) amb allò que els hagués passat si no hi haguessin participat (escenari contrafactual). Aleshores sí que estariem comparant situacions idèntiques; l'únic que les diferenciaria és la presència o absència de la intervenció en qüestió. Per tant, tota diferència que observéssim entre els resultats dels alumnes i les escoles en l'un i l'altre escenari podríem atribuir-la causalment al programa (Blasco i Casado, 2009).

És clar que l'escenari contrafactual no és directament observable: uns mateixos individus no poden ser i no ser objecte d'un mateix programa al mateix temps. Això fa que el principal repte de tota avaluació d'impacte consisteixi a proposar estratègies que permetin identificar la millor hipòtesi contrafactual possible, el que a la pràctica vol dir identificar un grup de control o comparació⁵ el més semblant possible al grup que participa en el programa. Quan comparem alumnes-PAE amb alumnes-no-PAE, centres-PAE amb centres-no-PAE, països-PAE amb països-no-PAE, contextos escolars sense PAE que passen a tenir PAE, cal garantir que allò que principalment distingeix les unitats tractades i les unitats control és que les primeres han rebut el programa i les segones no; en la resta de característiques, observables i no observables, unes i altres unitats haurien de ser el màxim d'equivalents. En el camp educatiu, com en general en l'àmbit de les ciències socials, això no és sempre fàcil d'assolir.

I tanmateix són diverses les estratègies que permeten avançar en la identificació d'escenaris contrafactuals, de grups de control creïbles. La més robusta passa per assignar aleatòriament les unitats en disposició d'accedir al programa (en el nostre cas, escoles o alumnes) entre grup de tractament i grup de control. Aquest procediment és la pedra angular del disseny experimental en avaluació, disseny que alguns autors han qualificat com el "gold standard" quan es tracta d'aportar evidències sobre què funciona (veure, per exemple, Duflo, Glennerster, i Kremer, 2007). Quan l'aplicació d'aquest disseny, per uns o altres motius, no és viable, poden aleshores entrar en joc determinats dissenys d'avaluació quasi-experimental amb contrafactual –en particular, aquells dissenys susceptibles de minimitzar els biaixos de selecció abans esmentats.

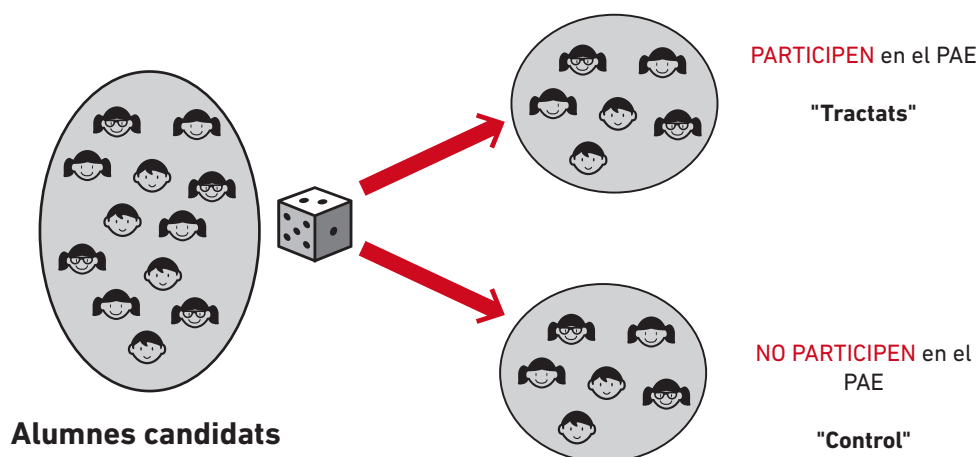
Les notes que segueixen presenten alguns dels principals trets característics d'uns i altres mètodes d'avaluació. Aquestes notes ni volen ser, ni poden ser exhaustives. El seu propòsit és proporcionar elements informatius que facilitin la interpretació dels exemples d'avaluacions exposats en apartats posteriors i, alhora, les seves potencials aplicacions a casa nostra.

2.4.1 NOTES SOBRE L'AVALUACIÓ EXPERIMENTAL

Avaluar l'impacte d'una política educativa mitjançant un disseny experimental és, des d'una perspectiva metodològica, molt similar a aplicar la lògica que segueixen els assajos clínics per testar l'efectivitat d'un fàrmac. La diferència rau en què allò que avaluem no són els efectes d'un medicament sobre l'estat de salut, sinó la capacitat d'una intervenció educativa d'impactar positivament sobre una determinada problemàtica. En ambdós casos, però, la clau de volta la trobem en l'aleatorització de la participació, això és, en l'aplicació d'un mecanisme de sorteig com a procediment per assignar, del conjunt d'unitats elegibles per al tractament, quines acaben efectivament rebent-lo i quines s'incorporen al grup de control.

En el nostre exemple, si fixéssim l'atenció en la unitat escola, es tractaria de triar aleatòriament, entre el conjunt de centres que han sol·licitat desenvolupar el PAE i reuneixen les condicions formals requerides per la convocatòria, quins s'adhereixen al programa (grup de tractament) i quins no (i passen a formar part del grup de control). El procediment seria el mateix en el cas que la unitat d'observació la fixéssim en l'alumne: centre per centre, triaríem de forma aleatòria quins alumnes, entre tots aquells que sol·liciten participar i reuneixen els requisits establerts, reben el programa i quins són assignats al grup de control (veure figura 1).

Figura 1. El disseny experimental. Programa vs. No programa



Font: elaboració pròpia

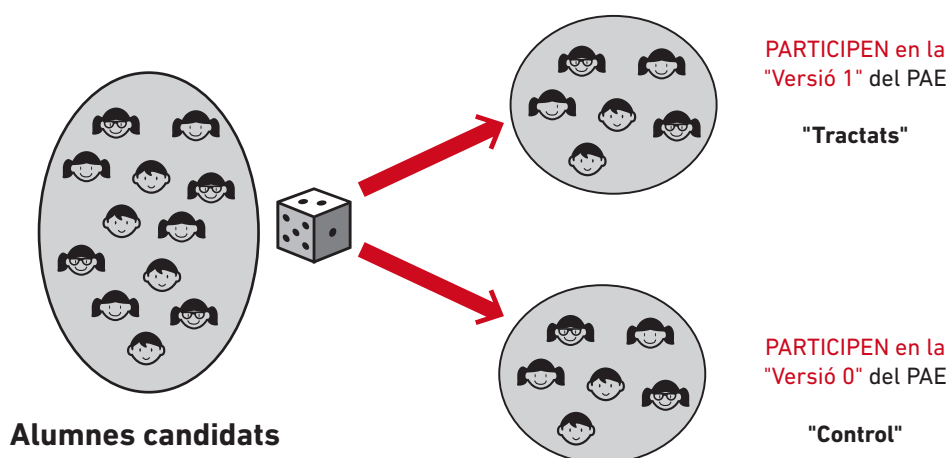
Quan el nombre de casos és prou rellevant, l'aleatorització garanteix que el grup d'individus o unitats control i el grup de tractament acabin essent equivalents en totes aquelles característiques, observables i no observables, que poden estar relacionades amb els resultats considerats. L'únic que aleshores els diferenciarà serà el fet d'haver participat o no en el programa, causa de tota possible diferència en aquests resultats. Més particularment, el fet que l'assignació aleatòria es realitzi entre sol·licitants actius d'un determinat recurs o servei permet comparar conjunts d'individus amb un mateix nivell de motivació, si més no aquella

motivació vinculada a l'acte de demandar-lo. Que la definició de les característiques que han d'acomplir els individus sol·licitants de l'ajut (població elegible) sigui més o menys restringida no té perquè qüestionar la robustesa dels resultats de l'avaluació dels seus impactes.

Val a dir que l'ús del disseny experimental en l'avaluació de polítiques públiques i socials és objecte de crítiques diverses. Dues d'elles han tingut un ressò especial. La primera d'elles qüestiona el sostrat ètic de l'experimentació social: resulta injust privar determinats individus (els del grup de control) dels beneficis que suposa una nova política o servei utilitzant un mecanisme tan "arbitrari" com l'aleatorització. El fonament d'aquest argument és, però, criticable. En primer lloc, la presumpció que s'està privant alguns individus de quelcom beneficiós deixa de tenir sentit si pensem que és precisament l'absència d'evidències sobre l'efectivitat del programa el que justifica l'experiment. En segon lloc, en situació d'escassetat de recursos, quan un determinat servei o programa no pot cobrir al conjunt de la població a la qual s'adreça (en el cas del PAE: conjunt d'alumnes a l'inici de l'ESO en tots els instituts i escoles de Catalunya), aleshores l'assignació aleatòria del recurs entre la població que el sol·licita i que n'és igualment elegible és segurament el mecanisme de distribució més just que puguem dissenyar.

En tot cas, quan un programa es basa en evidències prometedores i quan es disposa de recursos per dotar-lo d'una cobertura relativament àmplia, aleshores poden entrar en joc dos dissenys experimentals segurament més acceptables. Un primer disseny passaria per aleatoritzar la participació en diferents versions del programa (veure figura 2). D'aquesta manera, no es mesura l'impacte d'un programa en referència a una situació d'absència de programa, sinó fins a quin punt una versió del programa és més efectiva que una altra.

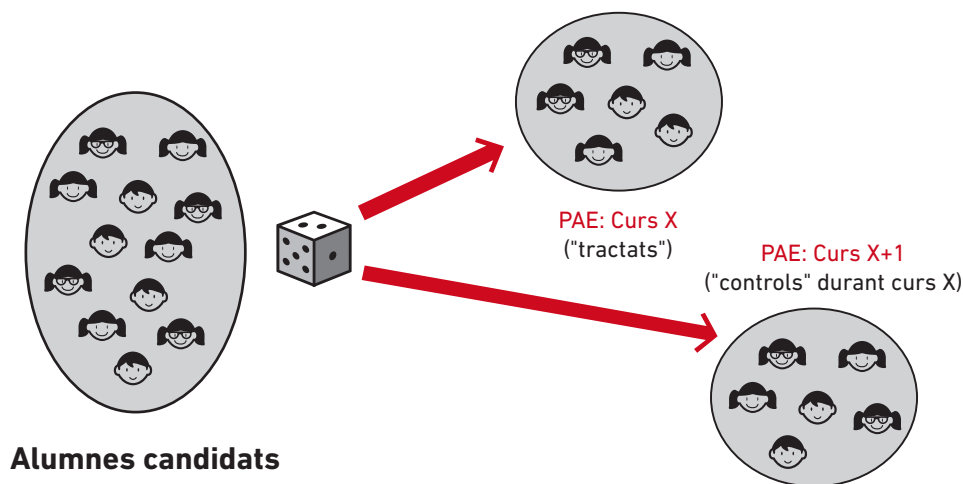
Figura 2. El disseny experimental. Programa 1 vs. Programa 0



Font: elaboració pròpia

Una segona alternativa optaria per aleatoritzar el moment i ordre d'entrada el programa (veure figura 3). En aquest cas, un sorteig decideix, entre tots els candidats a participar en el programa, qui hi accedeix primer i qui després (al cap d'un trimestre, d'un semestre o d'un any acadèmic). Els individus que són assignats a la segona edició del programa, mentre romanen a l'espera, actuen com a control d'aquells altres assignats a participar-hi en primera instància.

Figura 3. El disseny experimental. Programa en moment "X" vs. Programa en moment "X+1"



Font: elaboració pròpia

La segona crítica habitual als experiments socials remet a la manca de "validesa externa" dels seus resultats, això és, al fet que les conclusions que se n'obtenen, tot i ser vàlides en referència a la població, lloc i moment implicats en l'experiment (això és, tot i tenir "validesa interna"), poden no ser extrapolables a contextos diferents. Aquest és un argument fort, que, tanmateix, pot contrapesar-se en dues direccions. En primer lloc, són habituals els denominats pilots multi-site, en els quals un mateix programa s'avalua en llocs diversos (per exemple, diferents municipis), amb l'objectiu d'analitzar fins a quin punt els resultats d'impacte varien en funció dels contextos. D'altra banda, quan el nombre de rèpliques experimentals d'un determinat tipus d'intervenció és suficientment important, es pot dur a terme el que s'anomena una metanàlisi dels resultats obtinguts, és a dir, un exercici quantitatiu de síntesi que pretén establir si programes similars al considerat resulten efectius amb caràcter general, més enllà de les poblacions, llocs i moments en què s'apliquen.

És evident que les avaluacions experimentals plantegen reptes importants tant de caràcter

metodològic com de viabilitat tècnica. Tanmateix, el sorprenent és que un disseny avaluatiu com aquest, cada cop més utilitzat arreu i internacionalment considerat el mètode més robust de l'avaluació d'impacte, es trobi pràcticament absent de l'avaluació de polítiques públiques a casa nostra (Casado, 2012)⁶.

2.4.2 NOTES SOBRE ELS DISSENYS QUASI-EXPERIMENTALS

Els mètodes quasi-experimentals d'avaluació d'impacte comparteixen amb el disseny experimental la definició d'un grup de control o comparació (no participants en el programa) als efectes d'estimar l'outcome contrafactual dels participants. En canvi, a diferència del disseny experimental, que elimina el biaix de selecció mitjançant l'aleatorització de la participació, la resta de mètodes només assoleixen aquest objectiu si es compleixen certs supòsits relatius al mecanisme que governa la participació en el programa.

Són diversos els mètodes quasi-experimentals que s'utilitzen per avaluar l'efectivitat de programes i intervencions educatives. Tres d'ells, la regressió discontinua, els models de dobles diferències i els efectes fixos, disposen d'aplicacions interessants en aquest àmbit.

Regressió discontinua

És probable que aquesta sigui la tècnica d'avaluació no experimental que més projecció ha tingut en l'àmbit educatiu durant els darrers anys (Schlotter, Schwerdt, i Woessmann, 2010; Schneider, Carnoy, Kilpatrick, Schmidt, i Shavelson, 2007). La regressió discontinua es pot aplicar quan la participació en el programa en qüestió depèn del fet que una variable prengui un conjunt de valors determinats; això és, quan la possibilitat de rebre'l s'estableix a partir d'un tall en una variable numèrica. Més en particular, aquesta hauria de ser una variable presumiblement vinculada als outcomes d'interès.

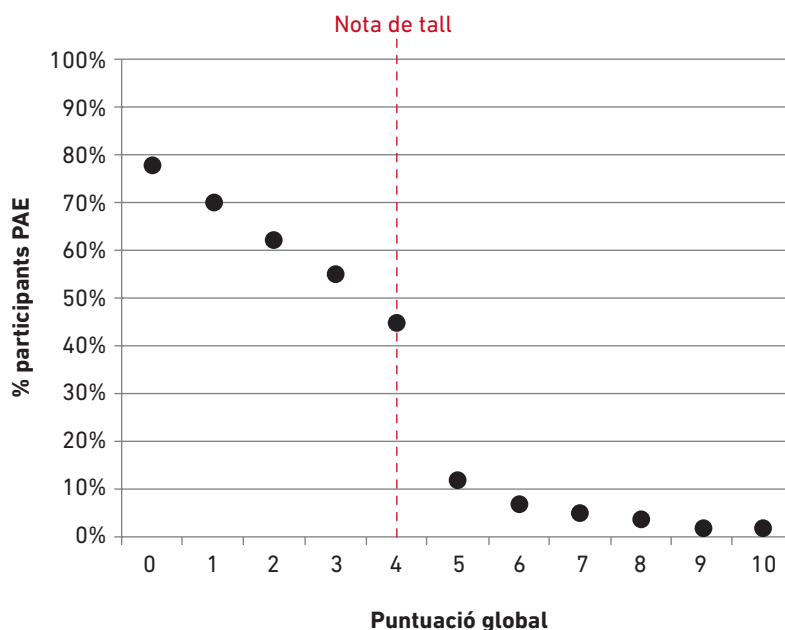
Per exemple, podria donar-se el cas que, per tal d'avaluar l'existència de consens en la decisió dels centres de sol·licitar el PAE, es requereix un mínim d'un 75% de vots del consell escolar en favor d'aquesta decisió. O es podria establir que fossin elegibles per rebre el programa únicament els instituts amb, com a mínim, un 25% d'alumnes a l'inici de l'ESO amb dèficits acreditats en les matèries instrumentals. Més encara, l'administració podria recomanar els centres PAE que prioritzessin la participació en el programa d'aquells estudiants que haguessin puntuat per sota d'un determinat llindar en les proves de competències bàsiques de 6è de primària.

En tant que la definició d'aquests llindars s'estaria establint de forma exògena (per determinació de l'administració educativa), hauríem d'esperar que els centres i els alumnes que es troben just per sota i just per sobre del punt de tall en qüestió fossin essencialment equivalents en tots els atributs rellevants, observables i inobservables.

Per exemple, imaginem que el Departament d'Ensenyament adjudica un nombre limitat de places PAE a cadascun dels centres implicats i els demana que prioritzin l'atenció d'aquells alumnes que van obtenir menys de 4 punts de qualificació global (en una escala de 0 a 10) en les proves de competències bàsiques de 6è de primària; recordem, tots ells alumnes que han sol·licitat accedir al programa.

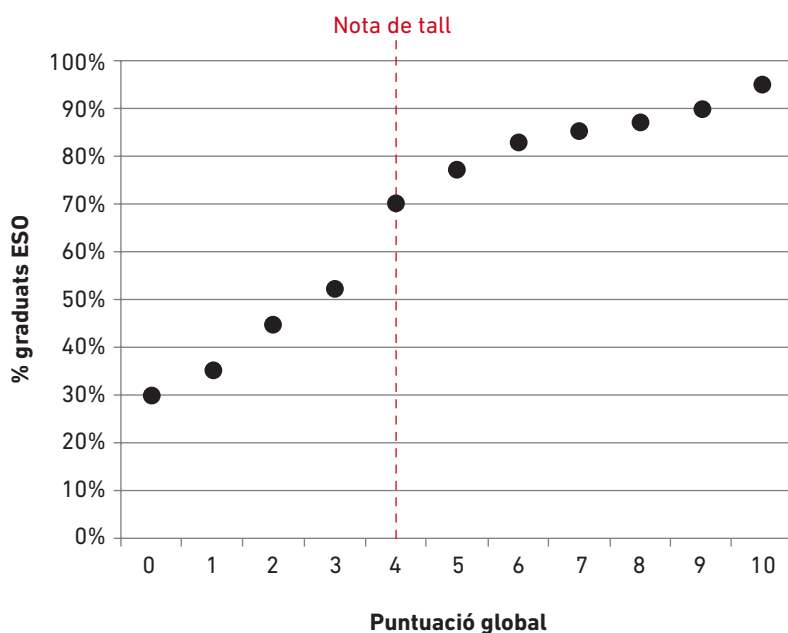
Hauríem, aleshores, de constatar que el percentatge d'alumnes PAE fa un salt a partir que aquests superen aquest llindar de qualificació (gràfic 1). Doncs bé, si el programa és efectiu, hauríem de poder observar que aquest salt en el percentatge de participació es correspon amb una discontinuïtat en els resultats dels alumnes amb puntuacions de partida al voltant d'aquest llindar (gràfic 2). Aquest segon salt reflectiria l'impacte del programa.

Gràfic 1. Proporció d'alumnes PAE segons puntuació global en proves de competències a 6è de primària



Font: elaboració pròpia (dades no reals)

Gràfic 2. Proporció d'alumnes que obtenen el GESO segons puntuació global en proves de competències a 6è de primària



Font: elaboració pròpia (dades no reals)

El mètode de regressió discontinua proporciona bones estimacions de l'impacte dels programes sobre la població que es troba a prop del llindar de referència; no en va, és la proximitat al llindar allò que garanteix la semblança entre participants i no participants⁷.

Variables instrumentals

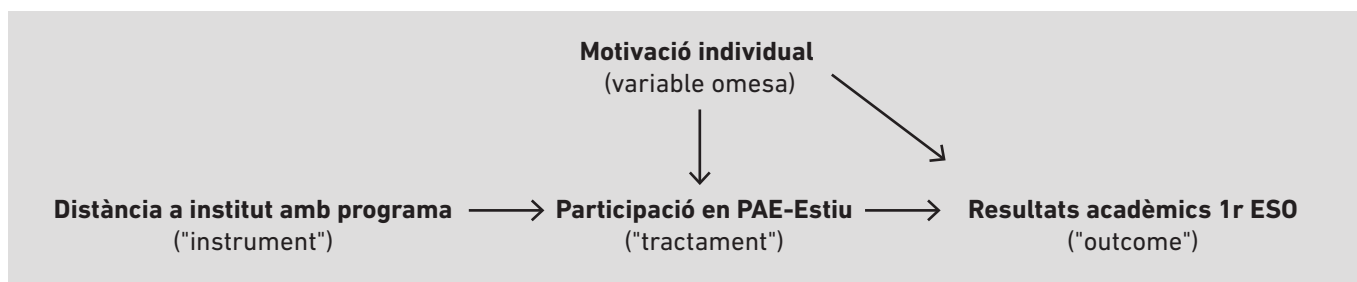
En el marc de l'avaluació d'impacte, una variable instrumental (o instrument) és un factor que es troba clarament vinculat a la probabilitat de participar en un programa, però que no manté cap associació directa amb els indicadors d'outcome que es consideren; en altres paraules, l'instrument és quelcom exogen al model explicatiu en qüestió (Imbens, 2014; Porter, 2012).

Són diverses les variables instrumentals que han estat utilitzades per avaluar l'impacte de polítiques i programes educatius (Schlotter et al., 2010). Alguns estudis han considerat canvis legislatius o normatius que modifiquen la probabilitat d'accedir a un determinat programa o de trobar-se afectat per una determinada política (canvis sense relació directa amb els outcomes d'interès): per exemple, un canvi en la durada de l'etapa d'escolarització obligatòria, la qual provoca variacions en els anys d'escolarització dels individus de diferents cohorts que van més enllà de les seves habilitats o disposicions escolars.

En altres contextos, s'han instrumentat factors "naturals" o de tipus geogràfic. Aquest és, per exemple, el cas de la distància geogràfica entre la residència dels individus i el lloc on s'ofereix el programa. La hipòtesi de partida és que, a igual distància entre residència i localització del programa en qüestió, el fet que uns individus s'hi adrecin i uns altres no ho facin no és aleatori. Sobretot quan estem davant de programes voluntaris, aquest fet ens parla de diferències en els atributs d'uns i altres individus sovint difícils d'objectivar (motivació, disponibilitat, iniciativa, etc.) i, tanmateix vinculats als outcomes en qüestió. En canvi, assumim que a iguals característiques (observables i no observables), els individus que viuen a prop de la seu (o seus) on s'ofereix el programa tindran una probabilitat més alta de participar-hi que no pas aquells que hi viuen lluny. No en va, la distància geogràfica pot implicar dificultats objectives manifestes d'accés al programa.

