Tesi doctoral per compendi de publicacions

PRÀCTICA CONTINUADA I FEEDBACK AUTOMÀTIC EN L'APRENENTATGE DE MATEMÀTIQUES EN LÍNIA: UN ESTUDI DES DE LA PERSPECTIVA DE LES ANALÍTIQUES D'APRENENTATGE

Josep Figueroa Cañas

Directora: Dra. Teresa Sancho Vinuesa

Programa de doctorat d' Educació i TIC (e-Learning)

Universitat Oberta de Catalunya

2021

A la memòria de la meva mare Felisa,

al meu fill Robert i a la meva filla Clàudia,

i sobretot, a l'Emma, la dona amb qui des de fa més de 30 anys anem fent camí.

AGRAÏMENTS

Són molts els motius i moltes les persones a les quals estic agraït, que d'alguna o altra manera m'han conduït a l'assoliment d'aquesta important fita per a mi, finalitzar la tesi doctoral.

Vull donar les gràcies, en primer lloc, a la meva família i amics. A la iaia Esperanza perquè em va cuidar de petit i perquè em va ensenyar a llegir i escriure. A la meva mare Felisa, pel seu esperit de superació, la seva actitud de no rendir-se mai, per pensar sempre, de vegades massa, en els altres abans que en ella mateixa, i sobretot per haver-me estimat. Al meu pare Fidel, per haver treballar tan dur per tirar la família endavant i fer-se càrrec del finançament dels meus estudis. A la meva germana Elena, per donar-me tranquil·litat i serenor quan les coses no van del tot bé. A la tia Elena, per estimar-me com si fos el seu fill. A la mare i el pare de l'Emma, l'Aurèlia i el Bartomeu, per haver-me fet sentir com un membre de la seva família i per haver-nos ajudat (a l'Emma i a mi) en tot allò que estava al seu abast. Al meu fill Robert i la meva filla Clàudia, perquè la seva existència em fa sentir enormement feliç. A l'Emma, per la seva paciència tendint a l'infinit, i per damunt de tot per estimar-me i haver-me deixat estimar-la. Als amics i amigues de la colla (el David, el Joan, el Jordi, la Julie, la Maria Àngels, el Marcel, la Pepa i la Sílvia) per haver compartit vivències especials que formen part de la nostra història.

En segon lloc, també vull mostrar el meu agraïment a persones amb què he tingut relació dins l'àmbit acadèmic. A molts i moltes mestres de l'EGB del Col·legi Bages, professors i professores de BUP i COU de l'Institut Lluís de Peguera, professors i professores de la Universitat Autònoma de Barcelona i de la Universitat Oberta de Catalunya, pels seus ensenyaments i per haver-me fet créixer com a persona. Vull donar les gràcies molt especialment a la Teresa (Dra Teresa Sancho) pel seu expert guiatge en el desenvolupament de la tesi, així com pel seu suport emocional, que en moments complicats m'ha empès a continuar avançant. També vull expressar el meu profund agraïment a Mr. Paul Garbutt per la seva col·laboració en la revisió de l'anglès dels articles.

ÍNDEX

CAPÍTOL 1

INTRODUCCIÓ

Ensenyar i aprendre matemàtiques esdevé un gran desafiament en entorns d'aprenentatge en línia i asíncrons (Trenholm et al., 2015; Engelbrecht i Harding, 2005). En sentit ampli, en un entorn d'aquest tipus s'utilitzen mitjans i dispositius electrònics per a accedir a les activitats i continguts formatius i la comunicació entre professorat i estudiant no ocorre en temps real, sinó en temps diferit (Holden i Westfall, 2007). En educació superior, l'ensenyament formal de les matemàtiques s'organitza en assignatures que agregades entre sí i amb les corresponents a d'altres àmbits de coneixement conformen les titulacions de graus i màsters. Per exemple, el grau d'enginyeria informàtica de la Universitat Oberta de Catalunya (UOC) conté les assignatures d'anàlisi matemàtica i estadística que conviuen amb les pròpies de l'àrea de computació. En general, els estudiants d'aquest grau senten poca motivació per les assignatures de matemàtiques atès que, d'una banda, no les consideren com a fonamentals en la seva formació com a futurs/es enginyers/eres informàtics, i d'altra, el seu contingut els resulta força abstracte i la notació matemàtica, molt rigorosa. Aquesta baixa motivació afegeix dificultat a l'hora d'aprendre matemàtiques, en un entorn en línia i asíncron, en què el treball personal esdevé determinant.

El relat d'aquest capítol d'introducció segueix el fil de la frase d'inici: "ensenyar i aprendre matemàtiques esdevé un gran desafiament en entorns d'aprenentatge en línia i asíncrons". En primer lloc, es presenta el model d'interaccions que permet emmarcar teòricament el procés d'ensenyament/aprenentatge (E/A) en un entorn d'aprenentatge en línia eminentment asíncron, així com identificar-ne els aspectes més rellevants. A continuació, s'aborda un element clau d'aquest procés, l'avaluació, incloent-hi la retroacció o *feedback*, necessària en el marc d'una avaluació formativa i eix vertebrador de les assignatures analitzades en aquesta tesi doctoral. Tot seguit, i amb el propòsit de situar aquest treball en un context concret, es detallen algunes de les especificitats de les assignatures de matemàtiques en línia i es presenta l'estratègia docent que ha inspirat el disseny de les assignatures de matemàtiques de les titulacions d'enginyeria a la UOC. Posteriorment, s'explica l'origen d'aquesta estratègia, consistent en la realització continuada d'activitats amb provisió de feedback automàtic, així com l'anàlisi de l'experiència en diferents assignatures en cadascuna de les quals s'hi ha fet una implementació particular. Els resultats obtinguts suggereixen continuar treballant en aquesta línia on docència, innovació i recerca es retroalimenten. Atès que aquesta tesi continua l'estudi d'aquesta estratègia docent a assignatures de matemàtiques de la UOC, ara, mitjançant l'aproximació de les analítiques d'aprenentatge ("learning analytics"), la introducció presenta el cicle de analítiques d'aprenentatge, així com la modificació que hi hem afegit. Cloem aquest capítol introductori amb l'exposició dels objectius i preguntes de recerca de la tesi.

1.1 Interaccions

El procés d'ensenyament/aprenentatge en línia se sustenta sobre una xarxa complexa d'*interaccions* (Anderson, 2003). En aquest context d'ensenyament/aprenentatge i en una primera aproximació, la *interacció* pren el significat d'intercanvi d'informació bidireccional entre dues o més persones, per exemple estudiants i professorat (Berge, 2002). Per donar cabuda a altres elements d'aquest procés, com per exemple els continguts objecte d'aprenentatge, el terme interacció inclou també la transmissió unidireccional o bidireccional d'informació entre persones i recursos relacionats amb continguts del tema que cal aprendre (Berge, 2002). Igualment són considerades interaccions els fluxos d'informació entre persones i recursos no relacionats amb continguts (Sabry i Baldwin, 2003), així com, entre persones i la interfície del sistema informàtic que permet tota la resta d'interaccions precedents (Hillman et al., 1994).

En aquesta secció presentem els diversos tipus d'interaccions que intervenen en el procés d'E/A en línia. La proposta que constitueix el primer referent teòric en relació a la categorització de les interaccions (Wang et al., 2014) correspon al model de tres interaccions de Moore (1989): **aprenent-contingut**, **aprenent-instructor** i **aprenent-altres aprenents**. Moore (1989) pren l'aprenent com a centre de les interaccions, entenent-lo com a principal agent responsable del seu aprenentatge. Les **interaccions entre aprenent i continguts** tenen lloc quan l'aprenent accedeix als continguts d'allò que vol aprendre. En són exemples la lectura que fa l'estudiant d'informació textual, la visualització de vídeos instruccionals, l'ús de software estadístic o la realització de qüestionaris d'avaluació (Abrami et al., 2011). Dins d'aquestes interaccions cal incloure-hi el diàleg intern que ocorre dins l'estudiant, essencial per aprendre, quan aquest accedeix a la informació, la interpreta i en crea coneixement (Moore, 1989). Quant a les **interaccions entre l'aprenent i l'instructor**, cal interpretar-les com el diàleg que s'estableix entre aquests dos agents, així com la provisió de feedback per part de l'instructor. Aquests diàlegs poden esdevenir mitjançant comunicacions de correu electrònic, converses a través de xats, o bé amb intercanvi de posts de lectura i escriptura en fòrums de discussió (Abrami et al., 2011). Per últim, les **interaccions entre l'aprenent i altres aprenents** corresponen, com en el cas aprenent-instructor, als diàlegs entre els dos agents, i per tant a l'intercanvi d'informació, si bé en aquest cas, entre iguals. Els exemples de comunicacions mitjançant correu electrònic, xats o fòrums hi són igualment aplicables (Abrami et al., 2011).

La base de la proposta del model d'interaccions de Moore (1989) s'ha anat mantenint al llarg d'aquests darrers vint anys, tot i que s'ha afinat amb la incorporació d'alguns nous tipus d'interaccions. Veiem-ne tres. El primer tipus correspon a les **interaccions estudiant-interfície** (Hillman et al., 1994). Aquestes interaccions engloben les transferències d'informació entre l'estudiant i el sistema informàtic o entorn virtual d'aprenentatge (Agudo-Peregrina et al., 2014; Joksimović et al., 2015). Hillman et al. (1994) considera que, en entorns d'aprenentatge en línia, les interaccions estudiants-interfície mereixen una diferenciació respecte de la resta d'interaccions, en tant que, en aquest tipus d'entorns d'aprenentatge la resta d'interaccions estan condicionades per les tecnologies de la informació i la comunicació. Aquestes tecnologies poden tant facilitar com dificultar altres interaccions, per exemple, l'accés dels estudiants als continguts dipositats a l'entorn virtual d'aprenentatge es veu influenciat per com de fàcil d'utilitzar ("user-friendly") resulta la interfície. Alguns autors (Agudo-Peregrina et al., 2014; Joksimović et al., 2015) empren el terme interaccions **estudiant-sistema** en lloc d'estudiant-interfície en referir-se als accessos dels estudiants als entorns virtuals d'aprenentatge. Com a segon tipus d'interaccions afegides al model de Moore (1989), Sabry i Baldwin (2003) va proposar les **interaccions estudiant-informació** per incloure les interaccions que es corresponen a fluxos d'informació, cap als estudiants, de temàtica no relacionada directament amb continguts sobre el tema a aprendre. Per exemple, accedir a la informació continguda en el pla docent d'una assignatura correspondria a una interacció estudiant-informació. De fet, amb una interpretació àmplia del terme "contingut", es podria considerar que aquesta interacció ja estava inclosa dins del model de Moore (1989) en la interacció estudiant-contingut. Tanmateix, l'aportació de Sabry i Baldwin (2003) permet separar nítidament la interacció estudiant-contingut (que correspon estrictament a contingut de la matèria objecte d'estudi) de la interacció estudiant-informació (que correspon estrictament a informació no relacionada amb contingut). Com a tercer tipus d'interacció, Anderson (2003) afegeix les **interaccions professorat-contingut** des de la perspectiva del professorat, principal agent responsable de l'ensenyament. Cal considerar com a exemples d'aquesta interacció, la tria de recursos didàctics per a l'assignatura o bé pujar el calendari de proves d'avaluació a l'entorn virtual d'aprenentatge.

L'esquema de la Figura 1 mostra de manera gràfica les interaccions esmentades anteriorment, indicades amb traç rectilini blavós i amb el nom de l'investigador principal a qui se li atribueix la seva aportació al model. També hem afegit, amb traç curvilini verdós, les **interaccions professorat-interfície** i **professorat-informació no relacionada amb contingut**, de manera anàloga a la proposta de la interacció professorat-contingut d'Anderson (2003).

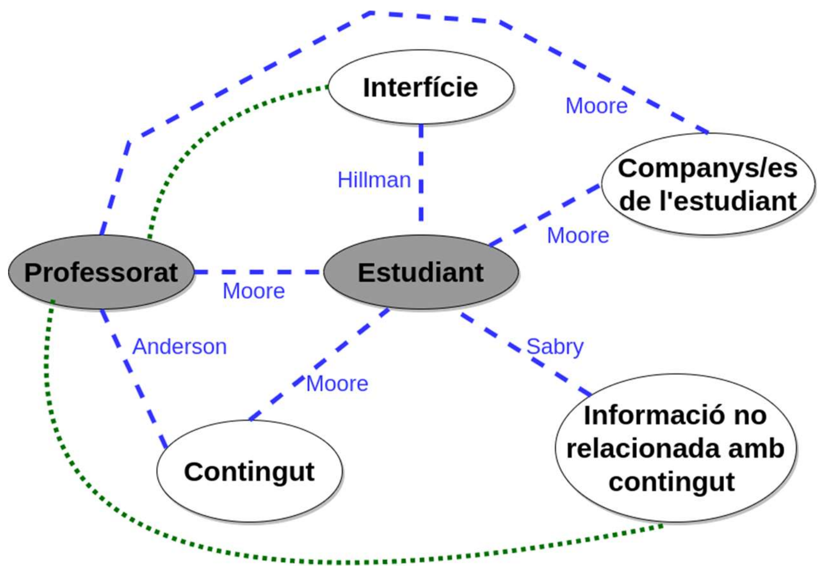


Figura 1. Model d'interaccions del procés d'ensenyament/aprenentatge en entorns en línia.

A tall d'exemple, quan el professorat accedeix a l'entorn virtual d'aprenentatge per penjar una determinada activitat d'aprenentatge es produeix una interacció entre professorat i interfície. Quan el professorat rep o transmet informació no relacionada directament amb els continguts, escrivint un post de presentació personal al fòrum o notificant una ampliació de termini de lliurament d'una activitat, de manera que també es produeixen interaccions entre professorat i informació no relacionada amb contingut.

A l'hora d'aplicar aquest model teòric d'interaccions (Figura 1) per a l'anàlisi del procés d'ensenyament i aprenentatge sorgeixen certes dificultats. El problema principal rau en el fet que les interaccions no sempre ocorren de manera independent (Kanuka, 2011; Swan, 2001) i, per tant, en un mateix esdeveniment poden convergir diversos tipus d'interaccions que s'acoblen entre sí. Per exemple, quan dos estudiants intercanvien posts a través d'un fòrum de discussió podem considerar que es produeixen interaccions entre estudiant i companys perquè es produeix una comunicació bidireccional entre estudiant i companys. Si aquests intercanvis inclouen comentaris sobre continguts, també podem pensar que es produeixen interaccions entre estudiant i contingut (Swan, 2001) atès que l'estudiant rep informació sobre la matèria objecte d'aprenentatge. Així, el fet que dos estudiants intercanviïn posts en un fòrum de discussió pot ser analitzat sota la perspectiva d'interaccions entre estudiant i companys exclusivament, d'interaccions entre estudiant i contingut únicament, o bé d'ambdós tipus d'interaccions conjuntament, segons on es pretengui posar el focus de l'anàlisi.

En definitiva, aquest model permet interpretar una gran varietat d'esdeveniments del procés d'E/A mitjançant les interaccions. En aquest sentit, Joksimović et al. (2015) considera que accions clau del procés com escriure un missatge de correu electrònic, afegir un post en un fòrum, fer un "login" en l'entorn virtual d'aprenentatge, matricular-se d'una assignatura, visualitzar un recurs didàctic de l'assignatura o realitzar un qüestionari d'avaluació són interaccions. D'aquesta manera, el model permet interpretar el procés d'E/A a través d'interaccions diferents i abordar la seva anàlisi de manera coherent i consistent.

1.2 Avaluació i feedback

En general, el propòsit d'un procés d'E/A en el marc de l'educació formal és l'assoliment, per part de l'estudiant, d'uns objectius d'aprenentatge determinats pel professorat. Aquest procés inclou un aspecte, l'anomenada avaluació, que d'una banda té la finalitat d'atribuir a l'estudiant el grau d'assoliment dels objectius d'aprenentatge (Archer, 2017), i d'una altra, d'ajudar-lo a incrementar el grau assolit (Archer, 2017).

L'avaluació juga un paper clau en el procés d'E/A atès que: només a través de l'avaluació podem esbrinar si una seqüència d'activitats d'aprenentatge han donat com a resultat els objectius d'aprenentatge planificats (Wiliam, 2011, p.3). De fet, les activitats d'aprenentatge realitzades i en general el comportament dels estudiants pel que fa a l'aprenentatge ("learning behaviour") corresponen a decisions influenciades per la pròpia avaluació (Biggs i Tang, 2011; Sadler, 2010). En aquest sentit, Keppell et al. (2006) sosté que els estudiants adapten les seves estratègies d'aprenentatge, en particular la selecció dels continguts a treballar, d'acord amb l'avaluació. Ho argumenta amb el fet que l'avaluació transmet missatges implícits i explícits informant sobre el que realment és important. També en la mateixa línia, Baleni (2015) apunta que els estudiants decideixen com gestionar el temps, per exemple quanta estona dediquen a l'estudi, segons la manera com són avaluats. Així doncs, cal dissenyar el sistema d'avaluació per tal que sigui un suport i no un impediment per a l'assoliment dels objectius d'aprenentatge. Podem distingir dos tipus d'avaluació, segons l'objectiu i el resultat. Quan l'objectiu rau en determinar la situació de l'estudiant en referència als objectius d'aprenentatge, parlem d'avaluació sumatòria ("summative assessment") (Bloom, 1971, citat per Newton, 2007). El resultat d'aquesta avaluació sumatòria es visualitza mitjançant una qualificació que valora el grau d'assoliment dels objectius i/o els aprenentatges adquirits (Gikandi et al., 2011). Si, en canvi, l'objectiu és oferir el suport i guia a l'estudiant que li permeti avançar en el seu aprenentatge, parlem d'avaluació formativa ("formative assessment") (Bloom, 1971, citat per Newton, 2007). El resultat de l'avaluació formativa és la informació que s'ofereix a l'estudiant, anomenat feedback, que li ha de permetre avançar. Però aquest progrés només serà possible si aquesta informació és proporcionada durant el procés d'aprenentatge i no al final (Spector et al., 2016; Tempelaar et al., 2018). Més endavant aprofundirem sobre el concepte de feedback. Abans de continuar és necessari fer notar que la frontera real entre ambdós tipus d'avaluacions resulta imprecisa. Tal com apunta Taras (2005), per tal d'ajudar l'estudiant a avançar (avaluació formativa) cal efectuar prèviament un judici del seu estat d'assoliment d'objectius (avaluació sumatòria), de manera que l'avaluació formativa inclou implícitament l'avaluació sumatòria. Per altra banda, si l'avaluació sumatòria es realitza durant el procés d'E/A, el seu resultat en forma de qualificació ja ofereix informació a l'estudiant (avaluació formativa), en tant que el fa coneixedor de la situació en què es troba i que pot ser-li d'ajuda per millorar.

Atès que l'avaluació constitueix un element essencial d'aquesta tesi, ens interessa fer-la explícita dins del model d'interaccions del procés d'E/A. Separem l'element *contingut* del model de la Figura 1 en dos elements: *recursos didàctics* i *sistema d'avaluació*, subelements continguts en la proposta de Moore (1989). Al mateix temps, substituïm el nom de l'element *informació no relacionada amb contingut* pel de *recursos no didàctics*. El resultat és el model adaptat d'interaccions representat en la Figura 2

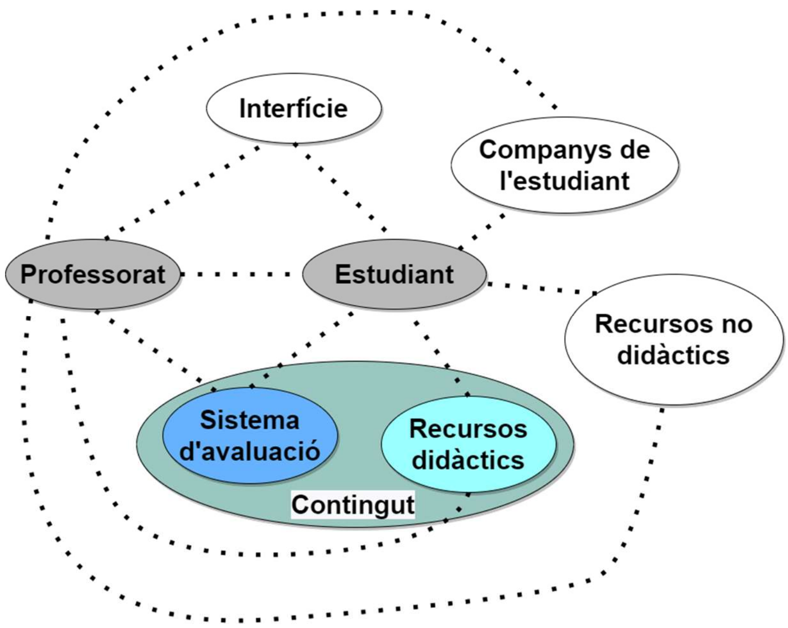


Figura 2. Model adaptat d'interaccions del procés d'ensenyament/aprenentatge en entorns en línia, en què hem dividit el component Contingut en Sistema d'avaluació i Recursos didàctics.

El que resta de secció el dediquem a desglossar els diversos tipus de feedback. Tal com hem esmentat anteriorment, l'avaluació formativa dona com a resultat un feedback, que en el context d'ensenyament/aprenentatge, s'entén com aquella informació proporcionada per un agent com a resposta a una activitat (Hattie i Timperley, 2007). Aquests autors presenten una classificació del feedback en quatre nivells: (i) **nivell d'activitat**, en què s'informa l'estudiant com ha realitzat l'activitat, (ii) **nivell de procés**, en què s'informa l'estudiant de com s'ha de realitzar i entendre l'activitat, (iii) **nivell d'autoregulació**, en què s'informa l'estudiant sobre com ajustar el seu aprenentatge a través, per exemple, d'una autoavaluació de l'activitat realitzada, (iv) **nivell del jo** ("self level"), en què no s'informa l'estudiant sobre cap activitat en concret, sinó que se li proporcionen missatges personals de suport i guia, per exemple "Estàs treballant molt bé". Shute (2008) opta per una classificació del feedback segons dues variables, el tipus d'informació subministrada i l'instant en què és emesa. Quant al **tipus d'informació**, Shute (2008) n'estableix tres tipus. En el primer tipus, s'informa a l'estudiant únicament del resultat (**KR**, "knowledge of result") de l'activitat en sentit binari, és a dir, si és correcta o no és correcta. En el segon tipus, es proporciona la resposta correcta de l'activitat (**KCR**, "knowledge of correct response"). Els dos primers blocs, que se situen dins del nivell d'activitat de Hattie i Timperley (2007), es consideren feedback correctiu. Finalment en el tercer tipus, es proporciona una explicació més extensa (**EF**, "elaborated feedback"), que pot consistir en informar sobre quin era el procediment per arribar a la resposta correcta, i/o informar sobre quines incorreccions conté el procediment presentat en la resolució de la pregunta, i/o oferir suggeriments de com millorar la resposta. Aquest EF supera el nivell de d'activitat de Hattie i Timperley (2007) i s'inclou dins dels nivells de procés i/o d'autoregulació. Independentment del tipus, Sadler (2010) considera que la informació subministrada en el feedback ha d'ajudar l'estudiant a entendre millor el resultat obtingut en la realització de la tasca, així com a reduir la distància entre la resolució actual i l'òptima de la mateixa. Es fa difícil amb la perspectiva de Sadler (2010) acceptar que un feedback només tipus KR, és a dir només un missatge correcte/incorrecte, ajudi a reduir la distància envers una resolució òptima de l'activitat. Pel que fa a la classificació del feedback segons **el moment de lliurament de la informació**, Shute (2008) distingeix feedback immediat de feedback retardat. Es considera immediat aquell en què la informació es subministrada just després de la resposta d'una pregunta/problema, o d'un qüestionari/prova (Shute, 2008). Es parla de feedback retardat en la resta de casos que no sigui immediat, que inclou un extens rang que va des d'uns segons fins a dies (Van der Kleij et al., 2015). Cal afegir que en entorns específicament en línia, es factible tecnològicament la correcció automàtica de les activitats d'avaluació, fet que facilita a aquests entorns poder proporcionar feedback en temps reduïts o de manera immediata (Van der Kleij et al., 2015).

