Capítol 6

Conclusions i Propostes

En aquest capítol farem balanç de la recerca realitzada fins al moment.

En primer lloc, repassaré per capítols quines han estat les preguntes que ens plantejàvem i quines són les conclusions a les que hem pogut arribar. En segon lloc, comentaré la metodologia que hem seguit argumentant la idoneïtat, avantatges i limitacions de les anàlisis i modelitzacions estadístiques utilitzades .

He considerat interessant recollir en una secció a part les principals reflexions de caràcter pedagògico-educatiu que han anat sorgint al llarg del treball.

Per últim, faré esment de les preguntes que resten obertes i de les perspectives de recerca en el futur.

6.1 Conclusions

En referència al primer Capítol

El primer capítol d'aquesta tesi aborda l'estudi de la variabilitat de la nota mitjana que cada estudiant obté a les PAAU —anomenada *nota PAAU*—, així com la determinació de les possibles fonts de variabilitat associades. Ens interessava poder detectar si es donaven diferències significatives entre els resultats d'un centre a un altre. Les preguntes plantejades es concreten en:

- Es pot parlar de diferències significatives entre centres pel que fa als resultats que els seus alumnes obtenen a les PAAU?

- Si la resposta és afirmativa, a més de donar una mesura d'aquestes diferències, i d'identificar els centres que en un sentit o en un altre s'aparten del comportament mitjà, voldrem saber a què es deguda aquesta

diferència: a la composició de l'alumnat; a la diferent preparació que ofereixen els centres; a diferències en l'avaluació realitzada pels centres ...?

- Per últim, és estadísticament possible discernir una causa d'una altra amb les dades de què disposem? En tot cas, quina informació addicional ens caldria per poder discernir?

Resumim les respostes:

- Mentre la *nota PAAU* varia significativament entre centres, no passa el mateix amb la *nota COU*. Un 20%, aproximadament, de la variació total de la *nota PAAU* correspon a variació entre centres. Aquesta diferenciació entre centres que presenta la *nota PAAU* es manté —fins i tot s'incrementa—, al fer la regressió de la *nota PAAU* respecte de la *nota COU*.

- En conseqüència, no podem dir que les diferències existents entre centres siguin degudes només a la composició del seu alumnat. La nostra conclusió és que els centres es regeixen per diferents estàndards en la preparació dels seus alumnes. I podria ser que, a la vegada, els criteris d'avaluació també fossin diferents d'un centre a un altre. Sembla que els profesors (i els centres) estarien classificant i ordenant els seus alumnes sense tenir en compte un referent extern, introduint cada professor el seu propi *biaix*. Els centres estarien puntuant amb criteris i escales diferents encara que com a resultat s'obtinguin distribucions d'aprovats per la *nota COU* similars d'un centre a un altre.

- La *nota PAAU* revela que la preparació dels alumnes no és tan similar com la *nota COU* donaria a entendre. Aquest fet avala la precaució amb que cal prendre tant la *nota COU* com la *nota expedient* i, a la vegada, qüestiona l'afirmació recollida en la *Memoria del Consejo de Universidades* de 1993:

  > "Y hay datos concluyentes[1] de que el expediente es mejor predictor del rendimiento en los estudiantes universitarios

[1] Entenem que per *datos concluyentes* es refereixen als resultats d'un estudi longitudinal, de seguiment —realitzat per un equip d'investigadors dirigit per Tomás Escudero— de 417 estudiants, que el curs 1975-76 superaven les proves d'accés i accedien a les Universitats de Navarra i de Saragossa. El mateix autor, a part d'insistir en les limitacions de l'estudi, diu: " En definitiva, la prueba de madurez académica —seguida de cerca por el expediente secundario— aparece como el mejor predictor del rendimiento universitario cuando se trata con la muestra total". I és que al nostre entendre, el fet més rellevant de l'estudi de T. Escudero (vegeu (Escudero, 1987), on es resumeixen els dos documents

> que las pruebas realizadas (Escudero, 1981, 1986). Por ello no estaría en absoluto justificada la minusvaloración de aquél en el cálculo de la calificación global de acceso".

Lamentablement, el *Ministerio de Educación* es basa en aquests arguments per defensar un major pes de la *nota expedient* en el còmput de la *nota d'accés*.

- El model de regressió de dos nivells de la *nota PAAU* versus la *nota COU* que contempla el gènere, si l'alumne ha repetit o no el COU, les opcions de COU i el tipus de centre (públic o privat) ens porta a una sèrie de conclusions en quant al paper predictiu d'aquestes variables que són coincidents amb altres estudis realitzats a nivell estatal (Muñoz-Repiso et al., 1991). La novetat del nostre enfoc és troba en la utilització d'un model que permet determinar l'*efecte* degut al centre en la *nota PAAU* individual, a la vegada que té en compte la *nota COU* de l'estudiant a més d'altres característiques individuals i de centre.

- Es comprova que els repetidors de COU obtenen resultats comparativament per sota dels seus companys. Les noies obtenen resultats per sota del que s'esperaria segons el seu expedient a COU. Ambdós factors (gènere i repetició de COU) es mantenen significatius al llarg dels tres anys estudiats, 1993-95. En canvi, el factor opció de COU no manté aquesta estabilitat. Una possible explicació d'aquest darrer fet es trobaria en la diferent dificultat de les proves d'un any a l'altre (la variabilitat observada de les mitjanes per matèries de tota la població al llarg dels anys planteja també aquest dubte sobre la dificultat).

---

citats en la *Memoria de ...*) és que posa en evidència diferències segons estudis universitaris: pels estudiants de Medicina (127 de 417, un 30% de la mostra) el rendiment a la universitat apareix molt més associat a la *nota expedient* (correlació de 0.40) que a la *nota PAAU* (correlació de 0.05); pels estudiant d'Enginyeria Superior (35 de 417, un 8%) les correlacions són 0.44 i 0.56, respectivament.

Un altre treball empíric al que sovint hem vist que alguns analistes fan referència és (Touron, 1987). J. Touron estudia el rendiment de 165 estudiants de primer curs de Medicina a la Universitat de Navarra (juny de 1984) i troba que la nota mitjana de les quatre assignatures de ciències a secundària és millor predictora de l'èxit a primer de Medicina que la *nota PAAU*. Llàstima que l'autor no disposés de les notes de les quatre matèries específiques a les PAAU. Potser, les conclusions del treball haurien estat unes altres.