Posem per cas que el Departament d'Ensenyament decideix plantejar una extensió del PAE al període de vacances d'estiu, centrant-se en l'atenció als alumnes que començaran l'ESO al setembre i que ho faran amb un baix nivell en competències instrumentals. El reforç escolar el realitzaran estudiants universitaris voluntaris a raó de dues hores diàries al llarg del mes de juliol. Imaginem també que, per uns o altres motius, el Departament decideix concentrar aquest recurs arreu del territori: així, el programa s'oferirà només en alguns instituts a cada municipi, si bé hi podrà accedir qualsevol alumne resident al mateix municipi, amb independència del centre de secundària que tingui assignat. En aquest cas, la distància entre el lloc de residència dels alumnes i els instituts amb PAE-Estiu podria arribar a condicionar el sol fet d'adreçar-s'hi. Per contra, si la decisió d'ubicar el programa en un o altre institut satisfà certes condicions⁸, hauríem d'esperar que el fet de viure més o menys lluny d'aquest institut, a igualtat d'altres característiques sociodemogràfiques, no mantindrà cap relació significativa amb els futurs resultats acadèmics dels alumnes (figura 4). Així, amb l'ajut d'aquest instrument inferiríem l'impacte del programa controlant per aquella probabilitat de participar-hi que depèn de variacions en la distància entre residència i institut amb PAE Estiu.

Figura 4. L'avaluació del PAE, amb la distància com a variable instrumental: disseny general



Font: elaboració pròpia

Com de robustes són les estimacions d'impacte basades en l'ús de variables instrumentals dependrà de la credibilitat de l'instrument mateix, això és, de la mesura en què aquest sigui capaç d'acomplir amb la doble condició de vinculació amb la participació i no vinculació amb els outcomes.

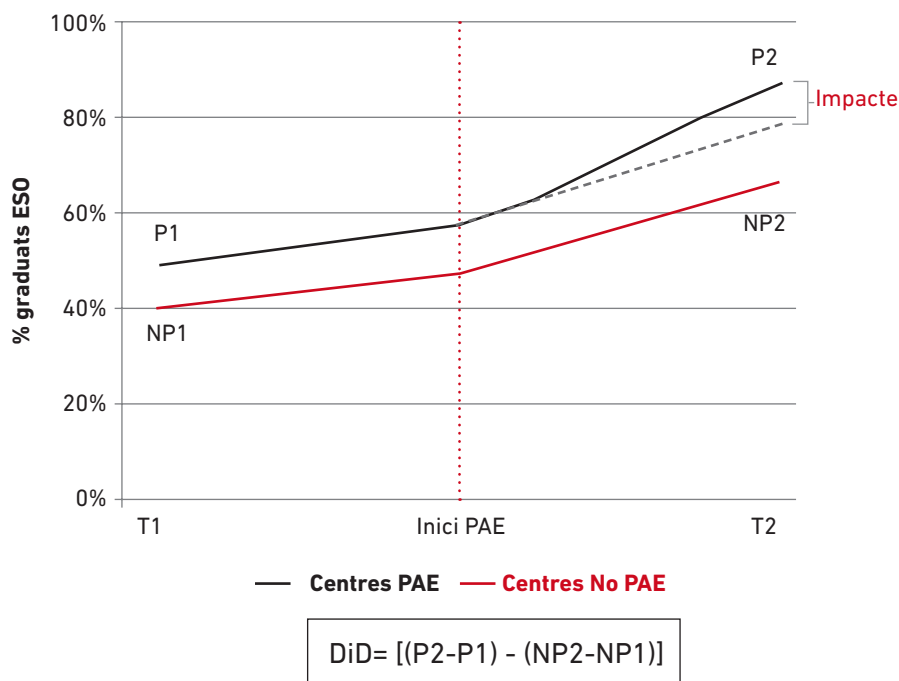
Dobles diferències

Els models de dobles diferències representen un pas més enllà del disseny d'anàlisi "abans-després". A diferència d'aquest disseny, la tècnica de les dobles diferències (o diferències-en-diferències) tracta de controlar els efectes dels factors contemporanis al programa comparant allò que els succeeix als individus o unitats que hi participen amb allò que succeeix a aquells que no hi participen en un mateix període d'observació.

L'aplicació més habitual dels models de dobles diferències és aquella que estableix la comparació entre dos grups d'individus (o d'escoles, regions, països...) al llarg del temps: en un primer moment, cap dels dos grups participa en el programa; en un segon moment, en canvi, un grup hi participa (tractament) i l'altre no (comparació). L'estratègia d'estimació de l'impacte que proposen aquests models es basa en la comparació de les diferències en els outcomes de cada grup abans i després de la implementació del programa. Assumint que el biaix de selecció entre ambdós grups roman constant en el temps, la diferència entre aquestes dues diferències assenyalaria l'impacte del programa.

Aplicat al nostre exemple (veure gràfic 3), i fixat al nivell de l'escola, la tècnica de les dobles diferències consistiria a: i) calcular la variació entre els resultats dels alumnes PAE abans i després de començar el programa, així com la variació entre els resultats d'alumnes d'iguals característiques (curs, nivell acadèmic, extracció socio-econòmica, etc.) en centres no-PAE també abans i després de la posada en marxa del programa; ii) calcular la diferència entre aquestes dues variacions. De nou, suposarem que qualsevol diferència pre-programa entre uns i altres centres, sigui aquesta més gran o més petita, s'hagués mantingut constant encara que el programa no hagués existit.

Gràfic 3. Evolució de la proporció d'alumnes-PAE i no-PAE que obtenen el GESO



Font: elaboració pròpia (dades no reals)

El principal punt feble dels models de dobles diferències és la possible existència de factors inobservables que variïn al llarg del temps, i que ho puguin fer amb intensitats diferents entre participants i no participants. Així, si la motivació d'alumnes-PAE i no-PAE varia al llarg del temps, i no podem garantir que ho faci en igual mesura entre uns i altres, no podem estar segurs que diferències en els seus resultats després del programa no siguin causades per aquestes variacions. Si volem que els resultats d'una avaluació d'impacte que utilitzi un disseny de dobles diferències resultin creïbles, caldrà oferir arguments que permetin descartar l'existència d'aquesta limitació.

Taula 1. Quatre mètodes d'avaluació d'impacte: característiques bàsiques

MÈTODE	TRETS PRINCIPALS	AVANTATGES	DADES REQUERIDES	LIMITACIONS
Aleatorització – disseny experimental	<p>Requereix l'assignació aleatòria de la població diana al grup de tractament i al grup de control</p>	<ul style="list-style-type: none"> • Si s'implementa correctament, proporciona les estimacions d'impacte més robustes, ja que elimina d'arrel el biaix de selecció • Els resultats que proporciona són transparents, fàcils d'entendre i difícils de qüestionar • Existeixen diferents dissenys que es poden adaptar al context de la política i les circumstàncies de la intervenció 	<ul style="list-style-type: none"> • Requeriments bàsics per controlar l'assignació aleatòria a la participació • Registre estricta de les assignacions a cadascun dels grups • Conveniència de recollir dades basals • Es requereixen mesures de resultat per al grup de tractament i de control 	<ul style="list-style-type: none"> • Refús a formar part del grup de control • Consentiment dels participants no sempre fàcil d'obtenir • L'aleatorització pot influir en el perfil dels candidats que es postulen per a l'experiment • El fet que els participants coneguin el seu estatus com a tractament o control pot alterar el seu comportament i influir en els resultats • Reserves ètiques a l'aleatorització • Volum important de requeriments de disseny i planificació
Dissenys de regressió discontinua	<p>Els individus prenen part en la intervenció segons si la seva puntuació en una variable continua es troba per sobre o per sota d'un llindar predeterminat. Aquest punt de tall marca la distinció entre grup de tractament i comparació</p>	<ul style="list-style-type: none"> • Es pot aplicar tant quan la discontinuïtat entorn del punt de tall és nítida i dicotòmica (disseny "sharp") com quan és difusa (disseny "fuzzy") • Sota certes condicions, proporciona estimacions d'impacte robustes (no esbiaixades) 	<ul style="list-style-type: none"> • La determinació del punt de tall ha de ser independent dels valors que s'assignen als individus de la població diana • Es requereixen dades sobre els valors en la variable d'assignació i mesures de resultat per als individus del grup de tractament i comparació 	<ul style="list-style-type: none"> • El mètode no és vàlid sense l'existència d'una variable contínua d'assignació • Les anàlisis poden resultar complexes i en ocasions incertes quan les raons que governen l'establiment del punt de tall no són clares o quan la mostra d'individus al voltant del llindar és reduïda • Poden plantejar-se dificultats a l'hora d'interpretar els resultats de les estimacions i de generalitzar les conclusions extremes

MÈTODE	TRETS PRINCIPALS	AVANTATGES	DADES REQUERIDES	LIMITACIONS
Variables instrumentals	S'utilitza una variable o instrument com a font de variació en la participació en el programa. L'instrument ha de ser exogen, això és, no relacionat amb els resultats considerats	<ul style="list-style-type: none"> Sota certes condicions, proporciona estimacions d'impacte robustes (idea d'experiment natural), perquè resol el problema del biaix de selecció Es pot aplicar en avaluacions retrospectives Permet l'estimació de diferents tipus d'impacte 	<ul style="list-style-type: none"> Requereix dades basals, de resultat i valors relatius a la variable instrumental per als individus del grup de tractament i comparació L'instrument ha de tenir relació amb la probabilitat de rebre el tractament, i afectar els resultats només via la determinació d'aquesta probabilitat L'instrument no pot associar-se a cap altre determinant dels resultats 	<ul style="list-style-type: none"> Sovint és difícil arribar a identificar variables instrumentals vàlides No és fàcil explicar el mètode als no experts La interpretació dels resultats no és fàcil i directa La possibilitat de testar les assumpcions que justifiquen la selecció de l'instrument sovint és limitada
Models dobles diferències	Es comparen les mesures de resultat de participants i no participants abans i després de la intervenció	<ul style="list-style-type: none"> Es controlen determinats factors inobservables que poden generar diferències sistemàtiques entre participants i no participants Es pot utilitzar en conjunció amb tècniques d'emparellament ("matching") Es pot utilitzar amb dades de secció creuada ("cross-section") o de panell 	<ul style="list-style-type: none"> Requereix dades basals i de resultat de participants i no participants abans i després de la intervenció Per tal de validar el mètode, es poden requerir múltiples observacions pre-intervenció, tant per a participants com per a no participants 	<ul style="list-style-type: none"> S'assumeix que els resultats de participants i no participants segueixen uns patrons d'evolució comuns Les anàlisis acostumen a ser complexes i no sempre fàcils de comunicar No sempre és fàcil obtenir la diversitat de dades pre-intervenció requerides per validar el mètode No es pot utilitzar per estimar efectes múltiples de la intervenció

Font: adaptació de Morris, Tödling-Schönhofer, i Wiseman, 2013, pp. 24-25.

Notes:

- ² La semblança entre l'hipotètic programa PAE i l'existent PIM (Programa Intensiu de Millora, obert als instituts catalans de secundària des del curs 2012-2013) no és casual.
- ³ De forma general, existeix endogeneïtat quan la variable de tractament (rebre o no rebre el PAE) no és independent (exògena) de l'outcome considerat i del context que el condiciona. Són diverses les fonts possibles d'endogeneïtat. En el camp que ens ocupa, és habitual trobar-se amb dues d'elles: quan la variable de tractament es troba determinada pel mateix outcome (endogeneïtat per causalitat inversa), i quan el tractament té relació amb variables no observades igualment relacionades amb l'outcome (endogeneïtat per variables omeses) (Schlotter, Schwerdt, i Woessmann, 2010).
- ⁴ El "biaix de selecció" és un cas concret d'endogeneïtat generada per variables omeses, habitual en el marc de l'avaluació de polítiques i programes. El biaix de selecció es produeix quan existeixen diferències entre tractats i no tractats que són prèvies a l'inici del programa i que poden ser responsables de les diferències entre els resultats d'uns i altres un cop aquest finalitza.
- ⁵ L'ús del terme "grup de control" acostuma a reservar-se al marc de l'avaluació experimental; l'expressió "grup de comparació" és més pròpia dels estudis no experimentals.
- ⁶ A data de març de 2015, i al nostre coneixement, els únics experiments en marxa a Catalunya per avaluar l'impacte de programes formatius els trobem en l'avaluació del programa "Mobilitza't-Mobile" (Barcelona Activa) i la "Maleta de les famílies" (Diputació de Barcelona). "Mobilitza't" és un programa de formació i inserció laboral en el sector del mòbil adreçat a joves amb baix nivell d'estudis; el programa s'inicia a finals de 2013 i es perllongarà fins a finals de 2016. El programa la "Maleta de les famílies" consisteix en la programació de tallers grupals, tutories i seguiments individuals per a famílies amb fills a 4t d'ESO en risc d'abandonament educatiu; la seva implementació es realitza durant els cursos 2014-2015 i 2015-2016. Ambdós programes es plantegen en clau de pilots experimentals i estan essent avaluats per Ivàlua.
- ⁷ La validesa del mètode requereix la satisfacció de dos supòsits bàsics: que cap altra variable rellevant presenta discontinuïtats entorn del mateix punt de tall, i que els individus no trien lliurement en quin costat del punt de tall s'ubiquen. Per tot plegat, estudis basats en la regressió discontinua són sovint qualificats d'experiments "naturals".
- ⁸ Per exemple, que no hagi estat guiada per la pressió de col·lectius de pares i mares d'uns o altres barris especialment motivats.

3. ELS REPTES DE L'EXPERIMENTACIÓ: EXEMPLES

A nivell internacional, comptem a dia d'avui amb un repertori prou ampli d'experiments duts a terme en diferents àmbits de la política educativa. En el món occidental, bona part d'aquests experiments han estat desenvolupats als Estats Units. I tanmateix, molts d'ells permeten il·lustrar quins podrien ser els reptes i quines les potencialitats de l'ús d'aquest mètode en l'avaluació de polítiques educatives a casa nostra. La selecció d'exemples que tot seguit descrivim persegueix justament aquest propòsit.

3.1 L'AUTONOMIA ESCOLAR: ELS EXPERIMENTS DE LES "CHARTER SCHOOLS"

La delegació als centres escolars de majors graus d'autonomia ha passat a ocupar una posició central en el debat educatiu a nivell global. Però, de quina mena d'autonomia estem parlant? I què ens diu l'evidència internacional sobre l'efectivitat d'unes i altres formes d'autonomia?

Per una banda, un bon nombre d'estudis basats en proves internacionals com les de PISA, TIMSS o PIRLS⁹ arriben a la conclusió que el rendiment mitjà dels estudiants millora en aquells centres que disposen d'un ampli marge d'autonomia en l'administració dels recursos (en particular, en l'assignació del pressupost disponible), en la gestió del professorat (quins professors contractar i com incentivar-los) i en la selecció de mètodes pedagògics (Woessmann, 2003; Fuchs i Woessmann, 2007; Woessmann, Luedemann, i Schuetz, 2009). En canvi, s'observa que no sempre allò que pot ser positiu en termes de rendiment mitjà, ho és també en el terreny de l'equitat educativa (Schütz, West, i Woessmann, 2007). Així, el pes de l'origen social en l'explicació dels assoliments escolars s'incrementa en aquells països on una part important de les escoles poden intercedir en el procés de selecció i admissió dels alumnes –i, per tant, caure en la temptació del *cream-skimming*¹⁰). Altres estudis evidencien una clara interacció entre autonomia escolar i determinats mecanismes de rendició de comptes (*accountability*), en el sentit que l'autonomia funciona millor –en termes de nivell mitjà i d'equitat de resultats– quan les escoles són sotmeses a aquests mecanismes (Woessmann, 2005b). Finalment, alguns autors han constatat que els efectes positius que l'autonomia escolar té associats en els països rics desapareixen en entorns empobrits (Hanushek, Link, i Woessmann, 2013).

Sigui com sigui, hauríem de recordar que aquests estudis difícilment aconseguen neutralitzar els problemes d'endogeneïtat propis de la comparativa internacional mitjançant dades de caràcter transversal; amb la qual cosa, difícilment arriben a identificar relacions de causalitat entre política i resultats d'interès (*outcomes*).

Una manera d'eludir aquestes limitacions passa per aprofitar la variabilitat que aquest factor sovint presenta dins d'alguns països o regions. Particularment interessant és la possibilitat de treure partit del fet que alguns centres (habitualment aquells que disposen d'un major nivell d'autonomia) disposen de sistemes d'admissió específics, sovint incorporant un mecanisme de

sorteig. Cas paradigmàtic d'aquesta situació és el de les charter schools.

Les escoles charter comencen a implantar-se als Estats Units des de primers dels anys 1990. Si bé són centres d'iniciativa i gestió fonamentalment privades (sense ànim de lucre), són definides com a part integrant del sistema públic d'educació. Gran part del seu finançament procedeix de fons públics, a canvi que aquestes garanteixin la gratuïtat de la matriculació. Les charter disposen d'un nivell d'autonomia escolar (professorat, pressupost, currículum i pedagogia) molt superior al de les escoles públiques convencionals, i la seva elecció per part de les famílies no es troba subjecta a cap restricció d'assignació zonal; en cas de sobredemanda, l'accés es determina mitjançant un sorteig. Aquesta darrera circumstància brinda una clara oportunitat al disseny d'experiments adreçats a avaluar els impactes d'aquests centres (escoles amb un estatus especial d'autonomia) sobre uns o altres outcomes d'interès.

Un balanç de l'evidència acumulada a dia d'avui assenyalaria que l'efectivitat d'aquest tipus d'escola acostumen a ser força variables. Es constata que les charter funcionen en determinats contextos i no en altres, funcionen per a determinats grups d'alumnes i no per a d'altres (Betts i Tang, 2011; Di Carlo, 2011; CREDO, 2009; Zimmer et al., 2009).

Fixem-nos, per exemple, en l'estudi de Clark et al. (2014) sobre l'impacte de 33 escoles charter de secundària. Aquests centres, tots ells significativament sobredemandats, es distribueixen en tretze estats diferents dels EUA (urbans i rurals) i presenten característiques sociodemogràfiques diverses. Com tots els experiments sobre charter schools, els autors utilitzen els resultats del sorteig d'accés a aquestes escoles per identificar el grup de control corresponent; aquest és, els "perdedors" de la loteria. Aquest mecanisme d'aleatorització garanteix que aquest grup és equivalent al grup de tractament (els "guanyadors" de la loteria) en totes aquelles característiques que poden tenir a veure, per exemple, amb la voluntat d'escolaritzar-se en una escola especial com ho són les charter. En conjunt, l'estudi identifica 1.400 alumnes tractats i 930 controls, i segueix l'evolució dels seus assoliments escolars al llarg de dos cursos. Com a resultat general, es constata que, de mitjana, les charter analitzades no generen cap impacte significatiu (ni positiu ni negatiu) en el rendiment acadèmic global dels alumnes que s'hi escolaritzen.

En canvi, estudis com els d'Angrist et al. (2013) i Abdulkadiroglu et al. (2009) sobre les escoles charter de Boston, o com els de Hoxby et al. (2009) i Dobbie i Fryer (2011) sobre les charter de la ciutat de Nova York, tots basats en la mateixa estratègia d'identificació (guanyadors i perdedors de la loteria d'accés), aporten evidències sòlides de casos d'èxit d'aquest perfil d'escoles, principalment en la millora dels resultats acadèmics i de continuïtat educativa dels alumnes més desavantatjats¹¹.

Val a dir que aquesta variabilitat en l'efectivitat de les charter schools ha portat no pocs experts a sostenir l'argument que, charter o no charter, allò que importa és el que l'escola fa amb

l'autonomia de què disposa (Fryer, 2011a).

A CASA NOSTRA...

- A Catalunya, l'any 2005 es posà en marxa el Projecte per a la Millora de la Qualitat dels Centres Educatius, conegut com a Projecte d'Autonomia de Centres (PAC). Aquest projecte comportava per als centres que s'hi adherien una entrada important de recursos (principalment en forma de subsidi econòmic) i la delegació d'un alt grau d'autonomia en funcions relacionades amb els àmbits pedagògic, organitzatiu, i de gestió de recursos humans i materials. A canvi, els centres s'obrien a l'avaluació externa i a la rendició de comptes. Entre 2005 i 2009, 635 centres escolars de primària i secundària van participar en el PAC. El mateix any 2009, la Llei d'Educació de Catalunya (del 10 de juliol de 2009) s'encarregaria de consagrar l'autonomia escolar com un dels principis rectors del sistema educatiu, i el Decret d'Autonomia dels Centres Educatius (del 3 d'agost de 2010) de traslladar els criteris del PAC al conjunt de centres públics i concertats de Catalunya.
- En aquest marc, han estat diversos els avenços que s'han promogut en la delegació als centres d'un major grau d'autonomia en les seves diferents dimensions; en la dimensió pedagògica mitjançant el foment dels projectes d'innovació pedagògica (entre d'altres instruments); en les dimensions organitzativa i de gestió mitjançant el reforçament i professionalització de les direccions dels centres (entre d'altres mesures).
- No podem dir, però, que la introducció d'aquestes reformes hagi vingut acompanyada de provés empíriques sòlides de la seva efectivitat, al temps que no han acabat d'establir les condicions de dissenys que haurien de fer possibles la seva avaluació d'impacte. És especialment destacable la manca d'una avaluació rigorosa del PAC, sobretot considerant que el pla fou inicialment formulat com una prova pilot.

3.2 EDUCACIÓ EN LA PETITA INFÀNCIA: PERRY PRESCHOOL I HEAD START

Diferents estudis s'han aproximat als efectes de l'educació preescolar –entesa com a etapa no obligatòria de l'educació infantil– sobre el progrés educatiu dels alumnes tot comparant les realitats institucionals i els resultats educatius de diferents països i regions (Schütz, 2009; Schütz, Ursprung, i Woessmann, 2008). Recordem, però, les limitacions que té la comparació internacional com a mètode per estimar impactes; limitacions que es deriven de la potencial omisió de variables rellevants per a l'anàlisi. Per altra banda, al nivell individual, cal comptar amb la presència d'un biaix de selecció important entre els col·lectius que escolaritzen els seus fills a edats primerenques. Encara que aconseguim comparar famílies “usuàries” i “no usuàries” semblants en allò observable (per exemple, característiques sociodemogràfiques), res no ens assegura que unes i altres no siguin diferents en altres trets inobservables que podrien estar relacionats amb l'outcome d'interès (per exemple, quant al valor que atorguen a la institució escolar, o quant a la capacitat d'adoptar determinades estratègies educatives en l'àmbit domèstic).

Superant aquestes limitacions, un bon nombre d'estudis han tret partit del disseny experimental que ha guiat el pilotatge o desenvolupament de diferents programes d'educació infantil als Estats Units. Exemples clàssics d'aquestes iniciatives són el Perry Preschool Program (iniciat l'any 1961 a la ciutat d'Ypsilanti, Michigan) i l'Abecedarian Project (iniciat el

1972 a Carolina del Nord). Més recent en el temps, situaríem el projecte d'estudi experimental del Head Start (iniciat el 2002 a nivell federal). Ens referim breument al primer i últim d'aquests experiments.

