



ÚS DE QÜESTIONARIS EN LA DOCÈNCIA D'ECOLOGIA I LA SEVA ANÀLISI MITJANÇANT LA PLATAFORME MOODLE

Lluís Zamora Hernández
Facultat de Ciències
Departament de Ciències Ambientals
lluis.zamora@udg.edu

Lluís Benejam i Vidal
Facultat de Ciències
Departament de Ciències Ambientals
lluisbenejam@hotmail.com

Objectius

Els canvis o renovació metodològica que proposa l'EEES, ha de comportar també un canvi en els criteris i procediments d'avaluació sobre l'adquisició de les competències corresponents pel fet que el sistema establert per avaluar una matèria és, en darrera instància, el factor que modula la metodologia d'estudi que utilitza l'alumne (Arbós, 2010).

Durant el curs 2009/10, la Universitat de Girona ha implantat la plataforma virtual interactiva Moodle, adaptada a la formació i emprada pel professorat com a suport a la tasca docent. Aquesta plataforma ofereix la possibilitat de crear nous recursos, alguns dels quals enfocats a l'avaluació de l'aprenentatge. A l'igual que qualsevol pràctica educativa, aquests recursos han de ser sotmesos a processos d'anàlisi i revisió orientats a la seva millora, processos de metaavaluació (avaluar les avaluacions) els quals han de permetre detectar contradiccions i introduir canvis que contribueixin la millora dels recursos produïts i, de retruc, del procés d'avaluació.

En el marc de diferents assignatures impartides per professors de l'Àrea d'Ecologia, s'han generat un seguit de proves d'opció múltiple (qüestionaris) utilitzant la plataforma Moodle, com a eina per al seguiment i avaluació dels aprenentatges (avaluació contínua). Els objectius de la present comunicació són:

1. Realitzar una metaavaluació dels qüestionaris desenvolupats per a la docència de l'Ecologia, emprant les eines disponibles a la plataforma Moodle
2. Analitzar el nivell de dificultat i poder de discriminació dels diferents ítems que formen els qüestionaris dissenyats
3. Comparar els indicadors psicòmètrics en funció del tipus de qüestionari i de la implementació de retroaccions

Desenvolupament

S'han analitzat 9 qüestionaris amb preguntes d'opció múltiple. Les característiques de cada qüestionari es resumeixen a la taula 1. Tots els qüestionaris tenen per objecte avaluar els coneixements bàsics i la seva comprensió (primers nivells de la piràmide de Bloom). En el cas del qüestionari de tipus adaptatiu i més d'un intent, s'han analitzat els resultats del darrer intent. Totes les proves s'han realitzat el curs 2010/2011.

Els diferents ítems de cada qüestionari han estat analitzats a partir de tres indicadors psicomètrics objectius, per tal de valorar la seva qualitat, detectar possibles errors i enunciats ambigus, mesurar el grau de dificultat i el poder de cada pregunta per discriminar entre participants més eficients o menys eficients. S'han recollit els quatre indicadors que proporciona l'eina d'anàlisi de la plataforma Moodle: l'índex de dificultat, la desviació típica de les respostes, l'índex de discriminació i el coeficient de discriminació (Crocker, 1986; Backhoff *et al.*, 2000; Gil *et al.* 2009) tot i que només s'analitzen en aquest estudi els resultats per l'índex de dificultat i el coeficient de discriminació.

Taula 1. Característiques dels qüestionaris analitzats. ¹Assignatures: Eco(Bio)= Ecologia (Llicenciatura Biologia); Eco(CCAA)= Ecologia (Llicenciatura Ciències Ambientals); EcoPob= Ecologia de Poblacions (Llicenciatura Biologia). ²Tipus: NA= No adaptatiu; A= Adaptatiu.* Només s'analitzen els resultats d'un dels grups de pràctiques que l'han realitzat.

Núm.	Assignatura ¹	Nombre alumnes	Tipus ²	Nombre ítems	Nombre respostes per ítem
1	Eco (Bio) Grup A	63	NA	80	4
2	Eco (Bio) Grup B	52	A	80	4
3	Eco (Bio) Grup A	67	A	40	4
4	Eco (Bio) Grup B	57	NA	40	4
5	Eco (CCAA)*	29	NA	25	3
6	Eco (Bio)	103	NA	7	3
7	Eco (Bio)*	24	NA	25	3
8	EcoPob	35	A	50	4
9	EcoPob	36	A	68	4

Per tal de valorar si la incorporació de retroaccions en els qüestionaris podia tenir algun efecte sobre els resultats de la prova, es varen dissenyar dos qüestionaris per l'assignatura "Ecologia" (Llic. Biologia). De cadascun d'aquests, es va fer una versió sense retroaccions i una segona versió amb una retroacció incorporada a cada resposta explicant el perquè es considerava que era incorrecta o correcta. Es va dividir el total d'alumnes en dos grups (A i B, comprovant que no hi havia diferències significatives en la nota mitjana d'expedient) i cada grup va realitzar una de les proves. Es va fer també una enquesta de valoració a tots els estudiants sobre les dues proves.