En referència al segon Capítol

En el segon capítol, plantejat com una continuació de l'anterior, hem abordat l'estudi de la variació conjunta entre les dues notes que avaluarien d'una manera directa la preparació per accedir a la universitat: la *nota COU*, responsabilitat del centre de secundària on l'estudiant ha cursat el COU, i la *nota PAAU* obtinguda com a resultat d'una avaluació externa al centre. Les preguntes plantejades a l'inici del capítol eren les següents:

- Existeix associació positiva a nivell centre entre la *nota COU* i la *nota PAAU* o, el que és el mateix, els centres que en COU es mantenen per damunt de la mitjana també fan el mateix a les PAAU?

- Un cop separada la variació entre centres de la variació dins dels centres, quina informació ens aporta aquesta última? Hi ha homogeneïtat dins dels centres? I correlació?

- Si el patró de comportament dels centres no és el mateix, com recollir la diferència entre centres? Concretament, en quina o quines de les següents mitjanes que es proporcionen als centres, val la pena de posar més èmfasi entenent que reflexen diferències significatives per una part important de centres:

  - mitjana de la *nota COU* pels alumnes del centre
  - mitjana de la *nota PAAU*
  - diferència entre ambdues mitjanes

La descomposició de la variació total conjunta de les dues puntuacions mitjanes observades de cada alumne (*nota COU* i *nota PAAU*) en variació entre centres i variació dins dels centres, ens ha permès comprovar que:

- No solament la *nota PAAU* "separa" els centres més que la *nota COU*, sinó que també dins d'un mateix centre la *nota PAAU* discrimina millor que la *nota COU*. Aquests fets confirmen els resultats del primer paràgraf de conclusions del Capítol 1.

- Al mateix temps, l'associació entre l'*efecte centre* de la *nota COU* i l'*efecte centre* de la *nota PAAU* és lleu, molt més lleu que l'associació existent entre els *efectes individuals* per una i altra nota.

- Pel que fa a la "possibilitat" d'una ordenació de centres a partir dels resultats acadèmics dels seus alumnes, podem dir que l'ordenació més informativa, la que pot incloure tots els centres la dóna la diferència entre la mitjana de COU i la mitjana de PAAU de cada centre.

En referència al Capítol 3

Els objectius de la investigació que he presentat en el Capítol 3 es concretaven en:

- Obtenció de mesures de la qualitat de la correcció en un sentit global, per dues assignatures de les PAAU
- Avaluació, per les dues matèries en estudi, de l'efecte degut al procés de correcció en la *nota PAAU* de cada estudiant
- Investigació de les possibles fonts de la variabilitat de la correcció
- Experimentació d'un sistema de control estadístic integrat al desenvolupament de les proves.

Resumeixo les principals conclusions derivades d'aquesta recerca:

- L'estudi sobre la fiabilitat de la correcció ha posat en evidència la baixa qualitat de la correcció en Filosofia i, sorprenentment, també en Matemàtiques. Malgrat els indicadors globals de la qualitat a Matemàtiques hagin estat més "tranquil·litzadors" es donen suficients casos individuals en que la discrepància entre correctors és pot considerar massa gran.
- La *inconsistència* és la principal font d'error (la variància deguda a la *inconsistència* representa un 13% de la variància total a Matemàtiques i un 34% a Filosofia. A Matemàtiques no s'aprecia una diferència entre la *severitat* dels correctors. En canvi, a Filosofia sembla que es produeixen diferents graus de *severitat*.
- Les conseqüències de la discrepància en la correcció afecten a tots els alumnes. En particular, per aquells que es troben en la frontera de l'aprovat pot representar una modificació en el passa/no passa de la prova d'accés. En l'estudi present, 11 dels 363 estudiants participants haurien estat classificats en sentit oposat si s'haguessin tingut en compte les puntuacions que van donar els correctors segons.
- La variància de l'*error de mesura* introduït per la correcció és de 0.838 a Matemàtiques. El pes d'aquesta nota en el còmput de la *nota PAAU* és 0.125 (pels detalls sobre el càlcul d'aquesta nota, vegeu l'Apèndix A). L'error de mesura a Matemàtiques participa en la variància no explicada (per la diferència entre notes verdaderes dels estudiants) de la *nota PAAU* amb un component de variància 0.01309. En el supòsit que

els altres components fossin estimats amb exactitud, l'efecte de l'error a Matemàtiques en la *nota PAAU* tindria una desviació estàndard de 0.1144. Sota aquest supòsit, podem dir que per un 95% dels examinats l'efecte de l'error de mesura pot sobreestimar la *nota PAAU verdadera* fins a 0.2288 o subestimar-la en, com a màxim, -0.2288.

Anàlogament, la variància deguda al component Filosofia —que té el mateix pes que Matemàtiques (pels estudiants que l'hagin cursat) en la *nota PAAU*—, és de 0.02553 (la desviació estàndard corresponent seria de 0.1598). Suposant que la resta de matèries fossin corregides amb exactitud, l'efecte de l'error a Filosfia comportaria que, per un 95% dels examinats la *nota PAAU verdadera* seria sobreestimada o subestimada en com a màxim 0.3196 punts.

- Quan analitzem les raons per les quals els correctors han discrepat en les seves valoracions apareixen diverses fonts d'*incertesa* en el procés d'avaluació. Un nombre important d'aquestes fonts estan lligades al procés d'elaboració de l'examen:
  - el format de l'examen de preguntes de resposta oberta que comporta major subjectivitat i imprecisió
  - la opcionalitat A/B present en cada prova que "enfosqueix" el coneixement de la dificultat de les opcions i preguntes
  - la no exhaustivitat en el seguiment del programa de la majoria d'exàmens, que introdueix un factor "sort" en quant als temes que apareixen en cada convocatòria
  - el reduït nombre de persones, en alguns casos només el mateix coordinador, que participen en l'elaboració de l'examen, limitant la capacitat de detecció d'errades
  - l'absència de coneixement empíric sobre la dificultat i poder discriminador de les preguntes que explicaria, en part, la variabilitat de resultats d'un any a l'altre
- Altres fonts d'*incertesa* estan lligades al sistema de puntuació, com els arrodoniments "forçats" que generen pobresa de puntuació.
- D'altres són pròpies de la correcció, com les "penalitzacions", i poden ser limitades amb la preparació de pautes de correcció i amb l'entrenament dels correctors.
- La nostra impressió és que els correctors no són, en general, *experts*. També és cert que els coordinadors haurien de fer un esforç en saber

comunicar als correctors què es el que volien *mesurar* en l'examen i en cada pregunta en concret. A partir d'aquí les pautes de correcció serien més entenedores i útils.