Així, a inicis de la dècada dels seixanta, l'experiment del Perry Preschool s'articulà de la forma següent: 123 nens de tres i quatre anys procedents de famílies desfavorides van ser aleatòriament distribuïts entre el grup de tractament (58 nens) –inclou: programa d'escolarització diària i en grups reduïts durant almenys un curs escolar, visites d'especialistes a domicili i reunions familiars col·lectives periòdiques– i grup de control (65 nens) –no reberen cap dels serveis anteriors.

Diferents estudis han pogut comparar l'evolució dels resultats d'un i l'altre grup en el curt, mig i llarg termini (fins arribar als 40 anys d'edat dels participants). De forma general, aquestes avaluacions conclouen que el programa tingué impactes positius destacables en l'àmbit de les capacitats cognitives. Aquests impactes, però, es produïen bàsicament en el curt termini, i s'esvaïen durant els primers cursos de l'educació primària. Per contra, s'han pogut documentar efectes del programa en el llarg termini, en concret, beneficis en els terrenys laboral (més ingressos i més estabilitat) i social (menys pràctiques de risc i delictives) un cop arribada a l'adulthood (Heckman, Moon, Pinto, Savelyev, i Yavitz, 2010a, 2010b). Tot plegat explica els bons resultats que el programa ha obtingut sempre que ha estat sotmès a exàmens cost-benefici (Barnett, 1985; Belfield, Nores, Barnett, i Schweinhart, 2006).

Les conclusions a què arriben les avaluacions d'impacte del programa Head Start s'adiuen, en part, amb els resultats del projecte Perry School. Head Start és el programa de provisió de serveis d'educació infantil de més abast a nivell federal. Encetat l'any 1965, el 2002 es posa en marxa un desenvolupament experimental del programa, conegut com Head Start Impact Study. L'experiment implica 23 estats, 383 centres preescolars i 4.667 infants de tres i quatre anys d'edat (tots ells de nou ingrés). Entre les famílies interessades, s'assignà aleatòriament la possibilitat d'inscriure's en un centre Head Start; a la meitat d'elles se'ls oferí aquesta opció (infants tractats), a la resta no (infants controls). Les darreres avaluacions d'aquest experiment conclouen (Peck i Bell, 2014; Puma et al., 2012): a) el programa té impactes cognitius positius, sobretot en l'àmbit lingüístic, mentre dura la participació en el programa i fins a l'inici del primer curs de primària; b) els efectes no cognitius són poc rellevants en el curt termini, però en canvi, alguns d'ells (sobretot els que fan referència a l'àmbit socioemocional) es fan evidents en el moment de cursar el tercer curs de primària. En tot cas, allà on es detecten impactes (cognitius o socioemocionals), aquests són especialment remarcables entre els infants de famílies socioeconòmicament més vulnerables. Avaluacions quasi-experimentals d'edicions anteriors del Head Start indiquen que les cohorts que hi participaren veieren sensiblement incrementades les seves probabilitats d'arribar a la universitat i disminuïdes les seves taxes de criminalitat.

Què ho fa que l'escolarització en la petita infantesa pugui produir beneficis cognitius que s'esvaeixen en el curt termini i en canvi mostri impactes positius a llarg termini en diferents variables clau del cicle vital (nivells de graduació, condicions laborals, criminalitat, salut, etc.) és una qüestió encara no resolta (Duncan i Magnuson, 2013)¹². En canvi, la literatura sí sembla coincidir en la constatació que són els infants de famílies socioeconòmicament i culturalment més vulnerables els qui més acostumen a beneficiar-se de l'escolarització a edats primerenques. Les principals revisions sistemàtiques sobre la qüestió (Camilli, Vargas, Ryan, i Barnett, 2010; Gorey, 2001; Manning, Homel, i Smith, 2010; Nelson, Westhues, i MacLeod, 2003) afegeixen dos punts interessants. Primer, que l'abast dels impactes de l'educació infantil s'incrementa quan segueix determinats paràmetres de qualitat (per exemple, es compta amb un professorat especialista i es treballa en grups reduïts); segon, que l'efectivitat d'aquesta educació augmenta a mesura que s'amplia el temps d'exposició a aquest tipus d'escolarització (hores al dia, i anys de durada/edat d'inici).

A CASA NOSTRA...

- Entre els anys 2004 i 2008, el Govern català impulsà l'extensió del nombre de places d'escola bressol pública en el marc del que aleshores es denominà el "mapa de llars d'infants" de Catalunya. Finalitzat aquest període, el nombre d'infants matriculats s'havia incrementat en 21.347. Des d'aleshores, i fins el curs 2011-2012, la cobertura pública de places d'escola bressol no va deixar d'augmentar. A partir d'aleshores, aquesta cobertura inicia un descens significatiu (Blasco, 2015).
- Associat a aquest descens en la cobertura trobem les successives reduccions que la Generalitat ha fet en la seva aportació al finançament de les places públiques d'escola bressol. En tres cursos escolars (del 2010-2011 al 2012-2013), la Generalitat ha passat de finançar el 50% del cost de la plaça pública a l'escolarització 0-3 anys a finançar-ne el 25%, deixant en mans dels municipis i les diputacions la possibilitat de compensar aquesta reducció.
- Tant les mesures d'impuls de la xarxa d'escoles bressol, com les accions de contenció de la despesa en aquesta etapa educativa no han estat plantejades apel·lant a evidències empíriques sobre la major o menor transcendència que aquest recurs pot tenir a casa nostra. En concret, conèixer els impactes que aquesta escolarització pot tenir a curt i mig termini sobre l'evolució educativa d'infants i joves a Catalunya hauria de contribuir a informar tota anàlisi econòmica que es pogués arribar a plantejar en termes de relació cost-benefici dels recursos destinats a aquesta intervenció.

3.3 EL QUE ES PAGA ALS PROFESSORS: INCENTIVS ECONÒMICS EN ENTORNS COMPLEXOS

Estudis comparats acostumen a observar una associació positiva entre els salaris que els professors perceben als diferents països (ajustats per PIB/càpita) i els resultats de rendiment que obtenen els seus alumnes i escoles (Woessmann, 2005a; OECD, 2010; Dolton i Marcenaro-Gutierrez, 2011). Tanmateix, no insistirem en els límits que presenten aquests estudis a l'hora d'identificar relacions de causalitat entre política i outcomes d'interès.

Alhora, l'intent d'identificar els efectes de les diferències salarials entre docents aprofitant possibles variacions en aquest variable dins d'un mateix país no deixa de plantejar problemes metodològics importants. D'una banda, es fa difícil controlar la circumstància que diferències

en el salari dels professors puguin trobar-se associades a la seva experiència o qualificacions, o fins i tot a les condicions en què exerceixen la docència (per exemple, en alguns països és habitual que els professors ocupats en escoles especialment desfavorides rebin una remuneració superior a la resta). No és fàcil separar els possibles efectes d'aquestes variables dels que puguin ser netament atribuïbles al nivell salarial dels docents (Loeb i Page, 2000; Hanushek i Rivkin, 2007). Més encara, en esquemes variables de pagament per resultats (performance-based pay), en què el salari del professor ve determinat per les millores en el rendiment dels seus alumnes, ens podríem trobar amb un problema de causalitat inversa: no és que un millor salari indueixi a millors resultats acadèmics, sinó que és la consecució de millores de rendiment el que provoca millores salarials.

La superació d'aquestes limitacions torna a requerir la identificació d'una font de variació exògena a les relacions que es volen estudiar. I tornariem a situar com a dissenys d'avaluació més robustos aquells que compten amb un mecanisme d'assignació aleatòria al tractament; en el cas que ens ocupa, veure's afectat per algun tipus de canvi en el salari. En aquest sentit, podríem fixar-nos en els experiments adreçats a avaluar l'efectivitat de diversos esquemes d'incentius econòmics per al professorat.

Una part important d'aquests experiments s'ha adreçat a testar diferents modalitats de pagament per resultats, principalment incentius que anomenariem de "voluntat de guany" –el professor rep un increment salarial si aconsegueix millorar el rendiment dels seus alumnes en una certa mesura¹³. Segons apunta l'evidència acumulada, mentre que aquests esquemes poden tenir efectes positius sobre el rendiment dels alumnes en països en desenvolupament econòmic (Glewwe, Ilias, i Kremer, 2010; Sundararaman i Muralidharan, 2011; Esther Duflo, Hanna, i Rya, 2012), en contextos desenvolupats, els seus impactes acostumen a ser nuls (Fryer, 2011b; Springer et al., 2010).

Altres experiments avaluen programes d'incentius econòmics al professorat no condicionats a resultats. Un exemple interessant el trobem en l'experiment conduït entre 2009 i 2011 per Mathematica Policy Research adreçat a avaluar l'efectivitat d'un programa d'incentius econòmics a la mobilitat del professorat més efectiu, programa conegut com "Teacher Transfer Initiative – TTI" (Glazerman, Protik, Teh, Bruch, i Max, 2013). Aquesta iniciativa, finançada pel Departament d'Educació dels EUA, consistia a incentivar el trasllat de professors amb un destacat "valor afegit"¹⁴ a centres de baix rendiment del mateix districte, oferint-los una retribució de 20.000 dòlars a canvi de treballar-hi durant almenys dos anys. L'experiment es duqué terme implicant 115 escoles de primària i secundària situades en deu districtes escolars de set estats diferents. L'objectiu de l'avaluació era estimar els impactes de la intervenció sobre les dinàmiques organitzatives de les escoles i sobre el rendiment acadèmic dels seus alumnes (en lectura i matemàtiques), valorant alhora la mesura en què el programa aconsegueix retenir els professorat traslladat al cap dels dos anys de durada del període d'incentius.

El disseny de l'experiment comportà, en primer lloc, la identificació dels centres de baix rendiment amb places de professorat vacants¹⁵. Els centres foren llavors emparellats d'acord amb el curs i assignatura on es localitzava la vacant, i d'acord també amb les seves puntuacions acadèmiques mitjanes i proporció d'alumnes desavantatjats. Constituïts els blocs de parelles, un mecanisme d'aleatorització s'encarregà de distribuir a quins equips docents de cada bloc s'obria la possibilitat de cobrir la vacant amb un professor TTI i a quins no (i, per tant, l'havien de cobrir a través dels canals convencionals de selecció i contractació); els primers són els equips docents "tractats" (85 equips), els segons representarien els equips docents "controls" (80 equips).

Pel que fa als impactes organitzatius, l'experiment demostra que el fet d'incorporar un professor TTI provoca canvis en la configuració dels equips docents dels cursos implicats; en concret, les escoles tendeixen a equilibrar la composició dels equips docent reassignant professors menys experimentats a aquells equips on s'incorpora un professor TTI. Pel que fa als resultats acadèmics, es constata l'impacte positiu de la iniciativa sobre el rendiment en lectura i matemàtiques dels alumnes de primària, sobretot el segon any del programa –durant el primer any, els impactes són únicament positius per al cas dels alumnes dels professors TTI, no pas per al conjunt dels alumnes dels equips on aquests professors s'incorporen. En canvi, el programa no es mostra efectiu en els centres de secundària. Pel que fa a la ràtio de retenció del nou professorat contractat, via TTI o via convencional, es constata que aquesta no difereix entre uns i altres al cap dels dos anys del programa –en canvi, es mostra significativament superior entre els professors TTI al cap del primer any d'implicació en la iniciativa.

A CASA NOSTRA...

- La definició de les condicions salarials i d'incentivació econòmica dels docents és un element central de les polítiques de professorat, al costat d'altres àmbits d'atenció com poden ser els sistemes de selecció i accés a la funció docent, les condicions de treball i carreres professionals, la formació inicial i continuada, els mecanismes d'avaluació docent, el suport a la professionalització i el lideratge dels centres, etc. A Catalunya, bona part d'aquests àmbits han estat objecte d'intervencions i reformes al llarg dels darrers anys (Bonal i Verger, 2013).
- En el cas concret dels complements i incentius salarials al professorat, podríem fer referència a les mesures preses en els darrers anys amb relació als denominats estadis de promoció. Així, l'Acord GOV/29/2012, de 27 de març, i la posterior Ordre ENS/330/2014, de 6 de novembre, adequen i estableixen el procediment de promoció docent per estadis per al període 2012-2015, modificant la normativa de reconeixement dels sexennis docents vigent des de 1994. Segons les disposicions actuals, per assolir un estadi es requereixen deu crèdits. De forma general, sis d'aquests crèdits s'obtenen segons els anys de servei actiu (nou crèdits en el cas del primer estadi); 4 crèdits (1 crèdit en el primer estadi) s'aconsegueixen en funció de l'acumulació de determinats mèrits. Un d'aquests mèrits remet a la implicació del docent en el millorament dels resultats del centre. En aquest punt, una de les mesures a considerar passa per l'aplicació al professor dels resultats de les avaluacions dels centres (qualificacions ordinàries, proves de competències, nivells de graduació i continuïtat educativa, etc.).

- En concret, se li atorgarà fins a 1 crèdit per cada curs escolar en què el centre assoleixi una millora significativa en el seu rendiment o mantingui uns bons resultats en els seus indicadors. La superació d'un estadi representa un increment salarial mensual d'entre 100 i 130 euros, un import que s'afegeix al d'altres complements retributius (de destinació i específics, i triennis docents).
- Aquest sistema de mèrits introdueix una certa lògica de pagament per resultats en les condicions retributives del professorat. No podem dir, però, que el seu disseny s'hagi fonamentat en cap exercici rigorós d'avaluació ex-ante (per exemple, revisió d'evidències o fonamentació de mesures d'efectivitat docent), alhora que la seva formulació deixa poc marge al plantejament d'avaluacions retrospectives d'impacte.

3.4 LES TIC COM A EINA D'ENSENYAMENT-APRENTATGE: ACCESS TO ALGEBRA I

La literatura internacional sobre l'ús de les TIC com a eina d'aprenentatge acostuma a atribuir-li impactes generals nuls o modestos. En canvi, l'aplicació de les TIC incrementa notablement la seva capacitat d'impacte quan forma part d'una estratègia pedagògica globalment planificada i quan és supervisada per tutors o docents (Cheung i Slavin, 2012a, 2012b; Hew i Cheung, 2013).

Fixem-nos, per exemple, en l'avaluació que Heppen et al. (2012) fan del programa Access to Algebra I. Aquest programa consisteix en el desenvolupament d'un curs online d'àlgebra basat en tres recursos bàsics: continguts pràctics interactius i software de seguiment i autoavaluació; professorat especialista a distància, i responsables de centre encarregats de la supervisió dels alumnes i el contacte amb professors i famílies. La prova pilot d'aquest programa es plantejà seguint un disseny experimental. S'identificaren 68 centres candidats a desenvolupar el programa, tots ells centres de secundària inferior (middle schools) ubicats en àrees rurals dels estats de Maine i Vermont. Cap d'aquests centres oferia una assignatura específica l'àlgebra en l'últim curs d'aquesta etapa educativa (curs vuitè) –això és, els continguts bàsics d'àlgebra s'impartien en el marc de l'assignatura ordinària de matemàtiques. La meitat d'aquests centres van ser aleatòriament seleccionats per desenvolupar el programa Access to Algebra I dins el currículum dels seus estudiants de vuitè curs, mentre l'altra meitat (centres control) continuaren desenvolupant el seu curs presencial tradicional de matemàtiques. L'estudi compara els progressos acadèmics d'aquells alumnes que, abans del moment de l'aleatorització, havien estat identificats pel professorat d'uns i altres centres com a alumnes aptes per rebre el programa, i es fixa en dues mesures d'outcome: a) puntuacions dels alumnes en proves estandarditzades d'àlgebra i altres competències matemàtiques realitzades a final de curs, i b) nivell de seguiment de l'assignatura d'àlgebra en cursos posteriors (fins a desè). Doncs bé, l'avaluació demostra que els alumnes que participen en el programa obtenen un nivell competencial en àlgebra sensiblement superior als alumnes elegibles dels centres control, que aquest guany no es produeix a expenses d'altres aprenentatges matemàtics no algebraics (és a dir, els alumnes del programa assoleixen el mateix nivell de matemàtiques convencionals que els alumnes que les estudien en els centres control), i que tot plegat

estimula els estudiants participants a continuar cursant l'assignatura d'àlgebra en cursos més avançats.

A CASA NOSTRA...

- Dos anys després de la seva implementació, el curs 2011-2012 el Departament d'Ensenyament congelà el programa EduCat 1x1, de digitalització de les aules catalanes de secundària i cicle superior de primària. El programa contemplava la provisió a l'alumnat d'ordinadors portàtils com a eina personal d'aprenentatge i treball. En el moment d'ésser paralitzat, el programa es trobava actiu en més de sis-cents centres. Ni l'aposta pel programa, ni la decisió de congelar-lo foren recolzades en cap evidència empírica sobre la seva capacitat d'impacte en el procés d'aprenentatge dels alumnes.
- D'altra banda, des del curs 2013-2014, el Departament d'Ensenyament participa en la implantació del programa mSchools, desenvolupat des del centre mEducation del Mobile World Capital Barcelona. El programa mSchools compta amb el finançament de la fundació GSMA, per valor de 2,5 milions d'euros fins el 2018. El programa vol propiciar un desplegament curricular complet vinculat a la tecnologia mòbil i a les seves aplicacions en l'entorn social i econòmic, i es planteja com a objectiu de fons contribuir a millorar els resultats acadèmics dels alumnes. Una de les mesures estrella d'aquest programa consisteix en la introducció de l'assignatura optativa "Mobilitzem la informàtica" al currículum de 4t d'ESO. En aquesta assignatura els alumnes aprenen, de forma cooperativa i amb l'ajut de mentors on-line, a dissenyar aplicacions mòbils. El curs 2014-2015, més de 11.000 alumnes cursaran l'assignatura. A dia d'avui, desconeixem l'efectivitat de "Mobilitzem la informàtica", o del programa mSchools en el seu conjunt, en l'assoliment dels resultats educatius que es plantegen. Conèixer-ne els impactes hauria de permetre dimensionar millor la seva escalabilitat, apostant per aquells components d'innovació que poguessin mostrar-se més efectius i afinant aquells que ho fossin menys.

3.5 LES TUTORIES INDIVIDUALITZADES: TIME TO READ I SWITCH-ON READING

Parlem aquí d'un mecanisme específic d'atenció a la diversitat, de caràcter eminentment compensatori, les tutories individualitzades o 1x1. Aquest instrument persegueix millorar el rendiment competencial dels alumnes més desavantatjats a través d'un reforç intensiu i personalitzat al marge del grup-classe ordinari. Els programes de tutorització individual poden ser diversos, depenent de si es realitzen en horari lectiu o extra-escolar, del perfil del personal tutor (professors especialistes o no especialistes, para-professionals, voluntaris, etc.), de les matèries treballades, del perfil i edat de l'alumnat atès, de la freqüència i durada d'exposició al programa per part dels alumnes (dosatge), etc.

Alguns d'aquests programes han estat avaluats experimentalment. Per exemple, en l'àmbit del reforç en lectura, podríem fer referència al pilotatge experimental dels programes Reading Recovery (desenvolupats per professors especialistes) (May et al., 2014), Experience Corps (amb tutors voluntaris majors de 55 anys) (Lee, Morrow-Howell, Jonson-Reid, i McCrary, 2012) o Reading Partners (amb tutors voluntaris de perfil divers) (Jacob, Smith, Willard, i Rifkin, 2014), tots ells desenvolupats als Estats Units i adreçats a alumnat de primària amb dèficits en aquest àmbit competencial. També al Regne Unit trobem exemples d'avaluacions experimentals

d'aquest mateix tipus d'intervencions. És el cas dels programes Time to Read i Switch-on Reading.

L'any 2014 Time to Read es desplegava a 100 escoles de primària d'Irlanda del Nord, intervenint sobre més de 1.200 alumnes entre 8 i 9 anys amb dèficits en competència lectora. Mentors voluntaris s'emparellen amb aquests alumnes i hi treballen seguint un esquema de tutorització 1x1 que es desenvolupa al llarg d'un curs escolar. Fins l'any 2008, les sessions de tutoria, de 30 minuts de durada, es realitzaven a raó d'un cop per setmana, i al marge del grup classe de referència. Un primer pilot experimental del programa, conduït entre setembre de 2006 i juny de 2008 per investigadors del Centre for Effective Education (Queen's University Belfast), manifestà resultats decebedors en la majoria dels outcomes considerats (bàsicament, variables relacionades amb la competència lectora, les actituds vers l'escolarització i l'autoestima) (Miller i Connolly, 2012). A partir de 2008, i per indicació del mateix informe d'avaluació, el programa passà a doblar el dosatge de les tutories, situant-se ara en dues sessions setmanals de 30 minuts cadascuna.

Entre octubre de 2008 i juny de 2010 tingué lloc la segona avaluació experimental del programa, duta a terme pel mateix equip d'avaluadors que s'encarregà de la primera (Miller, Connolly, i Maguire, 2012). Aquesta segona prova experimental comptà amb una mostra de 512 alumnes en 50 escoles de primària. El procediment consistí a assignar aleatòriament, dins de cada escola i entre els alumnes elegibles identificats pel professorat, aquells alumnes que rebrien el programa (un total de 263 alumnes) i aquells que no el rebrien (249 alumnes), amb l'objectiu de comparar els resultats corresponents un cop transcorreguts dos anys de seguiment. Els resultats d'aquesta segona avaluació permeteren atribuir al programa impactes positius en outcomes com la capacitat de decodificació, velocitat i fluïdesa en la lectura. En canvi, el programa no es mostrà efectiu en l'àmbit de la comprensió lectora, o a l'hora de fomentar el gaudi per la lectura o incrementar la confiança en l'eficàcia lectora.

Per la seva banda, el programa anglès Switch-on Reading ofereix els alumnes participants (nois i noies a l'inici de l'educació secundària amb baix nivell competencial en lectura) un mínim de 40 sessions de reforç individualitzat de 20 minuts de durada cadascuna, realitzades fora del grup-classe ordinari i programades per un període de deu setmanes. El pilotatge experimental d'aquest programa va comptar amb el finançament de l'Education Endowment Foundation¹⁹, i es dugué a terme entre desembre de 2012 i abril de 2013 seguint un disseny un xic diferent al convencional (Gorard, See, i Siddiqui, 2014). En aquest cas, es va aleatoritzar el moment d'entrada al programa, no pas el fet de rebre'l o no rebre'l. En altres paraules, ningú es quedava sense participar-hi, només que uns ho feien abans i altres després. Així, dels 300 alumnes elegibles per a Switch-on Reading, s'assignà aleatòriament quins 150 hi entren en una primera fase (trimestre 1) i quins 150 quedaven a l'espera d'entrar-hi en una segona fase (trimestre 2). A tots ells se'ls administrà una prova estandarditzada de comprensió lectora abans de l'aleatorització (pre-test) i just després de finalitzada la primera fase del programa (post-test).

Així doncs, l'estimació de l'impacte té en compte el primer trimestre d'implementació del programa, període en què els alumnes en espera actuen com a control dels alumnes participants en primera instància.