1.3 Matemàtiques en línia

En les dues seccions anteriors hem desenvolupat aspectes relacionats amb ensenyar i aprendre en entorns d'aprenentatge en línia, sense concretar l'àmbit de coneixement dels aprenentatges. En aquesta secció posem el focus sobre l'àmbit en què s'ha centrat el nostre estudi, les matemàtiques.

El treball de Smith et al. (2008) recull el posicionament, quant als reptes que suposa la docència de matemàtiques en línia, de professorat d'aquesta matèria amb una àmplia experiència de docència en línia. Segons aquests autors, la totalitat del professorat consultat en el seu treball destaca com a principals reptes a l'hora d'ensenyar matemàtiques, d'una banda, la naturalesa dels conceptes que tracten, eminentment abstractes, i d'una altra, la seqüenciació dels continguts, és a dir, els contingut es construeixen sobre continguts previs. En entorns d'aprenentatge en línia i asíncrons, la comunicació entre professorat i estudiant s'efectua generalment a través de correu electrònic mitjançant informació textual. Això suposa per als estudiants un greu inconvenient en termes del temps que cal destinar-hi (Lowenthal et al., 2017). En el cas d'assignatures de matemàtiques, cal afegir-hi la dificultat per comunicar raonaments específics de l'especialitat (Smith et al., 2008), atesa la simbologia matemàtica, així com l'ús d'un llenguatge molt normativitzat (Smith et al., 2008; Karal et al., 2013). Reprenent el fil de l'existència de conceptes abstractes, aquesta característica fa que els i les estudiants hagin d'interactuar de manera continuada amb els continguts per tal d'assimilar-los (Smith et al., 2008). En aquest tipus d'assignatures, la interacció de l'estudiant amb els continguts ha d'estar enfocada a la realització d'activitats d'aprenentatge, és a dir a la pràctica mitjançant activitats (Wan Niu Voon et al., 2014). En el treball de Smith et al. (2008), el professorat expressa la dificultat que suposa aconseguir que els estudiants de matemàtiques en línia realitzin les pràctiques encomanades. Cal tenir present que en entorns en línia i asíncrons, l'estudiant pren les seves pròpies decisions sobre quan realitzar les pràctiques encomanades i sobre quin ritme de treball, en general, vol seguir (Engelbrecht i Harding, 2005). En canvi, en entorns presencials, l'horari de classes i una organització específica per a cada assignatura durant tot el curs, marquen rutines i ritmes de treball.

En el context de les matemàtiques (en línia) del grau d'enginyeria informàtica de la UOC, el professorat del curs 2010-11, amb el lideratge de la Dra. Teresa Sancho, defensava que l'estratègia docent més efectiva era la realització, de forma regular, d'exercicis i problemes on l'estudiant demostra el seu domini dels continguts i la seva capacitat d'aplicar-los en situacions pràctiques. Per altra banda, des dels seus inicis, el model pedagògic de la UOC inclou, com a un pilar fonamental, l'avaluació continuada (Duart i Sangrà, 2000). D'acord amb aquest model, a totes les assignatures de la universitat es proposa una seqüència d'activitats repartides al llarg de l'assignatura, amb l'objectiu principal d'ajudar l'estudiant en el procés d'aprenentatge, tot i que secundàriament pot contribuir també a la superació de l'assignatura (Duart i Sangrà, 2000). En aquest marc, el professorat de matemàtiques del curs 2010-11 del grau d'enginyeria informàtica de la UOC va proposar com a estratègia docent la realització regular d'activitats d'avaluació continuada per part de l'estudiant. Més tard, aquesta proposta va estar justificada pels resultats de diversos autors: d'una banda, en relació a l'efectivitat de les proves d'entrenament ("practice tests") (Dunlosky et al., 2013) i d'una altra, sobre el fet que la realització de proves amb regularitat promou l'estudi entre els estudiants (Roediger et al., 2011). Dur a terme aquesta proposta tenia diversos problemes, tant per estudiants com per professorat. L'estudiant havia de fer front a certes dificultats, com ara haver d'escriure les resolucions de les activitats amb processadors de textos que no facilitaven l'ús de simbologia matemàtica i, sobretot, si el professorat havia de corregir l'exercici, no podia rebre el retorn fins al cap, com a mínim, d'una setmana d'haver lliurat les activitats. El professorat havia de corregir entre 60 i 70 exercicis per setmana donant un retorn personalitzat. Davant d'aquestes consideracions, van decidir buscar un sistema que permetés una activitat realment continuada, poc feixuga d'escriure, amb resposta immediata i pistes per a aprofundir en allò que s'havia fet. Per tant, calia disposar d'una eina que permetés aquesta immediatesa, impossible d'oferir a través de l'acció d'un professor en una aula estàndard amb 70 estudiants, i al mateix temps, calia que cada estudiant tingués una activitat el més personalitzada possible i diferent de la del company. Tot això es va concretar en el projecte d'innovació docent *Avaluació automàtica de continguts matemàtics en una aula Moodle*, que es presenta en la següent secció.

1.4 Un canvi metodològic en la docència de matemàtiques a les enginyeries de la UOC

En el marc del projecte *Avaluació automàtica de continguts matemàtics en una aula Moodle* , es va trobar la resposta tecnològica al problema d'haver de proporcionar feedback automàtic i immediat a l'estudiant: els qüestionaris Wiris Quizzes[1]. Els Wiris Quizzes són qüestionaris que milloren les capacitats dels qüestionaris de la plataforma Moodle incorporant-los un editor i intèrpret de símbols matemàtics a través del programa de càlcul simbòlic Wiris Cas (Calm et al., 2013). A més, són auto-avaluables i per tant de resposta automàtica, fet que els permet generar un feedback de manera immediata (Calm et al., 2013). Un altre aspecte a destacar és que els enunciats contenen paràmetres. D'aquesta manera, una mateixa pregunta genèrica, per exemple, on s'ha de calcular la recta tangent, pot esdevenir un conjunt pràcticament infinit d'enunciats diferents a partir de les concrecions numèriques de la pregunta (Calm et al., 2013). Això fa possible que diferents estudiants puguin respondre diferents preguntes, i rebin feedback individualitzat. En col·laboració amb el personal de Wiris Quizzes, el professorat que va participar en el projecte va elaborar un conjunt de qüestionaris per a les assignatures d'iniciació a les matemàtiques per a l'enginyeria i anàlisi matemàtica, del grau d'enginyeria informàtica, el 2010.

Mostrem a continuació un exemple de pregunta, de resposta d'un estudiant i de feedback proporcionat en un qüestionari de l'assignatura d'anàlisi matemàtica.

La Figura 3 presenta la formulació d'una pregunta per calcular la recta tangent a una funció en un punt. Es pot observar que també hi consta la puntuació obtinguda en la resposta (Figura 4), així com el feedback generat, informant que la resposta és correcta. Es tracta, doncs d'un feedback tipus *resultat* o KR (Shute, 2008). Podem veure que la resposta de l'estudiant s'ajusta al format exigit per la pregunta (Figura 4) i que està introduïda mitjançant l'editor del programa de càlcul simbòlic Wiris Cas. Aquesta pregunta permet facilitar un feedback addicional (Figura 5), tipus *resposta correcta* o KCR (Shute, 2008) informant que la resposta correcta és y=-3x-1, així com feedback tipus *explicació extensa* o EF (Shute, 2008) amb explicacions de com calcular la recta tangent i indicacions de com arribar al resultat final, feedback que correspon al nivell de procés de Hattie i Timperley (2007).

[1] WIRIS: http://www.wiris.com/en/quizzes

Pregunta 5 Correcta Puntua 1,00 sobre 1,00	Determina la recta tangente a la función $f(x) = -x^3 - 3x^2 - 3x - 1$ en $x = 0$ Escribe la respuesta en la forma $mx + n$, donde m es la pendiente de la recta y n su ordenada al origen. Es decir, de manera que la ecuación explícita de la recta sea $y = m \cdot x + n$

Figura 3. Exemple d'una pregunta de resposta curta en un qüestionari d'entrenament, juntament amb la puntuació.

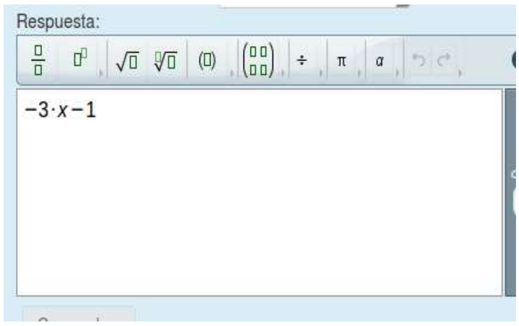


Figura 4. Exemple de resposta mitjançant l'ús del programa Wiris.

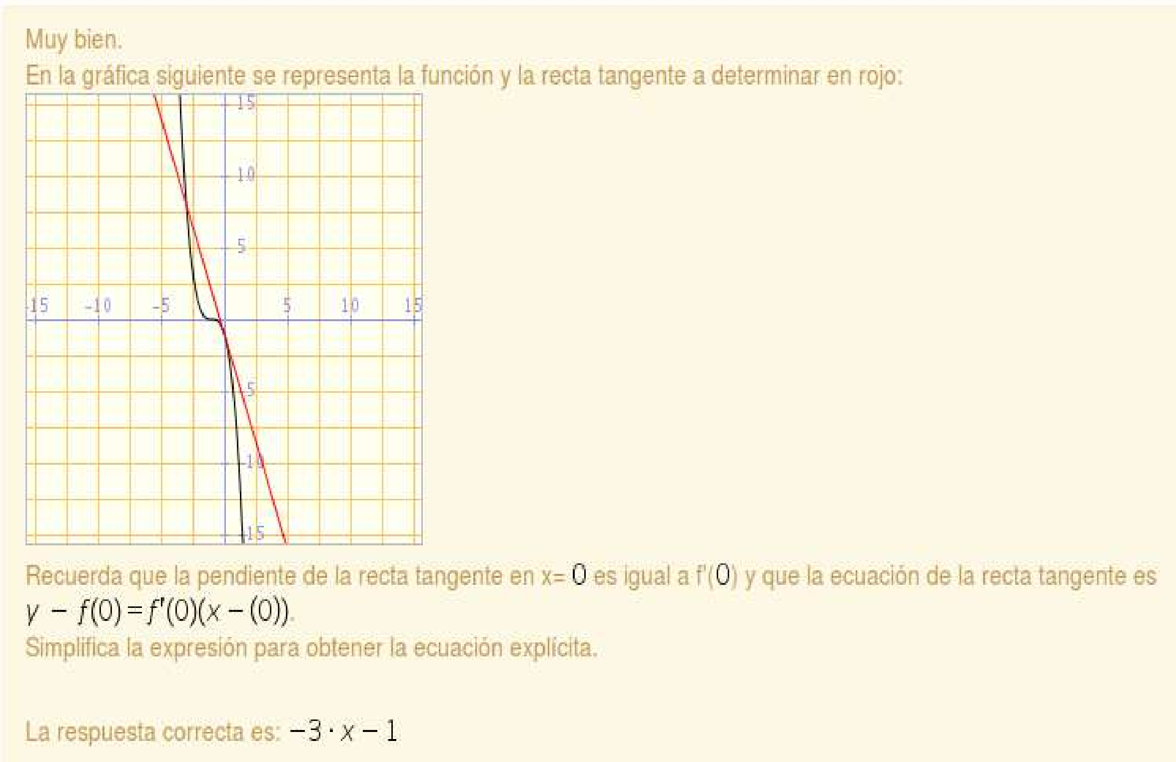


Figura 5. Exemple de feedback tipus resposta correcta (KCR) i explicació extensa (EF).

El projecte *Avaluació automàtica de continguts matemàtics en una aula Moodle* es va dur a la pràctica el primer semestre del curs 2010-11 a l'assignatura d'iniciació a les matemàtiques per a l'enginyeria (Sancho-Vinuesa i Escudero-Viladoms, 2012) i el segon semestre del curs 2010-11 a l'assignatura d'anàlisi matemàtica (Calm et al., 2013), ambdues com a experiència pilot. El projecte s'ha anat consolidant en les dues assignatures anteriors, mentre s'ha anat implementant a la resta d'assignatures de matemàtiques de diferents graus (enginyeria informàtica, enginyeria en aplicacions de les telecomunicacions, ciència de dades aplicada i multimèdia).

El canvi en la metodologia docent derivat del projecte no modificava ni els objectius d'aprenentatge ni els recursos didàctics de l'assignatura. Consistia en la realització continuada de qüestionaris (Wiris Quizzes), fent que les activitats d'avaluació prenguessin un paper destacat entre el conjunt d'activitats d'aprenentatge. Es creen dos tipus de qüestionaris (Wiris Quizzes) amb feedback automàtic i immediat. El primer tipus correspon als anomenats **qüestionaris d'entrenament**, que tenen com a finalitat practicar i preparar-se per al segon tipus de qüestionaris, **qüestionaris de mesura**. La finalitat d'aquests últims, que inclouen una pregunta oberta que obliga els estudiants a escriure és, a més de practicar, determinar el grau de coneixements adquirit en una part del contingut de l'assignatura. Els qüestionaris de mesura constitueixen la font principal per a determinar la qualificació de l'avaluació continuada i, per això, són emprats també en l'obtenció de la nota final de les assignatures. Podem dir, doncs, que els qüestionaris de mesura compleixen una doble funció. Per una banda, en tant que objectes que proporcionen feedback que ajuda l'estudiant en el seu aprenentatge, aquests qüestionaris compleixen la funció d'instruments d'avaluació formativa (Yorke, 2003). Per altra banda, atès que intervenen en l'obtenció de la nota final, també són considerats com a elements de l'avaluació sumatòria. A l'assignatura d'iniciació a les matemàtiques per a l'enginyeria, la realització continuada de qüestionaris es va concretar de manera que l'estudiant pogués realitzar com a mínim un qüestionari d'entrenament i un de mesura cada setmana.

Els plans docents de les assignatures d'iniciació a les matemàtiques per a l'enginyeria i d'anàlisi matemàtica preveien (curs 2010-11) i encara ho fan, l'existència d'una avaluació continuada, igual que a la resta d'assignatures de la UOC, en coherència amb el model pedagògic de la UOC (Duart i Sangrà, 2000). L'avaluació continuada no és obligatòria en aquest cas però es recomana que se segueixi per assegurar un progrés adequat i l'èxit a la prova final d'avaluació, en cas que existeixi. En general, sigui o no sigui obligatòria, consisteix en la realització de diverses activitats durant el semestre, planificades en el pla docent des de l'inici. Cadascuna d'elles s'anomena Prova d'Avaluació Continuada (PAC). El sistema d'avaluació d'iniciació a les matemàtiques per a l'enginyeria es basa, des dels seus inicis, en l'avaluació continuada, sense prova final. En canvi, a l'assignatura d'anàlisi matemàtica, el sistema d'avaluació contempla, també des dels seus inicis, l'avaluació continuada i la realització d'una prova final presencial obligatòria. L'avaluació continuada en aquest cas assegura un seguiment adequat de l'assignatura per part de l'estudiant i permet afinar la nota de la prova final. Aquest sistema d'avaluació s'ha mantingut fins la declaració de l'estat d'emergència el març del 2020.

Des del moment de la primera implantació el curs 2010-11, aquesta metodologia docent ha estat objecte de diverses anàlisis i reflexions que han conduït a una tesi doctoral (Escudero-Viladoms, 2012) i a diverses publicacions científiques (Calm et al., 2013; Escudero-Viladoms i Sancho-Vinuesa, 2016; Sancho-Vinuesa i Escudero-Viladoms, 2012; Sancho-Vinuesa et al., 2013; Sancho-Vinuesa et al., 2018), al marge de les pròpies d'aquesta tesi (Figueroa-Cañas i Sancho-Vinuesa, 2017; Figueroa-Cañas i Sancho-Vinuesa, 2019; Figueroa-Cañas i Sancho-Vinuesa, 2020; Figueroa-Cañas i Sancho-Vinuesa, 2021a; Figueroa-Cañas i Sancho-Vinuesa, 2021b). Sobre l'assignatura d'introducció a les matemàtiques per a l'enginyeria s'han efectuat estudis amb metodologia quantitativa (Sancho-Vinuesa, i Escudero-Viladoms, 2012), qualitativa (Escudero-Viladoms, 2012, Escudero-Viladoms i Sancho-Vinuesa, 2016) i mixta (Sancho-Vinuesa et al., 2013). El treball de Sancho-Vinuesa, i Escudero-Viladoms (2012) presenta una descripció, detallada tema per tema, de l'ús dels qüestionaris d'entrenament per part del conjunt d'estudiants, del rendiment acadèmic obtingut en els qüestionaris de mesura, així com la comparació del rendiment en qüestionaris de mesura segons l'ús dels qüestionaris d'entrenament, tot això durant el segon semestre d'implantació de la metodologia docent (2n semestre del curs 2010-11). Presenta, també, una distribució dels estudiants segons els resultats dels qüestionaris de mesura respecte l'ús de qüestionaris d'entrenament, tema per tema, així com una comparativa entre l'abandonament i l'ús de fòrum entre dos semestres amb implantació de la metodologia i dos semestres sense la implantació. En la seva tesi doctoral, Escudero-Viladoms (2012) analitza, en base a un estudi en profunditat sobre quatre estudiants del 2n semestre del curs 2010-11, l'efecte del feedback sobre la confiança matemàtica i sobre l'aprenentatge matemàtic, des d'una perspectiva qualitativa. Cal destacar-ne dos aspectes. En primer lloc, la confiança matemàtica incorpora elements actitudinals dels estudiants, com ara les seves creences sobre l'expectativa d'èxit en l'assignatura o sobre la seva capacitat per avançar sense interactuar amb companys i professorat. I en segon lloc, l'aprenentatge matemàtic no només té en compte els resultats acadèmics, sinó també, una valoració de dues competències matemàtiques (raonament i prova). Al treball de Sancho-Vinuesa et al. (2013) trobem una comparativa de l'abandonament entre nou semestres previs a la intervenció docent i el primer semestre amb la intervenció (1r semestre del curs 2010-11). D'aquest semestre, es presenta la distribució dels estudiants segons l'ús de qüestionaris d'entrenament i ús dels qüestionaris de mesura, tema per tema. També es presenten les distribucions dels estudiants segons l'ús de qüestionaris d'entrenament i la nota final de l'assignatura, així com segons l'ús de qüestionaris de mesura i la nota final de l'assignatura. El treball de Sancho-Vinuesa et al. (2013) inclou la perspectiva de sis estudiants pel que fa a factors emocionals relacionats amb el feedback rebut i l'ús intensiu de qüestionaris d'avaluació. A l'assignatura d'anàlisi matemàtica, Calm et al. (2013) presenta una comparativa del seguiment de l'avaluació continuada pel conjunt d'estudiants de diverses promocions prèvies a la implantació de la metodologia docent i la primera promoció amb la implantació (2n semestre del curs 2010-11). També presenta una evolució de les notes d'avaluació continuada en els dos primers semestres d'implantació. Sancho-Vinuesa et al. (2018), en les assignatures d'anàlisi matemàtica i matemàtiques II, analitza diverses variables (nombre d'aprovats de l'assignatura, nombre de presentats a l'examen final) entre promocions prèvies i posteriors a la implantació. En resum, en l'assignatura d'introducció a les matemàtiques per a l'enginyeria, trobem estudis quantitatius (Sancho-Vinuesa, i Escudero-Viladoms, 2012; Sancho-Vinuesa et al., 2013) que comparen dades d'estudiants d'una mateixa promoció, diferenciant-los segons l'ús de qüestionaris. Es tracta, doncs, d'estudis en què l'objecte d'anàlisi és l'estudiant d'una promoció. També trobem estudis que efectuen comparatives entre diverses promocions (Sancho-Vinuesa, i Escudero-Viladoms, 2012; Sancho-Vinuesa et al., 2013), prenent en aquest cas els estudiants de diverses promocions com a objecte d'estudi. En l'assignatura d'introducció a les matemàtiques per a l'enginyeria, els estudis quantitatius (Sancho-Vinuesa, i Escudero-Viladoms, 2012; Sancho-Vinuesa et al., 2013) empren procediments d'estadística descriptiva i no d'estadística inferencial. Pel que fa a l'assignatura d'anàlisi matemàtica, els treballs de Calm et al. (2013) i Sancho-Vinuesa et al. (2018) presenten estudis que exclusivament comparen diverses promocions, prenent els estudiants de diverses promocions com a objecte d'anàlisi.

1.5 Una perspectiva diferent: les analítiques d'aprenentatge

Aquesta tesi s'ha dut a terme amb la voluntat d'aprofundir en les anàlisis i reflexions prèvies sobre l'ús dels qüestionaris a l'assignatura d'anàlisi matemàtica des d'una perspectiva diferent: les analítiques d'aprenentatge. A més, hem ampliat el rang d'assignatures analitzades, afegint-hi l'assignatura d'estadística on, també, i d'acord amb els resultats obtinguts, hi hem dut a terme una intervenció docent.

Analitzar l'ús dels qüestionaris per part dels estudiants ens permet interpretar el nostre treball com un estudi sobre interaccions entre estudiant i sistema d'avaluació (Figura 6). Les interaccions es materialitzen en els accessos dels estudiants als elements d'avaluació (qüestionaris), a partir de dos enfocaments. El primer, corresponent al comportament de l'estudiant (Rienties et al., 2017), es concreta mitjançant mesures relacionades amb la quantificació de l'ús dels qüestionaris. El segon enfocament, corresponent al coneixement ("cognition") de l'estudiant (Rienties et al., 2017), es concreta mitjançant les qualificacions obtingudes en els qüestionaris.

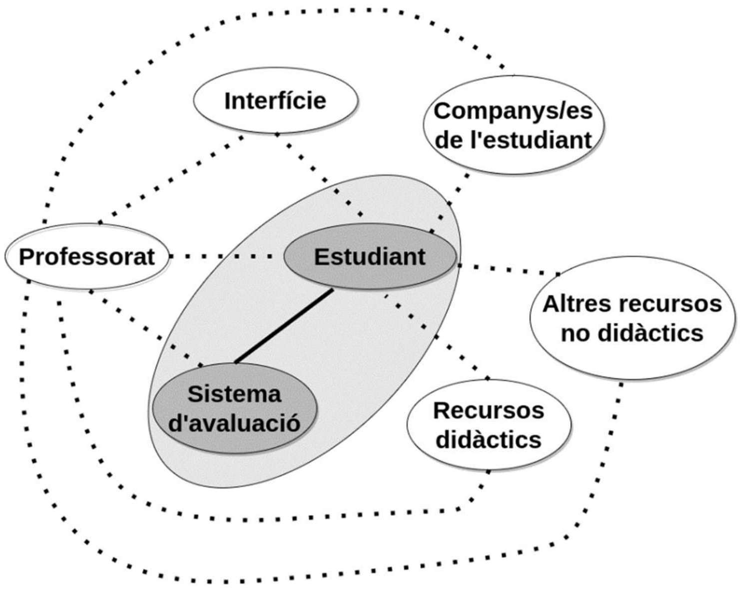


Figura 6. Model d'interaccions en què es destaquen les interaccions estudiant-sistema d'avaluació, principal focus de la tesi.

Pel que fa a l'assignatura d'anàlisi matemàtica, hem introduït una estratègia analítica que permet aprofundir en la comprensió del procés d'E/A. Les anàlisis efectuades estan sustentades en procediments d'estadística inferencial i s'ha considerat l'estudiant dins d'una mateixa promoció com objecte de l'anàlisi. A l'assignatura d'estadística, d'acord amb els resultats obtinguts, hi hem dut a terme una intervenció docent.

Tant en l'assignatura d'anàlisi matemàtica, com en la d'estadística, s'ha adoptat la perspectiva relacionada amb l'ús de les dades del procés d'ensenyament-aprenentatge per a la seva comprensió i millora: les analítiques d'aprenentatge ("learning analítics").

En 2011, el llavors concepte emergent d'**analítiques d'aprenentatge** es va definir com la mesura, recollida, anàlisi i transmissió de dades sobre aprenents i els seus contextos, amb el propòsit d'entendre i optimitzar l'aprenentatge i l'entorn en què ocorre (segons consta en la convocatòria de comunicacions del 1st International Conference on Learning Analytics and Knowledge, LAK 2011)[2]. Posteriorment, Clow (2012) presenta el cicle de les analítiques d'aprenentatge que, d'una banda, materialitza la definició anterior en quatre fases, ajudant així al seu ús, i d'una altra, suggereix la idea que les analítiques d'aprenentatge suposen una repetició infinita d'aquestes fases, amb l'objectiu de la millora contínua de l'aprenentatge (Figura 7).