Avaluació

Dificultat

La dificultat d'un ítem s'interpreta en funció de la proporció d'alumnes que responen correctament la pregunta. En realitat, mostra la "facilitat" de la pregunta pel fet que aquesta tindrà una dificultat menor si la proporció de respostes correctes és elevada (Wood, 1965). D'acord amb Backhoff *et al.* (2000), el nivell mig de dificultat d'un qüestionari de resposta múltiple hauria d'oscil·lar entre 50 i 60%, tot i que altres autors proposen que (a partir de tres opcions de resposta) s'hauria de situar entre 62 i 66% (Zimmaro, 2003). La distribució dels valors de FI hauria de ser : 5% de ítems fàcils, 20% mitjanament fàcils, 50% de dificultat mitjana, 20% mitjanament difícils i 5% difícils.

La taula 2 presenta el valor promig i les desviacions estàndards dels valors de l'índex de dificultat de les 9 proves analitzades. Es pot observar com tots els qüestionaris presenten un valor mig de dificultat superior a l'ideal de manera que han resultat "fàcils" per l'alumnat. El valor més elevat correspon al qüestionari 6, dissenyada com una prova simple sense impacte sobre la nota. Per contra, la prova 5 correspon a l'avaluació de la part de pràctiques de l'assignatura, resultant molt fàcil i posant en dubte la validesa d'aquesta prova com a sistema d'avaluació de l'aprenentatge. Respecte al comportament de la dificultat dels diferents ítems per cada prova, es pot observar que: (1) en general, tots els qüestionaris presenten un nombre excessivament alt de preguntes "fàcils" i "mitjanament fàcils", (2) la distribució d'aquests valors té generalment un biaix negatiu i (3) pràcticament no presenten preguntes

“difícils”. Sense considerar el qüestionari número 6, per la seva particularitat, els qüestionaris de tipus adaptatius presenten una dificultat menor (FI més alt) que els no adaptatius (ANOVA, $F_{1,406}:4,787$; $p<0,05$) probablement degut a la possibilitat de disposar d’una segona opció per respondre la pregunta tot i el grau de penalització.

Taula 2. Dificultat mitjana de les diferents proves. Mitjana i desviació típica (SD) calculades a partir del valor de l’índex de dificultat (FI) de tots els ítems de cada prova. La distribució dels valors de FI desitjada hauria de ser : 5% de ítems fàcils, 20% mitjanament fàcils, 50% de dificultat mitjana, 20% mitjanament difícils i 5% difícils

Núm.	Mitjana	SD	Assimetria	Curtosi	Nivell de dificultat desitjat				
					5%	20%	50%	20%	5%
1	65,8	22,6	-0,74	-0,17	5,7	36,4	56,8	0,8	0,4
2	71,9	18,7	-1,13	1,13	1,8	45,6	51,8	0,7	0,0
3	72,9	14,2	-0,29	-0,36	10,9	32,1	56,9	0,0	0,0
4	69,5	15,2	-0,29	-0,38	7,5	27,1	65,4	0,0	0,0
5	80,0	15,6	-1,01	1,56	26,0	45,8	28,1	0,0	0,0
6	92,4	6,3	-1,42	2,20	48,4	51,6	0,0	0,0	0,0
7	71,2	28,0	-0,91	-0,37	27,2	34,8	35,9	2,2	0,0
8	76,4	11,6	-0,72	-0,12	0,0	57,1	42,9	0,0	0,0
9	73,8	14,4	-0,63	-0,14	8,4	42,2	49,4	0,0	0,0

Discriminació

S’analitza el coeficient de discriminació (DC), mesura basada en la correlació entre les puntuacions de l’ítem i la puntuació total de la prova, per basar-se en el total d’alumnes que han realitzat la prova. El seu rang de valors és entre +1 i -1.

La taula 3 presenta els resultats de l’anàlisi dels diferents qüestionaris a partir del coeficient de discriminació. Segons Zimmaro (2003), les preguntes amb un valor igual o superior a 0,4 poden ser considerades com molt bons ítems i han de restar a la prova. Valors entre 0,2 i 0,4 indica preguntes amb poc poder de discriminació i cal ser revisades si es vol repetir la prova en un futur. Ítems amb valors per sota de 0,2 haurien de ser descartats pel fet que la majoria de respostes correctes corresponen als estudiants amb pitjor qualificació de manera que la pregunta discrimina poc. Quan tots els estudiants han respòs correctament la pregunta (FI=1), no es pot calcular l’índex DC tot i que s’hauria de descartar l’ítem per ser molt fàcil i no tenir poder de discriminació.