Aquest disseny neutralitza possibles reserves ètiques front l'assignació aleatòria del recurs, i alhora elimina eventuais efectes de desànim entre els individus control. La principal limitació del disseny és que no permet observar impactes més enllà del curt termini (primer trimestre d'implementació del programa). I val a dir que els impactes detectats són força positius. En efecte, la comparació de les puntuacions pre-test i post-test d'individus control i tractament en la primera fase del programa indica que aquest és efectiu per al conjunt d'alumnes, particularment per als alumnes acadèmicament i socialment més vulnerables.

Les conclusions dels experiments Time to Read i Switch-on-Reading corroboren la capacitat d'impacte que la literatura internacional atribueix a les fórmules de tutorització 1x1²⁰. Segons aquesta literatura, per tal que aquests impactes puguin perdurar en el temps, es requereix que les tutories parteixin d'un disseny d'activitats ben planificades i coordinades amb els processos d'aprenentatge que tenen lloc en el grup-classe ordinari (Borman et al., 2007; Correnti, 2009). Així mateix, l'efectivitat de la intervenció s'incrementa a mesura que augmenta el temps i freqüència de realització de les tutories, i quan són encarregades a professorat qualificat o a personal ben format i assessorat (Slavin, Lake, Davis, i Madden, 2011; Ehri, Dreyer, Flugman, i Gross, 2007).

A CASA NOSTRA...

- Enteses com a instrument compensatori o d'atenció a la diversitat, a Catalunya, les tutories individualitzades o tutories 1x1 es troben poc implantades en la vida ordinària dels centres. Val a dir, però, que no són poques les veus que, des de les mateixes escoles i instituts, assenyalen la conveniència d'arribar a articular aquest servei, cosa que acostuma a esdevenir difícil per manca de recursos humans i materials.
- Per la seva banda, en els darrers anys s'han posat en marxa programes de tutorització individual o en grups reduïts desenvolupats en horari extraescolar, sovint amb la col·laboració d'entitats locals, fundacions o universitats. És, per exemple, el cas del programa LEXIT: Lectura per a l'èxit educatiu, iniciat el curs 2011-2012, promogut per la Fundació "la Caixa", la Fundació Jaume Bofill i el Departament d'Ensenyament de la Generalitat de Catalunya, i adreçat a infants de 4t a 6è de primària¹⁶; del programa Èxit 1: Reforç escolar i activitats complementàries, iniciat el curs 2001-2002, impulsat per Consorci d'Educació de Barcelona, i adreçat a l'alumnat de 5è i 6è de primària i 1r i 2n d'ESO¹⁷; o dels Tallers d'Estudi Assistit, vinculats als Plans Educatius d'Entorn iniciats pel Departament d'Ensenyament el curs 2004-2005, i adreçats principalment a l'alumnat de secundària¹⁸.
- Tanmateix sabem poc sobre l'impacte net d'aquests programes, coneixement que ens permetria valorar la conveniència de potenciar-los i/o testar-ne el trasllat al marc ordinari de funcionament dels centres.

3.6 LA IMPLICACIÓ DE LES FAMÍLIES: MALLETE DES PARENTS I READY4KI

Ens fixem ara en l'avaluació d'aquells programes que tracten d'incentivar la implicació de les famílies en els processos i aprenentatges escolars dels fills. Aquests programes són diversos, i abasten tant aquelles actuacions que persegueixen involucrar les famílies en els processos educatius que tenen lloc en l'entorn domèstic, com aquelles intervencions que tracten de promoure la participació de pares i mares en la vida quotidiana de l'escola. Segons l'evidència internacional, quan el que hi ha en joc és la millora del rendiment i actituds dels alumnes, el primer tipus d'actuacions tendeix a ser més efectiu que no pas les destinades a estimular la participació escolar de les famílies (Jeynes, 2005, 2007; van Steensel, McElvany, Kurvers, i Herppich, 2011).

Les principals revisions de la literatura sobre la matèria remarquen, però, que una gran majoria dels estudis que s'han ocupat d'aquesta qüestió es basen en dissenys d'avaluació no del tot rigorosos (Jeynes, 2005, 2007; Mattingly, Prislín, McKenzie, Rodríguez, i Kayzar, 2002; Van Voorhis, Maier, Epstein, Lloyd, i Leung, 2013). I tanmateix, alguns d'aquests programes sí han pogut ser avaluats mitjançant dissenys experimentals.

Fixem-nos en l'avaluació d'un programa pilot de sensibilització dut a terme en escoles socialment desafavorides dels voltants de París. L'actuació s'inclou dins un programa més ampli adreçat a potenciar la relació família-escola (Mallette des Parents), tot i que s'implementa i avalua de forma independent. Consisteix en la programació de dues xerrades amb pares i mares d'alumnes en el darrer curs de l'escolarització obligatòria, alumnes que el professorat té identificats com en risc d'abandonament. Les xerrades corren a càrrec del director del centre, i tenen un component tant informatiu com de sensibilització –es remarca la importància de l'educació i d'escollir bé l'itinerari de continuïtat. En cada xerrada hi participa una mitjana de deu pares o mares. L'actuació cobreix una mostra de 37 escoles voluntàries. A través d'un sorteig s'escull quines classes dins d'aquestes escoles participen en el programa (reben les xerrades). Els alumnes de les classes no seleccionades (en escoles no seleccionades) conformen el grup de control. A partir d'aquí, es comparen els nivells de repetició, abandonament i rendiment dels alumnes tractats i dels alumnes control al llarg de dos cursos escolars, i es mesura si els nivells d'uns i altres és significativament diferent. Els resultats que se n'obtenen acaben identificant impactes significatius i duradors del programa sobre el outcomes educatius clau: els fills de les famílies participants es mantenen més temps que els alumnes control al sistema educatiu (principalment en opcions menys demandades de formació professional), i durant aquest temps rendeixen més i repeteixen menys (Goux, Gurgand, i Maurin, 2013).

Igualment interessant en aquest sentit és el pilotatge experimental de READY4KI!, a la ciutat de San Francisco, un programa que pretén induir millores en la competència lectora dels alumnes d'educació infantil, i que es basa en l'enviament de missatges telefònics SMS a les

seves famílies (York i Loeb, 2014). L'experiment en qüestió tingué lloc al llarg del curs 2013-2014 en 31 centres preescolars del principal districte escolar de la ciutat. Entre les 519 famílies amb fills escolaritzats en aquests centres que, a l'inici de curs, mostraren interès a participar en el pilot s'assignà aleatòriament aquelles que rebrien el protocol de SMS del programa i aquelles que rebrien SMS de tipus "placebo". El protocol de SMS de READY4K! inclou l'enviament de tres missatges de text setmanals durant vuit mesos del calendari lectius. Els continguts dels missatges combinen consells per a la pràctica de la lectura amb els fills, dades sobre la importància dels aprenentatges en la petita infància, i textos d'encoratjament a la tasca educativa de la família en l'àmbit domèstic. Per contra, els SMS "placebo" s'envien a les famílies que actuen com a grup de control a raó d'un cada dues setmanes, i inclouen continguts generals sobre els procediments d'inscripció a l'escola primària o sobre el calendari de vacunacions, entre d'altres. Els resultats de l'avaluació mostren que el protocol de SMS de READY4K! incrementa la implicació de pares i mares en la pràctica lectora dels fills en l'àmbit domèstic, així com la seva participació en la vida escolar. Al seu torn, es constata que increments d'aquesta major implicació reverteix en millores significatives en determinades àrees de la competència lectora dels infants (per exemple, coneixement de l'alfabet i fonètica). Més encara, els beneficis del programa són particularment rellevants entre les famílies negres i hispanes, i els seus fills.

A CASA NOSTRA...

- Catalunya disposa d'una llarga experiència en el desenvolupament d'actuacions dirigides a fomentar la implicació de les famílies en els processos escolars dels seus fills, actuacions que poden ser, i han estat, força diverses: accions informatives (per tal que les famílies coneguin el funcionament dels centres, l'estructuració del sistema educatiu, determinats recursos d'educació no formal, etc.), de sensibilització (adreçades a conscienciar les famílies sobre la importància de l'educació, i sobre la necessitat de recolzar el progrés escolar dels fills), de capacitació (tractant que pares i mares disposin de les competències necessàries per acompanyar els aprenentatges dels fills), de participació (aproximant pares i mares als canals més o menys institucionals de participació familiar en la vida escolar), etc. I són també diversos els agents responsables d'aquestes iniciatives: els mateixos centres escolars, a través dels seus tutors i equips directius; el Departament d'Ensenyament, mitjançant iniciatives informatives i de sensibilització²¹; els ajuntaments, a través dels seus programes i accions família-escola; fundacions i entitats del tercer sector, com a proveïdors d'intervencions públiques i en el marc d'iniciatives filantròpiques, etc.
- Sigui com sigui, no són poques les veus que, des de la comunitat educativa i des del món acadèmic, insisteixen en la conveniència de continuar apostant per mesures que involucrin les famílies en l'escolarització dels fills, en l'àmbit domèstic com en el de la participació escolar (Collet i Tort, 2011; Comas, Escapa, i Abellán, 2014). La defensa d'aquesta aposta, però, guanyaria força si poguéssim demostrar, de forma robusta, l'efectivitat dels programes família-escola a casa nostra; si més no, si estiguéssim en disposició d'incorporar els aprenentatges d'eventuals avaluacions d'impacte –a dia d'avui inexistents– en el seu disseny o reforma.

Notes:

⁹ PISA = Programme for International Student Assessment (OCDE); TIMSS = Trends in International Mathematics and Science Study (IEA); PIRLS = Progress in International Reading Literacy Study (IEA).

¹⁰ Pràctiques selectives adreçades bé a captar l'alumnat de més alt perfil acadèmic i/o socioeconòmic, bé a evitar l'entrada de l'alumnat acadèmicament o socioeconòmicament més vulnerable.

¹¹ Els autors d'aquestes avaluacions atribueixen l'èxit de les charter de Boston i Nova York al model educatiu No Excuses que bona part d'elles desenvolupen. Aquest model aposta per un alt nivell d'exigència i d'expectatives en l'àmbit de les actituds i dels resultats acadèmics dels alumnes, alhora que opta per l'extensió de l'horari lectiu i per sistemes de tutorització i monitorització més estrets i informats (Carter, 2000; Thernstrom i Thernstrom, 2004; Whitman, 2008).

¹² Una possible explicació del curt abast temporal dels impactes cognitius de programes com els aquí esmentats es troba en la baixa qualitat dels centres educatius als quals els infants hi accedeixen després de participar-hi, i contra la qual els programes en qüestió no aconsegueixen "immunitzar" (Currie i Thomas, 2000; Loeb i Lee, 1995). Per la seva banda, Heckman et al. (2012) sostenen que els impactes d'aquests programes observables en l'edat adulta es deuen principalment als canvis que aquests aconsegueixen en el terreny actitudinal i motivacional dels infants.

¹³ L'experiment conduït per Fryer et al. (2012) amb 150 professors de nou escoles a Chicago Heights permet comparar l'impacte d'esquemes de "voluntat de guany" amb l'efectivitat d'un esquema d'incentius basat en la "por a la pèrdua" (loss aversion), on els professors són bonificats a l'avançada, essent posteriorment obligats a tornar part d'aquesta bonificació si al final del període estipulat no s'aconsegueixen els resultats esperats. L'estudi conclou que l'esquema de por a la pèrdua, sobretot quan es defineix d'acord amb resultats-guany col·lectius, funciona millor que l'esquema tradicional d'incentius individuals.

¹⁴ Les mesures de "valor afegit" del professorat tracten de descriure la contribució que aquests fan als canvis en el rendiment acadèmic dels alumnes, mantenint constants altres variables que escapen al seu control; per exemple, l'extracció socioeconòmica dels alumnes o el seu rendiment previ (McCaffrey et al. 2004; Lipscomb et al. 2010). Pel que fa a l'experiment que ara ens ateny, es consideren elegibles a la TTI aquells professors la puntuació de "valor afegit" dels quals els situa entre el 20% de professors més efectius del districte.

¹⁵ Es consideren de baix rendiment les escoles que es troben entre el 20% de centres amb pitjor nivell acadèmic mitjà del districte.

¹⁶ Més informació a <http://www.lectura.cat/>

¹⁷ Més informació a http://www.edubcn.cat/exit/reforc_escolar_i_activitats_complementaries

¹⁸ Més informació a <http://ves.cat/hrLH>

- 19** *Finançament rebut en el marc de la primera convocatòria competitiva d'ajuts a projectes d'aquesta institució (any 2012).*
- 20** *Avaluacions experimentals d'altres programes de tutorització individualitzada apunten en aquesta mateixa direcció. Veure, per exemple, les conclusions de l'avaluació del programa 1x1 Catch Up Numeracy, per a alumnes d'educació primària amb dèficits en matemàtiques (Rutt, Easton, i Stacey, 2014), o els resultats del programa Becoming a Man, centrat en el treball conductual amb nois de secundària en risc d'abandonament escolar (Cook et al., 2014).*
- 21** *Veure, per exemple, el programa Família i Escola a <http://familiaiescola.gencat.cat/ca/>*

4. ELS REPTES DE L'AVALUACIÓ QUASI-EXPERIMENTAL: EXEMPLES

Com hem indicat més amunt, són diversos els mètodes quasi-experimentals que es poden utilitzar per avaluar els impactes de programes educatius. Si bé aquests mètodes no neutralitzen el biaix de selecció de forma tan neta com ho fa el disseny experimental, sota determinades circumstàncies, alguns d'ells poden oferir estimacions d'impacte certament robustes. Aquest és sobretot el cas de la regressió discontinua, les variables instrumentals i els models de dobles diferències. Seguidament il·lustrem la seva potencialitat en àmbits d'intervenció rellevants per a l'agenda educativa catalana.

4.1 DISSENYS DE REGRESSIÓ DISCONTÍNUA

4.1.1 DE QUÈ SERVEIX REPETIR CURS

No és fàcil arribar a identificar els efectes atribuïbles al fet de repetir un determinat curs. Així com ho acostuma a ser bona part de la política d'avaluació d'alumnes, ens trobem davant d'una decisió avaluativa – no promocionar un alumne – força idiosincràtica. Dins d'un mateix sistema educatiu, fins i tot existint criteris comuns quant a l'aplicació d'aquest instrument, podem acabar trobant interpretacions i usos de la repetició força diversos entre escoles. Així, pot perfectament ser que determinats centres tendeixin més que altres a evitar la pràctica de la repetició, per motius que poden ser ben diversos; pedagògics (confiança que la repetició no ajuda al progrés escolar dels alumnes), econòmics (estalvi dels costos associats a la repetició), o d'imatge (d'efectivitat del centre i/o de bon perfil d'alumnat).

L'avaluació dels impactes de la repetició de curs no es pot basar en la comparació dels resultats a curt, mig o llarg termini dels alumnes que repeteixen un determinat curs amb els resultats dels alumnes que el superen. Necessitem comptar amb mètodes que, d'entrada, eliminin els biaixos entre les característiques de repetidors i no repetidors. Sent difícil aplicar un disseny experimental a l'avaluació de la repetició, alguns estudis s'hi han aproximat mitjançant diferents dissenys de regressió discontinua.

L'estudi de Jacob i Lefgren (2009) ofereix un exemple interessant en aquest sentit. Els autors aprofiten la introducció l'any 1997, en les escoles públiques de Chicago, d'un test de competències extern i estandarditzat adreçat a determinar la promoció o repetició de curs dels alumnes en moments clau del seu itinerari escolar. En particular, mitjançant una aproximació de regressió discontinua, els autors comparen, a) els nivells de graduació en finalitzar l'educació secundària superior d'aquells alumnes que repetiren curs a 6è o a 8è (secundària inferior) obtenint en el test de competències una puntuació just per sota de la mínima exigida per passar de curs, amb, b) els nivells dels alumnes de les mateixes cohorts que promocionaren obtenint una puntuació just per sobre del llindar mínim requerit²². D'aquesta manera demostren que el fet de repetir 6è no afecta la probabilitat de finalitzar amb

èxit la secundària; en canvi, repetir 8è sí impacta negativament sobre aquesta probabilitat, particularment entre les noies afroamericanes.

Dos estudis de Marco Manacorda utilitzen aquest mateix mètode per avaluar els efectes de la repetició de curs a l'Uruguai (2008, 2012). En un d'aquests estudis (Manacorda, 2012), l'autor aprofita la instauració en aquest país, l'any 1996, d'una normativa que establia la repetició de curs automàtica dels alumnes de secundària amb més de tres assignatures suspeses en l'avaluació final²³. Aprofitant aquesta circumstància, i fent servir un model de regressió discontinua²⁴, es comparen els resultats educatius dels alumnes que es troben a banda i banda del llindar de les tres matèries pendents, considerant que el fet de situar-se just per sobre d'aquest punt de tall (i repetir) o just per sota (i promocionar) no reflecteix diferències rellevants en les característiques inicials d'uns i altres alumnes. D'aquesta manera, l'autor mostra com el fet de repetir redueix significativament la continuïtat dels alumnes a l'educació formal i, per tant, els seus nivells de graduació.

Els resultats de les avaluacions a les quals ens hem referit es troben en sintonia amb les conclusions a què arriben les principals revisions sistemàtiques de la literatura sobre aquesta qüestió (Allen, Chen, Willson, i Hughes, 2009; Holmes, 1989; Jimerson, 2001), a saber: que la decisió de fer repetir un alumne no tan sols és una decisió econòmicament costosa, sinó que no té impactes significatius (o quan se'n detecten són impactes negatius) ni sobre les actituds ni sobre el progrés educatiu dels alumnes repetidors.

A CASA NOSTRA...

- La política de repeticions és, a Catalunya i Espanya, una pràctica força més estesa del que ho és en molts altres països del nostre entorn, tant a l'educació primària com a la secundària (Dupriez, Dumay, i Vause, 2008; Goos et al., 2013).
- Sense anar més lluny, d'ençà de l'entrada en vigor de la Ley Orgánica General del Sistema Educativo (LOGSE, de 4 d'octubre de 1990), han estat diverses les mesures aplicades per tal de regular l'ús de la repetició per part dels centres. S'ha apostat per sistemes de promoció automàtica entre cursos (amb possibilitat de repetir en finalitzar cicles), com també per la repetició de qualsevol curs a partir d'un determinat nombre d'assignatures suspeses. Paral·lelament, han estat també diverses i canviants les vies obertes a recuperar les assignatures pendents (recuperacions durant el curs o al setembre).
- A Catalunya, el model actual als ensenyaments secundaris obligatoris estableix la repetició de curs per als alumnes amb tres o més matèries suspeses en l'avaluació final extraordinària pròpia de cada curs (Ordre EDU/295/2008). Excepcionalment, si l'equip docent així ho considera o si l'alumne ha exhaurit el nombre màxim de repeticions permeses a l'etapa (dues), es podrà eximir de la repetició a alumnes amb 3 o més assignatures suspeses. Per la seva banda, a partir del curs 2011-2012, l'Ordre ENS/56/2012 modifica el calendari de l'avaluació final extraordinària, restaurant els exàmens de setembre.
- Cap d'aquests models, però, ha estat empíricament contrastat, menys encara avaluat en termes dels seus possibles impactes sobre outcomes educatius clau – principalment, millora i recuperació del rendiment i assoliments formatius dels alumnes repetidors.

4.1.2 BEQUES I AJUTS ALS ESTUDIANTS

Els ajuts directes que les administracions, o algunes entitats filantròpiques, concedeixen als estudiants amb l'objectiu de facilitar la seva escolaritat són un instrument clau de política educativa. Tanmateix, avaluar l'impacte que aquest instrument pot arribar a tenir sobre les oportunitats educatives dels estudiants no sempre és una tasca senzilla.

En el pla de la comparativa internacional, acostuma a posar-se de manifest que no existeix una relació clara entre el nivell global de despesa en educació (mesurada en referència al PIB) i els resultats educatius mitjans dels països. Més encara, la debilitat d'aquesta associació és especialment evident entre els països desenvolupats (OECD, 2014). Tot sembla indicar que, a partir d'un determinat llindar, l'efectivitat de la despesa educativa es converteix més en una qüestió d'estratègia (com es gasta) que de volum (quant es gasta). En aquest punt, alguns estudis han evidenciat l'existència d'una associació positiva entre, d'una banda, el pes que, en cada país, té la despesa en ajuts als estudiants sobre el total de la despesa en ensenyaments no universitaris, i de l'altra, els seus assoliments agregats (nivells formatius assolits) (Alegre i Benito, 2014).

Novament hauríem de ser cauts en la interpretació d'aquests resultats, no només pels problemes d'endogeneïtat que plantegen, sinó també per la gran diversitat d'esquemes d'ajuts que acostumen a aplegar-se dins d'una única variable: beques i premis d'estudi, subsidis directes a famílies per al costejament de les despeses derivades de l'escolarització, préstecs preferencials, etc. Igualment, la comparativa entre països no sempre aconsegueix distingir entre les diferents modalitats d'accés a aquests ajuts, en funció de mèrits escolars (model "merit-based") i/o determinades mesures de deprivació econòmica (model "need-based").

Fins i tot quan es concentra l'interès en un determinat programa d'ajuts desenvolupat en un únic territori, es planteja el repte d'identificar grups de comparació creïbles, això és, estudiants no beneficiaris veritablement comparables als estudiants que efectivament reben l'ajut. Són nombrosos els biaixos que poden produir-se entre beneficiaris i no beneficiaris: biaixos competencials, en els models d'ajut "merit-based"; o biaixos socioeconòmics i culturals, en els models "need-based". Fins i tot si poguéssim controlar la incidència d'aquests biaixos, podria perfectament ser que beneficiaris i no beneficiaris diferissin en variables no observables relacionades amb els outcomes d'interès²⁵.

Una manera de superar aquestes limitacions passaria per treure partit dels llindars en els criteris d'elegibilitat que defineixen bona part dels programes de beques escolars, mitjançant l'aplicació de dissenys de regressió discontinua. En el cas d'aquelles beques la concessió de les quals exigeix tenir prèviament aprovades un mínim d'assignatures o haver obtingut unes qualificacions o puntuacions acadèmiques mínimes, es tractaria de comparar el que els passa als estudiants que en són beneficiaris i que es troben just per sobre del mínim exigít

d'assignatures aprovades o de qualificacions obtingudes (grup de tractament) amb el que els succeeix als estudiants no beneficiaris que es troben just per sota dels llindars acadèmics en qüestió (grup de comparació). En el cas de les beques que s'atorguen per renda, es compararien els alumnes becats els ingressos familiars dels quals es troben just per sota del màxim exigint (grup de tractament) amb els alumnes amb rendes familiars just per sobre d'aquest mateix màxim (grup de comparació). En qualsevol cas, és d'esperar que uns i altres alumnes, els que es troben just per sota i els que es troben just per sobre d'aquests llindars (acadèmics o de renda), s'assemblin no només en aquelles característiques observables que puguem controlar, sinó també en aquells inobservables que puguin estar relacionats amb la variable d'assignació.