A partir de la pregunta que es pretén respondre es defineixen quatre fases. En una primera fase se seleccionen els estudiants sobre els quals es farà l'anàlisi posterior. Després, en una segona fase, es generen o recullen les dades d'aprenentatge relacionades amb els estudiants. Posteriorment, en una tercera fase, es realitza el processament de les dades i posteriors càlculs, així com l'anàlisi dels resultats. I finalment, el cicle es tanca amb una quarta fase, que no sempre s'implementa, consistent a planificar i implementar una intervenció docent en base a l'anàlisi precedent, ja sigui sobre les mateixes persones que van iniciar el cicle o sobre d'altres.

[2] https://tekri.athabascau.ca/analytics/

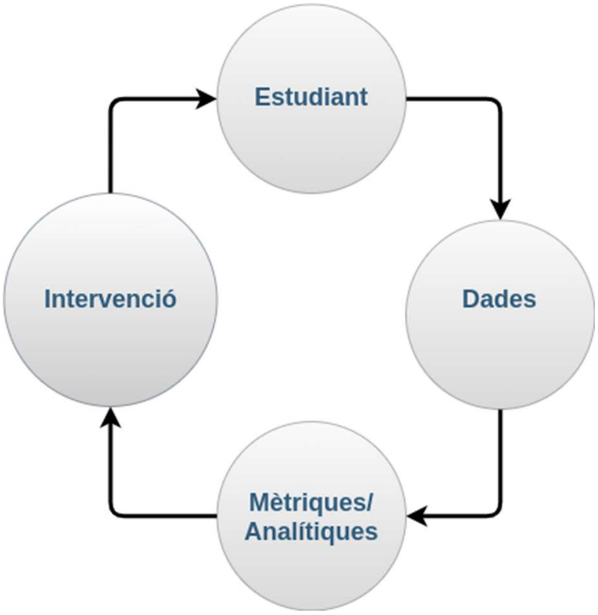


Figura 7. Cicle de les analítiques d'aprenentatge de Clow (2012).

Aquesta tesi amplia el cicle de quatre fases de Clow (2012) amb una cinquena fase de planificació, que a més, es considerada com la iniciadora del cicle. D'aquesta manera, queda explicitada la fase en què es planifiquen tota la resta de fases del cicle. En aquesta tesi doctoral hem definit i realitzat quatre cicles d'aquest tipus. (Figura 8).

En tots quatres cicles, el focus s'adreça a les interaccions fonamentals que s'esdevenen entre estudiant i sistema d'avaluació (taula 1) i la seva relació amb els resultats acadèmics. Tant el primer com el segon cicle l'estudi es duu a terme en el context de l'assignatura d'anàlisi matemàtica. La diferència rau que en el primer cicle s'estudien les interaccions en referència als resultats acadèmics dels qüestionaris de mesura, mentre que en el segon, s'estudien en referència al resultats acadèmics a l'examen final.

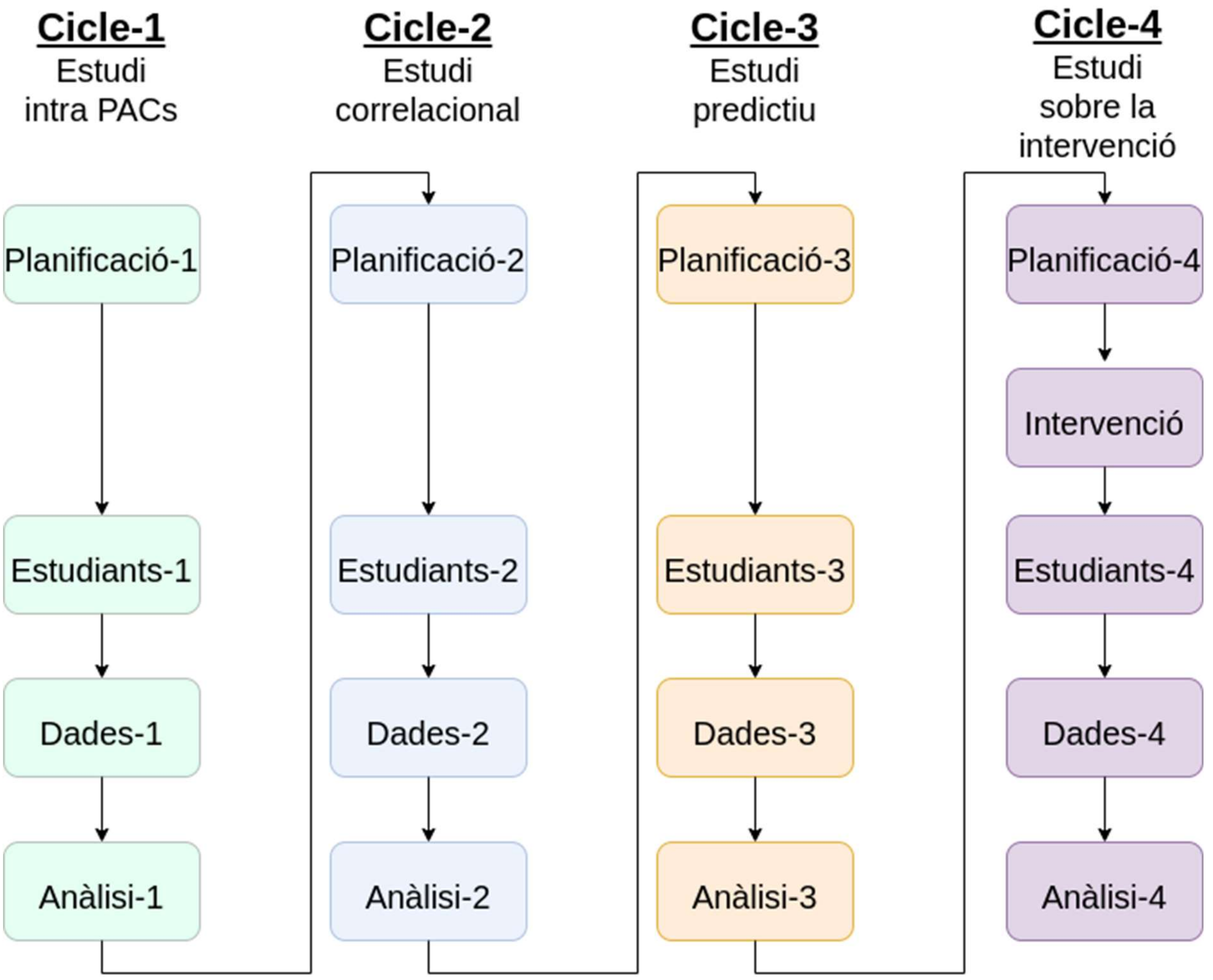


Figura 8. Cicles de les analítiques d'aprenentatge de la tesi

El tercer i quart cicle se situen a l'assignatura d'estadística, que comparteix amb la d'anàlisi matemàtica el fet de tenir un sistema d'avaluació amb examen final presencial obligatori i qüestionaris de mesura, si bé es diferencien en el fet que no disposa de qüestionaris d'entrenament. D'igual forma que al segon cicle, tant al tercer com al quart cicle, l'estudi es desenvolupa en relació als resultats acadèmics de l'examen final. El quart cicle inclou la implementació i avaluació d'una intervenció docent. El tercer cicle incorpora interaccions entre estudiant i altres recursos no didàctics així com estudiant i interfície. El quart cicle inclou interaccions entre estudiant-estudiant, estudiant-altres recursos no didàctics i estudiant-interfície. Tots quatre cicles s'estudien des d'una perspectiva quantitativa, i el quart es complementa amb una anàlisi qualitativa, a partir d'una entrevista a cinc estudiants.

Cicle	Assignatura	Interaccions	Resultats acadèmics analitzats
Cicle-1	Anàlisi matemàtica	(i) estudiant i sistema avaluació	Resultats acadèmics als qüestionaris de mesura
Cicle-2	Anàlisi matemàtica	(i) estudiant i sistema avaluació	Resultats acadèmics a l’examen final
Cicle-3	Estadística	(i) estudiant i sistema avaluació (ii) estudiant i altres recursos no didàctics (iii) estudiant i interfície	Resultats acadèmics a l’examen final
Cicle-4	Estadística	(i) estudiant i sistema avaluació (ii) estudiant i altres recursos no didàctics (iii) estudiant i interfície (iv) estudiant i estudiant	Resultats acadèmics a l’examen final

Taula 1. Resum de les interaccions i resultats acadèmics analitzats dels quatres cicles de la tesi.

1.6 Objectius

El propòsit d’aquesta tesi és contribuir a la millora de la qualitat docent de les assignatures de matemàtiques del grau d’enginyeria informàtica de la UOC, a través de l’estudi de les interaccions entre l’estudiant i el sistema d’avaluació. L'objectiu principal és aprofundir en la relació entre la pràctica (o entrenament) de l’estudiant amb qüestionaris en línia i els resultats acadèmics obtinguts. La recerca s’emmarca en l’estudi del comportament dels estudiants de les assignatures d’anàlisi matemàtica i estadística del grau d’enginyeria informàtica de la UOC. Per tal d’aconseguir l’objectiu principal, es plantegen quatre subobjectius:

Subobjectiu 1 (SO1): Determinar si, en l'assignatura d'anàlisi matemàtica, existeix relació entre la realització de qüestionaris d'entrenament i la qualificació de qüestionaris de mesura.

Subobjectiu 2 (SO2): Determinar si, en l'assignatura d'anàlisi matemàtica, existeix relació entre la realització de qüestionaris d'entrenament i de mesura que formen tot el conjunt de Proves d'Avaluació Continuada (PAC) i la qualificació en el examen final.

Subobjectiu 3 (SO3): Determinar si, en l'assignatura d'estadística, existeix relació entre la realització de qüestionaris de mesura que formen la primera part del conjunt de PAC i el resultat acadèmic en el examen final.

Subobjectiu 4 (SO4): Avaluar una intervenció docent duta a terme a l'assignatura d'estadística amb la finalitat de reduir la taxa d'abandonament.

Aquests subobjectius es concreten en sis preguntes de recerca:

Pregunta de recerca 1 (PR1): En l'assignatura d'anàlisi matemàtica, existeixen diferències de resultats en els qüestionaris de mesura entre els estudiants segons si opten o no per lliurar qüestionaris d'entrenament?

Pregunta de recerca 2a (PR2a): En l'assignatura d'anàlisi matemàtica, existeix relació entre la realització de qüestionaris d'entrenament i de mesura de l'estudiant i la seva qualificació en l'examen final?

Pregunta de recerca 2b (PR2b): En l'assignatura d'anàlisi matemàtica, existeix relació entre la qualificació mitjana del qüestionaris de mesura de l'estudiant i la seva qualificació en l'examen final?

Pregunta de recerca 3a (PR3a): En l'assignatura d'estadística és possible trobar un predictor en base a la realització i/o qualificació dels qüestionaris de mesura de la primera part del conjunt de PAC, per a classificar els estudiants anticipadament entre presentats i no presentats a l'examen final?

Pregunta de recerca 3b (PR3b): En l'assignatura d'estadística és possible trobar un predictor en base a la realització i/o qualificació dels qüestionaris de mesura de la primera part del conjunt de PAC, per a classificar els estudiants anticipadament entre aprovats i no aprovats en l'examen final?

Pregunta de recerca 4 (PR4): En l'assignatura d'estadística, permetre als estudiants realitzar algun dels dos primers qüestionaris de mesura, que no han lliurat dins de termini, redueix el nombre d'estudiants que no es presenten a l'examen final dins d'aquest grup d'estudiants?

CAPÍTOL 2

PRESENTACIÓ DE LES PUBLICACIONS

Les quatre publicacions que conformen el cos principal d'aquesta tesi responen a les sis preguntes de recerca plantejades i es corresponen amb els quatre cicles d'analítiques d'aprenentatge (figura 8). A continuació en detallem els cicles i les publicacions.

2.1 Exploring the efficacy of practicing with Wiris-Quizzes in online engineering mathematics

Figueroa-Cañas, J. i Sancho-Vinuesa, T. (2017). Exploring the Efficacy of Practicing With Wiris-Quizzes in Online Engineering Mathematics. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 12(3), 141-146. DOI: 10.1109/RITA.2017.2735499[3].

El 2017 la revista IEEE Revista Iberoamericana de Tecnologías del Aprendizaje va estar indexada a la categoria d'e-learning per SCOPUS amb un SJR dins del 3r quartil.

Aquest article presenta el treball corresponent al cicle-1 (**estudi intra PACs**) i respon a la pregunta de recerca PR1: En l'assignatura d'anàlisi matemàtica, existeixen diferències de resultats en els qüestionaris de mesura entre els estudiants segons si opten o no per lliurar qüestionaris d'entrenament?

[3] L'article publicat no coincideix, per errades en el procés de producció, amb l'article que va ser acceptat per a la seva publicació. El capítol 3 conté l'article realment acceptat.

L'estudi intra PACs s'inicia amb la **fase de planificació**, que estableix que cal esbrinar com es relacionen els qüestionaris d'entrenament (QE) amb els de mesura (QM). A l'assignatura d'anàlisi matemàtica, el pla docent estableix tres proves d'avaluació continua (PAC). La naturalesa de les PAC és diferent. Mentre que la primera PAC està formada per cinc paquets d'activitats, la resta en té un paquet cadascuna. Els paquets contenen qüestionaris, tant d'entrenament com de mesura (Figura 9). En aquest estudi ens focalitzem en les relacions que es produeixen a l'interior dels set paquets. Cerquem, doncs, si els dos instruments d'avaluació formativa (QE i QM) estan relacionats entre sí. Cal tenir en compte que els qüestionaris d'entrenament i mesura dins de cada paquet (a) contenen preguntes sobre una reduïda part de l'assignatura, (b) la tipologia i contingut de les seves preguntes mostren una considerable similitud, i (c) es realitzen en línia.

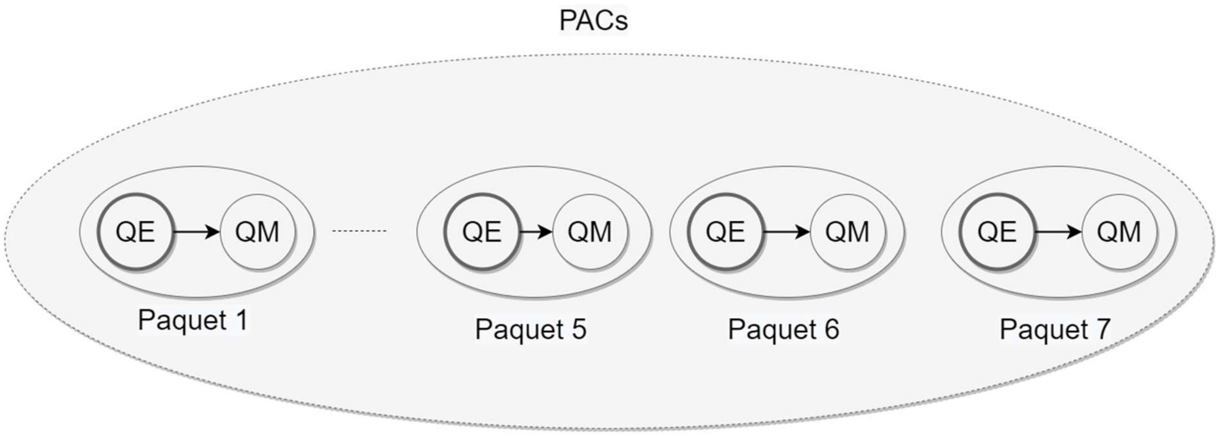


Figura 9 Diagrama de relacions entre qüestionaris d'entrenament (QE) i qüestionaris de mesura (QM), a l'assignatura d'anàlisi matemàtica.

El treball de Huisman i Reedijk (2012) és la principal referència metodològica d'aquest estudi intra PACs. Huisman i Reedijk (2012), en un entorn d'aprenentatge híbrid, comparen els resultats acadèmics de dos grups d'estudiants segons si realitzen qüestionaris d'entrenament (estudiants actius) o no en realitzen (estudiants inactius).

En finalitzar la fase de planificació, s'inicia la fase de **selecció dels estudiants**. Tot i que 131 estudiants es van matricular el primer semestre del curs 2015-16 a l'assignatura d'anàlisi matemàtica, només els 116 estudiants que van lliurar algun dels set qüestionaris de mesura van ser objecte d'estudi. D'aquests, de manera anònima i un cop finalitzats els set terminis de lliurament dels qüestionaris de mesura, s'obtenen les dades (fase Dades-1) relatives a la realització de qüestionaris d'entrenament i a les qualificacions obtingudes als qüestionaris de mesura. Es tracta d'un estudi ex-post facto atès que la realització de qüestionaris d'entrenament (variable independent) ocorre sense la interferència de l'investigador. Els cinc primers qüestionaris de mesura contenen dos tipus de preguntes. El primer tipus correspon a preguntes de format elecció múltiple o resposta oberta breu, amb correcció automàtica i comunicació de la qualificació immediata. Mentre que el segon tipus correspon a preguntes de format exposició raonada, amb correcció manual i comunicació de la qualificació diferida. Per a aquests cinc primers qüestionaris de mesura les qualificacions se separen en tres grups: (i) qualificacions exclusivament de les preguntes de correcció automàtica, (ii) qualificacions exclusivament de la pregunta d'exposició raonada, i (iii) qualificació global del qüestionari, que inclou les preguntes de correcció automàtica i la d'exposició raonada. A diferència dels cinc primers qüestionaris de mesura, els dos darrers qüestionaris de mesura (QM número 6 i 7) només contenen preguntes de correcció automàtica i per tant, només s'ha considerat la qualificació global dels qüestionaris.

Recollides les dades, es procedeix a la **fase d'anàlisi**. En primer lloc, en cadascun dels set paquets d'activitats, els estudiants se separen entre estudiants actius i inactius d'acord amb la categorització de Huisman i Reedijk (2012). Posteriorment, s'efectua un test de Yuen de mitjanes truncades al 25% per determinar si existeixen diferències en les qualificacions entre el conjunt d'estudiants actius i el d'inactius. Keselman et al. (2004) suggereix emprar aquest test per aconseguir robustesa en situacions en què les dades presenten dificultats per acceptar l'assumpció de normalitat, com és el cas del treball presentat.

Aquest estudi aborda una part de la problemàtica plantejada en la tesi. Ens ha permès respondre a la PR1 i afirmar que, a l'assignatura d'anàlisi matemàtica, **els estudiants que realitzen qüestionaris d'entrenament obtenen millors resultats als qüestionaris de mesura**.

Aquest **article** resulta **rellevant** per la novetat que constitueix determinar diferències de resultats acadèmics entre dos grups d'estudiants mitjançant el test de Yuen de mitjanes truncades.

2.2 Investigating the relationship between optional quizzes and final exam performance in a fully asynchronous online calculus module

Figueroa-Cañas, J. i Sancho-Vinuesa, T. (2021)[4]. Investigating the relationship between optional quizzes and final exam performance in a fully asynchronous online calculus module. *Interactive Learning Environments*, 29(1), 33-43. DOI: 10.1080/10494820.2018.155986.

El 2020 la revista Interactive Learning Environments ha estat indexada a la categoria d'Education & Educational Research pel JCR de Clarivate amb un JIF dins del 1r quartil.

Aquest article presenta el treball corresponent al cicle-2 (**estudi correlacional**) i respon a les preguntes de recerca PR2a: En l'assignatura d'anàlisi matemàtica, existeix relació entre la realització de qüestionaris d'entrenament i de mesura de l'estudiant i la seva qualificació en l'examen final? i PR2b: En l'assignatura d'anàlisis matemàtica, existeix relació entre la qualificació mitjana del qüestionaris de mesura de l'estudiant i la seva qualificació en l'examen final?

La **fase de planificació** estableix que cal determinar com es relacionen els qüestionaris d'entrenament i de mesura, instruments de l'avaluació formativa, amb l'examen final, l'instrument fonamental de l'avaluació sumatòria (Figura 10). És a dir, amplia la perspectiva del primer estudi (intra PACs), cercant relacions entre el conjunt de qüestionaris d'entrenament i de mesura, que formen les proves d'avaluació continuada (PAC), i l'examen final. Això suposa esbrinar si es pot establir alguna relació entre instruments d'avaluació tant diferents com els qüestionaris de mesura, que contenen preguntes de continguts parcials de l'assignatura i es responen en línia, amb l'examen final, en què les preguntes corresponen a qualsevol contingut de l'assignatura i es responen presencialment.

El treball d'Angus i Watson (2009) és la principal referència metodològica d'aquest estudi correlacional. Angus i Watson (2009), en un entorn d'aprenentatge híbrid, planteja una regressió

[4] Publicat en línia el desembre de 2018

lineal múltiple amb una transformació logística per determinar si la realització de qüestionaris està relacionada amb els resultats acadèmics a l'examen final.

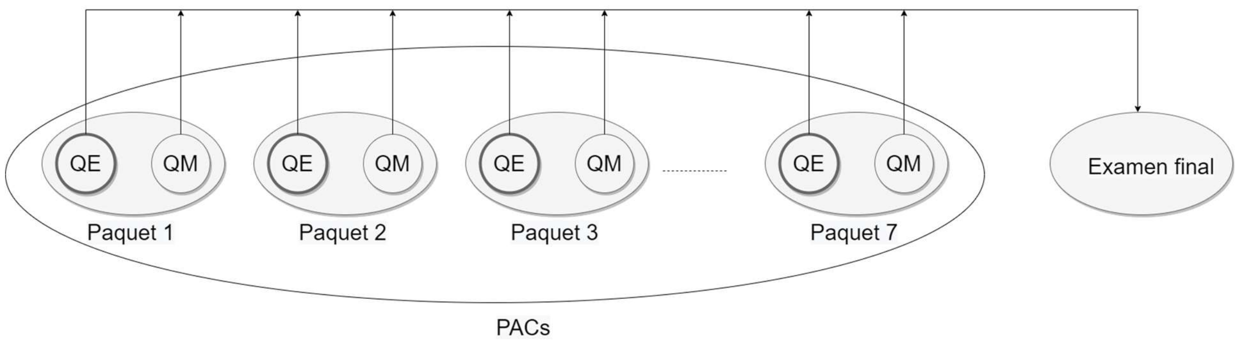


Figura 10 Diagrama de relacions entre qüestionaris d'entrenament (QE) i qüestionaris de mesura (QM) amb l'examen final, a l'assignatura d'anàlisi matemàtica.

En finalitzar la fase de planificació, es trien els **estudiants participants** en l'estudi. Aquests són els 176 estudiants que s'han matriculat en l'assignatura d'anàlisi matemàtica en el primer semestre del curs 2016-17 i que compleixen una doble condició. La primera és haver-se presentat a l'examen final i la segona haver lliurat algun dels set qüestionaris de mesura. D'aquests estudiants, de manera anònima, s'obtenen les dades en relació a si havien realitzat qüestionaris d'entrenament, les qualificacions obtingudes en els qüestionaris de mesura, la qualificació de l'examen final i altres dades referents a variables considerades de control. En aquest sentit, es recullen dades sobre en quina llengua (català o castellà) s'havien matriculat, i sobre si havien realitzat o no una prova inicial no obligatòria sobre coneixements d'anàlisi matemàtica de nivell d'estudis secundaris. En cas d'haver-la realitzada, també es recull la qualificació.