- L'experimentació d'una interpenetració entre tribunals ens ha permés conèixer millor les característiques del sistema i la possibilitat de futurs dissenys. Com dèiem, en general, a Catalunya en cada tribunal hi ha assignat un sol corrector per matèria. Els exàmens de tots els alumnes d'un mateix centre són corregits, doncs, pel mateix corrector. Els efectes d'un corrector especialment sever poden ser confosos amb una preparació dels alumnes d'un centre per sota de la mitjana. Per tal de separar l'efecte degut al corrector de l'efecte degut al centre cal interpenetrar correctors amb centres. Una possibilitat seria ajuntar 3 o 4 tribunals i repartir a l'atzar els exàmens entre els correctors assignats. Aquesta variant comportaria més dificultats d'organització. Però el guany seria considerable: tindria sentit parlar de diferències per tribunals, diferències entre centres,...

Algunes consideracions

El *Consejo de Universidades* en la Memòria de 1993, "Las Pruebas de de Aptitud para el Acceso a la Universidad: Problemática actual y propuestas de modificación", concretament en la propuesta 3.3 relativa a "Realización de una doble corrección del Comentario de Texto", assenyala:

> Dadas las características especiales de este examen, sería conveniente utilizar para el Comentario de Texto una doble corrección. La doble corrección es utilizada en distintos tipos de pruebas y ejercicios como garantía de ponderación y equilibrio en la calificación final que se otorga. [...] En el caso de la prueba de Comentario de Texto, que no se refiere a una materia específica, sino que tiene un carácter general, y dado que los profesores que han de corregirlo, al ser de cualesquiera de las otras materias que integran la prueba, pueden tener apreciaciones y criterios dispares, se hace más aconsejable el uso de este procedimiento. La calificación final será la media de la otorgada por ambos correctores siempre que no se diferencien en más de 2 puntos. Si la diferencia es mayor, los correctores procederían, en cada caso, a la revisión de las calificaciones efectuadas siguiendo meticulosamente los criterios de corrección específicos establecidos para el examen. La corrección de este examen se dará por concluida una vez todas las calificaciones se hayan obtenido promediando

> puntuaciones con diferencias que no superen el límite establecido (Modificación legal del R. D. 406/88. /Art. 4.1 (nuevo) y de la Orden 3-9-87/ Art. nuevo).

Aquesta proposta no es té en compte en l'actualitat en cap dels districtes universitaris amb els que hem entrat en contacte. Si, per una matèria com el Comentari de Text, es considera no admissible una diferència superior a 2 punts entre dues correccions, quina seria —ens preguntem— la diferència màxima que es podria admetre en altres matèries, com Filosofia o Matemàtiques, en que els correctors són *experts* en la matèria? Segons les dades del present estudi (Capítol 3), 14 dels exàmens de Matemàtiques (un 7%) i 77 de Filosofia (21%) reberen notes que diferien en més de dos punts.

Si volguèssim resumir en poques paraules el que fins ara hem pogut esbrinar sobre el paper avaluador de la *nota COU* i la *nota PAAU*, diríem que:

- La *nota COU* és el resultat d'una avaluació que, en principi, pot ser força completa. El professor té l'oportunitat de conèixer moltes facetes del coneixement i aprenentatge de l'alumne. Però, segons hem pogut veure, els criteris d'avaluació no són els mateixos d'un centre a un altre. Es pot parlar de *biaix*, introduït pel professor de l'assignatura o pel mateix centre. La *nota COU* està afectada de *biaix* relatiu al centre.

- La *nota PAAU* avalua, en principi, a tots els estudiants a partir dels mateixos *instruments de mesura*. L'estandardització està més assegurada en aquesta nota que en la *nota COU*. Però, aquests *intruments* tenen algunes imperfeccions. Hem pogut quantificar la influència de l'error de mesura degut al procés de correcció en dues matèries. L'estudi realitzat revela l'existència d'altres imperfeccions que, en general, afecten per un igual a tots els estudiants. La *nota PAAU* està afectada d'*error de mesura*. Una part d'aquest error és sistemàtic, derivat de l'existència de diferents graus de *severitat* entre els correctors.

- Per tal de comparar els efectes d'aquestes imperfeccions en el còmput de la *nota d'accés* cal tenir en compte que aquesta nota és la mitjana de la *nota expedient* i la *nota PAAU*.

  El *biaix* existent a la *nota COU* també apareix en la *nota expedient*. Via modelització estadística hem estimat la variabilitat (var. = 0.03085, D.E.= 0.1756) del *biaix* o *efecte centre* en la *nota expedient*. Respecte al comportament mitjà de la població, i prenent la variable *nota PAAU* com a variable de control, s'estima que els estudiants d'un mateix centre

de secundària veuen la seva *nota expedient* sobrevalorada o infravalorada en una quantitat comú (a tots els estudiants del centre) que hem anomenat *efecte centre* i que té una Desviació Estàndard de 0.1756.

- En l'actualitat, els alumnes poden sol·licitar una revisió de la correcció de l'examen i segons hem vist en les matèries estudiades la revisió podria comportar una rectificació a l'alça de 0.2288 per Matemàtiques i de 0.3196 a Filosofia. La pregunta és: per què no s'admet la rectificació de la *nota expedient* de tots els alumnes d'un mateix centre si hi ha indicis suficients com per pensar que aquest centre ha avaluat amb més rigor que la resta? Una reclamació d'aquest tipus podria representar un increment de 0.3512 en la *nota expedient*. La introducció d'un *ajust* per la severitat del centre hauria de ser automàtic, un cop estimats els paràmetres del model i calculada la variància dels *efectes centre*. Un *ajust* d'aquest tipus seria molt més just que l'actual sistema de revisió que només té en compte les revisisons sol·licitades pels alumnes.

Donat que els centres de secundària no estan utilitzant els mateixos estàndards en l'avaluació dels seus alumnes, i mentre que aquesta situació no millori (s'homogeneïtzi) és totalment injust que el pes de la *nota d'expedient* sigui el mateix que el de les proves PAAU en el còmput de la *nota d'accés*.

En referència al Capítol 4

En el Capítol 4 hem abordat l'estudi de l'estructura de correlació de les nou puntuacions que cada estudiant obté a les PAAU. Ens interessava explorar l'estructura de correlació tant a nivell individual com a nivell centre. Les preguntes que motivaven aquest estudi eren:

- Es pot identificar alguna raó per les baixes correlacions observades (vegeu taula de correlacions E.1 de l'apèndix E) entre els resultats a les PAAU de les diferents matèries?

- Quina informació aporten els "dos exercicis"? Són diferents?