Aquest procediment ha estat emprat en l'estimació d'impactes de diferents programes de beques universitàries (basats en rendiment o en renda), principalment als Estats Units (DesJardins, McCall, Ott, i Kim, 2010; Rubin, 2011; Scott-Clayton, 2011; Leeds i DesJardins, 2014; Curs i Harper, 2012; Bettinger, Long, Oreopoulos, i Sanbonmatsu, 2012). En aquest context, l'evidència acumulada tendeix a concloure que, mentre que els ajuts basats en la renda afavoreixen principalment l'accés a la universitat, les beques basades en el rendiment incrementen les taxes de permanència i graduació. En conjunt, la capacitat d'impacte d'unes i altres beques s'amplifica a mesura que augmenta el seu import i es simplifiquen els procediments administratius per a la seva sol·licitud (Dynarski i Scott-Clayton, 2013).

L'exemple que aquí exposem, però, remet a l'avaluació que fan Fack i Grenet (2015) del programa francès Bourses sur Critères sociaux, en concret, dels seus impactes en els nivells de matriculació, permanència i graduació dels estudiants que se'n beneficien. Aquest és el principal programa de beques universitàries a França, basat en un esquema "need-based". L'any 2010, una tercera part dels universitaris francesos rebien algun tipus d'ajut procedent d'aquest programa. En concret, el programa defineix sis nivells diferents de beca, l'accés als quals s'estableix segons llindars de "necessitat" (construïts en base a la renda familiar, el nombre de germans i la distància de la llar a la universitat). El nivell 1 comporta el pagament de les despeses de matrícula i una transferència de 1.500 euros anuals²⁶. A partir d'aquí, l'increment mitjà dels successius nivells de l'ajut correspon a un import de 600 euros anuals. Aprofitant l'exogeneïtat de la definició d'uns i altres llindars de necessitat, i utilitzant un disseny de regressió discontinua, els autors comparen els resultats dels estudiants que sol·licitaren beca els anys 2008 i 2009, i que acabaren trobant-se just per sobre i just per sota dels diferents llindars que donen accés als diferents nivells d'ajut.

Mitjançant aquest procediment, Fack i Grenet (2013) conclouen que rebre un ajut de 1.500 euros l'any (beca de nivell 1) incrementa fins a 4 punts percentuals la probabilitat d'accedir a la universitat i de romandre-hi dos cursos després; entre els estudiants de màster, s'observen impactes d'una magnitud similar sobre la probabilitat d'acabar graduant-se. Successius increments en la dotació de l'ajut (del nivell 2 de la beca en amunt) no reporten impactes positius significatius sobre els nivells de matriculació universitària, alhora que no s'aprecien

efectes diferencials dels ajuts en funció del sexe dels estudiants o del seu expedient acadèmic pre-universitari.

A CASA NOSTRA...

- Parlem de programes de beques de molt diversa índole: dels ajuts a l'alumnat amb necessitats educatives especial, beques de menjador o per a l'adquisició de material escolar, fins a les beques generals als ensenyaments secundaris postobligatoris o per als estudis superiors. A Catalunya, la gran majoria d'ajuts econòmics que es concedeixen en les etapes educatives obligatòries tenen la renda familiar com a principal criteri de prioritització. Aquest mateix esquema caracteritza bona part de les beques existents en la secundària postobligatòria i en els estudis universitaris, si bé aquí hi conviu amb altres ajuts basats en el rendiment acadèmic.
- El cas dels ajuts públics a la universitat és especialment paradigmàtic. Des del curs 2011-2012, a Catalunya es programen les beques "Equitat" i les beques "Excel·lència", ambdues per a estudiants de grau i d'ensenyaments no adaptats a l'Espai Europeu d'Ensenyament Superior. Les beques Equitat financen fins un 50% del preu públic de la matrícula, segons imports que es modulen per trams de renda. En canvi, les beques Excel·lència financen estades de mobilitat internacional a estudiants amb un alt rendiment acadèmic.
- A dia d'avui disposem de poca evidència sobre la capacitat d'uns i altres programes de beques de millorar les oportunitats d'aquells qui se'n beneficien en les diferents etapes educatives. Un dels pocs intents d'aproximació a aquesta qüestió el tenim en l'estudi de Mediavilla (2012, 2013). L'autor avalua els efectes de les beques i els ajuts a l'estudi implementats a Espanya el període 2005-2006. Utilitzant com a referència l'Enquesta de Condicions de Vida (2006), l'autor conclou: els alumnes que han estat perceptors d'una beca durant l'ensenyament obligatori tenen entre un 22% i un 25% més possibilitats de finalitzar amb èxit l'educació postobligatòria que alumnes de característiques similars que no n'han estat beneficiaris; aquesta probabilitat s'incrementaria entre les noies i els alumnes provinents de llars desfavorides²⁷.

4.2 L'ÚS DE VARIABLES INSTRUMENTALS

4.2.1 LA IMPORTÀNCIA DE LA RÀTIO D'ALUMNES PER CLASSE

El nombre d'alumnes per classe que estableixen els centres i/o les normatives marc corresponents és un dels aspectes de l'organització docent i escolar que més atenció acadèmica i política tradicionalment ha acaparat. Ara bé, determinar els efectes que aquest factor pot arribar a representar sobre el progrés acadèmic dels alumnes té les seves complicacions.

La simple comparació dels resultats que obtenen els alumnes en classes amb una ràtio més o menys elevada –es comparin alumnes dins un mateix centre, alumnes entre centres o entre països– planteja serioses limitacions metodològiques. Per exemple, ens podem trobar que les escoles tendeixin a tractar en classes reduïdes justament l'alumnat amb pitjor rendiment, o que les autoritats educatives optin per oferir a les escoles desfavorides recursos docents suplementaris, el que permetria al conjunt del centre ampliar el nombre de classes reduint-hi

l'ocupació. En aquest cas, l'assistència a grups-classe reduïts podria associar-se amb mals resultats, no tant pel nivell d'ocupació de l'aula, sinó pel perfil d'aquesta ocupació. Però podríem trobar-nos amb la situació inversa: les famílies amb més recursos i més preocupades per l'escolarització dels fills o bé escullen aquells centres que prioritzen els grups-classe reduïts, o bé pressionen els centres escollits per tal que així ho facin (Schlotter et al., 2010). En aquest cas, l'eventual associació entre classes reduïdes i bons resultats s'explicaria per l'extracció privilegiada dels alumnes que tendrien a concentrar-s'hi.

Diferents estudis han tractat de superar aquestes limitacions mitjançant estratègies metodològiques diverses. Dues d'aquestes estratègies, ambdues basades en l'ús de variables instrumentals, mereixen una breu explicació²⁸. En primer lloc, s'han aprofitat les fluctuacions en el moment del naixement com a font de variació exògena del volum d'alumnes que acaben component els diferents cursos i, per tant, les aules que els conformen. Així, s'entén que el fet de néixer abans o després del mes de tall que defineix l'adscripció a un curs o a un altre no depèn de factors relacionats amb l'outcome d'interès (Hoxby, 2000; Woessmann i West, 2006). Quan dins els centres es detecten diferències rellevants entre el número d'alumnes que componen un determinat curs i el número d'alumnes que componen el curs adjacent, quan s'exclou la possibilitat que aquestes diferències siguin degudes a dinàmiques de mobilitat dels estudiants, i quan aquestes diferències es tradueixen en un mateix nombre de classes per curs (les unes més nodrides, les altres menys), aleshores les diferències de rendiment entre els grups d'ambdós cursos poden ser atribuïdes a les diferències en la grandària de les classes respectives.

En segon lloc, alguns estudis han utilitzat l'establiment de normatives sobre el nombre màxim d'alumnes per classe com a instruments de la grandària de les aules (Angrist i Lavy, 1999; Woessmann, 2005a). L'aplicació d'aquest disseny troba justificació en la manca d'una relació lineal entre el número d'alumnes matriculats en un curs i el número d'alumnes de les classes que el componen. A partir d'aquí s'exploten les discontinuïtats que, dins de cada centre i entre centres, es produeixen al voltant del llindar màxim d'alumnes per aula, i es compara allò que els passa als alumnes en les aules de màxima ocupació amb el que els passa als alumnes que, per tal de no sobrepassar-se el llindar màxim, són assignats a aules subocupades del mateix curs.

Una visió general als resultats que obtenen els estudis basats en aquestes dues aproximacions ens duria a afirmar que els efectes de la grandària de les classes sobre els assoliments dels alumnes tendeixen a ser més aviat modestos. No obstant això, en alguns casos i sota certes circumstàncies, aquesta pràctica organitzativa sí aconsegueix efectes notables. Algunes revisions sistemàtiques sobre l'evidència acumulada entorn d'aquesta qüestió indiquen, per exemple, que les classes reduïdes (orientativament, amb menys de vint alumnes) poden ser particularment efectives per als alumnes més desfavorits, amb pitjor rendiment acadèmic i de menys edat, i sempre i quan es compti amb un professorat preparat per treure'n profit (McGiverin, Gilman, i Tillitski, 1989; Slavin, 1989; Hattie, 2005).

A CASA NOSTRA...

- El Ministerio de Educación Cultura y Deporte dictà per al curs 2012-2013 un increment general de les ràtios màximes d'alumnes per aula a les etapes de primària (passant-se d'una ràtio màxima de 25 a 30 alumnes) i a secundària (de 30 a 35 alumnes). Aquest increment es justificava com una mesura d'estalvi que no havia de posar en risc la qualitat del sistema i les oportunitats educatives dels alumnes. Des del Departament d'Ensenyament de la Generalitat, es denuncià la possible invasió de competències que aquesta ordre representava i s'optà per mantenir en la majoria d'escoles i instituts la mateixa ràtio màxima que el curs anterior. A Catalunya, el curs 2014-2015, la ràtio d'alumnes per aula continua establint-se en 25 alumnes a les escoles públiques (educació infantil i primària) i 30 als instituts (ESO).
- En definitiva, motivats per uns motius o altres, els canvis en la regulació d'aquestes ràtios han estat freqüents a Catalunya i Espanya en les darreres dècades. En cap cas, però, aquests canvis han recolzats en evidències empíriques sobre les seves possibles repercussions en el rendiment dels alumnes; menys encara, doncs, en estimacions econòmiques cost-efectivitat o cost-benefici de la seva aplicació.

4.2.2 LA LLIBERTAT D'ESCOLLIR CENTRE EDUCATIU

Ens preguntem aquí sobre els impactes que les polítiques d'obertura del marge d'elecció escolar poden produir en termes d'assoliments i oportunitats educatives. A parer d'alguns autors, aquestes polítiques representen un dels principals motors de canvi de bona part de les reformes educatives plantejades a escala mundial (Burch, 2009; Ball, 2012).

Per un costat, estudis basats en dades dels programes PISA, TIMSS o PIRLS, acostumen a observar una associació positiva entre, per una banda, el marge d'elecció escolar dels països i, per l'altra, el nivell mitjà i també d'equitat dels resultats que obtenen els seus estudiants (Schütz et al., 2007; Woessmann, Luedemann, Schuetz, i West, 2007). En el pla individual, però, quan aquestes associacions són corregides per l'extracció socioeconòmica i cultural dels alumnes i de les escoles on accedeixen, aleshores, el fet d'estudiar o no en un centre de lliure elecció acostuma a perdre tota rellevància (OECD, 2014).

Aquests estudis, però, presenten limitacions a l'hora de controlar l'endogeneïtat existent entre elecció escolar i resultats educatius²⁹. L'única manera de superar aquestes limitacions passa per aprofitar distintes fonts de variació en l'exposició a la política en qüestió. En el marc de determinats programes d'elecció escolar, una d'aquestes fonts pot arribar a ser un determinat mecanisme d'aleatorització. És el cas dels estudis que avaluen els impactes de programes de xecs escolars ("school vouchers") adreçats a ampliar el marge d'elecció de centre de les famílies socioeconòmicament més desafavorides; xecs que, en un context de sobredemanda, acaben sent distribuïts de forma aleatòria, mitjançant un sorteig. Aquests esquemes han estat habituals en diferents ciutats dels EUA³⁰, i han generat un volum important d'evidència experimental sobre la seva efectivitat³¹.

Alternativament, altres estudis han avaluat els efectes del marge obert a l'elecció de centre

mitjançant mètodes quasi-experimentals. És aquí on trobem aplicacions interessants de la tècnica basada en variables instrumentals.

En el cas de l'avaluació que fan Gibbons et al. (2008) dels impactes que tenen els contextos d'elecció de centre i competència entre escoles sobre el rendiment acadèmic dels alumnes anglesos de primària. En concret, es mesuren els efectes sobre els guanys competencials en llengua i matemàtiques que experimenten els alumnes entre el començament i el final de l'educació primària, prenent com a mostra l'alumnat de les escoles de l'àrea metropolitana de Londres.

En aquest estudi es defineix el marge d'elecció de centre de les famílies considerant el nombre de centres del districte escolar que les inclouen dins d'un determinat radi de proximitat³⁰. Així, es comprova que les famílies que resideixen a prop del límit del districte disposen d'un marge d'elecció menor que les que viuen en zones més centrals. Aquesta circumstància – concretament, la distància entre la residència de les famílies i els límits dels seus districtes escolars– és adoptada com a variable instrumental als efectes de l'avaluació d'impacte corresponent. El supòsit de fons és que, a igualtat en el nivell de qualitat de les escoles, la probabilitat que una família esculli una determinada escola disminuirà com més distant aquesta es trobi del seu domicili. I s'assumeix també que les famílies que s'instal·len lluny dels límits del districte mogudes per motius d'escolarització, ho fan no perquè valorin el marge d'elecció en ell mateix, sinó, en tot cas, perquè persegueixen ser a prop de determinades escoles. En conclusió, la distància entre residència i límits del districte es mostra com una variable vinculada a la probabilitat de trobar-se o no afectat per un major marge de tria de centre, i alhora no relacionada de forma específica amb la probabilitat d'èxit educatiu dels fills de les famílies que trien.

A través d'aquest mètode, els autors demostren que l'increment del marge de l'elecció de centre entre les famílies no comporta, per si sol, cap guany en el rendiment acadèmic dels seus fills, ni en llengua ni en matemàtiques. De fet, els efectes es mostren més aviat negatius, si bé en cap cas estadísticament significatius.

A CASA NOSTRA...

- De forma general, correspon a les Comunitats Autònomes concretar el procediment d'admissió de l'alumnat als centres sufragats amb fons públics, sobre la base de determinats criteris de tria i assignació escolar establerts a nivell estatal. Més enllà de les Comunitats Autònomes, els mateixos municipis tenen marge per substanciar alguns d'aquests procediments i criteris –és, per exemple, el cas de la definició de la proximitat als centres com a element de prioritització en l'accés.
- Tot plegat, a la pràctica, ha donat peu a l'existència de models autonòmics i a planificacions municipals de l'elecció escolar força diferents i alhora canviants en el temps. Cas interessant és, per exemple, la ciutat de Barcelona, que el curs 2006-2007 passà d'un model d'assignació zonal a un model d'escoles de proximitat. El nou model tractava d'oferir a totes les famílies un mínim de sis centres (tres públics i tres concertats) on obtinguessin màxima puntuació en el barem de proximitat. El curs 2011-2012 aquest mínim s'amplia a dotze centres (sis públics i sis concertats).

- Tota aquesta variabilitat entre territoris i en el temps respon poc a opcions preses en base a evidències empíriques sobre la seva efectivitat. Alhora, aquesta mateixa variabilitat ha estat poc aprofitada com a oportunitat per generar evidència, a casa nostra, sobre la capacitat d'impacte de les polítiques d'elecció de centre en el terreny del rendiment i l'equitat educativa³³.

4.3 L'ÚS DELS MODELS DE DOBLES DIFERÈNCIES

4.3.1 LA COMPRESIVITAT DEL SISTEMA EDUCATIU

El nivell de comprensivitat del sistema escolar –això és, el temps que els alumnes comparteixen un mateix itinerari formatiu, abans no s'inicia la separació entre la via acadèmica i la professional– és una de les característiques bàsiques de la seva configuració formal. Nombrosos estudis han tractat de determinar la seva incidència en els assoliments dels estudiants fent ús de proves internacionals de competències (Ammermüller, 2005; Dupriez et al., 2008; Duru-Bellat i Suchaut, 2005a; Gorard i Smith, 2004; Woessmann et al., 2009). Aquests estudis acostumen a evidenciar la presència d'una associació significativa entre el nivell de comprensivitat dels països i l'abast de les desigualtats educatives dels seus alumnes: com més precoç és la diferenciació (early tracking), més augmenten les distàncies de rendiment, i més aquestes distàncies reflecteixen les desigualtats socioeconòmiques i culturals dels estudiants. Alhora, tampoc sembla que els sistemes més diferenciats superin els més comprensius en mitjana global de rendiment.

La presència d'aquestes associacions, però, no implica necessàriament l'existència d'una relació de causalitat entre variables. En efecte, és esperable que el model de comprensivitat dels països es trobi correlacionat amb altres característiques no observades de caràcter contextual (nivell de riquesa o homogeneïtat cultural del país, característiques de les seves xarxes escolars, etc.) que, al seu torn, es trobin relacionades amb els resultats formatius dels seus estudiants.

Demostrar que l'early tracking és efectivament causant de desigualtat educativa i que, alhora, no origina guanys en termes de rendiment general requereix d'un pas més. Els dos exemples que esmentem a continuació fan aquest pas de la mà d'una estratègia d'identificació basada en l'ús de models de dobles diferències.

El primer exemple el trobem en l'estudi de Hanushek i Woessmann (2006) sobre els efectes de l'early tracking en els assoliments i desigualtats acadèmiques dels alumnes a diferents països de l'OCDE. L'estratègia dels autors consisteix a comparar les diferències entre els resultats que els alumnes obtenen a l'educació primària (sempre comprensiva) i els que obtenen a la secundària en països amb i sense early tracking. Aquest estudi il·lustra una aplicació

internacional de les dobles diferències on els països són observats no en dos moments diferents del temps, sinó, cadascun d'ells, en dues etapes educatives diferents. La conclusió d'aquesta recerca corrobora els arguments dels estudis correlacionals abans esmentats, a saber, que la diferenciació precoç incrementa les desigualtats socials de rendiment, sense provocar cap increment en la mitjana de resultats dels països.

Un segon exemple d'avaluació dels impactes de la comprensivitat basada en un model de dobles diferències ens el brinden aquells estudis que aprofiten la circumstància que aquesta reforma hagi estat desplegada gradualment arreu del territori –per exemple, regions dins d'un estat– segons una cronologia aliena a factors que podrien afectar els outcomes seleccionats (Meghir i Palme, 2005; Pekkarinen, Uusitalo, i Kerr, 2009). Aquest és l'enfocament que utilitzen Felgueroso, Gutiérrez i Jiménez (2013) per avaluar l'impacte de la reforma comprensiva continguda en la LOGSE sobre els nivells d'abandonament educatiu prematur dels joves. En efecte, després de la seva aprovació l'any 1990, la reforma fou implantada gradualment en les distintes Comunitats Autònomes, i ho feu seguint un calendari no determinat per raons susceptibles d'explicar el perquè d'uns o altres outcomes finals. Així, a través d'un model de dobles diferències, els autors observen com, en comparació amb el sistema previ, que oferia l'opció de cursar FP-1 als alumnes de 14 a 16 anys amb menys disposició cap als ensenyaments acadèmics, l'extensió de la comprensivitat fins als 16 anys representà un increment de l'abandonament entre els nois, no pas entre les noies. D'aquesta manera, l'estudi fonamenta un argument rellevant: que els beneficis que la comprensivitat pot aportar al rendiment acadèmic dels joves escolaritzats (i en especial al dels joves socialment més vulnerables), no tenen perquè associar-se a millores en els nivells de continuïtat escolar d'uns i altres, sobretot quan la política de comprensivitat és implementada sense marges de flexibilitat.

A CASA NOSTRA...

- Com de comprensiu ha de ser el sistema educatiu ha estat i continua essent una qüestió central en tota reforma educativa de cert calat. Així, a Espanya l'allargament de l'educació comprensiva fins als 16 anys –feta efectiva mitjançant la introducció de l'ESO– representà una de les fites principals de la LOGSE. D'igual forma, la reducció del tram comprensiu al primer cicle de l'ESO esdevé un dels punts nodulars de la reforma continguda en la Ley Orgánica para la Mejora de la Calidad Educativa (LOMCE, de 9 de desembre de 2013). En efecte, la LOMCE situa l'inici de la diferenciació d'itineraris a 4t d'ESO (considerat segon cicle de l'ESO), moment en què es distingeixen les opcions d'ensenyaments acadèmics d'iniciació al Batxillerat, ensenyaments aplicats d'iniciació a la Formació Professional, i els cicles de Formació Professional Bàsica (de dos anys de durada, comparables als antics Programes de Qualificació Professional Inicial). Si bé la introducció de la FP Bàsica és una realitat des del curs 2014-2015, la diversificació del segon cicle de l'ESO començaria a aplicar-se a partir del curs 2016-2017.
- També en aquest cas, però, caldria remarcar el fet que tant la reforma comprensiva que desplegà la LOGSE com la reducció de la comprensivitat operada per la LOMCE no han estat recolzades en evidències empíriques sòlides sobre l'efectivitat d'unes i altres opcions. I això malgrat l'evidència acumulada en aquest àmbit d'intervenció, com hem indicat més amunt, comença a ser abundant. A casa nostra aquest ha estat un debat marcadament polititzat, que ha viscut força d'esquenes a la generació i aprofitament de coneixement sobre què funciona i què no en aquesta qüestió.

4.3.2 LA DURADA DE LA JORNADA ESCOLAR

Estudis basats en els resultats de proves internacionals estandarditzades acostumen a detectar una associació positiva, tot i que generalment feble, entre el número d'hores lectives que ofereixen les escoles i el rendiment acadèmic que hi obtenen els estudiants (OECD, 2007; Hanushek i Woessmann, 2010). Continuem, però, mantenint la cautela necessària amb els resultats d'aquests estudis. Fins i tot quan comparem escoles dins d'un mateix país, hauríem de pensar que el fet que uns centres estableixin uns esquemes horaris i uns altres n'ofereixin de diferents no és aliè a altres aspectes relatius a la seva organització, als seus recursos, al seu professorat o al seu alumnat sovint difícils d'objectivar i, en canvi, previsiblement relacionats amb els resultats educatius dels estudiants.

Tractant de superar la incidència d'aquests possibles biaixos, alguns estudis han pogut analitzar l'efectivitat de mesures de reorganització dels horaris lectius mitjançant aproximacions quasi-experimentals rigoroses. En aquest terreny, els models de dobles diferències han tingut un recorregut interessant.

Posem com a exemple l'estudi de Lavy (2012) a l'estat d'Israel. L'autor avalua els efectes d'una reforma introduïda l'any 2004 a l'educació primària, la qual alterà el sistema de càlcul de l'assignació pressupostària dels centres. Resultat d'aquest canvi les escoles socioeconòmicament més desfavorides veieren incrementat el seu pressupost, mentre que aquest es reduí en les escoles més afavorides; la resta de centres es mantingueren igual com estaven abans de la reforma. També a conseqüència d'aquesta modificació, les escoles que incrementaren la seva dotació pressupostària tendiren a ampliar el seu nombre d'hores lectives per setmana, en especial les hores dedicades a assignatures troncales.