En la **fase d'anàlisi** es processen les dades de realització de qüestionaris d'entrenament i de mesura i es transformen en dues variables independents que informen sobre la realització global, per una banda, de qüestionaris d'entrenament i per, una altra, de qüestionaris de mesura. Concretament, es generen dues ràtios de realització de qüestionaris, una per als d'entrenament i una altra per als de mesura, que indiquen el percentatge de preguntes realitzades. Aquesta forma de mesurar la realització de qüestionaris afegeix matisos a la mesura binària (realitzat/no realitzat) emprada en l'estudi intra PACs. Les dues ràtios de realització es corresponen amb dues variables independents. En paral·lel, amb les dades de qualificació dels set qüestionaris de mesura es calcula la mitjana i es defineix així la tercera variable independent. Posteriorment s'efectua una transformació logística de la qualificació de l'examen final, que és emprada com a variable dependent. Amb les tres variables independents (ràtio de realització de qüestionaris d'entrenament, ràtio de realització de qüestionaris de mesura i mitjana de les qualificacions dels qüestionaris de mesura), la dependent (qualificació de l'examen final transformada logísticament) i les de control (llengua i realització/qualificació a la prova inicial de coneixements previs) es realitza una regressió lineal múltiple, basant-nos en la proposta d'Angus i Watson (2009). A més, per al grup d'estudiants que realitza la prova inicial de coneixements previs, es duu a terme un test de Welch de mitjanes aparellades entre les qualificacions de l'examen final i les de la prova inicial.

Aquest estudi aborda una part de la problemàtica plantejada en la tes. Ens ha permès respondre a la PR2a i PR2b i afirmar que, a l'assignatura d'anàlisi matemàtica, **la qualificació a l'examen final està relacionada amb el fet de realitzar qüestionaris d'entrenament i amb les qualificacions als qüestionaris de mesura**.

Aquest **article** resulta **rellevant** atès que cobreix la inexistència de treballs publicats sobre la relació entre realitzar qüestionaris en línia i els resultats acadèmics assolits, en un entorn d'aprenentatge de matemàtiques en línia i asíncron.

2.3 Early prediction of dropout and final exam performance in an online statistics course

Figueroa-Cañas, J. i Sancho-Vinuesa, T. (2020). Early Prediction of Dropout and Final Exam Performance in an Online Statistics Course. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 15(2), 86-94. DOI: 10.1109/RITA.2020.2987727.

El 2020 la revista IEEE Revista Iberoamericana de Tecnologías del Aprendizaje ha estat indexada a la categoria d'e-learning per SCOPUS amb un SJR dins del 4t quartil.

Aquest article presenta el treball corresponent al cicle-3 (**estudi predictiu**) i respon a les preguntes de recerca PR3a: En l'assignatura d'estadística és possible trobar un predictor en base a la realització i/o qualificació dels qüestionaris de mesura de la primera part del conjunt de PAC, per a classificar els estudiants anticipadament entre presentats i no presentats a l'examen final? i PR3b: En l'assignatura d'estadística és possible trobar un predictor en base a la realització i/o qualificació dels qüestionaris de mesura de la primera part del conjunt de PAC, per a classificar els estudiants anticipadament entre aprovats i no aprovats a l'examen final?

En la **fase de planificació** es decideix analitzar la mateixa metodologia docent emprada a l'assignatura d'anàlisi matemàtica en els dos estudis precedents (intra PACs i correlacional), però en aquest cas a l'assignatura d'estadística. El sistema d'avaluació d'aquesta assignatura es diferencia del de l'anàlisi matemàtica en tres aspectes. Primer, cada prova d'avaluació contínua (PAC) està formada per un únic paquet d'activitats. Segon, les PAC no contenen qüestionaris d'entrenament. Per ajudar els estudiants a entrenar-se, els qüestionaris de mesura permeten la realització de dos intents, prenent la qualificació del qüestionari com la màxima obtinguda entre els dos intents. I el tercer aspecte diferenciador rau en la pregunta d'exposició raonada, que en l'assignatura d'estadística sempre consisteix en activitats que s'han de resoldre fent servir el programa estadístic R i sempre es lliuren en un fitxer separat de la resta de preguntes del qüestionari. El sistema d'avaluació de l'assignatura d'estadística consisteix en un examen final presencial i sis PAC, cadascuna de les quals està formada per un qüestionari de mesura (QM) i per una pregunta sobre el programa R (RT). La fase de planificació determina que cal veure com es relacionen els primers qüestionaris de mesura del curs amb l'examen final (Figura 11), esbrinant si amb la informació relativa a l'ús i/o resultats dels primers qüestionaris de mesura es poden predir els resultats a l'examen final.

Els **estudiants seleccionats** són tots els 197 estudiants matriculats a l'assignatura d'estadística en el primer semestre del curs 2018-19. Això suposa ampliar la perspectiva d'estudiants considerats en els cicles precedents (estudi intra PACs i estudi correlacional), que només tenien en compte aquells que havien realitzat algun qüestionari, i per tant, que havien seguit la proposta docent de practicar amb qüestionaris amb feedback automàtic. En aquest estudi, els estudiants inclosos poden o no haver realitzat qüestionaris, i conseqüentment poden o no haver fet ús de la proposta docent.

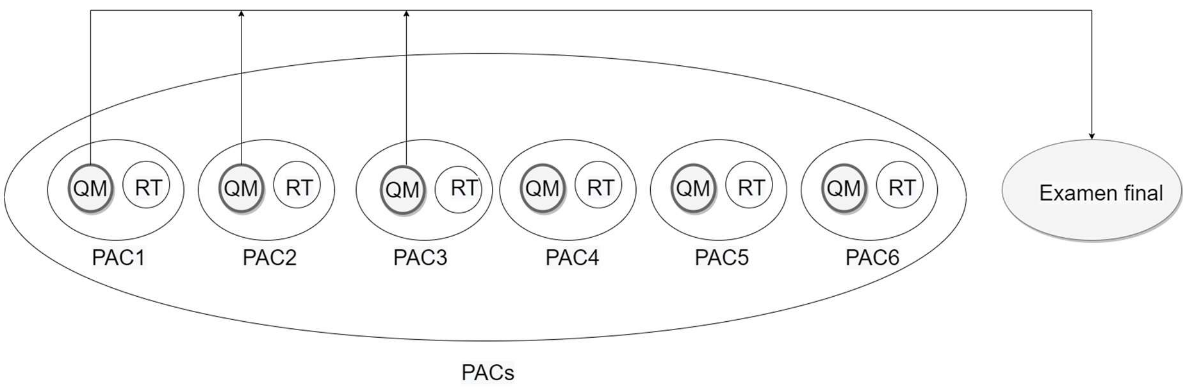


Figura 11 Diagrama de relacions entre els primers qüestionaris de mesura (QM) i l'examen final, a l'assignatura d'estadística.

Per a aquest article, les dades recollides de cada estudiant fan referència a quatre àmbits: de caràcter personal i invariant des de l'inici de la matriculació, indicadors de compromís ("engagement") envers l'activitat d'avaluació, indicadors d'assoliment ("achievement") i indicadors de compromís a l'aula virtual. Quant a les dades de caràcter personal, en concret, fan referència a si l'estudiant és repetidor/a i al nombre d'assignatures matriculades en el mateix semestre que està matriculat a l'assignatura d'estadística. Pel que fa al compromís en activitats d'avaluació, es recullen quins qüestionaris de mesura ha realitzat dins de les tres primeres PAC, quines de les tres primeres preguntes sobre R ha realitzat, i si ha realitzat una prova inicial opcional de coneixements previs de nivell d'ensenyament secundari. Finalment, quant al compromís a l'aula virtual, les dades fan referència a si ha visualitzat el pla docent i en quin moment, si ha accedit al tauler virtual de l'assignatura i en quins moments, si ha escrit posts en el fòrum de discussió i en quins moments; i si ha llegit posts del fòrum i en quins moments. A més, com a indicadors d'assoliment, s'han tingut en compte les qualificacions de la prova inicial, dels tres primers qüestionaris de mesura, de les tres primeres preguntes sobre R, i la qualificació a l'examen final.

En la **fase d'anàlisi**, en primer lloc s'han considerat tres intervals temporals. Cadascun d'ells s'inicia al començament del curs i acaba en cadascuna de les tres dates límit per al lliurament de les tres primeres PAC. En segon lloc, en cada interval temporal es processen les dades de manera que es construeixen atributs que contenen informació acumulada al llarg del propi interval, quant a realització de qüestionaris de mesura, preguntes sobre R, sobre accessos al pla docent, accessos al tauler i al fòrum. També es calculen les mitjanes de les qualificacions de tots els qüestionaris de mesura del període, així com de les qualificacions de les preguntes sobre R. A més, per una banda es transformen les qualificacions de l'examen final en dues categories: si l'estudiant s'hi ha presentat i si no s'hi ha presentat. I per una altra banda es duu a terme una altra transformació de l'examen final en dues noves categories: si l'estudiant ha aprovat l'examen final i si no l'ha aprovat (els no presentats també en formen part d'aquesta categoria). En tercer lloc, es generen predictors en forma d'arbres de decisió condicionals (Hothorn et al., 2006), per a cada interval temporal, mitjançant una validació creuada ("cross-validation") de 5 plecs ("folds") amb un submostratge ("undersampling") aleatori. A continuació, es calculen diverses mètriques de rendiment de la predicció: precisió ("precision"), sensibilitat ("recall") i mesura-F ("F-measure").

Aquest estudi aborda una part de la problemàtica plantejada en la tesi. Ens ha permès respondre a la PR3a i PR3b i afirmar que, a l'assignatura d'estadística, **no presentar-se a l'examen final o no reeixir-hi és predictible a partir de les qualificacions als primers qüestionaris de mesura**.

Aquest **article** resulta **rellevant** atès que l'activitat realitzada per l'estudiant a la meitat del semestre proporciona un predictor fàcilment interpretable, mitjançant un arbre de decisió. Això permet plantejar-se la possibilitat de planificar i implementar alguna intervenció docent amb l'objectiu que la predicció es pugui revertir.

Val a dir que l'article presentat en aquest apartat fins ara correspon a l'article ampliat de (Figueroa-Cañas i Sancho-Vinuesa, 2019) que figura a l'apèndix. En una primera etapa, l'estudi predictiu (cicle-3) es proposava respondre parcialment la pregunta de recerca PR3a, d'acord amb el següent redactat: En l'assignatura d'estadística és possible trobar un predictor en base a la realització dels qüestionaris de mesura de la primera part del conjunt de PAC, per a classificar els estudiants anticipadament entre presentats i no presentats a l'examen final? En aquesta primera etapa les dades recollides feien referència a tres dels quatre àmbits esmentats amb anterioritat: de caràcter personal i invariant des de l'inici de la matriculació, indicadors de compromís en l'activitat d'avaluació i indicadors de compromís a l'aula virtual. Quant a la fase d'anàlisi, es va considerar incloure també un quart període temporal, que s'iniciava al començament del curs i acabava en al data límit per al lliurament de la quarta prova d'avaluació continuada. En cadascun dels quatre intervals temporals es van processar les dades de manera que es van construir atributs de manera similar al cas exposat a l'inici d'aquest apartat 2.3. Igualment es van transformar les qualificacions de l'examen final en dues categories: si l'estudiant s'hi havia presentat i si no s'hi havia presentat. També per a cada interval temporal, es van generar predictors en forma d'arbres de decisió condicionals (Hothorn et al., 2006), tot i que a diferència de (Figueroa-Cañas i Sancho-Vinuesa, 2020), a partir d'una selecció aleatòria del conjunt de dades repartides en un 80% per al conjunt d'entrenament ("training set") i en un 20% per al conjunt de validació ("validation set"). Les mesures de rendiment triades van ser: exactitud ("accuracy"), precisió ("precision") i sensibilitat ("recall").

Aquest treball (Figueroa-Cañas i Sancho-Vinuesa, 2019) encaixa en la problemàtica plantejada en la tesi. Ens ha permès respondre parcialment a la PR3a i afirmar que, a l'assignatura d'estadística, **no presentar-se a l'examen final és predictible a partir del comportament als primers qüestionaris de mesura**.

2.4 Changing the recent past to reduce ongoing dropout: an early learning analytics intervention for an online statistics course

Figueroa-Cañas, J. i Sancho-Vinuesa, T. (2021). Changing the recent past to reduce ongoing drop-out: an early learning analytics intervention for an online statistics course. *Open Learning: The Journal of Open, Distance and e-Learning*. DOI: 10.1080.02680513.2021.1971963.

El 2020 la revista Open Learning: The Journal of Open, Distance and e-Learning ha estat indexada a la categoria d'e-learning per SCOPUS amb un SJR dins del 2n quartil.

Aquest article presenta el treball corresponent al cicle-4 (estudi sobre la intervenció) i respon a la pregunta de recerca PR4: En l'assignatura d'estadística, permetre als estudiants realitzar algun dels dos primers qüestionaris de mesura, que no han lliurat dins de termini, redueix el nombre d'estudiants que no es presenten a l'examen final dins d'aquest grup d'estudiants?

La **fase de planificació** estableix que en l'assignatura d'estadística s'implementarà una intervenció docent amb l'objectiu de reduir la taxa d'abandonament, prenent com a base els predictors trobats en el cicle-3. La intervenció consta de dos components. El primer component és informar, mitjançant un correu electrònic, els estudiants que no han realitzat els dos primers qüestionaris de mesura que això els fa estar en risc d'abandonar l'assignatura. El segon component és permetre que els estudiants contactats puguin disposar d'una segona oportunitat per realitzar aquell/s qüestionari/s de mesura no lliurat/s prèviament. La fase de planificació també determina com serà avaluada la intervenció, posant especial èmfasi en comprovar si assoleix l'objectiu de reducció de la taxa d'abandonament, en el nostre cas entès com la reducció del nombre d'estudiants que no es presenten a l'examen final.

La **fase d'intervenció** consta de la seva implementació real. Un cop finalitzat el termini per al lliurament de la segona PAC, es van enviar correus electrònics als 35 estudiants considerats en risc d'abandonament, a partir dels predictors del cicle-3. Se'ls van donar quinze dies per a poder realitzar els qüestionaris no lliurats prèviament. Passats aquests quinze dies el curs va continuar amb el calendari de PACs programat al pla docent abans del començament del curs. Es tracta, doncs, d'una intervenció molt focalitzada en un conjunt reduït d'estudiants i molt limitada en el temps.

Finalitzada la fase d'intervenció, s'inicia la **fase de selecció dels estudiants**. Aquesta determina que entre els 225 estudiants matriculats en l'assignatura d'estadística en el primer semestre del curs 2019-20, tots 35 estudiants que han rebut l'opció de poder realitzar qüestionaris de mesura en segona oportunitat formen part del conjunt objecte d'aquest estudi. També en formen part els 36 estudiants de la promoció d'estudiants d'estadística del primer semestre del curs 2018-19 que haurien estat considerats en risc d'abandonament seguint el mateix criteri aplicat als seus companys de la promoció del curs 2019-20.

Tant d'aquests 36 estudiants del curs 2018-19 com del 35 del curs 209-20, de manera anònima s'obtenen les dades (fase Dades-4) que indiquen si s'han presentat o no a l'examen final.

A mode de resum, la Figura 12 visualitza en primer lloc l'origen de les dades que condueixen a determinar quins estudiants són objecte de la intervenció (part esquerra), i en segon lloc, quina relació s'estableix entre la intervenció i l'examen final (part dreta).

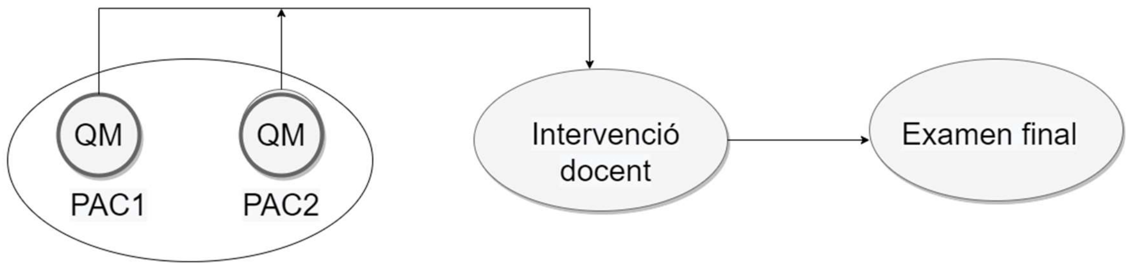


Figura 12 Diagrama de relacions entre la intervenció a l'assignatura d'estadística, els qüestionaris inicials i l'examen final.

En la **fase de dades** també recollim, dels 35 estudiants intervinguts, el resultat en els dos primers qüestionaris de mesura durant el termini ordinari de lliurament, si han realitzat o no els dos primers qüestionaris de mesura durant els quinze dies en què te lloc la intervenció, i si és el cas la qualificació obtinguda; així com si han realitzat o no i la qualificació en els qüestionaris 3, 4, 5 i 6. A més, arrepleguem dades sobre accessos a l'aula virtual i el pla docent durant tot el semestre. Per altra banda, recollim dades en relació als contactes d'aquests estudiants amb el seu professorat. A totes les dades anteriors, de caràcter quantitatiu, cal afegir-hi dades de caràcter qualitatiu per tenir en compte la perspectiva de l'estudiant, fet que diferencia aquest estudi de la resta d'estudis previs. Els 12 estudiants que van mostrar algun tipus de resposta a la intervenció, ja sigui havent respost al missatge rebut i/o havent realitzat algun qüestionari prèviament no realitzat, van ser convidats a participar en una entrevista semiestructurada. L'entrevista, que finalment va ser acceptada per 5 estudiants, contenia preguntes en relació a tres temes. El primer fa referència a com eren les condicions externes (familiars, laborals, ...) en el moment de decidir si realitzar o no els dos primers qüestionaris de mesura. El segon tema pren com a referència el qüestionari EFLA sobre instruments ("tools") per a efectuar posteriorment analítiques d'aprenentatge (Scheffel, 2017). En particular, el nostre *instrument* és el missatge enviat als estudiants. Les preguntes relacionades amb aquest segon tema pretenen esbrinar si el missatge, per una banda ha fet prendre consciència, els estudiants, de la situació en què es trobaven un cop finalitzades les dues primeres PAC i de les conseqüències del possible futur en l'assignatura. Per l'altra, si ha dut els estudiants a reflexionar sobre la seva manera d'actuar en els dos primers qüestionaris i com actuar en els qüestionaris següents. I el tercer tema, que està relacionat amb la càrrega cognitiva (Kalyuga, 2012), conté preguntes que prenen com a referència el qüestionari de Leppink et al (2013) com ara "eren complexes les preguntes dels qüestionaris de mesura?"

Recollides les dades, es procedeix a la **fase d'anàlisi**, que combina un treball de caràcter quantitatiu i un altre de qualitatiu. En el primer s'efectuen tres anàlisis. La primera consta d'una comparació, en relació a si s'han presentat o no a l'examen final, entre els 35 estudiants intervinguts de la promoció 2019-20 i els 36 estudiants de la promoció 2018-19 que compleixen les condicions per haver estat intervinguts. La segona anàlisi persegueix establir una relació entre com els estudiants responen a la intervenció i com havien actuat prèviament a ser intervinguts. La tercera anàlisi tracta de determinar com es relacionen l'assistència o no a l'examen final amb quina ha estat la resposta a la intervenció. Quant al treball qualitatiu, s'efectua una anàlisi a partir de les cinc entrevistes enregistrades.

Aquest estudi aborda una part de la problemàtica plantejada en la tesi. Ens ha permès respondre a la PR4 i afirmar que, a l'assignatura d'estadística**, la intervenció docent augmenta la probabilitat de reduir el nombre d'estudiants que no es presenten a l'examen final**.

Aquest **article** resulta **rellevant** per la novetat que constitueix implementar una intervenció que ofereix l'oportunitat de fer un pas enrere i modificar resultats dels comportaments del passat, en lloc d'oferir només la possibilitat d'introduir canvis en els comportaments futurs (Cambruzzi et al., 2015; Corrigan et al., 2015; Jayaprakash et al., 2014; Pardo et al., 2019).

CAPÍTOL 3

EXPLORING THE EFFICACY OF PRACTICING WITH WIRIS-QUIZZES IN ONLINE ENGINEERING MATHEMATICS[5]

The use of online self-assessed questionnaires Wiris-Quizzes as formative assessment tool is an increasingly widespread practice in a variety of subjects in higher education. Some previous studies have examined the impact of such questionnaires on learning outcomes in classroom environments or hybrid teaching. Our work presents an exploratory study that proves that in a course of mathematical analysis in a completely online environment, practices with Wiris-Quizzes improve learning outcomes. We have conducted an ex-post facto research from the results of the practice tests and continuous assessment tests carrying out an analysis using the trimmed means Yuen's test.

Index Terms- Formative assessment, Online engineering mathematics, Learning, Wiris-Quizzes

3.1 Introduction

Students of mathematics on an engineering degree at the Universitat Oberta de Catalunya (UOC) add to the lack of motivation and basic knowledge of the subject, the specific features of an online university: family and professional responsibilities that lead to a reduction in weekly study time, and the difficulties inherent to learning mathematics online [1], [2].

[5] En aquest capítol, que **conté el primer article del conjunt que conformen la tesi**, les referències numèriques a figures i taules corresponen a referències exclusivament d'aquest capítol. Les referències bibliogràfiques d'aquest capítol es mostren dins d'aquest mateix capítol.

In two subjects in this area, a teaching methodology based on the student's continuous activity has been implemented through an assessment and automatic feedback tool, with the aim of improving teaching quality [1]: the first semester of the academic year 2010-11, in the elective of initiation into mathematics for engineers module in the computer engineering degree, and in the first semester of the academic year 2011-12, in the compulsory of mathematical analysis module in the computer engineering degree and telecommunication engineering degree. The implementation in the latter has been carried out in a progressive way until completion in the first semester of the academic year 2015-16, during which this study has been conducted. More than 10 years of experience in this context led us to define a methodology that would help the student to perform, during the semester, a regular activity combined with immediate feedback. The learning resource that fitted our needs was the Wiris-Quizzes. They are quizzes designed in the Moodle environment, which are supported by the symbolic calculation program Wiris and its mathematical formula editor. Four main features justified its choice in the initial phase of the project: a) self-assessment and automatic response; b) immediate response and feedback; c) parameterised statements; and d) the possibility of the introduction of numbers and mathematical expressions that will be interpreted by the system. Thus, whenever a quiz is opened, the statement is different because the values of these parameters are different, and therefore there are a high number of variants [2]. The possibility of introducing mathematical expressions in a flexible way reduces the time spent by students writing the answers and permits that the type of questions proposed have a higher degree of complexity. As for the system response, the student receives corrective feedback [3] by providing the correct answer with a detailed resolution of the exercise, as well as a reference to the content of the didactic materials of the subject. Moreover, the qualification obtained is provided.

In this article we shall explore the learning characteristics of the subject mathematical analysis in the computer engineering degree through the use of quizzes with automatic feedback. Specifically, the main aim of this research is to determine, in the first approach given that this is an exploratory study, whether to take practice tests to pass the continuous assessment component of the subject mathematical analysis of the degree in computer engineering at the UOC is effective. We shall study this effectiveness through the analysis of the students' continuous assessment tests results. Specifically, we intend to verify the existence or non-existence of differences in results between two groups of students: a group with students who choose to submit practice tests (active students) and a group of students who choose not to do so (non-active students), taking as a reference the classification of active students of [5].

To the best of our knowledge, the context of our study, a mathematics subject in an engineering degree in a completely online environment, as well as the data analysis used, a trimmed means test, have not been used in other published papers.

3.2 Background

The use of online quizzes for formative purposes has been used in online higher education in areas such as mathematics in engineering [2] or in the social sciences [6]; but also in classroom teaching higher education, in so-called hybrid methodologies, in areas of study such as mathematics in economics [5], [7], [8], mathematics in engineering [9], biology [10], or economics [11].
There are several research papers devoted to determine the impact of online practice tests on learning. These studies coincide in affirming that taking these quizzes does improve learning, although they differ in the selected variable to measure the learning outcomes and in the methodology used to carry out the research, mainly in how the data analysis is conducted. Below we present the main characteristics of these studies.

As for the variable that measures learning, the final exam score is used as the dependent variable in [5], [7], [8], [9], [10]. The study of [11], however, uses the marks of didactic unit assessments and that of [2], the grade of the subject as a whole.