- Quines matèries o agregacions de matèries són les que millor "separen" els estudiants a nivell global? Quines matèries o agregacions presenten diferències entre centres? Quines matèries o agregacions presenten diferències entre estudiants dins dels centres?

- Com és la relació entre matèries a nivell global? Es veu afectada aquesta relació pel fet de descompondre la variació total en variació dins dels centres i variació entre centres?

- La relació entre variables es dóna de manera homogènia o es pot parlar de diferències per gènere, per opció de COU, per tipus de centre, pel fet d'haver repetit el COU,...

- Quines ponderacions són "raonables" en la situació actual i quines ponderacions serien aconsellables en el futur?

Les conclusions del Capítol 4 es resumeixen en:

- En quant al *rendiment* o *nivell* global mostrat pels alumnes en les PAAU i recollit per les anàlisis ACP, podem dir que els resultats són significativament diferents entre centres públics i privats, entre opcions de COU, entre nois i noies i si l'estudiant és repetidor o no. En canvi si estudiem el vector de notes del *primer exercici* la única diferència significativa, i que és present en les cinc proves, apareix quan l'estudiant és repetidor de COU. La resta d'efectes són més lleus en cada prova i, com dèiem, només es fan distingibles quan es calcula una nota agregada com la *nota PAAU*, el primer component principal en les anàlisis ACP,...

- Quan estudiem les notes del conjunt de proves PAAU, trobem diferències significatives entre les mitjanes dels centres. Aquestes diferències són en part explicades per la composició dels seus alumnes (sobretot per la variable repetició del COU). Però els centres no mostren un comportament uniforme al llarg de les matèries. L'anàlisi dels residus de nivell centre o *efectes centre* en cadascuna de les cinc proves del *primer exercici*, permet evidenciar una gran diversitat de situacions. No es troben, en general, centres que presentin un comportament per damunt (o per sota) de la mitjana global en totes les proves. Més aviat trobem alguns centres que per alguna de les proves presenta una discrepància important respecte del comportament general. Aquestes diferències rellevants entre centres són la causa de que, si es consideren models d'un sol nivell (sense admetre variació entre centres), s'observin algunes interaccions entre covariants. Per això sembla important informar —com ja s'està fent actualment a Catalunya i a altres districtes— a cada centre dels resultats obtinguts en cada prova, així com de la mitjana de tota la població, perquè cada centre investigui.

- Després d'ajustar per una sèrie de covariants i d'admetre descomposició de la variància residual en variació entre centres i dins dels centres, la correlació entre les matèries pels estudiants d'un mateix centre segueix prenent valors baixos. Un error de *mesura* important (fiabilitat al voltant del 60%, com la que s'ha estimat a Filosofia) explicaria en part

la lleu correlació observada entre les matèries a les PAAU. A l'admetre un error de mesura d'aquesta magnitud la correlació estimada entre matèries per les proves del *primer exercici* dins dels centres experimenta un increment important.

- L'error de *mesura* no és la única causa de la lleu correlació entre matèries. Les correlacions entre les puntuacions *verdaderes* també deuen de ser baixes. D'aquí que aquella primera consideració que ens fèiem sobre si valia la pena examinar de dues matèries si aquestes correlacionaven molt no té ara sentit. Segons les nostres anàlisis diríem que els exàmens de les diferents matèries tal com ara estan plantejats mesuren coses diferents i un examen no pot substituir un altre.

- El paper "separador" del *primer* i *segon exercici* varia segons l'opció de COU. Així, pels estudiants de les opcions A i B, el *segon exercici* apareix com a millor separador que el *primer*, mentre que pels estudiants de l'opció C el paper d'ambdós és similar, i pels estudiants de l'opció D el *primer exercici* és millor separador que el *segon*.

  Les matèries que més influeixen en l'ordenació dels alumnes a les PAAU són Matemàtiques I en les opcions A i B i Matemàtiques II i Història del món contemporani per l'opció C. A llarg dels anys aquestes matèries es troben antre les de mitjanes més baixes.

- Les matèries comunes[2] que millor "separen" els estudiant a nivell global són Català i Llengua estrangera (Anglès). Un cop fet l'ajust per les covariants esmentades i admetent la descomposició de la variació entre centres i dins dels centres, resulta: les mitjanes de Català i de Comentari de text són les que més varien d'un centre a un altre; dins dels centres les matèries que presenten més diversitat de puntuacions són Català i Llengua estrangera.

- En el sistema actual, el fet d'utilitzar una "mitjana de mitjanes" per calcular la *nota d'accés* dilueix, en gran part, els efectes de les imperfeccions detectades. Un sistema basat en un nombre inferior de proves i amb una ponderació diferenciada requeriria una elaboració més acurada dels exàmens, així com un major grau de fiabilitat en la correcció.

- A Catalunya[3] les proves PAAU es realitzen en dos dies no consecutius. No s'ha estudiat des del punt de vista estadístic la incidència de l'ordre

[2] Les matèries del *primer exercici* també s'anomenen "comunes".

[3] Els altres districtes universitaris de l'Estat dediquen tres dies a la realització d'aquestes proves.

de les proves en els resultats de les mateixes. Ara bé, crec que no hauríem de menystenir la possibilitat que alguns estudiants acusin un cansament acumulat que els faci baixar el rendiment en les últimes proves.

Malgrat que els pesos de cada component en la fórmula d'accés estan fixats *a priori*, caldria tenir present —com revelen les anàlisis realitzades— que el grau de fiabilitat en la correcció no és el mateix per totes les matèries, que la variància de les matèries difereix (per exemple, les notes extremes 0 i 10 són molt poc freqüents a Filosofia) i que, en conseqüència, el paper "separador" de les matèries no és el mateix.

Segons es despren del Capítol 4 i anteriors, matèries que intervenen amb el mateix pes en el còmput de la *nota PAAU* presenten una capacitat separadora (variància) i una precisió en lá correcció molt diferents. És el cas de Matemàtiques i Filosofia en l'opció A; Filosofia i Biologia en l'opció B.

En referencia al Capítol 5

En el Capítol 5 el nostre interès es centrava en donar resposta a algunes de les preguntes que amb anterioritat ens havíem plantejat quan estudiàvem el sistema d'avaluació PAAU-COU. A la vegada volíem contribuir a la reflexió sobre les novetats que introduïa el sistema PAAU-LOGSE i les repercusions que aquestes novetats podien tenir en el sistema d'accés a la universitat. Ens interessava la identificació d'aquells elements que requereixen un seguiment estadístic i la proposta d'intervencions per tal de garantir un sistema d'accés equitatiu. En concret, les preguntes formulades eren:

- Es pot parlar de diferències significatives entre centres pel que fa als resultats que els seus alumnes obtenen a les PAAU-LOGSE?
- Com és l'estructura de covariància entre la *nota PAAU* i la *nota expedient* de cada alumne? Quina informació pot ser rellevant per als centres?
- Hi ha indicis per dubtar que les escales de puntuació en els centres siguin homogènies?