Aprofitant aquest seguit de circumstàncies, l'autor planteja una avaluació dels impactes de la reforma (en particular, de l'increment dels recursos dels centres i de les hores lectives), mitjançant l'ús de dobles diferències. En concret, l'avaluació compta amb una mostra de 936 escoles i la possibilitat d'observar-ne els resultats entre els anys 2002 (pre-reforma) i 2005 (post-reforma). Un primer tipus d'estimacions s'adrecen a comparar l'evolució d'outcomes, abans i després de la reforma, de les escoles que passaren a incrementar els seus recursos (grup de tractament) amb els d'aquelles que no experimentaren canvis en la seva dotació pressupostària (grup de comparació). Un segon bloc d'estimacions contrasten els resultats de les escoles perjudicades per la reforma (grup de tractament) amb els de les escoles no afectades (grup de comparació). Cada bloc d'estimacions inclou càlculs específics de l'efecte dels diferents canvis ocasionats per la reforma: en pressupost, en hores lectives per setmana i en hores dedicades a assignatures troncales. Mitjançant aquests procediments, l'estudi demostra que una part significativa dels beneficis que l'increment dels recursos de les escoles té sobre el rendiment dels alumnes en matemàtiques i llengua anglesa són atribuïbles a l'ampliació del temps lectiu dedicat a aquestes assignatures.

Els resultats de l'avaluació de Lavy concorden amb les conclusions de les revisions de la literatura internacional realitzades per Patall et al. (2010), Redd et al. (2012) i Kidron i Lindsey (2014) sobre aquesta qüestió. Aquests estudis apunten, a més, l'especial efectivitat d'aquelles intervencions que omplen l'increment horari amb activitats instructives implementades per professors qualificats. Alhora, totes tres revisions coincideixen a assenyalar la manca d'avaluacions d'impacte rigoroses que caracteritza aquest àmbit d'intervenció.

A CASA NOSTRA...

- El curs 2006-2007, l'aleshores Departament d'Educació introduí l'anomenada "sisena hora" a la jornada escolar de totes les escoles públiques catalanes. Segons argumentaren els responsables d'aquesta mesura, així com els qui des d'espais diversos la defensaren, el seu principal propòsit era el d'incrementar les oportunitats educatives dels alumnes de l'escola pública, resolent alhora el "greuge horari" que patien respecte dels alumnes de l'escola concertada (que tradicionalment ha comptat amb una jornada lectiva de sis hores diàries). El supòsit rere d'aquest argument, això és, que més hores lectives equival a més aprenentatges i millor rendiment, no fou recolzat en cap revisió d'evidències o anàlisi ex-ante, igual com tampoc es contrastà empíricament en quina mesura el greuge horari de l'escola pública podia estar repercutint sobre el rendiment dels seus alumnes (o sobre altres outcomes no escolars).
- Cinc cursos després de la seva implantació, el 2011-2012, el govern optà per suprimir la sisena hora en la majoria de centres, al·legant-ne la manca de resultats educatius. El curs 2013-2014, la sisena hora es manté en aproximadament 400 centres de primària, escoles situades en entorns especialment desafavorits. En la resta de centres, la supressió de la sisena hora va ser substituïda pel recurs SEP (Servei Escolar Personalitzat), adreçat als alumnes amb més dificultats d'aprenentatge de cada escola. Igual com la introducció de la mesura no respongué a cap avaluació de la situació prèvia, tampoc la seva supressió selectiva ha estat avalada en cap evidència empírica.

Notes:

22 Concretament, els autors utilitzen un model de regressió per a discontinuïtat "difusa" ("fuzzy"). Aquest model s'aplica quan la discontinuïtat existent al voltant del punt de tall de la variable en qüestió no és nítida. En el disseny "fuzzy", l'assignació al tractament no és determinístic; així, l'impacte del tractament s'infereix en base a la probabilitat de ser tractat al voltant del punt de tall. Quan la discontinuïtat entorn del llindar és nítida i dicotòmica, aleshores s'opta per un model "sharp" de regressió discontinua.

23 La mateixa normativa introduïa també la repetició automàtica per als alumnes que comptabilitzessin més de 25 absències diàries no justificades al llarg del curs. Aquesta discontinuïtat és aprofitada per Manacorda en el segon estudi que l'autor adreça a l'avaluació d'impacte de la repetició de curs (Manacorda, 2008).

24 Com en l'estudi de Jacob i Legren (2009), Manacorda aplica un disseny de regressió discontinua "fuzzy".

25 Podríem imaginar que, a igualtat d'altres característiques, els alumnes i famílies més motivats i amb una major adhesió escolar s'esforcen més a aconseguir una beca que la resta, i que són justament aquests nivells de motivació i adhesió, no tant el fet d'haver aconseguit beca, allò que acaba explicant el seu èxit formatiu.

26 Els autors estimen que aquest import equival a una tercera part de les despeses mitjanes de manutenció (inclòs allotjament) d'un estudiant universitari al llarg d'un curs acadèmic. L'any 2010, a França, el cost de la matrícula universitària se situava entorn els 200 euros l'any (per a graus i màsters).

27 Val a dir, però, que la metodologia emprada en aquest estudi (basada en la tècnica del propensity score matching), per bé que equipara els alumnes tractats (beneficiaris) i els alumnes control (no beneficiaris) en tot un seguit de característiques observables rellevants, no assegura que uns i altres acabin de ser comparables. Pot perfectament ser que uns i altres difereixin en variables no observables que puguin estar relacionades amb l'outcome d'interès; principalment, disposicions individuals i familiars a sol·licitar una beca, i rendiment acadèmic previ.

28 La recerca sobre els efectes de la grandària de la classe compta amb un dels experiments més coneguts en l'àmbit de l'avaluació de la política educativa, el Project STAR (Student-Teacher Achievement Ratio), implementat a l'estat de Tennessee (EUA) entre 1985 i 1989. Tanmateix, la implementació de l'experiment i, per tant, la fiabilitat dels seus resultats ha estat posada seriosament en qüestió (Hanushek, 1999).

29 Fins i tot quan comparem famílies dins d'un mateix territori, cal comptar amb la més que probable incidència de biaixos entre participants (famílies que escullen de forma activa) i no participants (famílies que no ho fan). Per molt que s'assemblin en característiques sociodemogràfiques rellevants, una família que exerceix la tria de centre de forma activa segurament serà diferent a una que l'exerceix de forma passiva en qüestions que probablement afectaran els resultats acadèmics dels seus respectius fills.

30 Casos paradigmàtics són els programes de xecs de Milwaukee, Cleveland, Nova York, Dayton o Washington, D.C., tots ells iniciats durant la dècada dels 1990.

31 *Veure, per exemple, les avaluacions experimentals de Mayer et al. (2002), Barnard et al. (2003) o Chingos i Peterson (2012) sobre el programa de xecs escolars de la ciutat de Nova York.*

32 *El pes que els districtes i les escoles atorguen a la proximitat com a criteri de prioritització de l'accés escolar en cas de sobredemanda és força variable. Com també ho és la manera com aquest criteri és interpretat i mesurat. En tot cas, la definició que Gibbons et al. (2008) fan del radi de proximitat dels centres té en compte la mediana de les distàncies entre la residència de les famílies i les escoles que escullen en cadascun dels districtes.*

33 *A Catalunya, hauríem de referir-nos als estudis de Calsamiglia (2014) i Benito i González (2007), el primer sobre la incidència que tingué, l'any 2007, el canvi en el model d'assignació escolar de Barcelona sobre patrons de tria d'escoles de les famílies; el segon, sobre la influència de la zonificació escolar en els nivells de segregació escolar de diferents municipis catalans. D'aquesta manera, totes dues recerques estudien els efectes de variacions regulatives sobre outcomes que podríem considerar de naturalesa intermèdia.*

5. I A CASA NOSTRA, COM PODRÍEM AVANÇAR?

La preocupació per l'efectivitat de les intervencions educatives troba origen en l'aposta per promoure decisions i polítiques educatives ben informades, això és, basades en l'evidència empírica. La política educativa catalana, com l'espanyola, i com la de tants altres països del nostre entorn, són encara lluny d'aquesta situació. De fet, segons hem tingut ocasió d'exemplificar, les decisions sobre les polítiques educatives –sobre el seu llançament, el seu manteniment, la seva reforma, la seva supressió– rarament s'han pres considerant l'evidència empírica acumulada sobre l'efectivitat esperable d'unes o altres intervencions. Alhora, ben poques d'aquestes polítiques han acabat sent avaluades des del punt de vista de la seva capacitat d'impacte.

A la llum d'allò que podem extreure d'estudis i avaluacions dutes a terme en altres contextos, els arguments que es plantegen a continuació volen situar alguns espais d'oportunitat per fer avançar la pràctica de l'avaluació d'impacte de les polítiques educatives a casa nostra.

5.1 ATREVIR-SE A EXPERIMENTAR: APROFITAR LES PROVES PILOT...

Com hem indicat anteriorment, no hi ha disseny d'avaluació més robust que l'experimental. I si bé és cert que la utilització d'aquest disseny no sempre és viable, també ho és que són nombroses les seves possibles aplicacions. El repertori d'exemples que hem descrit en el capítol 3 ho posa de manifest.

El tipus de problemàtiques i els àmbits d'intervenció que repassàvem llavors certament connecten amb reptes importants que la política educativa catalana té plantejats avui dia. Alguns d'aquests àmbits tenen a veure amb elements força estructurals o contextuals de la política i el sistema educatiu (autonomia escolar, escolarització en la petita infància o els incentius econòmics al professorat); altres remeten a pràctiques més "micro", d'innovació o desenvolupament pedagògic (l'ús de les TIC a l'ensenyament, les tutories individualitzades o els programes d'implicació educatives de les famílies).

A Catalunya no són pocs els programes educatius que s'han iniciat en forma de prova pilot: el projecte d'autonomia de centres (des de 2005-2006), el programa de suport escolar personalitzat (encetat el curs 2011-2012), la introducció de la compactació horària a l'ESO i a la primària (des de 2012-2013), el programa intensiu de millora a l'ESO (des de 2013-2014), la modalitat "blended learning" al batxillerat (des de 2013-2014), el programa mSchools (des de 2013-2014), el pla experimental de plurilingüisme (des de 2013-2014), l'assignatura de serveis a la comunitat (des de 2014-2015), entre d'altres. En general, l'objectiu principal d'aquestes experiències pilot s'ha orientat, per sobre de tot, a testar determinats aspectes relatius als disseny i implementació de la intervenció, contrastar-ne la seva viabilitat tècnica i política, la seva acceptació social; tot plegat abans de procedir a la seva eventual generalització. En

canvi, rarament aquestes pilots han partit d'una estratègia avaluativa ben definida i adreçada a generar les condicions que haurien de permetre l'estimació posterior del seu impacte. El camí a recórrer en aquest terreny és, per tant, llarg. I alhora farcit d'oportunitats.

Per exemple, sembla clar que l'àmbit de les intervencions que designàvem com a més "micro" (innovacions pedagògiques, estratègies docents, modificacions curriculars, àdhuc iniciatives educatives més enllà de l'escola) ofereix un camp particularment fèrtil al desenvolupament de pilots d'impacte, en particular, de pilotatges experimentals. La mateixa dinàmica de la innovació estimula l'assaig i l'experimentació, una dinàmica que es podria articular mitjançant proves aleatoritzades de petita o mitjana escala. De fet, en contextos on tothom té clar que s'està "provant" una innovació educativa, d'abast limitat i amb una capacitat d'impacte encara a determinar, acostuma a acceptar-se que hi haurà candidats interessats (escoles, alumnes, famílies, territoris) que no accediran al programa, i que el criteri més just i operatiu per assignar la participació passa per l'aleatorització.

Alhora, és també en aquests contextos on la utilitat de l'avaluació s'observa més clarament. D'una banda, assaig experimental i acumulació d'evidències es combinen amb el propòsit d'aportar coneixement sòlid i útil per al conjunt del professorat i responsables de programes d'innovació educativa. D'altra banda, és l'obtenció d'aquest mateix coneixement el que permet fonamentar la conveniència (o no) d'escalar els programes pilot en un sentit o un altre.

Eventualment, la iniciativa d'aquests pilotatges pot néixer del territori i la comunitat educativa (municipis o conjunt de municipis, xarxes d'escoles, d'entitats educatives, moviments i federacions de mestres, etc.), i el seu desenvolupament disposar del suport i finançament de fundacions privades o organitzacions filantròpiques. En tot cas, entenem que aquestes iniciatives haurien de comptar amb el recolzament del Departament d'Ensenyament, principal responsable del possible escalat dels programes en qüestió.

El rol del Departament seria preminent quan el que hi ha en joc és el pilotatge d'aspectes estructurals de la política o el sistema educatiu. Segurament, els desafiaments i els problemes de viabilitat tècnica i política que els pilots experimentals poden plantejar en els àmbits d'actuació més nuclears són superiors als que plantegen en el terreny dels programes de caràcter més "micro". I això no vol dir que les polítiques d'autonomia escolar, d'elecció de centre, de professorat, d'organització de la jornada escolar, de planificació de l'educació infantil, de regulació de la repetició o promoció de curs, etc. no admetin la possibilitat de pilotatges experimentals. Pensaríem aleshores en la possibilitat de dur a terme proves pilot aleatoritzades orientades a mesurar l'impacte d'uns o altres components, aplicacions o reformes concretes que es puguin plantejar o formin part de cadascuna d'aquestes polítiques.

Per exemple, podria valorar-se la possibilitat de testar mitjançant pilots experimentals el desplegament d'aspectes específics de l'autonomia escolar com ara determinats esquemes

d'incentivació econòmica del professorat o determinats elements de professionalització dels equips directius. O es podrien dissenyar proves pilot per avaluar l'efectivitat de diferents models de provisió de l'educació 0-3 anys, o diferents modalitats d'ajuts a les famílies per accedir a aquests recursos. O pilotar de forma experimental diferents fórmules d'agrupació d'alumnes en distintes assignatures o cursos. De la mateixa manera, s'haguessin pogut desenvolupar autèntiques proves pilot –amb una mostra suficient de centres i amb un mecanisme d'aleatorització de la participació– adreçades a establir amb rigor els efectes de la jornada compacta a l'educació secundària i a la primària sobre les oportunitats educatives dels alumnes. Es poden trobar més exemples en la taula 2.

5.2 ... I L'EXCÉS DE DEMANDA

L'excés de demanda d'un programa pot ser "induït", quan la limitació de la seva cobertura és deliberada; cas típic d'aquesta situació és el dels programes pilot. Per contra, en l'àmbit educatiu, és habitual trobar serveis o programes amb un excés de demanda "real"; això és, els límits en la cobertura obeeixen a la incapacitat objectiva del servei o el programa d'atendre tothom qui el demanda.

Doncs bé, quan en determinats àmbits d'intervenció no es veu viable el plantejament de pilots aleatoritzats, però en canvi existeixen programes sobredemandats, caldria preservar la possibilitat d'aprofitar aquesta circumstància als efectes d'avaluar-ne experimentalment els impactes. Com hem dit en el seu moment, en situació d'escassetat de recursos, quan el programa no pot cobrir a la totalitat de la població a la qual s'adreça (alumnes, centres, famílies amb unes o altres característiques), aleshores l'assignació aleatòria del recurs entre els sol·licitants elegibles és amb tota probabilitat el mecanisme de distribució més just que puguem dissenyar. El mecanisme d'assignació més just i el que ofereix les millors condicions per a l'avaluació d'impacte.

De fet, l'accés a determinats serveis educatius amb excés de demanda ja es dirimeix actualment mitjançant un sorteig. Un clar exemple és el de l'accés a les escoles bressol públiques, un recurs sovint sobredemandat en molts municipis de Catalunya. En aquest cas, per exemple, es tractaria d'acabar comparant el progrés educatiu dels infants que, ens uns i altres municipis, "guanyen" la loteria d'accés a l'escola bressol (o a aquelles escoles bressol amb més sobredemanda) amb el progrés d'aquells altres que, amb els mateixos punts en la sol·licitud de preinscripció, "perden" el sorteig (controlant per les opcions educatives triades per les famílies d'aquests darrers infants). Quelcom similar podria arribar a plantejar-se per avaluar els impactes de la tria de centre en l'educació obligatòria: es tractaria de comparar els resultats acadèmics dels alumnes que entren en les escoles sobredemandades triades en primera opció amb els dels alumnes que, amb els mateixos punts en la baremació, prioritzen, sense èxit, les mateixes escoles.

Un procediment similar permetria avaluar de forma rigorosa aquells programes i serveis educatius més enllà de l'escola que arriben a trobar-se sobredemandats: activitats de reforç escolar, dispositius d'acompanyament educatiu, programes de lleure educatiu, etc. És habitual que les entitats proveïdores d'aquests serveis apliquin criteris de selecció de candidats que van més enllà dels requisits d'elegibilitat establerts formalment; criteris que poden recórrer a judicis sobre el nivell de motivació dels individus, capacitat per aprofitar el recurs, grau de vulnerabilitat, etc. Tot i en aquestes circumstàncies, si s'aconsegueixen baremar aquests criteris, podria encara existir marge a la introducció de mecanismes de sorteig allà on la sobredemanda es produeixi per part d'individus empatats a punts.

Sigui com sigui, cal preveure els sistemes d'informació necessaris per registrar el conjunt de candidats que demanden el recurs, programa o servei, superin o no en nombre el total de places disponibles, i establir els mecanismes que han de fer possible l'obtenció de dades i seguiment d'outcomes d'uns i altres candidats. Reprenem aquesta qüestió en el punt 5.5.

5.3 AVALUACIÓ QUASI-EXPERIMENTAL

No menystenim, però, els problemes de viabilitat (tècnica, política, social) que els estudis experimentals poden arribar a plantejar, sobretot quan es discuteix l'avaluació de polítiques o regulacions educatives de caire més sistèmic. Quan el disseny experimental no és viable, això és, quan l'accés al programa no es pot dirimir per la via de l'aleatorització, aleshores poden entrar en joc diferents alternatives d'avaluació quasi-experimental.

Més amunt hem recollit exemples d'avaluacions quasi-experimentals que aconsegueixen produir estimacions robustes d'impacte en àmbits d'intervenció que són també rellevants a casa nostra, i que podrien ser avaluats seguint aproximacions similars. Per exemple, es podria valorar la possibilitat d'aplicar dissenys de regressió discontinua a l'avaluació de programes de beques finançades i/o gestionades per la Generalitat de Catalunya que defineixen el criteri d'elegibilitat en funció d'un punt de tall, ja sigui en el nivell de renda (p.e. beques de menjador, beques a la secundària postobligatòria, o beques Equitat a la universitat) o de rendiment acadèmic (p.e. beques especials a la secundària, o beques Excel·lència a la universitat). Aquest mateix mètode es podria utilitzar per estimar els impactes de repetir curs, tant si acaba instaurant-se la revàlida externa que planteja la LOMCE en l'últim curs de l'ESO (els resultats de la qual serien vinculants als efectes d'obtenir o no el graduat a partir del curs 2017-2018), com en el marc actual, en què la repetició o la promoció de curs s'estableix d'acord al nombre de matèries suspeses en l'avaluació final de curs.

O es podria explorar l'aplicabilitat de variables instrumentals relacionades amb la distribució o densitat territorial de la xarxa d'escoles a l'avaluació de determinats canvis en la regulació de l'elecció escolar, principalment en municipis grans o mitjans. En un altre àmbit, si ens preocupen els efectes que la ràtio d'alumnes per aula pot tenir sobre els seus aprenentatges,

podríem arribar a testar les fluctuacions en el moment del naixement com a instrument de la grandària de les classes que componen els diferents cursos escolars. Aquest mateix instrument podria utilitzar-se per avaluar reformes conduents a incrementar de forma general la ràtio màxima d'ocupació de les aules. En el nostre cas es tractaria de comparar alumnes en grups-classe afectats i no afectats per la reforma a l'inici de l'escolarització, assumint que els grups afectats omplen la ràtio màxima per aula i que els grups no afectats no incrementen la ràtio en cursos successius.

Finalment, podria estudiar-se l'aplicació de models de dobles diferències per avaluar l'efectivitat de programes o mesures que han tingut un desplegament gradual al territori o a la xarxa de centres educatius. Aquest fou el cas de la jornada lectiva de sis hores a l'educació primària. Recordem que l'anomenada "sisena hora" fou introduïda el curs 2006-2007 per l'aleshores Departament d'Educació a tots els centres públics de primària; el curs 2011-2012, aquesta mesura fou retirada en un nombre significatiu d'escoles. Aquest fet obriria la porta a analitzar en quina mesura les diferències de rendiment que, a partir del curs 2011-2012, s'observen entre centres amb i sense sisena hora es corresponen amb les diferències que aquests mateixos grups de centres mostraven en el període en què tots ells disposaren d'aquest recurs. La comparació entre aquestes dues diferències ens acostaria a la mesura de l'impacte de la sisena hora.

5.4 ... PREVEIENT L'AVALUACIÓ EN EL DISSENY DEL PROGRAMA

La possibilitat d'optar per unes o altres alternatives d'avaluació quasi-experimental (regressions discontinües, variables instrumentals, dobles diferències, o altres mètodes aquí no considerats) dependrà del moment en què aquestes es considerin. Es dibuixen aquí dos escenaris: quan el disseny de la política contempla el disseny de l'avaluació (escenari ex-ante), o quan l'avaluació es dissenya un cop el programa es troba en marxa o ha ja finalitzat (escenari ex-post).

En el primer dels escenaris, la pròpia formulació de la intervenció –definició d'objectius, població destinatària, fases d'execució, activitats, recursos, etc. – incorpora la preocupació per les seves condicions d'avaluabilitat. Parlem aleshores de perspectiva ex-ante o prospectiva de l'avaluació d'impacte. Aquesta perspectiva, que és inherent als experiments socials, incrementa la probabilitat d'acabar obtenint bones avaluacions d'impacte també en absència d'aleatorització, principalment per dos motius: d'una banda, permet preveure fórmules d'implementació que facilitin la construcció de bons grups de comparació; d'altra banda, ofereix la possibilitat d'establir des de l'inici els instruments i les mesures que requerirà l'avaluació del programa (indicadors de resultat, sistemes d'informació, protocols de recollida d'informació, etc.). Més enllà d'això, com indiquen Casado i Todeschini (2013), el simple fet de pensar com avaluar l'impacte del programa en el mateix moment que aquest s'està dissenyant, "obliga als responsables del programa a precisar nítidament quins són els outcomes d'interès

sobre els quals la intervenció pretén tenir algun impacte, així com els mecanismes pels quals s'espera que aquests efectes es produeixin" (2013: 22). Tot plegat acaba millorant, no tan sols s l'avaluabilitat del programa, sinó també el propi disseny substantiu de la intervenció.

En el segon escenari, quan l'avaluació es dissenya ex-post, la seva viabilitat es troba totalment condicionada a les característiques de la implementació del programa i a les bases d'informació disponibles. En el pitjor dels casos, pot arribar a fer-se impossible la identificació de cap grup de comparació mínimament creïble i/o l'obtenció de la informació necessària per al seguiment dels resultats en qüestió.