In terms of methodology and results, reference [11] implements a test of differences between mean marks in the unit assessments of two groups of students: a group of students who have taken practice tests and a group of students who have not. Students have decided to be in one group or in one other; therefore it is a non-experimental design. It establishes that taking practice tests improves learning. The average marks of unit assessment's students who take practice tests is higher, with a statistical significance less than 5%, than the ones of the students who opt to not take them. Reference [6] also uses a test of differences between means, in this case in the final exam scores, between two groups of students. Here, it is a quasi-experimental design with a control group formed by students without the choice of taking practice tests, and a second group, the experimentation group, formed by students obliged to take practice tests. The study shows the improvement of learning by taking practice tests, taking into account that the average of the final exam of the subject is significantly higher in the group of students obliged to take practice tests, the experimentation group, than in the group of students without available practice tests, the control group. In reference [10], the data analysis consists of comparing the mean of the final exams between cohorts that have available practice tests and those that have not. The conclusion for [10] differs according to the volition in taking the practice tests. For a cohort with absolute volition, there is no significant improvement in learning, while in another cohort where the practice tests represent 20% of the total grade of the subject, it is observed that the average of the final exam grade of students is higher than that of cohorts without available practice tests. An ex-post facto design, along with an analysis of variance test (ANOVA) is used by [9], who reaches the same conclusion as [10], in the sense that only when conducting quizzes with direct repercussion on the final mark improvement in learning is noticed, whereas there is no improvement otherwise. Reference [9] shows that students who submit the highest number of quizzes obtain better grades in the final exam. Through a correlational study and a logistic regression, reference [7] proves the improvement of learning, establishing that there is an association between taking practice tests and the final exam grade. Another correlational study, in this case using a ridge regression [8], concludes the quizzes' effectiveness, proving the existence of a positive association between the number of quizzes attempted and the final exam grade. An ex-post facto design and a means difference test [5] conclude the improvement of learning in the students that take practice tests, active students, noting that their final exam means are higher than that of their non-active peers.

3.3 Educational context and pedagogical methodology

The Universitat Oberta de Catalunya, UOC, requires students to take a mandatory face-to-face final exam in all mathematics subjects in Computer engineering, as well as the possibility, recommended by teachers to the students, to participate in the continuous assessment. Each subject determines the detail of the continuous assessment, the purpose of which is both formative and summative. The qualification system of the subject under study integrates in a weighted way the continuous assessment qualification, CA, and the face-to-face final exam grade, FE. Thus, the final grade of the subject, FG, is obtained according to (3.1)

$$FG = \max(FE \; ; \quad 0,65 \cdot FE + 0,35 \cdot CA) \quad (3.1)$$

This system ensures the optional nature of the continuous assessment for those students who either choose to be evaluated exclusively by the final exam (CA = 0) or, for those whose continuous assessment qualification obtained throughout the learning process, would lower the final exam score by incorporating it into the weighted average with a weight of 35%. In short, continuous assessment will only be an element of summative assessment, by contributing quantitatively to the final grade [4] whenever it favors the student, while keeping its formative character, given that the information provided the student is intended to contribute to their learning.

The continuous assessment consists of seven continuous assessment quizzes (CAQ), in which students have only one attempt to respond. The first five quizzes, planned to be taken weekly, focus on basic aspects of mathematical analysis, many of which have already been studied in secondary school or in the subject initiation into mathematics for engineers. These are Wiris-Quizzes questionnaires of six questions, where the first five are of the multiple choice or short answer type. Correction is done automatically and instantly, giving the student immediate feedback. The sixth question pertains to the type of open response in which the student must respond in a reasoned manner. Correction corresponds to the teacher and the response time can be up to one week. The last two continuous assessment quizzes, scheduled to be held monthly, focus on advanced aspects of mathematical analysis, not previously worked on in other subjects.

Prior to the beginning of each continuous assessment quiz, students have the possibility of taking one or more practice tests (PT), consisting of 10 parameterised questions of the multiple choice or short answer type, see for example the one shown in Fig. 1. Once the answers have been sent, as in the case of Fig. 2 by introducing mathematical expressions in the Wiris program, the system automatically and immediately provides corrective feedback informing of the correct answer, offering an aid to comprehension (Fig. 3), numerically evaluating the response (Fig.1), as well as the whole set of ten questions. The purpose of these quizzes is for the student to get confidence with this type of activity and to permit them flexibility in performing the tests. The number of models of practice tests differs according to the continuous assessment quiz.

Question No. 5 **Correct** **Score 1.00** **out of 1.00**	**FIND THE EQUATION OF THE TANGENT LINE TO THE FUNCTION** ***F(X)* = $-x^3-3\cdot x^2-3\cdot x-1$ AT *X* = 0.** **Write the answer in slope-intercept form $m\cdot x+n$, where m is the slope of the line and n is the y-intercept.**

Fig. 1 Example of a short answer question in a Practice Test, including the score

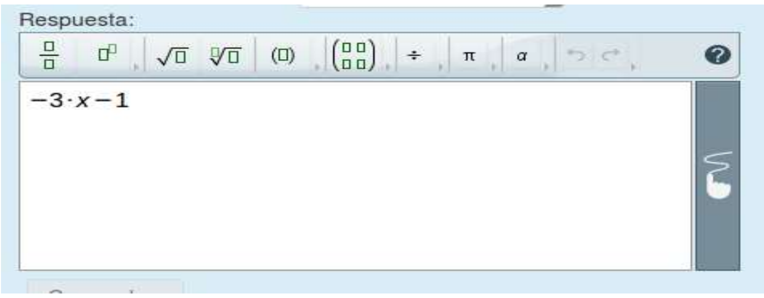
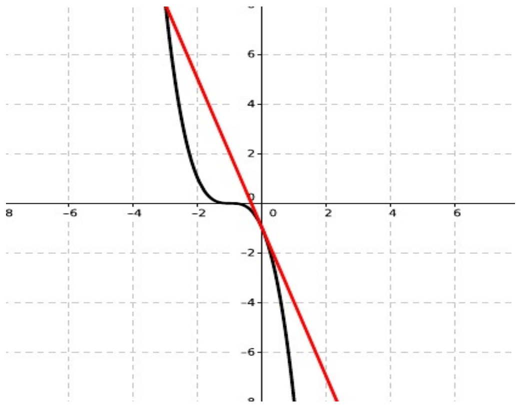


Fig. 2 Example of response introduced using the Wiris program.

Very well.
The next graph represents the function in black, and the tangent line in red



Recall that the slope of the tangent line at x=0 is equal to f'(0) and the equation of tangent line is y-f(0) = f'(0) (x-0). Simplify the expression to get the slope-intercept form.

The correct answer is y = 3x-1

Fig. 3 Example of correct answer and aid in understanding

3.4 Method

The methodological approach of this study is quantitative. It is an ex-post facto retrospective design since submitting practice tests is a voluntary decision of each student, on which the authors of the research have not had any control. The data come from the Moodle activities register of the completed subject, which mainly contains the marks of practice tests and continuous assessment quizzes. As for the type of analysis, a trimmed means test was performed.

3.4.1 Participants

The 131 students in the mathematical analysis subject of the first semester of the academic year 2015-16 in computer engineering degree at the UOC participated in this study. There is no sample design, since the study includes the entire student population that submit continuous assessment quizzes.

3.4.2 Instruments

The qualifications of the practice tests and the continuous assessment quizzes constitute the main data source of the present study. The Moodle activity register of the subject mathematical analysis provides the quizzes' overall qualifications and individual question qualifications as well.

3.4.3 Procedure

In order to maintain the confidentiality of the data, personal references have been conveniently encrypted. In addition, a filtering of the data with the program Libreoffice Calc has been carried out, so that each continuous assessment quiz keeps the code of each student and its qualifications (per question and overall). The information stored from the practice tests is the student code and the total grade of each submission. The set of 17 quizzes, including continuous assessment quizzes and practice tests, on which the filtering described previously has been done, is converted into 17 files that are the basis for later statistical treatment.

3.4.4 Data pre-processing

According to [5], for each continuous assessment quiz we separate the students into two groups according to their activity level: active students and non-active students. We consider active student in a continuous assessment quiz a student that has behaved in an active way, submitting the practice tests associated with that quiz, evidenced in the Moodle register of the subject. Each student decides to belong to the active or non-active group that may be different depending on the quiz. In the case of continuous assessment quizzes with a single model of associated practice tests, active behaviour is proved by having submitted at least one test. In the case of the continuous assessment quizzes with more than one model of the associated practice test, active behaviour is proved by having submitted at least one practice test from each model of practice tests associated with its continuous assessment quiz. In the first five continuous assessment quizzes, there are three dependent variables studied: the overall qualification, the automatic correction question qualification, and the reasoned exposure question qualification. In the last two continuous assessment quizzes, the dependent variable studied is only the overall qualification.

3.4.5 Data analysis

The box and whisker plots of the variables defined for each the seven continuous assessment quizzes show asymmetry, indicating the distribution is skewed and so, non-normal. A good example is shown in Fig.4.

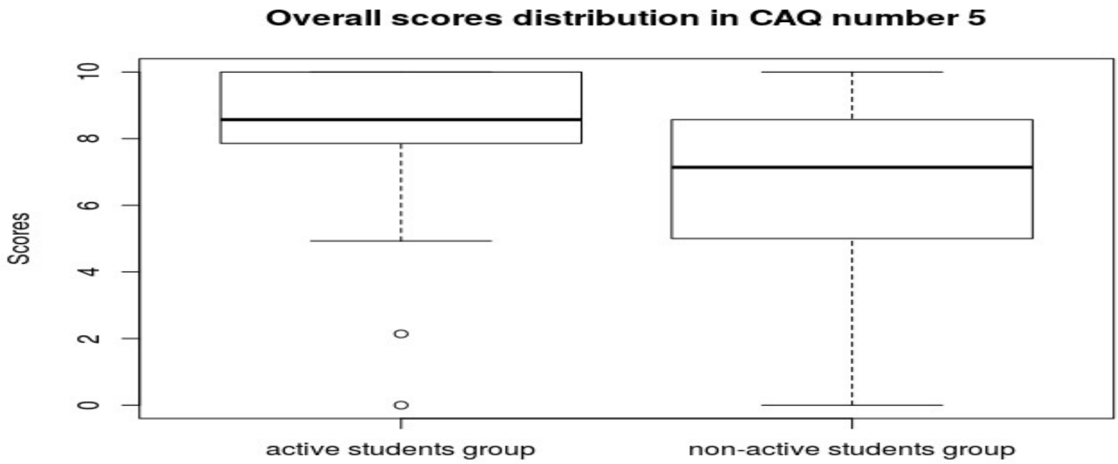


Fig. 4 Box and whisker plot of the overall score distribution in continuous assessment quiz num. 5.

Applying a Brown-Forsythe test we cannot assume the homogeneity of variances between the two groups, quite frequent in studies in the field of education. In cases where the distribution of the data is skewed, the mean may not measure appropriately the central tendency given its sensitiveness to extreme values. In contrast to the mean, the 25 percent trimmed mean, which removes the 25 percent of the largest and the smallest values, is able to reduce the effect of extreme values and better measure the central tendency. The use of trimmed means to make comparisons between groups, and in particular the Yuen's test, is suggested by [12], in order to achieve robustness, given the existence of non-normality and heteroscedasticity. Yuen's test is a modified t-test based on trimmed means and it is pertinent to use it in this study because the data, as mentioned earlier, show non-normality and heteroscedasticity.

Several authors cited by [12] consider different percentages of trimming that go from 10% to 25%, for both upper and lower values. In our study we will consider a 25% trimming. In this way, we are calculating the average of the data that constitutes the interquartile range, that is, the average of 50% of the central qualifications. For the Yuen's test we set a confidence level of 95%. The null hypothesis considers that the trimmed means are the same for the active students group as for the non-active students group, and the alternative hypothesis, that the trimmed means are different for the two groups. The calculations were performed with the R program using the yuen() function, which performs the Yuen's test to determine the existence of significant differences between two groups; and the function yuen.effect.ci(), which provides the explanatory measure of effect size, ξ [13]. Although, since it is not experimental nor quasi-experimental design, and therefore no cause-effect relationship can be established, it seems opportune to show the size of the effect as complementary information.

3.5 Results

In general, for each continuous assessment quiz, active students score better on the overall grade and on the score of the automatic correction questions, but not on the reasoned exposure questions.

A. Active students obtain better overall grades.

The trimmed mean of the active students' overall qualifications is significantly higher, according to the results shown in Table I, in four of the seven continuous assessment quizzes, since the p-value is less than 0.05. Thus, we can say that the overall scores of the active students are higher by more than 50% (4 out of 7) of the continuous assessment quizzes. In three of the quizzes, differences, always greater than 1 point out of 10, and in number 6 greater than 2 points, are observed. The size of the effect in all cases shows values close to 0.5, which is considered as high on a low, medium, high scale

[13]. It is also observed that it is in the extreme quizzes (numbers 1, 2 and 7) that there are no significant differences in the overall score.

Table I. Yuen's test results (Overall scores in CAQs)

Continuous Assessment Quiz (CAQ)	Differences in 25% Trimmed Means (A)	p-value	Effect size
No. 1	+0.30	0.2721	-------
No. 2	+0.29	0.4282	-------
No. 3	+1.64 (0.92 ;2.35)	0.0000*	0.53
No. 4	+0.84 (0.03 ; 1.64)	0.0430	0.44
No. 5	+1.95 (0.98 ; 2.92)	0.0003	0.56
No. 6	+2.08 (0.63 ; 3.53)	0.0057	0.53
No. 7	+0.58	0.1690	-------

* p-value < 0.0001. Alternative hypothesis: true difference in trimmed means is not equal to 0.

(A) Differences between active student group and non-active student group. 95% confidence interval given in parenthesis

B. Active students perform better on automatically-corrected questions.

Only in those continuous assessment quizzes where the trimmed mean of active students' overall qualification is better, is the trimmed mean of the active students' automatically-corrected questions better. This continues to show that in more than 50% of the quizzes (3 out of 5) the scores are better for active students. Table II shows that in Quizzes 3, 4 and 5 there is a difference in the score of the

Table II Yuen's test results (Automatic correction questions scores in CAQs)

Continuous Assessment Quiz (CAQ)	Differences in 25% Trimmed Means (A)	p-value	Effect size
No. 1	+0.21	0.5124	-------
No. 2	+0.17	0.6482	-------
No. 3	+2.11 (1.14 ;3.07)	0.0000*	0.66
No. 4	+1.30 (0.55 ; 2.05)	0.0011	0.49
No. 5	+2.72 (1.53 ; 3.92)	0.0000*	0.61

* p-value < 0.0001. Alternative hypothesis: true difference in trimmed means is not equal to 0.

(A) Differences between active student group and non-active student group. 95% confidence interval given in parenthesis automatic correction questions, as shown in Table I for the overall score. In all three cases the average increase in the automatic correction questions score is higher than the one shown in the overall score. The most remarkable increase corresponds to number 5, with a difference of 2.72 points out of 10 (Table II) in the automatic correction questions score, which also showed a maximum difference of 2.08 points in the overall score (Table I). In terms of the size of the effect, the values are somewhat higher than those obtained in the global scores, so they remain equally high.

C. Active students do not perform better on reasoned exposition questions.

In none of the five continuous assessment quizzes with a reasoned exposition questions are differences between active and non-active students observed (Table III). In three quizzes the difference between the trimmed means does not exceed 0.15 points out of 10, and even in number 4 the difference is 0. This unanimity in the quizzes results contrasts with what was shown in the previous sections.

Table III. Yuen's test results (Reasoned exposition questions scores in CAQs)

Continuous Assessment Quiz (CAQ)	Differences in 25% Trimmed Means (A)	p-value	Effect size
No. 1	+0.14	0.3096	-------
No. 2	+0.75	0.3017	-------
No. 3	+0.31	0.4656	-------
No. 4	0	-----	-------
No. 5	+0.07	0.8815	-------

* p-value < 0.0001. Alternative hypothesis: true difference in trimmed means is not equal to 0.

(A) Differences between active student group and non-active student group. 95% confidence interval given in parenthesis

D. The number of active students is significantly reduced at the end of the semester.

Table IV shows that the number of students submitting all models of practice tests, active students, is higher than those that are not active students in the continuous assessment quizzes numbers 2, 4 and 5, those with a single model of practice tests and of weekly frequency. In numbers 1 and 3, with more than one model and also weekly frequency, the dominance is reversed between active and non active students, the number of active students becoming less than non-active students. In the last two continuous assessment quizzes the ratio of non-active students to active students is close to 3:1

Table IV. Number of active and non-active students

Continuous assessment quiz (CAQ)	Active students	Non-active students
No. 1	44	69
No. 2	71	43
No. 3	53	61
No. 4	64	42
No. 5	65	46
No. 6	27	74
No. 7	25	69

3.6 Discussion

The proportion of students that submit practice tests, PT, calculated on the total number of enrolled students, corresponding to the first five continuous assessment quizzes, CAQ, shows values between 49% and 54%, in tests with a single model, and between 34% and 41%, in tests with multiple models. These last values correspond to students who have submitted at least one practice test of each model. All these figures ranged between the participation of 33% in the study presented by [10] and the participation of 68% in reference [11]. Nevertheless, in the last two CAQs, the percentage drops to 20%.

A number of possible explanations for the high participation rate in the first five CAQ can be supplied. First, the level of knowledge, basic as indicated by the teaching plan, helps the student to feel confidence in responding to the questions posed to him/her. Second, the fact that the frequency of CAQs is weekly makes it help the student to be regular in his/her dedication. Third, given that the student has two days to respond to the CAQ, s/he prefers to practice with the PTs and to get feedback, which allows him/her to reduce the gap between his/her level of knowledge and that required to pass the CAQ, thus increasing the chance to succeed on the CAQ, which ultimately has a direct impact on the final grade. The participation rate decreases in the cases of multiple models of PT associated with CAQ (numbers 1 and 3). This may be due to the increase in weekly workload that entails submitting more than one type of practice test.

The decline in participation rate in the last two continuous assessment questionnaires could have several explanations. First, the advanced level of content and/or the difficulty of the questions. There is a very significant increase in the number of students who do not submit the PT, despite having downloaded it, which may be due to the difficulty students face in solving them. Second, the increase in the time allowed for the completion of the CAQ, which can be up to two weeks, may cause some students to consider that it is sufficient time to prepare and revise the CAQ well, and prefer to focus only on the completion of this type of quizzes, since they are the only ones that have a direct effect on the final qualification in the subject. In studies [5] and [9], in the final grade of the subject the online quizzes have been weighted with a percentage of 10% and 25%, respectively. Finally, it is also possible that students do not sufficiently counterbalance the benefit of the possible feedback they would receive from the cost of submitting PTs.

From the point of view of the possible benefits that the active students obtain in submitting PTs, we can say that the feedback that they receive brings knowledge of possible similar questions in content and typology. This could impact on the observed association between overall CAQ scores and associated PT submission. Reference [9] proves that in the final exam questions, the content of which had previously been covered in the online quizzes, higher scores are obtained. The fact that the results in terms of the association of global qualifications extend to that of association of qualifications of automatic correction questions may be due to the typology of the PT, constituted solely by that type of questions. The latter may also be the reason why there is no association between the performance of PT and the qualifications of the questions of reasoned response. Thus, as the PTs are planned, their completion does not seem to help to establish differences among students in their reasoned exposition of mathematical contents.

Other alternative explanations in relation to conducting practice tests could be that the active students come from a previous level of prior knowledge and/or that the motivation is a differential element with respect to the rest of the students. Reference [5], however, discards motivation as an explanatory factor. Given that there is a difference according to the type of questions (automatic correction and reasoned exposition) we believe that the observed differences in the overall scores between active and non-active students cannot be attributed exclusively to the level of previous studies. In one way or another, the realization of the PT influences the existence of these differences.

In the CAQs at the beginning and end of the course, the same associations are not established as in the CAQ in the middle. It could be that in the latter, students with good previous results are already thinking more about focusing efforts on evaluation and giving up practice. With regard to the second assessment quiz, it might be that the similarity between the questions in its associated PT is less than in the rest, that the questions have a higher level of difficulty and/or that the feedback provided has not been as useful as in other PTs. Finally, CAQ no. 1 has three PT models to be submitted in a single week, which makes interactions between them and overall scores more complex.

3.7 Conclusions

According to the aim of the present study, we can conclude, in a first approximation, that the learning acquired by the students that submit Wiris-Quizzes practice test is higher than that by the students who choose not to submit them. Our conclusion establishes an association, statistically significant by means of the Yuen's test, between continuous assessment questionnaire scores as an indicator of learning and submitting of practice tests. This same relation coincides with that presented by [11] in an in-class environment and only with a practice test and a continuous assessment test. In our case, it has been verified in four of the seven continuous assessment quizzes. With final exam scores as a learning indicator, [5] and [9] in an in-class setting, complemented by the use of online quizzes, it is concluded that there are statistically significant differences between the learning of students that submit practice tests and those who do not. The improvement of the learning is verified in the questions of short answer or multiple choice, to which corresponds an automatic correction. In contrast, the specific analysis of the reasoned exposition question of the continuous assessment quizzes indicates that the submission of practice tests does not result in a significant difference in the marking of this type of question.

The practice tests are recommended because it allows one to maintain regularity in work, and above all because it affords one feedback that facilitates the process of self-regulation, of great importance in online studies. On the contrary, it is not enough to improve the ability of expressing mathematical reasoning.

The present study contributes to extend results obtained in previous studies in face-to-face environments to fully online environments in the field of higher mathematics education. In addition, it uses the comparison between groups by contrasting trimmed means as an analytical strategy of the data. The retrospective ex-post facto design carries with it limitations in the conclusions of cause-effect relationships, which can be corrected to some extent whenever mechanisms are introduced that control the confounding variables, such as establishing homogeneous subgroups [12]. In this sense, and with the purpose of deepening the problem, new variables could be introduced that include emotional aspects, such as motivation, or variables that inform the effective use of feedback received, or that establish levels of difficulty of the quizzes. In addition, it would be of interest to extend the comparison of results to the grades of the final exams by grouping the students according to their behavior on the practice tests and/or continuous assessment quizzes. It would also be interesting to move towards correlational designs to establish closer associations between variables such as those proposed by [7] and towards quasi-experimental designs that allow the definition of equivalent groups in terms of knowledge of the subject, in order that the differences found be due exclusively to the intervention: the conduct of practice tests. In our case, we cannot ensure equivalence between the two groups of students, since they have been self-selected. Finally, the relevance of the analysis presented here as evidence of the effectiveness of the continuous practice in the learning of mathematics through an automatic assessment tool should be highlighted. This could be a key element in new proposals of mixed teaching methodologies that include face-to-face with online environments.

Acknowledgements

This research has been funded by the Spanish government through "ICT-FLAG" (Enhancing ICT education through Formative assessment Learning assessment, Learning Analytics and Gamification) project (TIN2013-45303-P).

References

[1] T. Sancho-Vinuesa and N. Escudero-Viladoms, "¿Por qué una propuesta de evaluación formativa con feedback automático en una asignatura de matemáticas en línea?," *Revista Universidad y Sociedad del Conocimiento*, vol. 9, nº. 2, pp. 59–79, 2012.

[2] R. Calm, R. Masià, C. Olivé, N. Parés, F. Pozo, J. Ripoll, and T. Sancho-Vinuesa, "Wiris Quizzes: un sistema de evaluación continua con feedback automático para el aprendizaje de matemáticas en línea," *Teoría de la Educación; Educación y Cultura en la Sociedad de la Información*, vol 14, nº 14, pp 452-472, 2013

[3] J. Hattie and H. Timperley, The power of feedback," *Review of Educational Research*, vol. 77, nº. 1, pp. 81-112, 2007.

[4] M. Yorke, "Formative assessment in higher education: moves towards theory and the enhancement of pedagogic practice," *Higher Education*, vol. 45, nº. 4, pp. 477–501, 2003.

[5] R. Huisman and H. E. Reedijk, "The impact of individual online tests in addition to group assignments on student learning," *Icicte 2012*, pp. 1–16, 2012.

[6] B. M. Klecker, "The impact of formative feedback on student learning in an online classroom," *Journal of Instructional Psychology*, vol. 34, no. 3, pp. 161–165, 2007.

[7] S. D. Angus and J. Watson, "Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set," *British Journal of Educational Technology*, vol. 40, nº. 2, pp. 255–272, 2009.

[8] E. Fitkov-Norris and B. Lees, "Online formative assessment: Does it add up to better performance in Quantitative modules?" in *Proceedings of the 11th ECRM*, 2012, pp. 115–121

[9] L. Chirwa, "A case study on the impact of automated assessment in engineering mathematics," *Engineering Education.*, vol. 3, nº. 1, pp. 13–20, 2008.