Taula 3. Poder de discriminació de les diferents proves. Mitjana i desviació típica (SD) calculades a partir del valor del coeficient de discriminació (DC) de tots els ítems de cada prova. Es mostra el percentatge de preguntes de cada qüestionari en funció de diferents intervals de valor (veure text per més informació). ND= no es pot calcular l’índex (FI=1, SD=0).

Núm.	Mitjana	SD	Assimetria	Curtosi	% preguntes amb valor de l’índex				
					ND	<0,2	0,2-0,4	0,4-0,6	>0,6
1	0,42	0,15	-0,31	0,14	1,25	7,50	30,00	52,50	8,75
2	0,47	0,14	-0,37	-0,36	1,25	3,75	23,75	48,75	22,50
3	0,36	0,11	-0,68	1,04	0,00	7,50	50,00	42,50	0,00
4	0,42	0,14	-0,86	0,94	0,00	10,00	30,00	55,00	5,00
5	0,31	0,35	-0,81	-0,06	4,00	28,00	20,00	32,00	16,00
6	0,68	0,10	0,31	-2,19	0,00	0,00	0,00	14,29	85,71
7	0,31	0,19	-0,09	0,03	16,00	24,00	36,00	20,00	4,00
8	0,69	0,09	-0,34	0,62	0,00	0,00	0,00	8,00	92,00
9	0,42	0,19	-0,49	-0,09	0,00	10,29	30,88	32,35	26,47

S’observa que el poder discriminatori mitjà de tota la prova oscil·la entre 0,31 i 0,69 valors que es troben entre el que seria desitjable. Però si analitzem els ítems de cada prova, destaca que un percentatge molt elevat (entre el 4 i 28%) hauria de ser descartat pel seu poc poder discriminatori i aproximadament un 30% de les preguntes haurien de ser revisades en cas de voler repetir el qüestionari. En general, només la meitat de les preguntes que formen

cadascuna de les proves pot ser considerada com a “bona” o “molt bona” per la seva capacitat de discriminació.

Si el qüestionari vol ser utilitzat per avaluar els coneixements de l'alumnat, la presència d'un elevat nombre d'ítems amb un baix poder de discriminació pot resultar contraproduent (Sim & Rasiah, 2006). D'acord amb els resultats obtinguts, el 90% dels qüestionaris analitzats han de ser revisats parcialment i un 25% d'aquests s'haurien de modificar en la seva totalitat. S'ha realitzat una simulació per analitzar els possibles canvis sobre les qualificacions després de retirar del qüestionari les 5 preguntes amb un coeficient de discriminació més baix (figura 1). En general, s'observa que els alumnes menys eficients (nota més baixa d'expedient) obtindrien puntuacions més baixes i, en canvi, permetria incrementar la nota dels més eficients.

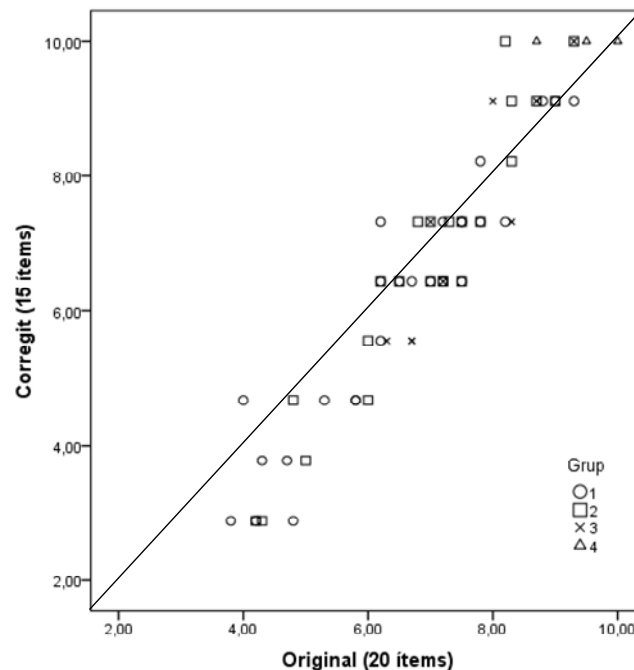


Figura 1. Comparació de les qualificacions originals (amb 20 ítems) i les obtingudes si es retiren les 5 preguntes amb menys capacitat discriminatòria (corregit, amb 15 ítems) de la primera prova parcial de l'assignatura Ecologia. Els alumnes s'han classificat per grups en funció de la nota mitjana del seu expedient acadèmic, si aquesta és <1,25 (1), <1,5 (2), <2 (3) o >2 (4).