Resumeixo les respostes i reflexions generades per les anteriors preguntes:

- La variància total de la *nota PAAU* pels estudiants del batxillerat LOGSE és superior a la corresponent variància estimada pels estudiants de COU en anys anteriors. A la vegada, les diferències entre centres són superiors en les PAAU-LOGSE que en les PAAU-COU. Un

20% de la variació total de la *nota PAAU* correspon a variació entre centres (resultat similar al que obteniem en el Capítol 1 pels estudiants de COU).

- Com ja succeïa en el sistema anterior, els nois obtenen millors resultats en les proves PAAU que les noies, a igualtat de condicions pel que fa a la resta de variables. Els estudiants que han perdut algun curs amb anterioritat obtenen resultats significativament més baixos que la resta.

- Les diferències entre centres en la *nota PAAU* experimenten un increment quan ajustem (per regressió) aquesta nota amb la *nota expedient*.

- S'observen diferències significatives entre opcions de PAAU, tan pel que fa a la *nota expedient* com a la *nota PAAU* però en sentit contrari. Els estudiants de les opcions C: Humanitats i E: Arts que, en mitjana obtenen els resultats més baixos a secundària, són els que, comparativament, obtenen la millor *nota d'accés*.

- El fet que l'accés a alguns estudis universitaris no estigui restringit a determinades opcions, afegit a la desigualtat generada per la no aplicació de criteris uniformes de puntuació entre assignatures a les PAAU, provoca una situació de clara injustícia.

  Aquesta situació de desequilibri pot facilitar l'aparició d'estratègies d'accés no desitjables.

  A la vegada, es posa en qüestió el sentit de l'actual fórmula d'accés, concretament les ponderacions assignades a cada matèria.

  És del tot urgent, assegurar un tracte equitatiu en l'accés a estudis universitaris que no demanen prerequitis concrets. No sembla adequat examinar de matèries que no es consideren rellevants pels estudis universitaris que vol seguir l'alumne si, a la vegada, no es pot garantir l'aplicació d'un criteri d'avaluació uniforme.

- La lleu diferència observada entre els resultats a les PAAU dels alumnes segons l'any en que el seu centre es va incorporar al sistema LOGSE (els que obtenen millors resultats són els del 94 i 97) s'explicaria en gran part per la diferent preparació d'aquests alumnes —reflectida en la seva *nota expedient*.

- Les diferències observades entre centres, fins al moment —encara que la major part no hagin estat significatives—, alerten sobre la necessitat de seguiment dels resultats prenent en consideració una sèrie de

característiques: localització del centre (món rural/urbà), grandària, privat/públic, modalitats de batxillerat que s'imparteixen.

6.2 Sobre la metodologia

- La naturalesa de les dades generades per les proves PAAU (estructura jeràrquica: estudiants agrupats en escoles) i el tipus de preguntes que ens fèiem a l'inici d'aquesta tesi recomanaven una anàlisi via models de nivell múltiple. Crec que aquest enfoc s'adiu molt al context en estudi, perquè utilitza informació de cada individu; a la vegada, té en compte les agrupacions existents (dades correlacionades); i estima amb eficiència paràmetres en ambdós nivells d'anàlisi: estudiants i centres.

  En la majoria del models utilitzats s'ha pressuposat homogeneïtat de la variància —models (1.3) i (4.1), per exemple—, tan pel que fa a la variància entre centres com a la variància dins dels centres. En el Capítol 1 s'apuntava la necessitat de revisar aquesta hipòtesi.

  En el futur seria interessant investigar la possibilitat de modelitzar la variància amb algunes de les variables que semblen introduir diferències. Entenc que caldria treballar amb una mostra més gran de centres.

- Els estimadors *encongits*, utilitzats per exemple en l'estimació dels *efectes centre*, incorporen eficiència. El seu paper "corrector" fa que resultin més estables que altres indicadors més simples: els residus o *efectes centre* en la *nota PAAU* provinents de la regressió amb la *nota COU*, analitzats en el Capítol 1, resultaren ser més estables que les diferències entre mitjanes d'ambdues notes. Per aquestes raons, crec que els estimadors *encongits* poden tenir molta aplicació en educació en el càlcul de mesures relatives a col·lectius.

- La modelització estadística presentada en el Capítol 3 ha permès el càlcul d'una sèrie d'indicadors de la qualitat de la correcció. Malgrat els resultats obtinguts no poden generalitzar-se a d'altres matèries, anys o regions, són un punt de partida per les anàlisis en anys posteriors. El seguiment de la qualitat del procés demana una recollida sistemàtica de dades que aportin informació sobre el procés. En aquest sentit, creiem que la modelització presentada (descomposició de la variància de la nota observada en diferents fonts) és adequada perquè es pot aplicar a totes les matèries, la seva simplicitat permet divulgar-ne els resultats entre els no-estadístics, i admet ampliacions que recullin altres fonts d'error no incorporades en l'estudi actual. L'opció d'examen seria una

variable a introduir en el model, a la vista de la incidència que ha tingut en els resultats de l'any 95.

- Per dissenyar l'experiment del Capítol 3 ens va caldre conèixer l'estructura de l'organització. Insistim en la necessitat de que una part de la recerca s'ha de fer en el mateix moment en que es realitzen les proves i en la necessitat d'aconseguir una interpenetració de tribunals i correctors. L'actual estructura en que tots els exàmens d'una mateix centre són corregits per un mateix corrector —que en la majoria de casos està adscrit a un sol tribunal—, facilita les tasques de control d'exàmens que, no cal dir-ho, són primordials però dificulta el seguiment estadístic de la labor dels correctors i centres.

- Monitoritzar una investigació lligada a la ejecució té, no solament el valor de proporcionar informació veraç, sinó que permet intervenir per realitzar "ajustos" a temps, si es dóna el cas. En aquest sentit seria interessant estudiar des del punt de vista estadístic algunes "fórmules d'ajust" de les imperfeccions observades. Penso en l'ajust pel grau de *severitat* del corrector (Longford, 1995), en l'ajust per l'opció d'examen escollida, en l'ajust per la "fidelitat" del centre als estàndard fixats per l'Administració,...