Per tant, l'aposta ha de ser clara en favor de propiciar el primer dels escenaris sempre que sigui possible. Val a dir que aquesta aposta es deixa notar en una part important de les convocatòries per a programes educatius d'institucions internacionals com el Banco Iberoamericano de Desarrollo o la mateixa Comissió Europea, les quals supediten la concessió de finançament al disseny ex-ante d'avaluacions d'impacte.

Taula 2. Oportunitats per a l'avaluació d'impacte de polítiques i programes educatius a Catalunya

ÀMBIT D'INTERVENCIÓ	OPORTUNITATS NO APROFITADES	ESPAIS D'OPORTUNITAT
SISTEMA EDUCATIU		
Comprensivitat vs. diferenciació precoç	<ul style="list-style-type: none"> Reforma comprensiva, LOGSE (G.E., 1990): no fonamentada empíricament (ex-ante), avaluació retrospectiva limitada (ex-post) Reforma de la secundària, LOMCE (G.E., 2013): no fonamentada empíricament (ex-ante) 	<ul style="list-style-type: none"> Possibilitat d'explotar la introducció gradual de les reformes comprensives en el territori com a estratègia d'identificació de contextos "tractats" i "control" (models dobles diferències)
Educació en la petita infància	<ul style="list-style-type: none"> Reforma comprensiva LOGSE (G.E., 1990): no fonamentada empíricament (ex-ante), avaluació retrospectiva limitada (ex-post) Expansió de places d'escola bressol (G.C. i G.L., entre 2005 i 2010): no fonamentada empíricament (ex-ante), avaluació retrospectiva limitada (ex-post) Reducció aportacions a escoles bressol (G.C., des de 2010): no fonamentada empíricament (ex-ante), sense avaluació d'impacte (ex-post) 	<ul style="list-style-type: none"> Possibilitat d'explotar el mecanisme d'aleatorització davant l'excés de demanda en l'accés a escoles bressol (disseny experimental) Possibilitat d'explotar la introducció gradual de les reformes comprensives en el territori com a estratègia d'identificació de contextos "tractats" i "control" (models dobles diferències)

QUASI-MERCAT EDUCATIU		
Elecció de centre	<ul style="list-style-type: none"> Mecanismes d'accés/assignació escolar (G.C. i G.L., anys diversos): no fonamentats empíricament (ex-ante), avaluació retrospectiva limitada (ex-post) 	<ul style="list-style-type: none"> Possibilitat d'explotar el mecanisme d'aleatorització davant l'excés de demanda en l'accés a determinades escoles en determinats entorns (disseny experimental) Possibilitat d'aprofitar desequilibris en la distribució territorial de la xarxa escolar (variables instrumentals)
Autonomia escolar	<ul style="list-style-type: none"> Pla d'Autonomia dels Centres (G.C., 2005): prova pilot no aleatoritzada, sense avaluació d'impacte (ex-post) Decret d'Autonomia dels Centres (G.C., 2010): no fonamentat empíricament (ex-ante), sense avaluació d'impacte (ex-post) 	<ul style="list-style-type: none"> Possibilitat de desenvolupar proves pilot per testar l'efectivitat de determinades mesures d'autonomia contemplades en el Decret de 2010 (disseny experimental)
INCENTIUS I AJUTS ECONÒMICS		
Retribució al professorat	<ul style="list-style-type: none"> Canvis en estructures retributives, promoció i incentius (G.E. i G.C., anys diversos): no fonamentats empíricament (ex-ante), sense avaluació d'impacte (ex-post) 	<ul style="list-style-type: none"> Possibilitat de desenvolupar proves pilot per testar l'efectivitat de determinats esquemes d'incentius monetaris o no monetaris (disseny experimental)
Beques i ajuts als estudiants	<ul style="list-style-type: none"> Introducció i canvis en beques escolars (G.E. i G.C., anys diversos): no fonamentats empíricament (ex-ante), avaluació retrospectiva limitada (ex-post) Reducció beques escolars (G.E. i G.C., des de 2010): no fonamentada empíricament (ex-ante), sense avaluació d'impacte (ex-post) Programa de beques universitàries: Equitat i Excel·lència (G.C., des de 2011): no fonamentat empíricament (ex-ante), sense avaluació d'impacte (ex-post) 	<ul style="list-style-type: none"> Possibilitat d'explotar variabilitat territorial en la cobertura de beques i en la introducció de canvis en dotacions com a estratègia d'identificació d'alumnes "tractats" i "control" (variables instrumentals) Possibilitat d'aprofitar discontinuïtats en la percepció al voltant del punt de tall en el nivell de renda o acadèmic establert per a accedir a determinades beques (regressions discontinües)
ASPECTES ORGANITZATIUS		
Alumnes per aula	<ul style="list-style-type: none"> Establiment de les ràtios màximes alumnes per classe, primària i secundària (G.E. i G.C., anys diversos): no fonamentat empíricament (ex-ante), sense avaluació d'impacte (ex-post) Ampliació ràtios màximes alumnes per classe, primària i secundària (G.E. i G.C., 2012-2013): no fonamentada empíricament (ex-ante) 	<ul style="list-style-type: none"> Possibilitat d'explotar variacions en les ràtios motivades per fluctuacions en la grandària de les cohorts d'alumnes i/o per mesures d'increment de ràtios (variables instrumentals)

<p>Horaris i jornada escolar</p>	<ul style="list-style-type: none"> • Introducció 6a hora a la primària pública (G.C., 2006-2007): no fonamentada empíricament (ex-ante), sense avaluació d'impacte (ex-post) • Suspensió 6a hora en majoria de centres (G.C., 2011-2012): no fonamentada empíricament (ex-ante), sense avaluació d'impacte (ex-post) • Compactació jornada secundària pública (G.C., 2012-2013): fonamentada en prova pilot no aleatoritzada • Introducció "setmana blanca" (G.C., 2011-2012): no fonamentada empíricament (ex-ante), sense avaluació d'impacte (ex-post) • Supressió "setmana blanca" (G.C., 2012-2013): no fonamentada empíricament (ex-ante), sense avaluació d'impacte (ex-post) <ul style="list-style-type: none"> • Possibilitat de desenvolupar proves pilot per testar l'efectivitat de mesures d'ampliació o compactació horària, o canvis en el calendari escolar (disseny experimental) • Possibilitat d'aprofitar discontinuïtats en l'adhesió a la reforma al voltant del punt de tall en el nivell de consens mínim exigít per part del consell escolar (regressions discontinües) • Possibilitat d'aprofitar la introducció (i retirada) parcial de canvis horaris en la xarxa escolar com a estratègia d'identificació de contextos "tractats" i "control" (models dobles diferències)
<p>DECISIONS I PRÀCTIQUES PEDAGÒGIQUES</p>	
<p>Repetició de curs</p>	<ul style="list-style-type: none"> • Canvis en models repetició-promoció automàtica (G.E., G.C. i centres, anys diversos): no fonamentats empíricament (ex-ante), sense avaluació d'impacte (ex-post) <ul style="list-style-type: none"> • Possibilitat d'aprofitar discontinuïtats en la probabilitat de repetir al voltant del nombre mínim d'assignatures aprovades (o del llindar mínim de puntuació en una revàlida vinculant) requerit per passar de curs (regressions discontinües) • Possibilitat de comparar alumnes amb el mateix rendiment acadèmic en cursos consecutius, uns afectats i els altres no per canvis en els criteris de repetició (variables instrumentals)
<p>Innovació i desenvolupament pedagògic</p>	<ul style="list-style-type: none"> • Introducció i canvis en programes diversos (TIC, enfocament competències, sistemes de tutoria, aprenentatge cooperatiu...) (G.C. i centres, anys diversos): no fonamentats empíricament (ex-ante), sense avaluació d'impacte (ex-post) <ul style="list-style-type: none"> • Possibilitat de desenvolupar proves pilot per testar l'efectivitat de determinades innovacions curriculars i didàctiques (disseny experimental)
<p>MÉS ENLLÀ DE L'ESCOLA</p>	
<p>Fora escola Relació família-escola</p>	<ul style="list-style-type: none"> • Introducció i canvis en programes diversos (plans d'entorn, projectes comunitaris, educació no formal...) (G.C., G.L. i centres, anys diversos): no fonamentats empíricament (ex-ante), avaluació retrospectiva limitada (ex-post) <ul style="list-style-type: none"> • Possibilitat de desenvolupar proves pilot per testar l'efectivitat de programes d'implicació de les famílies en l'escolarització dels fills, així com de programes fora escola (disseny experimental) • Possibilitat d'aprofitar discontinuïtats en la participació al voltant del punt de tall en el nivell de renda o acadèmic establert per a accedir a determinats recursos fora escola (regressions discontinües)

5.5 SISTEMES D'INFORMACIÓ I ACCÉS A LES DADES

Tota avaluació d'impacte que aspi a ser rigorosa requereix bones dades en les quals basar les seves estimacions; bones dades no tan sols relatives a les característiques del programa en qüestió, sinó també per construir les variables independents de context requerides per l'anàlisi (característiques dels alumnes i de les seves famílies, de les seves classes, professors, escoles i entorns, etc.) i les seves variables d'outcome (indicadors de rendiment, de trajectòria educativa, o variables relacionades amb expectatives, actituds o disposicions escolars). Difícilment es pot avançar en l'estudi de l'efectivitat de les polítiques i programes educatius quan, 1) els registres educatius són incomplets (manca d'informació rellevant) o inconnexos (sense vinculació entre ells i amb altres registres administratius), i quan, 2) l'accés a la informació es troba significativament limitat o restringit. A casa nostra tenim força camí per recórrer, en tots dos nivells.

En primer lloc, en els registres educatius manca informació rellevant sobre els alumnes; principalment, el nivell socioeconòmic i instructiu de les seves llars. Aquesta informació podria agregar-se a nivell d'escola i proporcionar indicadors del perfil social dels centres més acurats que els que actualment es fan servir. Altres informacions es troben poc sistematitzades, poc tractades o disperses; succeeix, per exemple, en el cas de la informació sobre les beques escolars (incloses les de menjador) o de la relativa al perfil i condicions del professorat dels centres. D'aquesta manera es dificulta la possibilitat d'establir creuaments entre variables d'interès; per exemple, entre les seves qualificacions acadèmiques (ordinàries i en proves de competències) i el fet que siguin o hagin estat beneficiaris d'una beca escolar, o d'alguna adaptació curricular, o hagin passat per una escola bressol o llar d'infants, o hagin tingut uns determinats professors o uns altres, etc. I es fa també difícil connectar els registres educatius disponibles amb altres registres administratius, com per exemple, el de la Seguretat Social (per conèixer la trajectòria i situació laboral de pares i mares, i també dels mateixos alumnes un cop finalitzen els estudis) o el de l'Agència Tributària (per conèixer la situació i evolució dels nivells de renda de les famílies i els alumnes en qüestió).

La construcció d'un autèntic sistema de registres individuals hauria de permetre completar i integrar aquest conjunt d'informacions al voltant d'un nexa central, l'alumne, i a partir d'aquí poder procedir a les agregacions i desagregacions de dades que cada anàlisi requereixi (per classe, per professor, per centre o xarxa escolar, barri, etc.). D'aquesta manera es facilitaria alhora la possibilitat de resseguir la caracterització de l'estudiant al llarg de tota la seva trajectòria educativa (fins i tot més enllà).

A manca d'aquest sistema, determinats tipus d'avaluacions d'impacte podrien cobrir-se amb dades de panell, susceptibles d'introduir un seguiment longitudinal de resultats amb prou

recorregut en el temps, i sempre i quan aquests panells incloguessin informació completa i rellevant sobre les variables d'interès. Aquesta eina esdevé útil, per exemple, per recollir i ampliar informació sobre outcomes no cognitius, però vinculats amb el progrés acadèmic (disposicions i expectatives vers l'educació, autoestima, autonomia, capacitat crítica, etc.).

En segon lloc, bona part de la informació que en l'actualitat inclouen els registres educatius no és de fàcil accés per als recercadors. En efecte, els grups de recerca sovint es troben amb impediments per poder disposar de dades desagregades (anonimitzades i dissociades³⁴) dels alumnes o els centres als quals assisteixen, tant sobre outcomes educatius (p.ex. resultats en avaluacions ordinàries i en proves externes de rendiment) com sobre determinades característiques sociodemogràfiques.

En resum, si volem arribar a disposar d'evidències sòlides sobre el grau d'efectivitat de programes i polítiques educatives, el primer que hem d'aconseguir és millorar les seves condicions d'avaluabilitat, entre d'altres, aquelles que atenyen als sistemes, fonts i tipus d'informació que posem en joc.

5.6 QUÈ FUNCIONA I PER QUÈ FUNCIONA: LA IMPORTÀNCIA DE FILAR PRIM

Essent simplistes, diríem que les avaluacions d'efectivitat dels programes poden proporcionar dos tipus d'evidències: evidència que el programa funciona o evidència que el programa no funciona (té impactes negatius o no en té). I tanmateix, això es quedar-se curt. Vegem per què.

Primer, és habitual que els programes educatius es fixin més d'un objectiu estratègic. Si més no, acostumen a ser diversos els outcomes en base als quals poder valorar els seus efectes. Així, és freqüent que un programa aconseguixi impactar positivament sobre part dels seus outcomes i no sobre una altra part. Ens trobaríem, per tant, davant d'un cas de programa amb "efectes mixtos", un matís que massa sovint les categories generals de "funciona" o "no funciona" tendeixen a eclipsar en el debat públic. Òbviament, des del punt de vista de l'opció política que emmarca la intervenció, no tots els resultats són igualment rellevants; alguns tenen a veure amb objectius prioritaris, d'altres no. Per tant, no tot impacte té el mateix "valor". I no obstant això, el fet de poder diferenciar els efectes assolits sobre uns i altres outcomes aporta un coneixement de gran importància de cara a possibles reorientacions de la intervenció en qüestió.

Segon, les intervencions educatives, per específiques i focalitzades que siguin, acostumen a incloure actuacions diverses, articulades com a components d'un mateix programa i dissenyades sobre la base d'uns objectius finals comuns. Per tant, pot perfectament ser que un programa no funcioni, i en canvi, algun dels seus components particulars sí ho faci; o a la inversa, que un programa que globalment funciona contingui components inefectius (o fins i tot perjudicials).

Tercer, les intervencions educatives, novament, per específiques i focalitzades que siguin, s'adrecen a una població que no deixa de presentar característiques diverses. Això vol dir que un programa que s'observa que no funciona en considerar-se els resultats mitjans del conjunt de la població atesa, sí es mostri positiu quan es para atenció al que els succeeix a uns col·lectius determinats. Igualment a la inversa, un programa en conjunt efectiu pot estar concentrant els seus impactes positius en uns subgrups concrets, i en canvi resultar innocu (àdhuc negatiu) per a d'altres subgrups.

Quart, la gran majoria de programes educatius necessiten temps, en dos sentits. Per una banda, en el terreny de la implementació, les intervencions requereixen el que podríem denominar un temps de "rodatge" o "estabilització" abans no puguem considerar que el seu desenvolupament és mitjanament fidel al seu disseny inicial (això és, a allò que es pretén avaluar) i alhora ha pogut acomodar-se a les lògiques i rutines particulars dels agents implicats en cada cas. Això és particularment cert en el cas d'aquelles intervencions que representen canvis significatius en les dinàmiques i en les pràctiques d'aquests agents. D'altra banda, en el terreny de l'avaluació, diríem que els programes requereixen temps abans no és raonable començar a observar-ne possibles impactes. Més enllà que la intervenció pugui definir objectius mesurables a curt, mig o llarg termini, i més enllà del temps de rodatge que pugui requerir la seva implementació, determinats canvis en la realitat tractada arriben amb el temps. Aquesta darrera possibilitat té el seu revers: programes que es mostren efectius en el curt termini, però els impactes dels quals es van amortint amb el pas del temps (fade-out effects).

Traslladem aquests quatre arguments al nostre exemple imaginari, el Programa d'Acceleració Educativa (PAE). Suposem que una avaluació rigorosa de la seva efectivitat acaba conclouent que el programa, en conjunt, no funciona, que no produeix impactes significatius. Recolzaria aquesta constatació una eventual decisió de retirada total del programa? La resposta seria, no necessàriament.

En primer lloc, podria donar-se el cas que el PAE no aconsegueixi impactes positius sobre el nivell competencial en matemàtiques, però en canvi sí comporti petits guanys en l'àmbit lingüístic o en outcomes no observats (per exemple, millores actitudinals i en el terreny socio-emocional). En segon lloc, una avaluació de l'impacte diferencial de cadascun dels seus components podria detectar que allò que fa que el programa, en conjunt, no funcioni és l'ingredient d'agrupació per nivell que aquest incorpora; en canvi, podríem observar que l'increment del temps lectiu dedicat a les assignatures instrumentals sí mostra impactes positius significatius en el progrés acadèmic dels alumnes. En tercer lloc, podria ser que el PAE resulti inefectiu quan ens fixem en la totalitat de l'alumnat atès, però en canvi sí tingui un impacte específic en el cas concret d'aquells participants que provenen de famílies amb un capital socioeconòmic i educatiu més reduït. Finalment, si poguéssim disposar d'una finestra d'observació prou àmplia, podria ser que detectéssim que el programa no és efectiu durant els seus primers dos anys de funcionament, i que en canvi, comença a generar impactes positius

a partir del tercer any. Més encara, podria observar-se que determinats guanys competencials no afloren sinó en el mig termini, quan l'alumne es troba finalitzant l'ESO; guanys que podrien interpretar-se com a resultat indirecte d'altres millores en el terreny actitudinal i emocional.

En síntesi, és important completar la conclusió sobre l'efectivitat global d'un programa amb evidències sobre el detall dels impactes, saber quins són els components del programa que funcionen i quins els que no ho fan, saber per a qui funcionen i per a qui no, a partir de quin moment i durant quant temps. Arribar a aquest nivell de detall no és sempre fàcil. No sempre es disposa de grups de comparació específics per als diferents nivells de participació en les diferents fases o components del programa; no sempre és senzill destriar sobre quins subgrups de participants té sentit analitzar l'ocurrència de possibles efectes heterogenis; no sempre és possible començar el procés d'avaluació quan millor convé, ni disposar d'una finestra temporal prou àmplia per a l'observació dels resultats considerats. Però si volem desentrellar els mecanismes d'efectivitat (o inefectivitat) de les intervencions educatives, saber què ho fa que un programa funcioni o no, cal impulsar avaluacions d'impacte que vagin en aquesta direcció.

Notes:

³⁴ S'entén, sense risc d'incórrer en l'incompliment de les normes bàsiques referents al secret estadístic.

REFERÈNCIES

- Abdulkadiroglu, A., Angrist, J., Dynarski, S., Kane, T. J., i Pathak, P. (2009). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. National Bureau of Economic Research, NBER Working Paper No. 15549. Recuperat a partir de <http://www.nber.org/papers/w15549>
- Alegre, M. A., i Benito, R. (2014). Youth Education Attainment and Participation in Europe: the role of contextual factors and the scope of education policy. *European Journal of Education*, 49(1), 127-143.
- Allen, C. S., Chen, Q., Willson, V. L., i Hughes, J. N. (2009). Quality of Research Design Moderates Effects of Grade Retention on Achievement: A Meta-Analytic, Multilevel Analysis. *Educational Evaluation and Policy Analysis*, 31(4), 480-499.
- Ammermüller, A. (2005). Educational opportunities and the role of institutions. ZEW Discussion Paper No. 05-44. Recuperat a partir de <http://www.econstor.eu/handle/10419/24135>
- Angrist, J., Cohodes, S.R., Dynarski, S.M., Pathak, P.A., i Walters, C.D. (2013). Charter Schools and the Road to College Readiness: The Effects on College Preparation, Attendance and Choice. The Boston Foundation and NewSchools Venture Fund. Recuperat a partir de <https://www.tbf.org/~media/TBF0rg/Files/Reports/Charters%20and%20College%20Readiness%202013.pdf>
- Angrist, J. D., i Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, 114(2), 533-575.
- Ball, S. J. (2012). *Global Education Inc.: New Policy Networks and the Neoliberal Imaginary*. London: Routledge, Taylor & Francis Group.
- Barnard, J., Frangakis, C. E., Hill, J. L., i Rubin, D. B. (2003). Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City. *Journal of the American Statistical Association*, K(462), 299-323.
- Barnett, W. S. (1985). Benefit-cost analysis of the Perry Preschool Program and its policy implications. *Educational evaluation and policy analysis*, 7(4), 333-342.
- Belfield, C. R., Nores, M., Barnett, S., i Schweinhart, L. (2006). The High/Scope Perry Preschool Program Cost-Benefit Analysis Using Data from the Age-40 Followup. *Journal of Human Resources*, 41(1), 162-190.
- Benito, R., i González, I. (2007). *Processos de segregació escolar a Catalunya*. Barcelona: Mediterrània, Col. Polítiques 59.

Bettinger, E. P., Long, B. T., Oreopoulos, P., i Sanbonmatsu, L. (2012). The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment. *The Quarterly Journal of Economics*, 127(3), 1205-1242.

Betts, J. R., i Tang, Y. E. (2011). The Effect of Charter Schools on Student Achievement. A Meta-analysis of the Literature. University of Washington and Center on Reinventing Public Education. Recuperat a partir de http://www.oxydiane.net/IMG/pdf/Charter_NCSRP_BettsTang_Oct11.pdf

Blasco, J. (2015). Infància a Catalunya. Mesures contra la pobresa infantil: ampliació selectiva d'escoles bressol i extensió de la tarifació social. Barcelona: UNICEF, Docs Infància a Catalunya. Recuperat a partir de http://www.unicef.es/sites/www.unicef.es/files/docs_infancia_cat_escoles_bressol_.pdf

Blasco, J., i Casado, D. (2009). Guia pràctica 5. Avaluació d'impacte. Ivàlua. Recuperat a partir de http://www.ivalua.cat/documents/1/01_03_2010_11_33_12_Guia5_Impacte_Setembre2009_revfeb2010_massavermella.pdf

Bonal, X., i Verger, A. (2013). L'Agenda de la política educativa a Catalunya: una anàlisi de les opcions de govern (2011-2013). Barcelona: Fundació Jaume Bofill, Col. Informes Breus 45.

Borman, G. D., Slavin, R. E., Cheung, A. C. K., Chamberlain, A. M., Madden, N. A., i Chambers, B. (2007). Final Reading Outcomes of the National Randomized Field Trial of Success for All. *American Educational Research Journal*, 44(3), 701-731.

Burch, P. (2009). *Hidden markets: The new education privatization*. New York: Routledge.

Calsamiglia, C., i Güell, M. (2014). The Illusion of School Choice: Empirical Evidence from Barcelona. IZA DP No. 8202. Recuperat a partir de <http://ftp.iza.org/dp8202.pdf>

Camilli, G., Vargas, S., Ryan, S., i Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *The Teachers College Record*, 112(3), 579-620.