[10] S. Voelkel, "Combining the formative with the summative: The development of a two stage online test to encourage engagement and provide personal feedback in large classes," *Research in Learning Technology*, vol. 21, nº. 1063519, pp. 1–18, 2013.

[11] L. Sly, "Practice tests as formative assessment improve student performance on Computer Managed Learning Assessments," *Assessment & Evaluation in Higher Education*, vol. 24, no. 3, pp. 339–343, 1999.

[12] H.J. Keselman, A.R. Othman, R.R. Wilcox, and K. Fradette, "The new and improved two-sample T test", Psychological Science, vol 15, no. 1, pp. 47-51, 2004.

[13] R.R. Wilcox and T.S. Tian, "Measuring effect size: a robust heteroscedastic approach for two or more groups", *Journal of Applied Statistics*, vol 38, no. 7, pp. 1359-1368, 2011.

[14] Brewer, Khun, and N. J. Salkind, "Encyclopedia of Research Design Volume 1," in *SAGE Publications*, 2010, pp. 124–130

CAPÍTOL 4

INVESTIGATING THE RELATIONSHIP BETWEEN OPTIONAL QUIZZES AND FINAL EXAM PERFORMANCE IN A FULLY ASYNCHRONOUS ONLINE CALCULUS MODULE[6]

Most teachers of mathematics think that regular practice is essential for success. In face-to-face instruction settings, regular practice requires doing homework, which has to provide students with feedback in order to be useful. Online homework allows teachers to assume the workload involved in providing feedback to a large number of students enrolled in higher education. Several studies have established an association between completing online homework quizzes and learning achievement in face-to-face environments. In a fully asynchronous online environment, where students have to work autonomously, offering them a set of optional interactive quizzes with automated feedback may be a good teaching strategy to support them in their learning process. Within this framework, this paper investigates the relationship between optional quizzes and final exam performance in a calculus module. By means of multiple linear regression, this study found that an association exists between the learning achievement measured by final exam marks and participation in online quizzes together with the learning achieved through completion of these quizzes.

[6] En aquest capítol, que **conté el segon article del conjunt que conformen la tesi**, les referències numèriques a taules corresponen a referències exclusivament d'aquest capítol. Les referències bibliogràfiques d'aquest capítol es mostren dins d'aquest mateix capítol.

Keywords: Learning achievement; online mathematics; online homework quizzes; automated feedback; e-assessment; asynchronous online environment.

4.1 Introduction

Mathematics is a "challenging discipline to teach" (Trenholm, Alcock, & Robinson, 2015, p. 1198) in a fully asynchronous online environment. In face-to-face environments, the study of learning mathematics has been, and still is, a great challenge because of its abstract content, its specific language and logic and its high level of formalism and rigour. It becomes still more challenging in a fully asynchronous environment, in which all student–teacher and student–student interactions take place in different times and spaces. For students taking the one-semester calculus module as part of the Computer Engineering degree at the Open University of Catalonia (UOC), this challenge is even greater, as most of the students are adults with demanding family commitments and heavy workloads (Calm et al., 2013). Engelbrecht i Harding (2005) consider that no well-defined, consensual teaching methodology exists in the field of online mathematics. However, it is widely accepted by mathematics teachers that continuous practice is essential for success in the subject, mainly in terms of acquiring procedural skills. Motivated by improving students' learning and academic results, engineering mathematics teachers at the UOC implemented a teaching methodology based on the aforementioned consensus (Sancho-Vinuesa & Escudero-Viladoms, 2012). At the core of this methodology is the completion of periodic interactive online Wiris-quizzes, which are optional for students. The most remarkable features of the Wiris-quizzes are: (a) self-assessment and automatic response, (b) immediate feedback, (c) parameterised questions; and (d) numerical and algebraic recognition of answers. We decided that the teaching methodology needed to be examined to determine its usefulness. The present study is guided by the research question: does the teaching methodology based on the completion of optional quizzes support students in their learning process when taking online mathematics modules in a fully asynchronous environment?

Doing homework offers students the opportunity to practice mathematical skills (Leong & Alexander, 2014; Palocsay & Stevens, 2008; Roschelle, Feng, Murphy, & Mason, 2016). Cooper (1989) considers that homework is the set of tasks assigned to students by school teachers that are meant to be carried out during non-school hours. In literature we find several studies (Angus &

Watson, 2009; Chirwa, 2008; Fitkov-Norris & Lees, 2012; Hannah, James, & Williams, 2014; Huisman & Reedijk, 2012) on taking online quizzes that are designed to be completed during non-school hours and are assigned by teachers. Accordingly, as per Cooper's (1989) interpretation, these online quizzes can be considered examples of homework – specifically, as online homework. We therefore consider that the online homework research field includes research into online quizzes and, thus, both research fields are relevant to our study.

For practice by means of online quizzes to be really useful to students, feedback has to be provided (Butler, Pyzdrowski, Goodykoontz, & Walker, 2008) and, if possible, given immediately (Palocsay & Stevens, 2008). Immediate feedback is beneficial, above all, in the sense that it enhances procedural skills, whereas delayed feedback is better suited for the acquisition of conceptual knowledge (Shute, 2008). Every task that provides information to help students in their learning process belongs to the category of formative assessment instruments (Yorke, 2003). Taking into account that feedback provided after completing online quizzes aims to help students in their learning process, online quizzes followed by feedback are considered, in all cases, as formative assessment instruments. Regardless of this, whenever the online quiz score is included in the final mark, this online quiz is considered as a summative assessment instrument.

Several studies (Johnson & Mckenzie, 2013; Wooten, 2013; Zerr, 2007) have analysed the usefulness of requiring students to complete online homework, by comparing the learning outcomes of students whose online homework is a summative assessment instrument with those whose online homework is not. Asan example, in Zerr's (2007) study, the online homework score is included as 10% of the final mark. Disparate results were obtained. Zerr (2007) did find evidence that students who were required to do online homework achieved better learning outcomes than students who were not. Wooten (2013), on the other hand, did not find significantly better outcomes.

According to studies with an experimental design (Roschelle et al., 2016), a quasi-experimental design (Babaali & Gonzalez, 2015) and designs in which student groups have been made by self-selection (Chirwa, 2008; Figueroa-Cañas & Sancho-Vinuesa, 2017; Huisman & Reedijk, 2012), the learning outcomes of students who complete online quizzes are significantly higher than of those who do not complete these quizzes. Although correlational designs do not allow researchers to establish strict cause-and-effect relationships in the same way that experimental designs do, these can be acceptable alternatives, especially in the field of education, if the variables are appropriately controlled.

The final exam mark is a variable that is widely used as a measure of learning achievement in studies with a correlational design (Angus & Watson, 2009; Fitkov-Norris & Lees, 2012; Hannah et al., 2014; Johnson & Mckenzie, 2013; Palocsay & Stevens, 2008). However, Trautwein and Köller (2003) pointed out that the measurement can be influenced by the teacher's marking style, and is therefore not reliable enough. The studies of Fitkov-Norris and Lees (2012), Hannah et al. (2014) and Johnson and Mckenzie (2013) provide evidence that a statistical association or relationship exists between the online quiz scores attained by students and the final exam marks they are awarded. Different measures of participation in online quizzes are also analysed in correlational designs (Angus & Watson, 2009; Fitkov-Norris & Lees, 2012; Hannah et al., 2014). Those studies have established an association between participation in online quizzes and final exam marks. Finally, using the students' prior mathematical competence as a control variable ensures that the significant relationship found between completing online quizzes and final exam performance is only attributable to the online quizzes (Angus & Watson, 2009). Johnson and Mckenzie (2013) and Palocsay and Stevens (2008) have also included this control variable.

In light of the previous references concerning the relationship between completing online quizzes and exam performance, and taking into account our context of learning online mathematics in a fully asynchronous environment, it seems reasonable to state the following hypotheses:

- Hypothesis 1: A statistical association or relationship exists between the average score obtained in optional online quizzes and final exam marks.
- Hypothesis 2: A statistical association or relationship exists between the participation in optional online quizzes and final exam marks.

The aim of the present study is to confirm or reject these hypotheses within our context. To do so, we have implemented a correlational design by means of multiple linear regression. As far as we know, no studies have been conducted to examine the relationship between completing optional online quizzes and learning achievement in a fully asynchronous online environment.

4.2 Methodology

4.2.1 Participants and learning context

The participants in this study were students enrolled in the first semester of the 2016/17 calculus module forming part of the Computer Engineering degree at the Open University of Catalonia. The teaching plan for this calculus module allowed students to complete two different types of optional quizzes: a trial test and a real test, which are similar to the tests used by (Croft, Danson, Dawson, & Ward, 2001). Trial tests involved students completing exercises and acquiring procedural skills. They contained both multiple-choice and short fill-in-the-blank questions. An unlimited number of attempts and time were allowed. After each attempt, immediate and automated feedback was provided. Ten trial test templates were assigned throughout the semester, the scores of which were not included when calculating the final mark. Real tests differed from trial tests mainly because in real tests: (a) only one attempt was allowed, (b) time to answer the questions ranged from one to two weeks, (c) seven templates were assigned, (d) the scores attained could be included in the final mark, and (e) a constructed-response question was included. The constructed-response question had to be answered by giving a detailed explanation, which required students to show in-depth conceptual knowledge and adequate written mathematical communication as well as procedural skills. Moreover, it required manual teacher correction, and feedback was therefore delayed by a week. Since feedback was provided after completing both types of optional quiz, these had to be considered as formative assessment instruments.

The module included two assessment instruments: (a) a compulsory in-person final exam, and (b) a non-compulsory and essentially formative online continuous assessment throughout the semester. The final mark for the module was mainly based on the final exam mark, which could be moderately modified by the continuous assessment mark. The real test scores formed part of the continuous assessment mark.

In addition, during the first week teachers assigned an initial test to ascertain the students' prior knowledge of secondary-education calculus (including topics such as derivatives, primitives, etc.). In order to encourage participation, students who voluntarily completed and submitted the test obtained a bonus, which also formed part of the continuous assessment mark.

Now we can be more precise. The participants were the 176 calculus students who sat the final exam and submitted, at least, one real test. This group of students will be referred to hereafter as the "whole_group". There were 115 students inside the whole_group who also submitted the initial test. Hereafter, it will be referred to as the "initial_test_group", which represented 65% of the whole_group.

4.2.2 Measure and data collection

The unit of analysis used in this study is the student. Teacher records were the source of information for collecting the final exam marks attained by students, and also the mean of the real test marks. The Final_Exam_Mark variable, ranging from 0 to 10, has been considered the variable that measures the total learning achievement for the module, as was the case in the studies by Angus and Watson (2009), Fitkov-Norris and Lees (2012) and Hannah et al. (2014). The Mean_Real_Test_Marks variable, also ranging from 0 to 10, measures the learning achieved by doing real tests.

Participation in online quizzes, as mentioned in hypothesis 2, has been interpreted as a completion rate, specifically, the proportion of the number of questions answered by the student out of the total number of questions sent by teachers. This measure fine-tuned other participation measures (Angus & Watson, 2009; Fitkov-Norris & Lees, 2012; Hannah et al., 2014). We have considered two separate completion rate variables: Real_Test_Completion_Rate, and Trial_Test_Completion_Rate, each one ranging from 0 to 1. The Moodle activity log was the source of information for collecting the required data to calculate both completion rates and the three control variables that follow. The first control variable was the language chosen by the student to communicate with his/her teacher. The dummy variable Language was included to prevent the confounding effects of using one language or another in the student–teacher communication. So far, all variables are common for both the whole_group and the initial_test_group. The second control variable was only taken into account for the whole_group: the dummy variable Initial_Test_Attempted, which indicated whether a student had submitted the initial test or not and, therefore, whether the student belonged to the initial_test_group or not. It was desirable to include the variable to avoid possible bias since participation in the initial test – the first test – may indicate a high degree of participation in other online quizzes. The third control variable is only available for the initial_test_group: the Initial_Test_Mark, ranging from 0 to 10. Since the Initial_Test_Mark can be interpreted as a proxy for prior mathematical competence, the confounding effects of previous mathematical knowledge has been controlled, as previously considered by Angus and Watson (2009), Fitkov-Norris and Lees (2012) and Hannah et al. (2014).

Data used for the present study are split into two data sets ([Dataset1] and [Dataset2]). [Dataset1] contains data concerning the whole_group, whereas [Dataset2] is related to the initial_test_group.

4.2.3 Statistical analysis

Two main statistical techniques have been applied in this study: (1) a two-paired means test, and (2) multiple linear regression. Regarding the first technique, we carried out a Welch test to contrast whether the mean of the Initial_Test_Mark was the same as the mean of the Final_Exam_Mark. As for the second technique, multiple linear regression allows a relationship or association to be established between one dependent variable and one or more independent variables (Lee-Thomas, Kaw, & Yalcin, 2011). One model including the dependent and independent variables has been proposed for the whole_group and a different model for the initial_test_group, both of which are detailed below. The existence of a statistical association between the dependent and one of the independent variables will be accepted if two conditions are complied with: (a) the overall model is statistically significant[7], and (b) the regression coefficient between the two variables is statistically significant. In order to prevent errors in estimating the regression coefficients, the assumptions of normality, homoscedasticity and no autocorrelation must be complied with (Fahrmeir, Kneib, Lang, & Marx, 2013).

[7] With the expression "statistically significant" we mean that, regarding the test to which the expression is referring, we reject the test null hypothesis (H_0), considering the significance level (α). The criterion for rejecting H_0 has been the obtaining of a p-value that is lower than the significance level (α).

4.2.3.1 Model for the whole_group

The first proposed model for the whole_group is expressed as (1)

$$\text{Final\_Exam\_Mark}_i = b_0 + b_1\cdot\text{Mean\_Real\_Test\_Marks}_i + b_2\cdot\text{Real\_Test\_Completion\_Rate}_i + b_3\cdot\text{Trial\_Test\_Completion\_Rate}_i + b_4\cdot\text{Language}_i + b_5\cdot\text{Initial\_Test\_Attempted}_i + e_i \quad (1)$$

whereby e_i is the error term.

We transformed the Final_Exam_Mark variable into the Transformed_Final_Exam_Mark variable (see Appendix A.1), in order to comply with the normality assumption (Angus & Watson, 2009). This transformation is monotonous, that is, increasing intervals for the Final_Exam_Mark corresponds to increasing intervals for the Transformed_Final_Exam_Mark, and vice versa. Therefore, the statistically significant association between the Transformed_Final_Exam_Mark and any independent variable can be extrapolated to the Final_Exam_Mark and the same independent variable; moreover, the association sign is maintained. Later, collinearity between the two variables related to real tests, Mean_Real_Test_Marks_01[8] and Real_Test_Completion_Rate was detected, and the latter was eliminated from the model. Finally, using a stepwise method, the best fit model (see Appendix A.2 for a complementary explanation) for the whole_group is expressed as (2):

$$\text{Transformed\_Final\_Exam\_Mark}_i = c_0 + c_1\cdot\text{Mean\_Real\_Test\_Marks\_01}_i + c_3\cdot(\text{Trial\_Test\_Completion\_Rate}_i)^2 + c_4\cdot\text{Language}_i + c_5\cdot\text{Initial\_Test\_Attempted}_i + e_i \quad (2)$$

4.2.3.2 Model for the initial_test_group

The first proposed model for the initial_test_group is expressed as (3):

[8] Mean_Real_Test_Marks_01 is the former Mean_Real_Test_Marks rescaled in the interval [0,1].

$$\text{Final\_Exam\_Mark}_i = b_0 + b_1 \cdot \text{Mean\_Real\_Test\_Marks01}_i + b_2 \cdot \text{Real\_Test\_Completion\_Rate}_i + b_3 \cdot \text{Trial\_Test\_Completion\_Rate}_i + b_4 \cdot \text{Language}_i + b_5 \cdot \text{Initial\_Test\_Mark}_i + e_i \qquad (3)$$

Following a procedure that is equivalent to the one described for the whole_group (see Appendix A.3), the model for the initial_test_group is expressed as (4):

$$\text{Transformed\_Final\_Exam\_Mark}_i = c_0 + c_1 \cdot \text{Mean\_Real\_Test\_Marks\_01}_i + c_3 \cdot (\text{Trial\_Test\_Completion\_Rate}_i)^4 + c_4 \cdot \text{Language}_i + c_5 \cdot \text{Initial\_Test\_Mark\_01}_i + e_i \qquad (4)$$

whereby Initial_Test_Mark_01 is the former Initial_Test_Mark rescaled in the interval [0,1].
All the statistical calculations have been carried out using the R program, version 3.4.2.

4.3 Results

In the second part of this section, we present the results obtained using the techniques described in the methodology section (subsection 4.2.3). First, we highlight some results from the statistical summary for the whole_group and for the initial_test_group, shown in Table 1 and Table 2, respectively.

Table 1 Whole_group variable statistic summary

	Final_ Exam_ Mark	**Mean_ Real_ Test_ Marks**	**Real_Test_ Completion_ Rate**	**Trial_Test_ Completion_ Rate**	**Language**	**Initial_ Test_ Attempted**
Mean	5.5	8.6	0.95	0.43	0.58	0.65
Standard deviation	2.2	1.9	0.11	0.35	---	---
Minimum	0.8	1.0	0.43	0.00	---	---
Maximum	10.0	10.0	1.00	1.00	---	---

Table 2 Inital_test_group variable statistic summary

	Final_ Exam_ Mark	Mean_ Real_ Test_ Marks	Real_Test_ Completion_ Rate	Trial_Test_ Completion_ Rate	Language	Initial_ Test_ Mark
Mean	5.8	9.0	0.97	0.52	0.56	6.0
Standard deviation	2.0	1.6	0.08	0.33	---	2.5
Minimum	0.8	1.0	0.43	0.00	---	0.6
Maximum	10.0	10.0	1.00	1.00	---	10.0

When comparing the variables between the two groups, we observe that there are no remarkable differences (e.g., the mean of the Final_Exam_Mark is 5.5 for the whole_group, and 5.8 for the initial_test_group). When comparing intragroup values, the high values of the Mean_Real_Test_Marks stand out (8.6 out of 10 for the whole_group and 9.0 for the initial_test_group), and the mean of the Real_Test_Completion_Rate is higher than its counterpart for the trial tests (greater than 0.94 for the former and less than 0.53 for the latter, in both groups).

A. Results for the whole_group

The overall model for the whole_group, shown by expression (2), is statistically significant (α = 5%) taking into account the p-value = 1.33×10^{-6} obtained in an F-test. The adjusted R^2 has a value of 0.15. In addition, the model complies, with a significance level of 0.05, with the assumption of normality (p-value = 0.10 in a Shapiro-Wilk test), homoscedasticity (p-value = 0.26 in a Breusch-Pagan test), and no first-order autocorrelation (p-value = 0.85 in a Durbin-Watson test).

As was explained in subsection 4.2.3, the statistically significant association between the Transformed_Final_Exam_Mark and the independent variables is extrapolated to the untransformed variable and the same independent variables. Due to the complex transformation (see Appendix A.1), the magnitude of the regression coefficients is not easily interpretable, although its signs are. A positive sign indicates a direct relationship, so when the independent variable rises, the dependent one also rises. A negative sign indicates an inverse relationship that corresponds to the opposite behaviour. So, despite Table 3 showing the estimated regression coefficients taking the Transformed_Final_Exam_Mark as the dependent variable, the results that follow are concerned with the Final_Exam_Mark as a dependent variable.

Table 3 Regression results for the whole_group

	Mean_ Real_ Test_ Marks_01	**Trial_Test_ Completion_ Rate^2**	**Language**	**Initial_ Test_ Attempted**
Estimated regression coefficient ($\hat{c}_i$)	$\hat{c}_1$=0.58**	$\hat{c}_2$=0.21*	$\hat{c}_3$=-0.10	$\hat{c}_4$=0.10
t-Statistics	3.33	1.99	-1.62	1.49

* p-value < 0.01, ** and p-value < 0.001. Adjusted R^2 = 0.15

First, a statistically significant direct association has been found between the Mean_Real_Test_Marks and the Final_Exam_Mark, since the estimated regression coefficient, $\hat{c}_1$, is positive (0.58) and significant (α=1%). Thus, keeping the rest of variables constant, the higher the Mean_Real_Test_Marks, the higher the Final_Exam_Mark. Second, a statistically significant direct association has also been found between the Trial_Test_Completion_Rate squared and the Final_Exam_Mark because $\hat{c}_2$ is positive (0.21) and significant (α=1%). Thus, keeping the rest of the variables constant, the higher the Trial_Test_Completion_Rate, the higher the Final_Exam_Mark. And third, there are no statistical associations regarding the control variables: Language and Initial_Test_Attempted.

B. Results for the initial_test_group

Although Table 4 shows the estimated regression coefficients taking the Transformed_Final_Exam_Mark as the dependent variable, the results that follow are concerned with the Final_Exam_Mark as a dependent variable.

Table 4 Regression results for the initial_test_group

	Mean_ Real_ Test_ Marks_01	**Trial_Test_ Completion_ Rate[4]**	**Language**	**Initial_ Test_Mark_01**
Estimated regression coefficient ($\hat{c}_i$)	$\hat{c}_1$=0.63**	$\hat{c}_2$=0.24*	$\hat{c}_3$=0.12	$\hat{c}_4$=0.46**
t-Statistics	-2.71	-1.99	1.65	-3.25

* p-value < 0.01, ** and p-value < 0.001. Adjusted $R^2 = 0.24$

First, a statistically significant direct association has been found between the Mean_Real_Test_Marks and the Final_Exam_Mark, since the estimated regression coefficient, $\hat{c}_1$, is positive (0.63) and significant (α = 1%). Second, a statistically significant direct association has also been found between the Trial_Test_Completion_Rate to the fourth power and the Final_Exam_Mark because $\hat{c}_2$ is positive (0.24) and significant (α = 1%). Thus, keeping the rest of the variables constant, the higher the Trial_Test_Completion_Rate, the higher the Final_Exam_Mark. Third, a statistically significant direct association has been found between the Initial_Test_Mark and the Final_Exam_Mark, since the estimated regression coefficient, $\hat{c}_3$, is positive (0.46) and significant (α = 1%). Thus, keeping the rest of the variables constant, the higher the Initial_Test_Mark, the higher the Final_Exam_Mark. Fourth, there are no statistical associations regarding the Language control variable.

As for the Welch test technique, no differences have been found between the mean of the Initial_Test_Mark and the mean of the Final_Exam_Mark. The confidence interval for the differences between the two means was (-7.5, 4.5), with a significance level of 0.05. Therefore, we could not reject the null hypothesis according to which the two means are equal.

4.4 Discussion

In the present study, we have investigated the relationship between final exam performance and the completion of online optional quizzes (real and trial tests). Given the results presented, the main findings are: (1) there is a direct association between final exam marks and real test marks (hypothesis

1 confirmed), and (2) there is a direct association between final exam marks and the trial test completion rate, but we could not establish any association concerning the real test completion rate (hypothesis 2 partially confirmed).

The relationship between final exam marks and online quiz marks is in line with the studies of Fitkov-Norris & Lees (2012) and Hannah et al. (2014) in face-to-face environments. Interpreting the relationship in terms of learning, in our fully asynchronous environment, the learning achieved at the end of the module and the learning achieved by completing real tests are statistically associated. It is worth pointing out that students complete quizzes even when it is optional to do so and, moreover, score high marks. Zerr (2007) also found high scores, but in compulsory online quizzes. It seems that students consider completing quizzes to be useful for them, and that they try to make the most of them, conscious that real test marks can help them to improve their final mark and even to pass the module.

In regard to the relationship between final exam marks and the online completion rate, it should be recalled that the completion rate has been used as the participation indicator. The relationship is in line with studies that consider different participation indicators (Angus & Watson, 2009; Fitkov-Norris & Lees, 2012; Hannah et al., 2014). This relationship could only be found in the trial tests due to the elimination of the Real_Test_Completion_Rate to overcome the collinearity with the Mean_Real_Test_Marks. Fitkov-Norris & Lees (2012) also found collinearity between variables, but decided to keep variables since ridge regression was used. It is notable that participation, considered separately from scoring, is associated with the learning attained in the final exam. Participation — the completion rate — in real tests is clearly higher than in trial tests. It seems students prefer to focus on the real test, maybe because the perceived benefit of completing this kind of optional test is higher. The real tests may have the benefit of improving the final mark, as well as the guaranteed feedback, whereas the only benefit of trial tests is that of feedback. It seems that the drawback of being limited to just one attempt per real test is not enough to outweigh the benefits.