- L'opció d'examen és un factor a tenir en compte. Un ajust per l'opció demanaria que l'examen constés d'una part comuna per tal poder equiparar resultats. En la teoria de *Test equating* es coneix amb el nom de *anchor* aquest element comú que permet calcular probabilitats condicionades i dóna peu a fórmules d'equiparació (Holland and Rubin, 1982).

- Un tema educatiu pendent i que donaria lloc a un estudi interessant també des del punt de vista estadístic és el d'establir quin aspecte de la *maduresa* intenten avaluar en comú les cinc proves del *primer exercici* i quin(s) aspecte(s) de la *maduresa* avalua cada prova de manera aïllada. Un cop aclarit aquest punt es podria passar a fer-ne el corresponent seguiment estadístic. Podríem descompondre la nota observada en una mesura de l'habilitat general més una mesura de la(les) habilitats específiques.

- L'error de *mesura* incrementa la variació observada entre estudiants dins de les escoles. Si es redueix l'error de *mesura* probablement observaríem una correlació intra-centres més gran en cada matèria (diferències reals entre centres majors del que ara observem) així com una correlació entre matèries també més gran (major associació entre les notes verdaderes de les matèries (Woodhouse, 1996)).

- El càlcul de la qualitat de la correcció l'any 97 ha posat en evidència les limitacions del coeficient de correlació mostral com a mesura de la fiabilitat de la correcció. En el cas que es donin diferents graus de *severitat* entre els correctors, el coeficient de correlació mostral dependrà del disseny (assignació d'exàmens a correctors). Atès que en algunes matèries de les PAAU hi ha indicis suficients com per suposar que coexisteixen diferents graus de *severitat*, es recomanable estimar la fiabilitat via la modelització presentada en el Capítol 3.

- La modelització del vector de notes PAAU proposada en el Capítol 4 permet estimar el paper d'una sèrie de covariants en el rendiment de cada matèria; a la vegada, descompon la variància residual en variació entre centres i variació dins dels centres. Entre altres fets, la desigual distribució de repetidors de COU en centres públics i privats demana ser tenida en compte en l'anàlisi de la variació. Com a resultat de l'estimació, s'obtenen els residus o *efectes centre* en cada matèria que venen a ser "el valor afegit de cada centre" relatiu a la matèria. L'estimació conjunta d'aquests *efectes* via el model multivariant proposat (4.1) guanya en eficiència respecte al model univariant sempre i quan existeixi correlació entre centres. En el nostre cas, l'associació entre efectes a nivell centre és molt lleu i per tant els resultats no variarien gaire d'un model a un altre. Jo, m'atreveixo a dubtar de la "no associació". Més aviat em decanto per pensar que el sistema de puntuació a les PAAU no és prou precís i no recull el que hauria de recollir. Per això mantinc que em sembla més adequat el model multivariant i que cal treballar en la millora de les puntuacions (i de tot el procés d'avaluació).

  La modelització (4.1) admet vàries generalitzacions: introducció de la nota de COU corresponent a la matèria (en totes les matèries excepte Comentari de text); modelització dels efectes aleatoris,...

6.3 Monitorització de les proves PAAU: Propostes

En l'article *Monitoring the university admissions process in Spain* es trobem recollides moltes de les reflexions que ens hem anat fent a mesura que avançava la recerca així com propostes de millora de tot el procés, algunes de les quals ja s'han posat en pràctica[4]. Les àrees d'actuació serien

[4] *Monitoring ...* ha estat acceptat per a la publicació en *Higher Education in Europe.* La redacció es va iniciar durant una estada —finançada amb una beca CIRIT per a la recerca

1. Elaboració i correcció dels exàmens: homogeneització de la correcció i intervenció per corregir discrepàncies.

   - Considerar en cada matèria la possibilitat de substituir l'examen actual, o una part del mateix, per una prova de preguntes de resposta tancada
   - De manera sistemàtica, i pels exàmens de preguntes de resposta oberta, en cada convocatòria i per cada matèria, s'hauria de seleccionar una mostra d'exàmens i realitzar-ne una doble correcció per tal de mesurar la *fiabilitat* de la correcció i detectar fonts de discrepància.
   - Investigació i aplicació d'un "ajust per la severitat" en aquelles matèries que ho requereixin.
   - Per matèries amb una *fiabilitat* baixa, realitzar la doble correcció de tots els exàmens, com ja recomanava el *Consejo de Universidades* el 1993 i utilitzar la mitjana de les dues notes en l'avaluació.
   - Per matèries amb una fiabilitat mitja-baixa i amb un nombre reduït de correctors realitzar la doble correcció d'un nombre n (uns 30) d'exàmens de cada corrector. En el cas de detectar diferents graus de *severitat* es recomana l'aplicació de l'ajust corresponent.
   - Proposar i experimentar (abans d'aplicar de manera generalitzada) un "ajust per la desviació del centre respecte dels estàndards" en la *nota expedient*.
   - Revisar què es vol avaluar en cada matèria i construir exàmens amb preguntes el més tancades possible. Un examen amb moltes preguntes i de resposta tancada o breu permet revisar moltes parts de la programació fet molt desitjable, sobretot en assignatures bàsiques com les Matemàtiques per a un estudiant d'Enginyeria.

2. Formació dels coordinadors i correctors

   - L'estudi de la qualitat de la correcció s'ha d'estendre a totes les matèries fent partícips als coordinadors perquè un cop es disposi de dades incorporin els seus suggeriments a la millora de les proves.
   - Investigació de la dificultat d'exàmens realitzats amb anterioritat

a l'estranger— al Departament of Medical Statistics, de Montfort University, Leicester, la tardor de 1996. L'article es va presentar com una comunicació al 10th European Meeting of the Phychometric Society, Santiago, 1997; i a la European Conference on Educational Research, ECER 97, a Francfurt.

- Confecció de bancs d'exàmens i ítems dels quals es conegui la dificultat i la capacitat discriminadora entre d'altres característiques perquè han estat provats empíricament.
- Consensuar les pautes de correcció específiques de cada examen amb una mostra de correctors
- Investigar l'impacte de l'opció d'examen. Proposar i experimentar un "ajust per l'opció d'examen". A la vegada, anar reduïnt al màxim l'opcionalitat dels exàmens.
- Separar la feina de "vigilant" de la de corrector
- Preparar els correctors per la seva tasca
- Proposar una interpenetració de tribunals i correctors

3. Creació d'un sistema d'informació

   - Si es vol comparar els resultats de les proves al llarg dels anys, donat que l'estudi alerta sobre la no homogeneïtat dels exàmens d'una mateixa assignatura, caldria pensar en proves que permetin estudis comparatius amb una part que faci d'*anchor*.
   - Fer un seguiment de les proporcions d'estudiants que van superant cada fase del procés d'avaluació i investigar la estabilitat d'aquests indicadors

6.4 Consideracions d'àmbit educatiu

Entre les consideracions pedagògiques que es desprenen del present estudi, destaquem les següents:

- De les anàlisis estadístiques es desprèn la recomanació de limitar al màxim l'opcionalitat en els exàmens. Ara bé, la possibilitat d'escollir entre dues o més opcions està clarament justificada en algunes matèries en que el context en que es presenten les preguntes o exercicis pot resultar discriminatori. Per exemple, a l'intentar avaluar l'expressió en una llengua és preferible oferir a l'alumne diversos temes de manera que puguem avaluar la seva capacitat independentment del tema escollit.