Carter, S. C. (2000). *No Excuses: Lessons from 21 High-Performing, High-Poverty Schools*. Washington, D.C.: The Heritage Foundation.

Casado, D. (2012). Per què no avaluem les polítiques públiques com els fàrmacs? Una aposta per l'experimentació social. Avaluació per al Bon Govern. Recuperat a partir de www.avaluació.cat

Casado, D., i Todeschini, F. A. (2013). Guia pràctica 10. Avaluar l'impacte de les polítiques actives d'ocupació. Ivàlua. Recuperat a partir de http://www.ivalua.cat/documents/1/17_12_2013_11_58_01_Guia10_PAO_cat.pdf

Cheung, A. C. K., i Slavin, R. E. (2012a). Effects of educational technology applications on reading outcomes for struggling readers: A best-evidence synthesis. Best Evidence Encyclopedia. Recuperat a partir de http://www.bestevidence.org/word/tech_strug_read_Jul_18_2012.pdf

Cheung, A. C. K., i Slavin, R. E. (2012b). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, 7(3), 198-215.

Chingos, M. M., i Peterson, P. E. (2012). The Effects of School Vouchers on College Enrollment: Experimental Evidence from New York City. Brown Center on Education Policy at Brookings Institution. Recuperat a partir de http://www.hks.harvard.edu/pepg/PDF/Impacts_of_School_Vouchers_FINAL.pdf

Clark, M. A., Gleason, P. M., Tuttle, C. C., i Silverberg, M. K. (2014). Do Charter Schools Improve Student Achievement? *Educational Evaluation and Policy Analysis*. Online first.

Collet, J., i Tort, A. (2011). Famílies, escola i èxit: millorar els vincles per millorar els resultats. Barcelona: Fundació Jaume Bofill, Col. Informe Breus 35.

Comas, M., Escapa, S., i Abellán, C. (2014). Com participen mares i pares a l'escola? Diversitat familiar i d'implicació en educació. Barcelona: Fundació Jaume Bofill, Col. Informes Breus 49.

Cook, P. J., et al. (2014). The (Surprising) Efficacy of Academic and Behavioral Intervention with Disadvantaged Youth: Results from a Randomized Experiment in Chicago. National Bureau of Economic Research, NBER Working Paper No. 19862. Recuperat a partir de <http://www.povertyactionlab.org/publication/surprising-efficacy-academic-and-behavioral-intervention-disadvantaged-youth-results-ran>

Correnti, R. (2009). Examining CSR program effects on student achievement: Causal explanation through examination of implementation rates and student mobility. Ponència presentada a la 2nd Annual Conference of the Society for Research on Educational Effectiveness. Washington, DC., March 2009. Recuperat a partir de http://www.lrdc.pitt.edu/BOV/documents/Correnti_ExaminingCSRProgramEffects_033012.pdf

CREDO (2009). Multiple Choice: Charter Schools Performance in 16 States. Center for Research on Education Outcomes (CREDO), Stanford University. Recuperat a partir de http://www.btu.org/sites/default/files/research/credo_standrod_charter_school_performance_full.pdf

Currie, J., i Thomas, D. (2000). School Quality and the Longer-Term Effects of Head Start. *The Journal of Human Resources*, 35(4), 755-774.

Curs, B. R., i Harper, C. E. (2012). Financial aid and first-year collegiate GPA: a regression discontinuity approach. *The Review of Higher Education*, 35(4), 627-649.

DesJardins, S. L., McCall, B. P., Ott, M., i Kim, J. (2010). A Quasi-Experimental Investigation of How the Gates Millennium Scholars Program Is Related to College Students' Time Use and Activities. *Educational Evaluation and Policy Analysis*, 32(4), 456-475.

Di Carlo, M. (2011). The Evidence on Charter Schools and Test Scores. The Albert Shanker Institute Policy Brief. Recuperat a partir de <http://shankerblog.cdjd.info/wp-content/uploads/2011/12/CharterReview.pdf>

Dobbie, W., i Fryer Jr, R. G. (2011). Getting beneath the veil of effective schools: Evidence from New York City. National Bureau of Economic Research, NBER Working Paper No. 17632. Recuperat a partir de <http://www.nber.org/papers/w17632>

Dolton, P., i Marcenaro-Gutierrez, O. D. (2011). If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance. *Economic policy*, 26(65), 5-55.

Duflo, E., Glennerster, R., i Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of development economics*, 4, 3895-3962.

Duflo, E., Hanna, R., i Rya, S. P. (2012). Incentives Work: Getting Teachers to Come to School. *The American Economic Review*, 102(4), 1241-1278.

Duncan, G. J., i Magnuson, K. (2013). Investing in Preschool Programs. *Journal of Economic Perspectives*, 27(2), 109-132.

Dupriez, V., Dumay, X., i Vause, A. (2008). How Do School Systems Manage Pupils' Heterogeneity? *Comparative Education Review*, 52(2), 245-273.

Duru-Bellat, M., i Suchaut, B. (2005). Organisation and Context, Efficiency and Equity of Educational Systems: what PISA tells us. *European Educational Research Journal*, 4(3), 181-194.

Dynarski, S., i Scott-Clayton, J. (2013). Financial Aid Policy: Lessons from Research National Bureau of Economic Research, NBER Working Paper No. 18710. Recuperat a partir de <http://www.nber.org/papers/w18710>

Ehri, L. C., Dreyer, L. G., Flugman, B., i Gross, A. (2007). Reading Rescue: An Effective Tutoring Intervention Model for Language-Minority Students Who Are Struggling Readers in First Grade. *American Educational Research Journal*, 44(2), 414-448.

Fack, G., i Grenet, J. (2015). Improving College Access and Success for Low-Income Students: Evidence from a Large French Need-based Grant Program. *American Economic Journal: Applied Economics*, 7(2), 1-34.

Felgueroso, F., Gutiérrez-Doménech, M., i Jiménez-Martín, S. (2013). ¿Por qué el abandono escolar se ha mantenido tan elevado en España en las últimas dos décadas? El papel de la Ley de Educación (LOGSE). *Economic Reports*, 02-2013, 1-26.

Fryer Jr, R. G., Levitt, S. D., List, J., i Sadoff, S. (2012). Enhancing the efficacy of teacher incentives through loss aversion: A field experiment. National Bureau of Economic Research, NBER Working Paper No. 18237. Recuperat a partir de <http://www.nber.org/papers/w18237>

Fryer Jr, R. G. (2011a). Creating « No Excuses»(Traditional) Public Schools: Preliminary Evidence From an Experiment in Houston. National Bureau of Economic Research, NBER Working Paper No. 17494. Recuperat a partir de <http://www.nber.org/papers/w17494>

Fryer Jr, R. G. (2011b). Teacher incentives and student achievement: Evidence from New York City public schools. National Bureau of Economic Research, NBER Working Paper No. 16850. Recuperat a partir de <http://www.nber.org/papers/w16850>

Fuchs, T., i Woessmann, L. (2007). What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics*, 32(2-3), 433-464.

Gibbons, S., Machin, S., i Silva, O. (2008). Choice, competition, and pupil achievement. *Journal of the European Economic Association*, 6(4), 912-947.

Glazerman, S., Protik, A., Teh, B., Bruch, J., i Max, J. (2013). Transfer Incentives for High-Performing Teachers: Final Results from a Multisite Randomized Experiment. U.S. Department of Education, IES-NCEE 2014-4003. Recuperat a partir de <http://ies.ed.gov/ncee/pubs/20144003/pdf/20144003.pdf>

Glewwe, P., Ilias, N., i Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, 2(3), 205-227.

Goos, M., Schreier, B. M., Knipprath, H. M. E., De Fraine, B., Van Damme, J., i Trautwein, U. (2013). How Can Cross-Country Differences in the Practice of Grade Retention Be Explained? A Closer Look at National Educational Policy Factors. *Comparative Education Review*, 57(1), 54-84.

Gorard, S., See, B. H., i Siddiqui, N. (2014). Switch-on Reading. Education Endowment Foundation. Recuperat a partir de http://educationendowmentfoundation.org.uk/uploads/pdf/FINAL_EEF_Evaluation_Report_-_Switch-on_-_February_2014.pdf

Gorard, S., i Smith, E. (2004). An international comparison of equity in education systems. *Comparative education*, 40(1), 15-28.

Gorey, K. M. (2001). Early childhood education: A meta-analytic affirmation of the short-and long-term benefits of educational opportunity. *School Psychology Quarterly*, 16(1), 9.

Goux, D., Gurgand, M., i Maurin, E. (2013). The effect of school and peers on dropout behavior. Ivàlua. Recuperat a partir de http://www.ivalua.cat/documents/1/28_06_2013_07_44_49_Marc_Gurgand.pdf

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis*, 21(2), 143-163.

Hanushek, E. A., Link, S., i Woessmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics*, 104, 212-232.

Hanushek, E. A., i Rivkin, S. G. (2007). Pay, working conditions, and teacher quality. The future of children, 17(1), 69-86.

Hanushek, E. A., i Woessmann, L. (2006). Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries. *The Economic Journal*, 116(510), C63-C76.

Hanushek, E. A., i Woessmann, L. (2010). The economics of international differences in educational achievement. National Bureau of Economic Research, NBER Working Paper No. 15949. Recuperat a partir de <http://www.nber.org/papers/w15949>

Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research*, 43(6), 387-425.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., i Yavitz, A. (2010a). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, 94(1-2), 114-128.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P., i Yavitz, A. (2010b). A new cost-benefit and rate of return analysis for the Perry Preschool Program: A summary. National Bureau of Economic Research, NBER Working Paper No. 16180. Recuperat a <http://www.nber.org/papers/w16180>

Heckman, J. J., Pinto, R., i Savelyev, P. A. (2012). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. National Bureau of Economic Research, NBER Working Paper No. 18581. Recuperat a partir de <http://www.nber.org/papers/w18581>

Heppen, J. B., Walters, K., Clements, M., Faria, A.-M., Tobey, C., Sorensen, N., i Culp, K. (2012). Access to Algebra I: The Effects of Online Mathematics for Grade 8 Students. U.S. Department of

Education, IES-NCEE 2012-4021. Recuperat a partir de http://ies.ed.gov/ncee/edlabs/regions/northeast/pdf/REL_20124021.pdf

Hew, K. F., i Cheung, W. S. (2013). Use of Web 2.0 technologies in K-12 and higher education: The search for evidence-based practice. *Educational Research Review*, 9, 47-64.

Holmes, C. T. (1989). Grade level retention effects: A meta-analysis of research studies. A L. A. Shepard i M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp. 16-33). London: The Falmer Press.

Hoxby, C. M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *The Quarterly Journal of Economics*, 115(4), 1239-1285.

Hoxby, C. M., Murarka, S., i Kang, J. (2009). How New York City's charter schools affect achievement. Cambridge, MA: The New York City Charter Schools Evaluation Project 2009. Recuperat a partir de http://users.nber.org/~schools/charterschoolseval/how_NYC_charter_schools_affect_achievement_sept2009.pdf

Imbens, G. W. (2014). Instrumental Variables: An Econometrician's Perspective. National Bureau of Economic Research, NBER Working Paper No. 19983. Recuperat a partir de <http://www.nber.org/papers/w19983>

Jacob, B. A., i Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, 1(3), 33-58.

Jacob, R. T., Smith, T. J., Willard, J. A., i Rifkin, R. E. (2014). Reading Partners: The Implementation and Effectiveness of a One-on-One Tutoring Program Delivered by Community Volunteers. MDR Policy Brief. Recuperat a partir de http://readingpartners.org/wp-content/uploads/2014/06/Reading-Partners_final.pdf

Jeynes, W. H. (2005). A Meta-Analysis of the Relation of Parental Involvement to Urban Elementary School Student Academic Achievement. *Urban Education*, 40(3), 237-269.

Jeynes, W. H. (2007). The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement A Meta-Analysis. *Urban Education*, 42(1), 82-110.

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30(3), 420-437.

Kidron, Y., i Lindsay, J. (2014). The Effects of Increased Learning Time on Student Academic and Nonacademic Outcomes: Findings from a Meta-Analytic Review. REL 2014-015. IES-NCEER / REL Appalachia. Recuperat a partir de <http://eric.ed.gov/?id=ED545233>

Lavy, V. (2012). Expanding School Resources and Increasing Time on Task: Effects of a Policy Experiment in Israel on Student Academic Achievement and Behavior. National Bureau of Economic Research, NBER Working Paper No. 18369. Recuperat a partir de <http://www.nber.org/papers/w18369>

Leeds, D. M., i DesJardins, S. L. (2014). The Effect of Merit Aid on Enrollment: A Regression Discontinuity Analysis of Iowa's National Scholar's Award. Mimeo. Recuperat a partir de <https://aefpweb.org/sites/default/files/webform/aefp40/The%20Effect%20of%20Merit%20Aid%20on%20Enrollment.pdf>

Lee, Y. S., Morrow-Howell, N., Jonson-Reid, M., i McCrary, S. (2012). The Effect of the Experience Corps® Program on Student Reading Outcomes. *Education and Urban Society*, 44(1), 97-118.

Loeb, S., i Lee, V. E. (1995). Where do Head Start attendees end up? One reason why preschool effects fade out. *Evaluation and Policy Analysis*, 17(1), 62-82.

Loeb, S., i Page, M. E. (2000). Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics*, 82(3), 393-408.

Manacorda, M. (2008). The cost of grade retention. CEP Discussion Paper No. 878. London: Centre for Economic Performance, London School of Economics and Political Science. Recuperat a partir de <http://cep.lse.ac.uk/pubs/download/dp0878.pdf>

Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics*, 94(2), 596-606.

Manning, M., Homel, R., i Smith, C. (2010). A meta-analysis of the effects of early developmental prevention programs in at-risk populations on non-health outcomes in adolescence. *Children and Youth Services Review*, 32(4), 506-519.

Mattingly, D. J., Prislín, R., McKenzie, T. L., Rodríguez, J. L., i Kayzar, B. (2002). Evaluating evaluations: The case of parent involvement programs. *Review of Educational Research*, 72(4), 549-576.

Mayer, D. P., Peterson, P. E., Myers, D. E., Tuttle, C. C., i Howell, W. G. (2002). School choice in New York City after three years: An evaluation of the school choice scholarships program. *Mathematica Policy Research*, MPR No. 8404-045. Recuperat a partir de <http://www.mathematica-mpr.com/~media/publications/PDFs/nycfull>

May, H., et al. (2014). Evaluation of the i3 Scale-up of Reading Recovery. CPRE, No. RR-79. Recuperat a partir de http://lfws.literacy.org/sites/default/files/researchreport/1488_readingrecoveryreport.pdf

McGiverin, J., Gilman, D., i Tillitski, C. (1989). A meta-analysis of the relation between class size and achievement. *The Elementary School Journal*, 90(1), 47-56.

Mediavilla, M. (2012). Les beques com a factor d'èxit escolar: Una avaluació d'impacte a partir d'una metodologia quasi-experimental. A M. Martínez i B. Albaigés (Eds.) *L'estat de l'educació a Catalunya. Anuari 2011* (pp. 167-182). Barcelona: Mediterrània, Col. Polítiques 75.

Mediavilla, M. (2013). Heterogeneidad en el impacto de la política de becas en la escolaridad secundaria postobligatoria en España: un análisis por subgrupos poblacionales. *Estudios de Economía*, 40(1), 97-120.

Meghir, C., i Palme, M. (2005). Educational reform, ability, and family background. *The American Economic Review*, 95(1), 414-424.

Miller, S., i Connolly, P. (2013). A Randomized Controlled Trial Evaluation of Time to Read, a Volunteer Tutoring Program for 8- to 9-Year-Olds. *Educational Evaluation and Policy Analysis*, 35(1), 23-37.

Miller, S., Connolly, P., i Maguire, L. K. (2012). The effects of a volunteer mentoring programme on reading outcomes among eight- to nine-year-old children: A follow up randomized controlled trial. *Journal of Early Childhood Research*, 10(2), 134-144.

Morris, S., Tödling-Schönhofer, H., i Wiseman, M. (2013). Design and commissioning of counterfactual impact evaluations guidance to help employers and workers to manage the transition to the new classification, labelling and packaging system. Luxembourg: European Commission, DG Employment, Social Affairs and Inclusion.

Nelson, G., Westhues, A., i MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment*, 6(1), 31a.

OECD. (2007). *PISA 2006 Science Competencies for Tomorrow's World*. Vol. 1. Paris: OECD Publishing. Recuperat a partir de http://www.navarro.cl/web/educacion/docs_comision/Documentos%20e%20informes%20academicos/PISA.pdf

OECD. (2010). *PISA 2009 Results: What Students Know and Can Do*. Vol. 4. Paris: OECD Publishing. Recuperat a partir de http://www.llv.li/rss/pdf-llv-sa-pisa_2009_oecd-bericht_englisch__band_1_-_was_schueler_wissen_und_koennen.pdf

OECD. (2014). *PISA 2012 Results. What Makes Schools Successful? Vol. 1. How Resources, Policies and Practices are Related to Education Outcomes*. Paris: OECD Publishing. Recuperat a partir de <http://www.oecd.org/pisa/keyfindings/Vol4Ch1.pdf>

Patall, E. A., Cooper, H., i Allen, A. B. (2010). Extending the School Day or School Year: A Systematic Review of Research (1985-2009). *Review of Educational Research*, 80(3), 401-436.

Peck, L. R., i Bell, S. H. (2014). The Role of Program Quality in Determining Head Start's Impact on Child Development. Third Grade Follow-Up to the Head Start Impact Study. OPRE Report No. 2014-10. Recuperat a partir de http://www.acf.hhs.gov/sites/default/files/opre/hs_quality_report_4_28_14_final.pdf

Pekkarinen, T., Uusitalo, R., i Kerr, S. (2009). School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics*, 93(7), 965-973.

Porter, S. R. (2012). Using Instrumental Variables Properly to Account for Selection Effects. Mimeo. Recuperat a partir de <http://eric.ed.gov/?id=ED531905>

Puma, M., et al. (2012). Third Grade Follow-Up to the Head Start Impact Study: Final Report. OPRE Report No. 2012-45. Recuperat a partir de <http://eric.ed.gov/?id=ED539264>

Redd, Z., Boccanfuso, C., Walker, K., Princiotta, D., Knewstubb, D., i Moore, K. (2012). Expanding time for learning both inside and outside the classroom: A review of the evidence base. *Child Trends*. Recuperat a partir de http://childtrends.org/wp-content/uploads/2013/03/Child_Trends-2012_08_16_RB_TimeForLearning.pdf

Rubin, R. B. (2011). The Pell and the Poor: A Regression-Discontinuity Analysis of On-Time College Enrollment. *Research in Higher Education*, 52(7), 675-692.

Rutt, S., Easton, C., i Stacey, O. (2014). Catch Up® Numeracy. Education Endowment Foundation. Recuperat a partir de <http://www.nfer.ac.uk/publications/EFCU01/EFCU01.pdf>

Schlotter, M., Schwerdt, G., i Woessmann, L. (2010). Econometric Methods for Causal Evaluation of Education Policies and Practices: A Non-Technical Guide. IZA DP No. 4725. Recuperat a partir de http://www.eenee.de/portal/page/portal/EENEEContent/_IMPORT_TELECENTRUM/DOCS/EENEE_AR5.pdf

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., i Shavelson, R. J. (2007). Estimating causal effects: Using experimental and observational designs. American Educational Research Association, Washington, DC. Recuperat a partir de http://69.8.231.237/uploadedFiles/Publications/Books/Estimating_Causal_Effects/ECE_Front-TOC.pdf

Schütz, G. (2009). Does the quality of pre-primary education pay off in secondary school? An international comparison using PISA 2003. Ifo Working Paper 68. Recuperat a partir de <http://www.cesifo-economic-studies.de/pls/guest/download/Ifo%20Working%20Papers%20%28seit%202005%29/IfoWorkingPaper-68.pdf>

Schütz, G., Ursprung, H. W., i Woessmann, L. (2008). Education policy and equality of opportunity. *Kyklos*, 61(2), 279-308.

Schütz, G., West, M., i Woessmann, L. (2007). School accountability, autonomy, choice, and the equity of student achievement: International evidence from PISA 2003. Paris: OECD Publishing. Recuperat a partir de <http://www.oecd.org/edu/39839422.pdf>

Scott-Clayton, J. (2011). On Money and Motivation A Quasi-Experimental Analysis of Financial Incentives for College Achievement. *Journal of Human Resources*, 46(3), 614-646.

Slavin, R. E. (1989). Class size and student achievement: Small effects of small classes. *Educational Psychologist*, 24(1), 99-110.

Slavin, R. E., Lake, C., Davis, S., i Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, 6(1), 1-26.

Springer, M. G., et al. (2010). Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT). RAND Education i National Center of Performance Incentives. Recuperat a partir de http://www.rand.org/content/dam/rand/pubs/reprints/2010/RAND_RP1416.pdf

Sundararaman, V., i Muralidharan, K. (2011). Teacher Performance Pay: Experimental Evidence from India. *The Journal of Political Economy*, 119(1), 39-77.

Thernstrom, A., i Thernstrom, S. (2004). No excuses: Closing the racial gap in learning. New York: Simon and Schuster.

Van Steensel, R., McElvany, N., Kurvers, J., i Herppich, S. (2011). How effective are family literacy programs? Results of a meta-analysis. *Review of Educational Research*, 81(1), 69-96.

Van Voorhis, F. L., Maier, M. F., Epstein, J. L., Lloyd, C., i Leung, T. (2013). The impact of family involvement on the education of children age 3 to 8. MDRC Report. Recuperat a partir de http://www.mdrc.org/sites/default/files/The_Impact_of_Family_Involve ment_ES.pdf

Whitman, D. (2008). Sweating the small stuff: Inner-city schools and the new paternalism. Washington, DC.: Thomas B. Fordham Institute

Woessmann, L. (2003). Schooling resources, educational institutions and student performance: the international evidence. *Oxford Bulletin of Economics and Statistics*, 65(2), 117-170.

Woessmann, L. (2005a). Educational production in Europe. *Economic policy*, 20(43), 445-504.

Woessmann, L. (2005b). The effect heterogeneity of central examinations: evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics*, 13(2), 143-169.

Woessmann, L., Luedemann, E., i Schuetz, G. (2009). *School accountability, autonomy and choice around the world*. Cheltenham: Edward Elgar Publishing Ltd.

Woessmann, L., Luedemann, E., Schuetz, G., i West, M., R. (2007). *School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003*. Paris: OECD Publishing. Recuperat a partir de <http://www.oecd.org/edu/39839361.pdf>

Woessmann, L., i West, M. R. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review*, 50(3), 695-736.

York, B. N., i Loeb, S. (2014). *One Step at a Time: The Effects of an Early Literacy Text Messaging Program for Parents of Preschoolers*. National Bureau of Economic Research. NBER Working Paper No. 20659. Recuperat a partir de <http://www.nber.org/papers/w20659>

Zimmer, R. W., Gill, B., Booker, K., Lavertu, S., Sass, T. R., i Witte, J. (2009). *Charter schools in eight states: Effects on achievement, attainment, integration and competition*. Santa Monica: RAND Corporation.