Retroaccions

L'anàlisi dels qüestionaris 1 a 4 (mateixes preguntes però amb retroacció o sense retroacció) mostra que les qualificacions són més altes en el cas de les proves de tipus adaptatiu que incorporen les retroaccions (ANOVA, $F_{1,107}:5,119$; $p<0,05$), tot i controlar l'efecte de la nota d'expedient acadèmic. En canvi, no hi ha diferències significatives entre les notes de l'examen parcial, dels alumnes que han realitzat un tipus o un altre de qüestionari (ANOVA, $F_{1,107}:1,169$; $p=0,282$) el que indicaria que la realització de la prova amb retroaccions o sense no té un efecte directe sobre el rendiment acadèmic de la prova escrita presencial. No obstant, aquest fet no implica que la realització del qüestionari no pugui ser considerada una activitat d'aprenentatge, però probablement influeixen més altres aspectes com l'eficiència de l'estudiant o el temps d'estudi entre la realització del qüestionari i l'examen parcial. Les valoracions de l'alumnat sobre la realització dels qüestionaris són positives, arran dels resultats de les enquestes. El 87% considera que la prova els ha servit per entendre millor els continguts i el 68% ha estat d'utilitat per resoldre dubtes. Les crítiques més freqüents són per la limitació del temps i el nombre elevat de preguntes.

Conclusions

- Les proves d'opció múltiple són un recurs útil per l'avaluació de l'aprenentatge de la docència en ecologia, especialment per l'elevat nombre d'alumnes i la concentració en el temps d'activitats avaluable i exàmens parcials que formen part de l'avaluació contínua.
- És molt important realitzar una metaavaluació dels qüestionaris per tal de detectar errors, mancances i mesurar la qualitat del banc de preguntes. El disseny del qüestionari és important per tal d'assolir els objectius pels quals ha estat ideat, ja sigui com a complement que faciliti l'aprenentatge dels continguts o bé com a prova d'avaluació.
- Tots els qüestionaris analitzats en aquest estudi presenten un nivell de dificultat inferior al desitjat i en alguns casos han resultat especialment senzills malgrat que la seva funció era la d'avaluar els coneixements de l'alumnat. No es pot tenir, per tant, la certesa de que les qualificacions obtingudes pels estudiants reflecteixin el seu grau de coneixement dels continguts.
- La majoria dels qüestionaris analitzats presenten un valor elevat d'ítems (al voltant del 30%) que han de ser revisats o fins i tot suprimits del banc de preguntes, d'acord amb el corresponent valor del coeficient de discriminació. La modificació d'aquestes preguntes, té una conseqüència directa sobre la nota obtinguda, afavorint els resultats dels alumnes més eficients.
- L'ús de les retroaccions en els qüestionaris no sembla tenir una repercussió directa sobre una millora de la qualificació de l'estudiant a la prova escrita presencial tot i que no es pot concloure que no sigui una eina facilitadora de l'aprenentatge. Els alumnes que han realitzat els qüestionaris consideren majoritàriament que han servit per entendre millor els continguts i per resoldre dubtes.
- Malgrat que la plataforma Moodle presenta una eina d'anàlisi dels qüestionaris, caldria incorporar nous indicadors que permetessin valorar, per exemple, la validesa (com ara el coeficient de Spearman-Brown) o l'homogeneïtat dels ítems.

Bibliografia

- Arbós, A. (2010). Primeras Jornadas Internacionales sobre EEES: Evaluación. Universitat Internacional de Catalunya.. Barcelona.
- Backhoff, E., Larrazolo, N. i Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa*, 2 (1). Recuperat el dia 12 de febrer de 2010, a <http://redie.uabc.mx/vol2no1/contenido-backhoff.html>
- Crocker, L. (1986) Introduction to classical and modern test theory. Ed. Fort Worth: Harcourt Brace Jovanovich College, Londres.
- Gil, I., Bilbao, J. i F. Jiménez (2009). Análisis de cuestionarios para evaluar conocimientos matemáticos a través de la plataforma Moodle. MoodleMoot Euskadi. Universidad de Deusto. Recuperat el 9 de febrer de 2010, a <http://www.slideshare.net/euskalmoot/analisis-cuestionarios-inmaculada-gil>
- Sim, S. i Rasiyah, R.I. (2006). Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a Para-clinical multidisciplinary paper. *Annals Academy of Medicine*, 35 (2): 67-71.
- Wood, D. (1965). Elaboración de tests : desarrollo e interpretación de los tests de aprovechamiento. Ed. Trillas, Mèxic.
- Zimmaro, D. (2003). Analyzing Multiple-Choice Item responses. Research, Evaluation and Assessment Resources. Division of Instructional Innovation and Assessment. The University of Texas at Austin. Recuperat el 3 d'abril de 2010, a: <http://www.utexas.edu/academic/mec/scan/analysis.html>