- L'examen de Filosofia ha generat més *inconsistència* que el de Matemàtiques. A la vegada en la prova de Matemàtiques es dóna més variabilitat de notes que en la de Filosofia. Potser la raó es trobi en la naturalesa d'aquestes assignatures. Ens preguntem:

  - Es pot posar un examen de Filosofia amb preguntes més objectives (de resposta tancada, curtes,...) que l'actual?
  - En cas afirmatiu, com serien els resultats? Es guanyaria amb fiabilitat?
  - Es just que en l'accés a la Universitat d'un estudiant de Ciències, per citar un cas, tingui el mateix pes la nota de Matemàtiques que la de Filosofia, si, com acabem de veure, són assignatures amb diferent capacitat discriminadora i diferent qualitat en la correcció?

- L'anàlisi dels resultats de la prova de Matemàtiques I, posa en evidència les dificultats dels estudiants en respondre a preguntes sobre probabilitats. Es important assegurar una formació actualitzada dels professors de matemàtiques —i cito aquesta assignatura perquè és la que millor coneixo, sense menystenir el que pugui fer referència a la resta— en tots els temes del programa i, en particular, en els temes de probabilitat i anàlisi estadística. L'actualització dels continguts de matemàtiques és un tema de discussió en molts altres països[5].

- El fet que comparativament les noies obtinguin pitjors resultats a les PAAU que els nois a igualtat de condicions, planteja un seguit de preguntes que caldria adreçar a educadors i psicòlegs:

  - Per què les noies tenen més dificultat que els nois en demostrar la seva preparació a les PAAU?
  - Es el tipus de prova (molts exàmens en poc temps) que no les afavoreix?
  - Se senten les noies més pressionades per un sentit propi de responsabilitat, de projecte personal de futur?
  - O potser és que les noies són avaluades a la secundària amb més "generositat" que els nois?

  Segurament l'avaluació a la secundària té en compte alguns aspectes formatius, com la realització de treballs, que no es tenen en compte en les PAAU. En aquest sentit, la *nota expedient* i la *nota PAAU* no estarien avaluant el mateix. El tema és prou interessant com per requerir un anàlisi en profunditat. Però voldria apuntar alguns elements de discussió:

[5] Vegeu en (Boland and Nicholson, 1996) una breu discussió sobre els darrers canvis introduïts en el curriculum de secundària en països del món anglosaxò i de les mesures d'adaptació que aquests països han anat prenent al respecte.

  - Cal introduir la recerca a secundària? Jo crec que si, però ha de comptar amb una bona orientació per part del professor.
  - Què en pensen els professors? Tenen clars els objectius dels treballs que proposen? Hi poden dedicar el temps que voldrien?

- L'Estadística ens ha permès constatar l'existència de diferències en els resultats a les PAAU[6] d'un centre a un altre —diferències explicades només en part per la composició d'alumnes—. Les anàlisis revelen que hi ha indicis suficients com per pensar que els centres de secundària no estan utilitzant els mateixos estàndards en la formació i/o l'avaluació dels seus alumnes. Es evident que si algun centre destaca per "massa generós" cal que la Inspecció intervingui (com es fa actualment) per tal que no torni a passar. Però, a la vegada, seria bo poder corregir el *biaix* introduït en el càlcul de la *nota d'accés* de la promoció d'estudiants afectats sense haver d'esperar al curs següent. De totes maneres, caldria actuar en l'arrel del problema: en la formació dels estudiants. D'aquí la necessitat d'orientar els professors i els centres de secundària.

- Per un sistema educatiu és molt important disposar de dades fiables. En aquest sentit, les proves PAAU, com a examen extern (l'únic, actualment) als centres i estàndard, es revela com a un instrument de gran utilitat (vegeu (Martí-Recober et al., 1997)).

- Les dades en educació —i entre elles, els resultats de les proves PAAU— són dades observacionals. El seguiment de tot procés educatiu i la necessària comparació de resultats requereixen rigor i contextualització. Si es garanteixen aquests dos elements, la prevenció que existeix actualment envers la "comparació de centres" no tindria sentit de ser.

- Cal fer un pas endavant en la "cultura" de realitzar estudis que siguin útils a la Administració i que, al mateix temps, serveixin de referència i contrast per als centres.

- Per això, és tan important fer partícips als professors de secundària de la necessitat que ells mateixos analitzin els resultats dels seus alumnes. Proposem, en aquesta línia, que la Oficina de *Coordinació* [...] subministri als centres un fitxer amb les dades dels seus alumnes (resultats a les PAAU i altres informacions de què disposa) juntament amb les informacions relatives a tota la població (dades agregades per regions, comarques,...) i un software fàcil de manejar per tal que cada professor

[6] Ens referim als dos sistemes: PAAU-COU i PAAU-LOGSE. En aquest darrer, fins i tot, es podria parlar d'un increment.

pugui situar les dades del seu centre en el marc de la població. Més que un sistema de control dels centres, estem proposant un sistema que ajudi als centres en el reconeixement dels seus *punts forts* i els seus *punts febles*, orienti la discussió i en permeti fer una diagnosi.

- La necessària estandardització del nivell amb què els estudiants haurien d'accedir a la universitat contrasta amb el nou batxillerat LOGSE que no solament no preveu un curs orientat a la universitat, com era fins ara la funció del COU, sinó que proposa programacions genèriques sense una concreció clara dels continguts. Aquestes mancances poden fàcilment produir una diferenciació entre la preparació dels alumnes d'un centre a un altre major encara que en el sistema que fineix (BUP-COU). Per tot això és ara molt important que la universitat "tuteli" els centres de secundària. Altrament, en pocs anys, pot trobar-se amb què l'alumnat que accedeixi a la universitat no tingui la preparació suficient. Seria bo que els professors de secundària disposessin de materials i propostes d'exercicis del nivell que s'espera que un estudiant tingui a la fi del batxillerat i del qual hauria de respondre en un examen d'accés a la universitat.