So far, discussion has been concerned with all students: the whole_group and the initial_test_group. We will now focus on the latter. Firstly, it has to be highlighted that the association between final exam marks and the completion of online quizzes has been established, taking into account that the initial test marks have been controlled. Therefore, interpreting the initial test marks as a proxy for prior mathematical competence, the association cannot be attributable to the students' previous knowledge. Moreover, the association found between the final exam marks and the initial test marks can also be interpreted as the prior mathematical competence being related to the learning achieved at the end of the module. Angus & Watson (2009), Johnson & Mckenzie (2013) and Palocssay & Stevens (2008) also found that previously acquired skills are related to the final exam mark. The Initial_Test_Mark variable causes the overall model for the initial_test_group to fit better than that of the whole_group, as is shown by its higher value of adjusted R^2. Nevertheless, the fact that adjusted R^2 is 24% shows that other variables are necessary in order for the model to fit better. Leong and Alexander's (2014) paper stated that attitude variables are also related to academic results. Secondly, the average of the initial test mark and the average of the final exam mark are not significantly different. On average, the level of previous knowledge is maintained at the end of the course. It is worth interpreting this positively, considering the fact that the initial test was unsupervised, with no constructed-response questions, while the final exam was supervised and only consisted of constructed-response questions. Maintaining the entry level can be considered as a strong point in the teaching methodology analysed, based on completing optional quizzes, given the intrinsic difficulty of learning online mathematics (Trenholm et al., 2015).

4.5 Conclusion and future research

In summary, the present paper extends the results obtained in previous studies on the efficiency of online homework and online quizzes in face-to-face environments (Angus & Watson, 2009; Fitkov-Norris & Lees, 2012; Hannah et al., 2014) to a fully asynchronous online environment. We can conclude that completing optional quizzes with automatic correction and feedback supports students' attainment of the learning objectives in the fully asynchronous online calculus module given as part of the Computer Engineering degree programme at the Open Universitiy of Catalonia (UOC).

In this study we have performed a quantitative analysis, based on a correlational design, of the learning results of 176 students in a specific module, over one semester. The main limitation of this design is that it does not allow any causation to be inferred. So, even though we found a relationship between the real test marks and the final exam mark, we could not infer that the former caused the latter.

We suggest some possible lines of future research as follows. In order to gain an in-depth understanding of the effectiveness of completing optional online quizzes, new variables and perspectives need to be incorporated into the analysis carried out so far. Moreover, according to Leong and Alexander (2014), the extent to which the use of this tool interacts in some sense with the beliefs, attitudes and emotions of the students towards mathematics should be analysed, as well as whether this leads to a better learning experience. Another area for research is teacher experience (Trenholm et al., 2015) in designing, implementing and using this methodology, and whether this experience can really transform the whole concept of teaching practice. In order to have a more complete vision, it also would be interesting to explore the experience in other mathematical modules with specific features, such as introduction to mathematics for engineering, statistics, probability and stochastic processes.

Acknowledgements

We would like to thank Dr. Ramon. Masià., Dr. Laura. Calvet, Dr. Miquel. Ferrer. and Mr. Paul. Garbutt. for their valuable contributions in helping to improve this study.

References

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. Open Journal of Statistics, 5(7), 754–767.

Angus, S. D., & Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. *British Journal of Educational Technology*, *40*(2), 255–272.

Babaali, P., & Gonzalez, L. (2015). A quantitative analysis of the relationship between an online homework system and student achievement in pre-calculus. *International Journal of Mathematical Education in Science and Technology*, *46*(5), 687–699.

Butler, M., Pyzdrowski, L., Goodykoontz, A., & Walker, V. (2008). The effects of feedback on online quizzes. *International Journal for Technology in Mathematics Education*, *15*(4), 131–136.

Calm, R., Masià, R., Olivé, C., Parés, N., Pozo, F., Ripoll, J., & Sancho-Vinuesa, T. (2013). Wiris Quizzes: un sistema de evaluación continua con feedback automático para el aprendizaje de matemáticas en línea. *Teoría de La Educación. Educación Y Cultura En La Sociedad de La Información*, *14*(2), 452–472.

Chirwa, L. (2008). A case study on the impact of automated assessment in engineering mathematics. *Engineering Education*, *3*(1), 13–20.

Cooper, H. M. (1989). Homework. New York: Longman.

Croft, A., Danson, M., Dawson, B., & Ward, J. (2001). Experiences of using computer assisted assessment in engineering mathematics. *Computers & Education*, *37*(1), 53–66.

[Dataset1] available at: https://figshare.com/s/2be802a9426c11fa28ef

[Dataset2] available at: https://figshare.com/s/5661f314b9990c2d700b

Engelbrecht, J., & Harding, A. (2005). Teaching undergraduate mathematics on the internet. *Educational Studies in Mathematics*, *58*(2), 253–276.

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression: models, methods and applications*. Berlin Heidelberg: Springer-Verlag.

Figueroa-Canas, J., & Sancho-Vinuesa, T. (2017). Exploring the Efficacy of Practicing with Wiris-Quizzes in Online Engineering Mathematics. *Revista Iberoamericana de Tecnologias Del Aprendizaje*, *12*(3), 141–146. doi:10.1109/RITA.2017.2735499

Fitkov-Norris, E., & Lees, B. (2012). Online formative assessment: Does it add up to better performance in Quantitative modules? In *Proceedings of the 11th ECRM* (pp. 115–121).

Hannah, J., James, A., & Williams, P. (2014). Does computer-aided formative assessment improve learning outcomes? *International Journal of Mathematical Education in Science and Technology*, *45*(2), 269–281.

Huisman, R., & Reedijk, H. E. (2012). The impact of online tests in addition to group assignments on student learning. *ICICTE 2012 Proceedings 654*, 654–667.

Johnson, J. A., & Mckenzie, R. (2013). The effect on student performance of web-based learning and homework in microeconomics. *Journal of Economics and Economic Education Research*, *14*(2), 115–126.

Lee-Thomas, G., Kaw, A., & Yalcin, A. (2011). Using Online Endless Quizzes as Graded Homework. In *ASEE Annual Conference & Exposition*.

Leong, K. E., & Alexander, N. (2014). College students attitude and mathematics achievement using web based homework. *Eurasia Journal of Mathematics, Science and Technology Education*, *10*(6), 609–615.

Palocsay, S. W., & Stevens, S. P. (2008). A Study of the Effectiveness of Web-Based Homework in Teaching Undergraduate Business Statistics. *Decision Sciences Journal of Innovation Education*, *6*(2), 213–232.

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online Mathematics Homework Increases Student Achievement. *AERA Open*, *2*(4), 1–12.

Sancho-Vinuesa, T., & Escudero-Viladoms, N. (2012). ¿Por qué una propuesta de evaluación formativa con feedback automático en una asignatura de matemáticas en línea? *Revista de Universidad Y Sociedad Del Conocimiento (RUSC)*, *9*(2), 59–79.

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, *78*(1), 153–189.

Trautwein, U., & Köller, O. (2003). The Relationship Between Homework and Achievement—Still Much of a Mystery. *Educational Psychology Review VO - 15*, *15*(2), 115.

Trenholm, S., Alcock, L., & Robinson, C. (2015). An investigation of assessment and feedback practices in fully asynchronous online undergraduate mathematics courses. *International Journal of Mathematical Education in Science and Technology*, *46*(8), 1197–1221.

Wooten, T. (2013). An Investigation Of Online Homework: Required Or Not Required? *Contemporary Issues In Education Research – Second Quarter*, *6*(2), 189–199.

Yorke, M. (2003). Formative Assessment in Higher Education: Moves Towards Theory and the Enhancement of Pedagogic Practice. *Higher Education*, *45*(4), 477–501.

Zerr, R. (2007). A Quantitative and Qualitative Analysis of the Effectiveness of Online Homework in First-Semester Calculus. *Journal of Computers in Mathematics and Science Teaching*, *26*(1), 55–73.

4.6 Appendix

A. Logistic transformation and best regression model procedure

The variant of the logistic transformation applied to the Final_Exam_Mark variable (A.1), as well as the procedure for treating the collinearity and obtaining the best regression models (A.2 and A.3) are described as follows:

A.1 First, the Final_Exam_Mark variable is rescaled in the interval [0, 1], resulting in the Final_Exam_Mark_01 variable. Then, the latter is transformed into the Transformed_Final_Exam_Mark variable, according to the variant of the logistic transformation (Angus & Watson, 2009). The specific expression is:

$$\text{Transformed\_Final\_Exam\_Mark}_i = \ln[(\text{Final\_Exam\_Mark\_01}_i+0.5) / (1- \text{Final\_Exam\_Mark\_01}_i+0.5)]$$

A.2 First, the Mean_Real_Test_Marks variable was rescaled in the interval [0, 1], resulting in the Mean_Real_Test_Marks_01 variable. Then, the collinearity between variables was checked by analysing the variance inflation factors (VIF). Values greater than 5 indicated high correlation between the variables (Akinwande, Dikko, & Samson, 2015). In our study, we established the cut-off at 4, and decided to eliminate a variable whose VIF exceeded that value. It was found that the only VIFs greater than 4 corresponded to the Mean_Real_Test_Marks_01 and the Real_Test_Completion_Rate variables, with values of 4.76 and 4.72, respectively. The Real_Test_Completion_Rate was then eliminated from the model. After that, no variables exceeded the VIF cut-off 4, so the collinearity levels among variables were assumable. Next, several sets of variables, including polynomial transformations of the Trial_Test_Completion_Rate were checked using the stepwise method, with the purpose of selecting the best set of variables. The final set was the one that lowered the Akaike Information Criterion (AIC) to its minimum value because smaller AIC values corresponded to a better model fit (Fahrmeir, Kneib, Lang, & Marx, 2013).

A.3 First, the Mean_Real_Test_Marks and the Initial_Test_Mark variables were rescaled in the interval [0, 1], resulting in the Mean_Real_Test_Marks_01 and the Initial_Test_Mark_01 variables, respectively. It was found that the only variance inflation factors (VIF) greater than 4 corresponded to the Mean_Real_Test_Marks_01 and the Real_Test_Completion_Rate variables, with values of 4.95 and 4.76, respectively. The Real_Test_Completion_Rate was then eliminated from the model. After that, no variables exceeded the VIF cut-off 4, so the collinearity levels among variables were assumable. Next, several sets of variables, including polynomial transformations of the Trial_Test_Completion_Rate were checked using the stepwise method, with the purpose of selecting the better set of variables. The final set was the one that lowered the Akaike Information Criterion (AIC) to its minimum value.

CAPÍTOL 5

EARLY PREDICTION OF DROPOUT AND FINAL EXAM PERFORMANCE IN AN ONLINE STATISTICS COURSE[9]

Higher education students who either do not complete the courses they have enrolled on or interrupt their studies indefinitely remain a major concern for practitioners and researchers. Within each course, early prediction of student dropout helps teachers to intervene in time to reduce dropout rates. Early prediction of course achievement helps teachers suggest new learning materials aimed at preventing at-risk students from failing or not completing the course. Several machine learning techniques have been used to classify or predict at-risk students, including tree-based methods, which, though not the best performers, are easy to interpret. This study presents two procedures for identifying at-risk students (dropout-prone and non-achievers) early on in an online university statistics course. These enable us to understand how classifiers work. We found that student dropout and course performance prediction was only determined by their performance in the first half of the formative quizzes. Nevertheless, other elements of participation on the virtual campus were initially considered. The classifiers will serve as a reference for intervention, despite their moderate performance metrics.

Index Terms: dropout prediction, performance prediction, decision trees, quiz completion, online university education.

[9] En aquest capítol, que **conté el tercer article del conjunt que conformen la tesi**. les referències numèriques a figures i taules corresponen a referències exclusivament d'aquest capítol. Les referències bibliogràfiques d'aquest capítol es mostren dins d'aquest mateix capítol.

5.1 Introduction

This article is an extended version of the one presented at the LASI2019 conference [1] in three aspects. First, the current study adds robustness to that presented at LASI2019: it includes a cross validation (CV) procedure in the classification performance evaluation, and employs a random undersampling approach to deal with the imbalanced data problem. Second, new student assessment attributes are considered. In the original version, the only attribute assessed was test completion. In this new version, assessment test grades are also considered. And third, a new response variable, *failure*, is intended to be predicted at an early stage, and this is added to the variable *dropout* studied in the original paper.

Students who do not complete a course or interrupt their studies indefinitely have been a matter of concern among practitioners and researchers for a considerable time. These students are usually called *dropout students*. In online courses, the high dropout rates justify the abundant research on this particular topic, as shown in the extensive review of [2], where 159 studies published between 1999 and 2009 were analysed. More recently, within the European framework, reducing student dropout rates in higher education is considered crucial to attaining the ambitious objective of at least 40% of people in their thirties having completed higher education studies by 2020 [3]. With the aim of improving our teaching practice and guided by European strategy, we decided to conduct research on non-successful students of an online statistics course at the Universitat Oberta de Catalunya.

In this study, we use a learning analytics (LA) approach. LA were initially defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" [4]. From a practical perspective, *analysis and data reporting* is carried out by developing predictive models. LA and educational data mining (EDM) approaches overlap in several aspects, for example, developing predictive models. Nevertheless, LA prioritizes a model's interpretability (in response to the "*for purposes of understanding learning*" requirement) over its optimization, in contrast to the aim of EDM [5]. Using an LA approach, predictive models must facilitate intervention with students to *optimize learning*; [6] and [7] are two extensive reviews of LA interventions that proceed from the creation of previous prediction models intended to improve learning. To cite an example, [8] shows a reduction of 14% in student dropout rates by means of a tutorial action plan after the early identification of dropout-prone students. In a higher education context, we find two levels of dropout: (a) the course-level dropout, and (b) the beyond course-level dropout. In the former, dropping out occurs within the course or subject [9], where teachers can intervene to prevent dropout at an early stage if they have the relevant information. With beyond course-level dropouts, students withdraw from their studies. The analysis requires other kinds of variables and intervention depends on other stakeholders, such as programme directors, as well as teachers. Students who either fail to complete a course or complete it unsuccessfully, that is, without passing, are prone to drop out of their studies.

Since we wish to understand the learning process and define timely interventions, we first proposed to identify as many at-risk students of an online statistics course as possible before the mid-term stage. We considered two types of at-risk students separately: (a) dropout-prone students, and (b) fail-prone students. Our procedures were based on the prediction/classification provided by binary decision trees generated at several time points during the term. The data involved in these procedures were mainly gathered from formative self-assessment test data (completion and grades) and indicators of engagement in the virtual classroom, such as participation in forums and number of noticeboard accesses.

Online formative self-assessments have been proved to be statistically significant predictors of student success in both blended [10] and online environments [11]. Both studies [10, 11] took into account every test taken throughout the whole course, during which the cycle of (1) studying content, (2) taking tests, (3) receiving feedback, and (4) restudying content is repeated throughout the term. Our study aimed to determine whether it is possible to predict which students will not perform well by studying only the cycles pertaining to the first half of the term.

5.2 Literature review

Extensive research has been conducted in relation to predictive models of student success [12], from which literature we reviewed 121 articles. Most authors define unsuccessful students as those who either drop out of or fail their courses.

In a higher education (HE) context, three levels of dropout can be distinguished: (a) course-level, studied by several authors ([13] in computer science and linguistics courses, [14, 15] in a computer science course, and [16, 17] in a maths course); (b) degree-level, studied by [18] (in a computer science degree); and (c) institutional-level, targeted by [19] (in an open university, UDIMA). In this paper, we focus on dropout at course level.

According to [2], there is an absence of consensus on the definition of course-level dropout. Several definitions are found. [20] directly associates course dropout with course failure: dropout students are those who do not obtain A, B, or C grades, that is, those who fail the course. [21] defines dropout students as those who fail to complete the course and whose tuition fees have not been refunded, while [22] considers dropout students as non-completers in a broader sense. So, in one sense, the terms "dropout" and "failed student" can be interchangeable at course level. The definition of dropout from the perspective of failure [20] is used in studies [9], [23] and [24]. The definition used in [21] is explicitly mentioned in [25], which adds a further requirement: the dropout student must have accessed the e-learning platform at least once throughout the course duration. That means the student must leave a trace in the information system before leaving the course in order to be considered a dropout student. For [8] and [26], students who do not sit the final exam are defined as dropout students. [27] does not give a precise definition of non-completer students.

In HE, predictive models are applied in (a) blended environments [9, 14, 16, 17], (b) entirely online environments [8, 15, 23, 24, 28, 29], and in massive online open courses (MOOCS) [30, 31, 32, 33] as well. Instructors in blended environments gather their own rich face-to-face information, which contributes to enhancement of the models. In contrast, in online environments and MOOCs, instructors experience serious difficulties in adding new information to the models.

The aim of the models is to predict retention or success, operationalized as variables in different formats. [17] and [29] predict final outcome into two categorical classes: passed versus failed. [25, 26] predict dropout into two classes: dropout and completer. [8, 14] predict final grades into binary format (passed-failed) and also into two classes, high-scoring (at least 90 out of 100) versus non-high-scoring (less than 90 out of 100). [13] predicts into academic standing codes: pass, fail, conditional fail, repeat the year/level, and repeat a single semester. [34] predicts final grades into three categories: high performance (at least 80.5 out of 100), medium performance (less than 80.5 and at least 57.5 out of 100), and low performance (less than 57.5 out of 100). [15] predicts final grade as a continuous interval variable.

With respect to the modelling attributes or features employed, three main categories are common: demographics, virtual learning environment (VLE) interactions (e.g. forum participation), and assessment activities or exam performance. The first category is formed by time-invariant data available at the beginning of the course, whereas the other two categories include time-varying data which are collected incrementally throughout the course. Demographic attributes such as gender and professional information are used by [23, 24, 25, 27]. Forum participation is a feature included in [9, 33, 35]. Finally, the grades achieved in running assessment activities are analysed in [8, 15, 23, 24, 25, 26, 27, 28, 31, 32].

In terms of machine learning techniques, [15] uses a decision tree regression because its response variable is the grade expressed as a numerical value. In cases where response variable values are divided into classes, classification methods are employed: (a) decision tree-based algorithms [13, 23, 24, 27, 29, 30, 35]; (b) algorithms based on neural networks [8, 23, 24, 25, 26, 27, 32, 35]; (c) support-vector machines [8, 23, 24, 25, 26, 27, 35]; (d) naive Bayes [13, 23, 24, 27, 30, 35]; and (e) logistic regression [8, 26, 27, 31, 35].

In predicting modelling, imbalanced data is a well-known problem. In general, models are biased towards the majority class. Some authors, including [28] and [14, 23], specifically target this issue by using sampling methods (data level approach). [28] also deals with the problem by setting class weights for the machine learning algorithms (algorithm level approach). Others, such as [30], ignore the issue while acknowledging the limitations of the study, while still others, such as [13] and [29], make no explicit mention of the problem.

The time point at which prediction is made differs among researchers. [14] makes the prediction at the end of the course, meaning that prediction cannot lead to any useful intervention with students. Predictions need to be disclosed before the course is half over in order to be of any real use [27]; if they are not, practitioners will not be in time to successfully intervene with students. [36] does disclose the prediction before the midterm stage, but as the attributes are static, they provide no actionable information for intervention purposes. Predicting at a single time point, before the midterm stage, is the option chosen by [24] and [27]. In contrast, multiple time points, albeit not the same ones, are proposed by [9], [8], [23], [25] and [26].

With respect to model performance metrics, a great variety of evaluation measures are used: precision [5, 8, 9, 13, 25, 29, 30]; recall [5, 8, 13, 25, 29, 30]; accuracy [5, 8, 13 17, 24, 25, 30, 35, 36]; and F-measure [13, 17, 23, 25, 29, 30, 31, 35]. AUC is used by [13, 30, 31] and Mean Square Error (MSE) by [15]. An ROC curve is considered by [16, 19]; a precision-recall curve by [37]; and a precision-recall AUC curve by [28]. [36, 37] report Cohen's Kappa coefficient.

This article analyses the construction of early predictive models for identifying at-risk students in an online statistics course at an online university. Although we have considered an extensive set of attributes, our model is based on decision trees dependent only upon the non-compulsory formative assessment grade average. This shows the efficacy of regularly participating and performing well in formative assessments, which is a strength of our research.

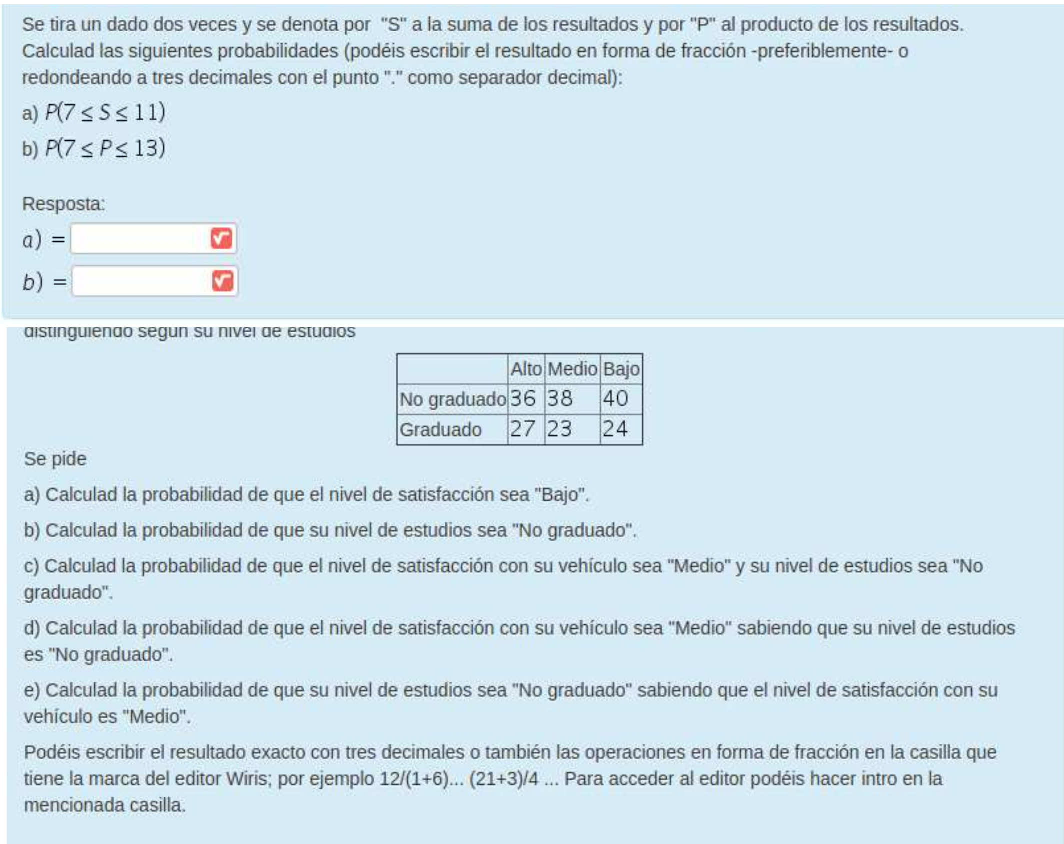
5.3 Methodology

5.3.1 Participants and learning context

The participants in this study are the 197 students enrolled in the first term of the 2018/19 online statistics course, which is part of the Computer Engineering degree at the Universitat Oberta de Catalunya.