- Les proves d'accés a la universitat són, per a la majoria d'estudiants, el primer examen extern al qual s'enfronten. Seria bo que amb anterioritat haguessin realitzat exàmens globals. Valoro positivament —per moltes raons—, la proposta d'introduir una prova externa de graduació a la fi del batxillerat LOGSE. A la vegada, és important que els estudiants, en el decurs de l'ensenyament secundari, hagin preparat i realitzat exàmens que abastin una part important de la programació.

- La àmpliament argumentada necessitat de proposar exàmens de "preguntes de resposta el més tancada possible" planteja un canvi en la pedagogia. Els professors de secundària haurien d'anar incorporant aquests tipus de proves com una eina de docència —limitar-les a l'avaluació és desaprofitar el seu potencial com a eina d'aprenentatge.

- Les habilitats en comunicació escrita són de gran importància en les proves de resposta oberta —hem parlat de limitar-les, no d'eliminar-les de les PAAU—. Donat que aquestes habilitats necessiten d'aprenentatge, està clar que hauria de ser una de les pràctiques prioritàries en secundària.

- Les dues darreres consideracions han sorgit de la reflexió sobre les proves d'accés a la universitat. La proposta que fèiem s'exten als estudiants

que realitzant el batxillerat LOGSE tinguin planejat la incorporació al món laboral sense seguir estudis universitaris. També a ells —i més aviat, encara que als seus companys—, se'ls demanarà capacitat de comunicació oral i escrita, de treball en grup, de concreció.

- Donat que hi ha indicis suficients com per dubtar que tots els centres de secundària que estan impartint el batxillerat LOGSE disposin de les condicions i orientacions necessàries per garantir una preparació adequada dels seus alumnes en la fase d'accés a la universidad, és important intensificar la coordinació amb els centres que imparteixen el batxillerat LOGSE orientant-los en els continguts i nivell en cada matèria. També en el sistema LOGSE, malgrat haver millorat el sistema de puntuació, s'acusa la manca de estàndards d'avaluació en el batxillerat. Entenem que és important que els professors dels centres tinguin present, quan avaluen els seus alumnes, el significat de cada puntuació en concret i la necessitat de mantenir aquest significat d'un any a l'altre. L'administració hauria d'arbritar sistemes d'assessorament, coordinació i inspecció en aquesta direcció.

Espero que les "Consideracions..." anteriors siguin enteses com el que són: fruit en part de les anàlisis estadístiques i fruit en part de l'experiència docent. Aquestes darreres s'han de prendre com a opinions: l'experiència de cada persona ja sabem que és reduïda.

6.5 Preguntes obertes i Perspectives de recerca

No hem abordat en aquesta tesi el tema de l'efectivitat de les escoles. En educació, els treballs que tracten aquest tema solen tenir en compte mesures de la preparació dels alumnes a l'inici d'un cicle i a la fi del mateix, juntament amb covariants que recullen característiques dels alumnes i/o de les escoles. Els models de nivell múltiple presentats en aquesta tesi han resultat també de molta utilitat en estudis sobre l'efectivitat de les escoles (Kreft (1987), Gray and Wilcox (1995), Fitz-Gibbon and Vincent(1994)). En el nostre país comença a haver-hi una valoració i interés per la recollida sistemàtica d'indicadors educatius. Cal citar en aquesta línia el treball iniciat per Santamaria, Ribera i Vilardell(1998) sobre eficiència relativa entre centres de secundària al Vallès Occidental. L'esforç de definició de les variables que poden ser rellevants en educació és molt important i, naturalment, forma part de l'establiment de sistemes d'informació i plans per a la millora de resultats. Per

exemple, el treball de Kreft (1987) posa en evidència la importància de les variables "estils en la docència", "clima en l'escola", "clima a la llar",... per davant de variables que només fa dues dècades semblaven les més influents en el rendiment dels alumnes (nombre d'alumnes per aula, recursos econòmics de l'escola,...).

Es d'esperar que les millores en l'escola redundin en una millor orientació i preparació dels alumnes, cap a la formació professional o universitària. Per això és tan important endegar projectes de recerca en l'àmbit de la secundària i divulgar-ne els resultats. Un coneixement més profund de la secundària facilitaria la confecció de les proves PAAU. A la vegada un coneixement dels posteriors resultats en la universitat aportaria elements d'interés en l'actualització de les proves. La major part dels treballs sobre rendiment a la universitat i puntuacions a les PAAU (dels que tinc coneixement) s'han efectuat des de la universitat, amb col·lectius reduïts, tenint en compte puntuacions mitjanes (*nota expedient*, *nota PAAU*,...).

En aquesta tesi m'he centrat en l'estudi de les notes de COU i de les PAAU, en la relació entre ambdues,... tenint sempre en compte el caràcter continu d'aquestes puntuacions. No he abordat altres tipus de dades, també interessants, com les que es deriven del fet de superar o no cada fase avaluadora. L'estudi de les proporcions d'alumnes que superen cada fase (estudi que ja es pot iniciar a la secundària) i les variables relacionades està pendent.

Un altre tema pendent es la introducció de l'error de mesura en la modelització de la variació de les notes PAAU i la seva relació amb algunes covariants. L'error de mesura detectat en la correcció de les proves PAAU té en algunes matèries un component sistemàtic, a més de l'aleatori. Per totes aquestes raons, crec que els models d'equacions estructurals per variables latents (Muthén and Satorra, 1989) —i en particular, l'Anàlisi de classes latents (LCA, vegeu l'article introductori de Hagenaars, 1998) —podrien ser models de mesura adequats a les variables en estudi i a les preguntes que susciten. Un altre enfoc que pot resultar interessant és el que ofereixen els models Log-lineals, amb variables directament observables o amb variables latents.

Un altre àmbit de recerca interessant seria el de la Teoria de Resposta a l'Ítem (IRT) (Lord, 1980). Concretament, a partir de l'estudi descriptiu de 1997 sobre la *dificultat* i la capacitat *discriminadora* de les preguntes, es podria proposar una modelització de l'encert/fracàs que tingués en compte els paràmetres citats a més del paràmetre *elecció* de l'opció d'examen.

Per últim, un àmbit de recerca on ja he començat a treballar i voldria continuar-hi és la proposta de fórmules d'*ajust* per tal de reduir l'impacte de la *severitat* dels correctors en les PAAU, de l'opcionalitat d'aquestes proves, dels diferents estàndards d'avaluació en els centres de secundària (*severitat* dels centres en la *nota expedient*).