The statistics course includes two generic assessment instruments: (a) a compulsory in-person final exam, and (b) non-compulsory online continuous assessment throughout the semester. The final grade for the course is mainly based on the final exam grade, adjusted with a cross-referenced table. For example, if a student obtains a continuous assessment grade of A+ and a final exam grade of

D+, the final mark is adjusted to a C. So students who fail the final exam but whose grade is close to the pass threshold can pass the course if their continuous assessment grade is high enough. This undoubtedly increases the incentive to participate in continuous assessment. The core of the continuous assessment is the set of six different pairs of tests (quizzes and R tasks) known as continuous assessment activities (CAA). Each quiz consists of multiple choice and short-response questions which are corrected and graded immediately, providing students with automated feedback. Fig. 1 depicts an example of two short-response questions extracted from the second quiz. Quizzes are produced using the Moodle quiz authoring suite with the WIRIS plugin. The items are different every time the student takes the quiz, so students are exposed to different questions. In each of the six continuous assessment tests, students have two attempts to solve a quiz, the grade being taken from the higher score of the two. R tasks are constructed-response questions that have to be solved using the statistical program R. They require manual teacher correction and therefore the feedback


Se tira un dado dos veces y se denota por "S" a la suma de los resultados y por "P" al producto de los resultados. Calculad las siguientes probabilidades (podéis escribir el resultado en forma de fracción -preferiblemente- o redondeando a tres decimales con el punto "." como separador decimal):

a) $P(7 \leq S \leq 11)$

b) $P(7 \leq P \leq 13)$

Resposta:

$a)$ =

$b)$ =

distinguiendo según su nivel de estudios

	Alto	Medio	Bajo
No graduado	36	38	40
Graduado	27	23	24

Se pide

a) Calculad la probabilidad de que el nivel de satisfacción sea "Bajo".

b) Calculad la probabilidad de que su nivel de estudios sea "No graduado".

c) Calculad la probabilidad de que el nivel de satisfacción con su vehículo sea "Medio" y su nivel de estudios sea "No graduado".

d) Calculad la probabilidad de que el nivel de satisfacción con su vehículo sea "Medio" sabiendo que su nivel de estudios es "No graduado".

e) Calculad la probabilidad de que su nivel de estudios sea "No graduado" sabiendo que el nivel de satisfacción con su vehículo es "Medio".

Podéis escribir el resultado exacto con tres decimales o también las operaciones en forma de fracción en la casilla que tiene la marca del editor Wiris; por ejemplo 12/(1+6)... (21+3)/4 ... Para acceder al editor podéis hacer intro en la mencionada casilla.

Fig.1 Screenshot of two questions of the second quiz provided is delayed. Both quizzes and R tasks can be mainly considered as formative assessment instruments since their main aim is to provide students with information to aid them in their learning process [38]. In addition, teachers assign an initial test to ascertain students' prior knowledge of secondary-education-level statistics during the first week. In order to encourage participation, students who voluntarily complete and submit the test get a bonus, which also forms part of the continuous assessment grade.

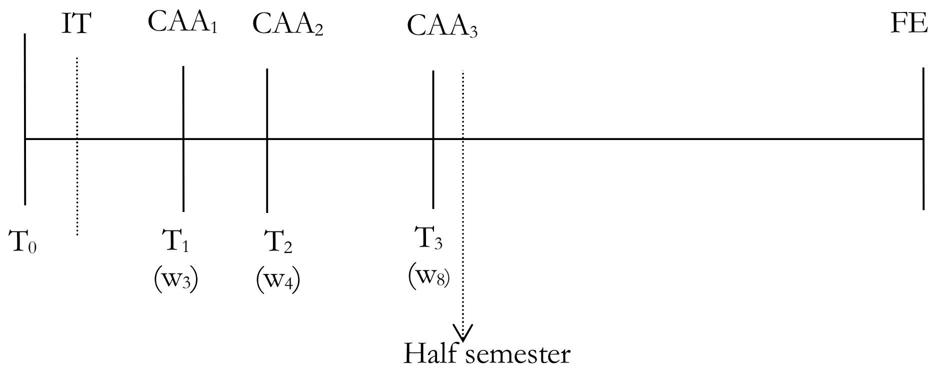
The Universitat Oberta de Catalunya provides a virtual campus where students perform all activities related to the courses in which they are enrolled. Each virtual classroom includes direct access to the teaching plan, which contains specific information about the teaching methodology and the assessment system. In addition, there are three different communication spaces: a communication tool (the forum), an information tool (the noticeboard) and a Moodle. The latter is used by students to perform and submit the continuous assessment tests and also to receive feedback about their performance. The noticeboard is used by teachers to upload course information. The forum allows students and teachers to interact with each other, asynchronously. All reading accesses (to the noticeboard, forum and teaching plan) and writing accesses (to the forum) are recorded by the Universitat Oberta de Catalunya's information system.

5.3.2 Measurement and data collection

Data was collected at three time points (T_1, T_2, T_3), which coincided with the first three continuous assessment activities (CAA) submission cut-offs, the only tests that occurred in the first half of the course (Fig. 2). The separation between submission deadlines was variable, ranging from 1 to 4 weeks. We defined three periods of time (*Period.1*, *Period.2*, *Period.3*) from previous submission deadlines (T_i) as

$$Period.i = (T_{i-1}, T_i]$$

The attribute selection for our study was based on [8], [23] and [13]. We considered two sets of attributes: *Set_Engagement* and *Set_Achievement* (Table I). *Set_Engagement* is formed by two initial student status attributes (*Repeating* and *EnrolledCourses*), three virtual classroom usage attributes for each of the three time points T_i (*TeachingPlan_T_i*, *BBoard_T_i*, *ForumWr_T_i* and *ForumRe_T_i*), one



T_i: Cut-off for the i-th Continuous Assessment Activity (CAA_i) being submitted. T_0 indicates the beginning of the semester.
w_i: i-th week.
IT: Initial Test
FE: Final Exam

Fig.2 Assessment instrument time distribution

assessment completion attribute (*T_InitialTest),* and two assessment completion attributes for each of the three T_i *(N_Quizzes_*T_i and *N_RTasks_*T_i). The remaining attributes, apart from those related to student status, serve as indicators of student engagement [12]. *Set_Achievement* shares two initial student status attributes and the three virtual classroom usage attributes with *Set_Engagement*, but replaces the assessment completion attributes with assessment achievement attributes (*S_InitialTest, A_Quizzes_*T_i and *A_RTasks_*T_i). [13] states that using scores as attributes can be acceptable in educational contexts where grades correspond to formative assessment, as is the case in our study.

During the first period (*Period.1*), we gathered students' registration data, such as the number of courses enrolled on in the semester and whether they were repeat students. This data, contained in the Universitat Oberta de Catalunya's information system and anonymously delivered to us, filled the instances of the attributes *Repeating* and *EnrolledCourses*. The Moodle activity log was the source of information to determine whether the student had completed and submitted the initial test and, likewise, the first continuous assessment test. With that data, the instances of the attributes *T_InitiaTest*, *N_Quizzes_*T_1 and *N_RTasks_*T_1 were filled. The Moodle activity log also provided the grades for the initial test, quizzes and R tasks, and these were used to calculate the *S_InitiaTest*, *A_Quizzes_*T_1 and *A_RTasks_*T_1. The virtual classroom activity log provided the dates and times of all accesses to the forum, noticeboard and teaching plan, which, after pre-processing, filled the

TABLE I. ATTRIBUTE SETS FOR PERIOD i [*]

Name	Description	Type and values
Repeating [(A,B)]	Indicates whether the student is repeating the subject	Type: Boolean Values: 1, 0
EnrolledCourses [(A,B)]	Indicates the total number of courses enrolled on in the semester	Type: Integer Values:{1, …}
T_InitialTest [(A)]	Indicates whether the student has taken the initial test	Type: Boolean Values: 1, 0
S_InitialTest [(B)]	Indicates the grade awarded in the initial test	Type: Real Values:[0,10]
N_Quizzes_T_i [(A)]	Indicates the number of quizzes taken until the T_i	Type: Integer Values:{0, …, i}
A_Quizzes_T_i [(B)]	Indicates the average mark awarded in the quizzes {1, …, i}	Type: Real Values:[0,10]
N_RTasks_T_i [(A)]	Indicates the number of Rtasks taken until the T_i	Type: Integer Values:{0, …, i}
A_RTasks_T_i [(B)]	Indicates the average mark awarded in the Rtasks {1, …, i}	Type: Real Values:[0,10]
TeachingPlan_T_i [(A,B)]	Indicates whether the student has viewed the teaching plan until the T_i	Type: Boolean Values: 1, 0
BBoard_T_i [(A,B)]	Indicates the number of periods in which the student has accessed the noticeboard until the T_i	Type: Integer Values:{0, …, i}
ForumWr_T_i [(A,B)]	Indicates the number of periods in which the student has written messages on the forum until the T_i	Type: Integer Values:{0, …, i}
ForumRe_T_i [(A,B)]	Indicates the number of periods in which the student has read messages on the forum until the T_i	Type: Integer Values:{0, …, i}

[*] with i = 1, 2,3 [(A)] Set_Engagement [(B)] Set_Achivement

instances of the attributes *BBoard_T_1*, *ForumWr_T_1*, *ForumRe_T_1* and *TeachingPlan_T_1*. All this data were transferred to the second period (*Period.2*) and incremented with the specific information collected in that period to fill the instances of the attributes ending in '_T_2', and so on, for the third period.

For the purposes of the present study, we defined two different outcome/response variables for at-risk students: (1) *dropout* and (2) *failure*.

(1) The *dropout* variable is defined based on [21]. A dropout student is a student who obtains a final grade of "Not Completed", which means the student has not taken the compulsory final exam (see

Fig. 3). This approach is similar to the one used in [8] and [29]. The Boolean response variable *dropout* indicates whether the student complies or not with the previous definition, that is, whether they belong to the dropout-student class (*dropout*) or to the completer-student class (*completer*). The Universitat Oberta de Catalunya's information system was used to fill the variable *dropout*. By combining attributes, such as predictors, with the response variable *dropout*, the data set [DS1] was created. It contains the instances of the set of attributes *Set_Engagement* (Table I) and those of the *dropout* variable for each different period. Likewise, the instances of all attributes included in the set of attributes *Set_Achievement*, and the response variable *dropout*, constitute most of the data set [DS2].

(2) The *failure* variable is based on [20]. A failed student is a student who obtained a final exam grade of less than 5 on a 0-10 scale (see Fig. 3). By this definition, all dropout students are also failed students. The Boolean response variable *failure* indicates whether the student passed the final exam, that is, whether they belong to the passed-student class (*passed*) or to the non-passed-student class (*failed*). The data set [DS2] was extended with the instances of the *failure* variable.

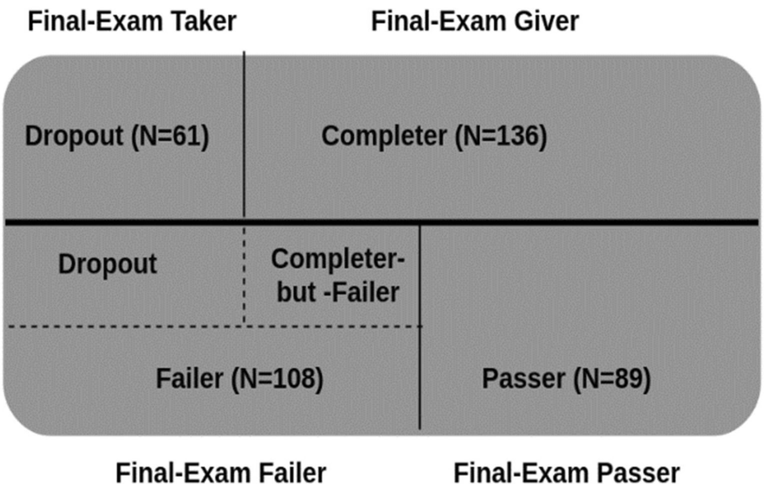


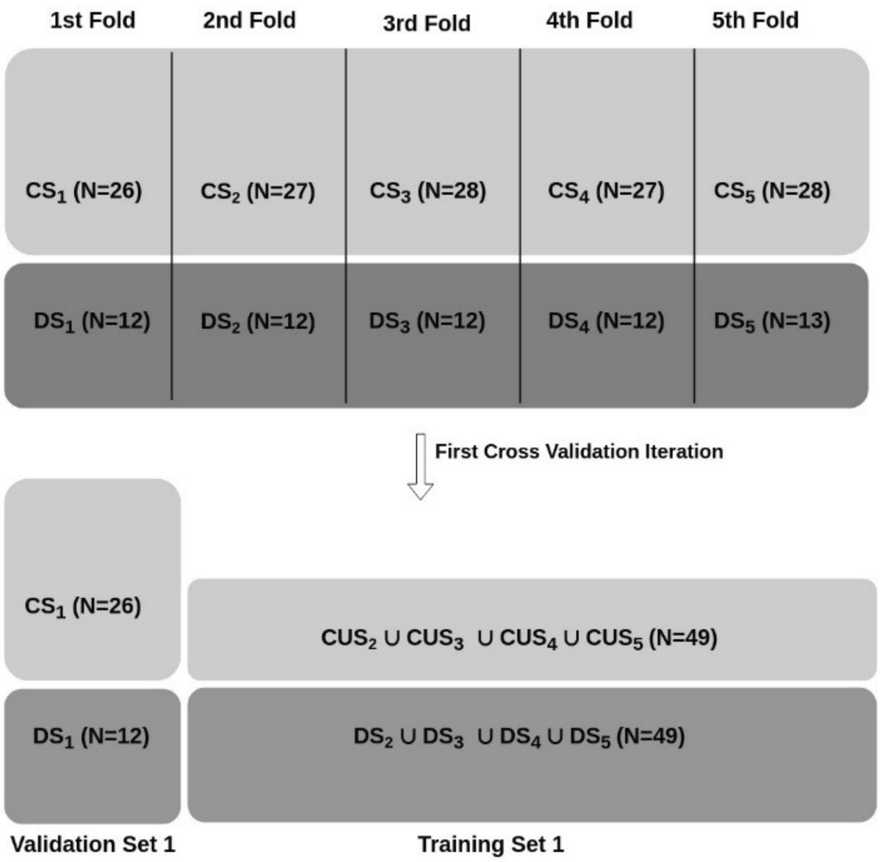
Fig.3 Sudents' distribution

5.3.3 Classification method

Our problem is one of classification/prediction, the result of which will be a binary classification model or binary classifier to predict whether or not a student will be classed as at-risk at the end of the semester. This study follows the learning analytics priority of interpretability over optimization [5], since the classifiers are intended to result in a future intervention. We have decided to employ tree-based methods because "tree-based methods are simple and useful for interpretation" [39, p. 303]. [19] also uses decision trees in its final stage of model generation, due to ease of interpretation. Our aim is not necessarily to find the best-performing classifier, but one which is easily interpreted.
A binary decision tree is fundamentally an oriented graph that starts with a node called a root, follows through arcs called branches, and ends in terminal nodes called leaves. Each non-terminal node, including the root, represents an attribute, a yes/no question regarding the value of a certain attribute. In our study, each leaf represents one of the two classes, either *dropout/completer* or *failed/passed*. The branches that stem from a node represent the values of the attribute associated with the node, the answer to the question related to the attribute [40].

As we can observe in Table I, not all attributes have the same number of possible values. In creating the nodes, a bias towards attributes with a large number of possible values is an issue detected in studies using decision tree models [41]. We chose conditional tree models because they mitigate that bias [41]. We used the ctree() function available in the statistics program R to grow our conditional tree models.
Given the data sets [DS1] and [DS2] for the dropout variable, the number of dropout students (minority class) was 61 and the number of completers (majority class) was 136 (Fig. 3). Therefore, the ratio of dropout over completer was 1:2.22, which meant we had an imbalanced data problem. In this situation, classifiers tend to bias towards the majority class (completer). To correct this, several techniques are used: sampling methods, cost-sensitive methods, kernel-based methods and active learning methods [42]. The application of sampling methods improves classification accuracy [42]. Among these, random undersampling has been used extensively. This method consists of randomly removing data from a previously selected majority sample to equal the size of an also previously selected minority sample. To address our imbalanced data problem, we decided to use random undersampling. In addition, we used five-fold cross validation (CV) to estimate the accuracy of the

performance of our predictive models. [43] states that the correct way to combine five-fold CV with random undersampling is to first split all the data into five folds and then reduce the size of the training set from the original training set. We randomly selected five folds (Fig. 4), maintaining the distribution of the whole group of students. Then, in the first iteration, we generated the training set as a *completer* subsample and a *dropout* subsample (Fig. 4), the *dropout* subsample being the union of the dropout samples contained in the remaining folds. The *completer* subsample was produced by randomly undersampling the *completer* class sample not contained in the first fold. The validation set was the whole of the first fold without any resampling.



CS_i: i-th fold completer sample
DS_i: i-th fold dropout sample
CUS_i: i-th fold undersampled completer sample

Fig. 4 Cross validation with random undersampling

By not resampling the validation set, we ensured that all data was included in the model performance assessment. The remaining four iterations were implemented in the same way. Six *dropout* classifiers were selected, one for each of the two data sets and each of the three periods (Fig. 2).

Given the data set [DS2] for the *failure* variable, the number of failed students (majority class) was 108 and the number of passers (minority class) was 89 (Fig. 3). Therefore, the ratio of passers over failed students was 1:1.21, which we did not consider a significant imbalance of data.

As above, we carried out a five-fold CV. We randomly split all the data into five folds, maintaining the distribution of the subpopulations: passers, dropouts and completer-but-failed students. In each iteration, one fold was taken as the validation set and the four remaining folds as the training set. Three classifiers were selected, one for each of the three periods (Fig. 2).

In predicting at-risk students, our primary concern was to reduce the number of actual dropout or failed students misclassified as completer or passed students, respectively. Since our aim is to use predictive models to implement interventions with at-risk students, we wished to identify as many of them as possible. Considering at-risk students as *positive* cases, our main aim was to achieve low *false negative* values. For this reason, we decided to measure our model's performance, mainly using the *recall* metric. Our secondary concern was to achieve low *false positive* values to prevent the model from targeting actual non-risk students as at-risk. Therefore, we also considered the *precision* metric; and the combination of both these metrics, their harmonic mean, *F-measure*, was also taken into account. [29, p. 3] uses the same metrics because "the goal is primarily to recognise at-risk students". We used the following definitions, according to [44]:

$$Precision = \frac{TruePositives}{TruePositives+FalsePositives} \quad (1)$$

$$Recall = \frac{TruePositives}{TruePositives+FalseNega} \quad (2)$$

$$F-measure = 2 * \frac{Recall*Precision}{Recall+Precisio} \quad (3)$$

5.4 Results

A. Dropout/completer student classifiers

Two groups of classifiers were trained for the variable *dropout*, considering the sets of attributes *Set_Engagement* and *Set_Achievement*.

A.1) Dropout classifiers using the Set_Engagement

The performance metrics for the three models created using assessment engagement (completion of assessment tests) were all below 70% (Table II), demonstrating a lack of predictive power with regard to [13]. Each classification performance metric was calculated as the average of the five metrics resulting from each of the five iterations corresponding to the five-fold CV.

Table II. Dropout classifiers' performance metrics (*Set_Engagement*)

Models	*Recall*	*Precision*	*F-measure*
After T_1	55.8%	66.6%	52.8%
After T_2	61.0%	65.9%	61.0%
After T_3	65.8%	62.3%	62.7%

A.2) Dropout classifiers using the Set_Achievement

When replacing assessment engagement with assessment achievement (assessment test grades), the model created after completion of the third continuous assessment test (*Model.Dropout.T3*) showed performance metrics over 70% (Table III). The remaining models created at time points T_2 and T_3 showed poorer performance than the third, at under 70%. Our result was situated at the lower edge of [8]'s, which reports recall values in the range 69.23% to 96.73% when predicting dropout at the midterm stage. [8]'s study takes place in a entirely online environment, using black-box machine learning algorithms, in contrast to our white-box decision tree algorithm. Differences in prediction accuracy, with lesser accuracy attributable to the decision tree algorithm, are in line with [39].

Table III. Dropout classifiers' performance metrics (*Set_Achievement*)

Models	*Recall*	*Precision*	*F-measure*
After T_1	57.2%	69.0%	61.9%
After T_2	60.9%	77.9%	67.5%
After T_3	75.1%	70.3%	72.0%

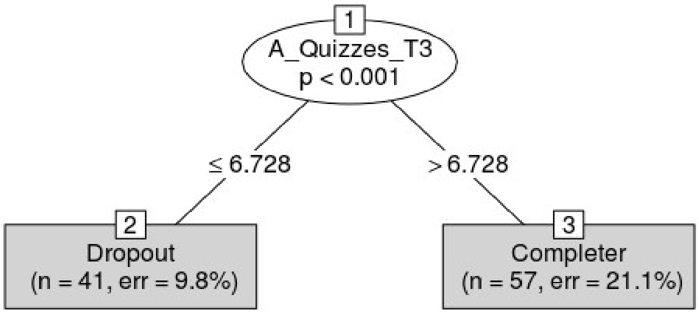


Fig. 5 Model.Dropout.T3

Model.Dropout.T3 is extremely simple, with just one node (Fig.5). The only attribute shown in the model is the average of the three first quizzes (*A_Quizzes_T3*), which shows that this is the attribute most strongly associated [41] with the response variable *dropout*.

B. Failed/passed student classifiers

The model trained after the third continuous assessment test (Model.Failure.T3) showed performance metrics above 75% (Table IV). The remaining models created at time points T2 and T3 showed poorer performance measures than the third, at under 70%. [23], in their study of an online computer science course, and presented F-measures ranging from 77% to 82% when predicting final outcome fail/pass at the midterm stage. At its lower edge, the decision tree algorithm J48 appears. Our F-measure (76.3%) is slightly lower than that of [23], which used assignment scores and virtual learning usage in the same way. [29] reports F- measures under 50%, also when predicting final outcome fail/pass final course at the midterm stage, using decision tree algorithm C4.5, the only method used in this particular study. [29] employs assignment scores and virtual learning engagement data from several completely online arts, maths and business courses. Our F-measure was nearly higher than that of [29].

TABLE IV. FAILURE CLASSIFIERS' PERFORMANCE METRICS

Models	*Recall*	*Precision*	*F-measure*
After T_1	63.0%	72.6%	66.9%
After T_2	66.7%	77.2%	70.6%
After T_3	75.9%	78.3%	76.3%

Model.Failure.T3 is also extremely simple, containing a single node (Fig. 6). Only the attribute related to quiz achievement (A_Quizzes_T3) has discriminatory power. The difference between a student being classified as dropout or failed is roughly 1 point on a 0-10 scale, according to Model.Dropout.T3 and Model.Failure.T3, which is not a great deal.

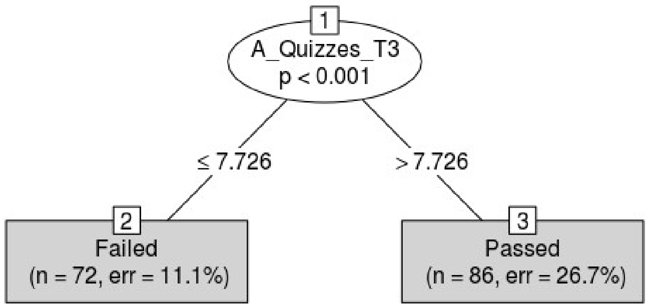


Fig. 6 Model.Failure.T3

5.5 Discussion

In this section, we will summarize the interpretation and utility of the models we have identified.

A. Dropout and performance prediction are determined by how tests are performed

As mentioned in section IV on predicting dropout, when attributes from the *Set_Engagement* were considered, the models' evaluation metrics were low. We therefore affirm that completing tests, which are attributes included in *Set_Engagement,* does not determine dropout prediction. Instead, when using attributes from the *Set_Achievement,* after three continuous assessment tests the model was found to have predictive power. Only one attribute in the model had discriminatory power: the *A_Quizzes_T_3*, which was directly related to quiz performance and therefore set apart from attributes related to use of forums. With regard to performance prediction, Fail/Pass classifiers, the same reasoning is applicable. The dominance of evaluative attributes was in line with [23], where the most important attribute was the midterm grades.

B. Quiz grades are more determinant than R tasks

R tasks also have an evaluative character, although they do not intervene in the prediction models as the quizzes do. The main difference between quizzes and R tasks is that the latter require students to write an extensive answer to communicate mathematics. Although we might think that grades in the more demanding assignments (R tasks) should be more determinant in predicting final performance, the figures tell us another story. The structure of the final exam, in which the percentage of questions involving R skills learnt is no higher than 25%, may affect the discriminatory power of the R task grades.

C. Dropout and performance prediction models are mutually coherent

To simplify, we classified dropout and failed students according to their final exam grade (where the dropout did not take the final exam and the failer did not achieve a grade of at least 5 out of 10). Hence the difference between the two types of students is the threshold established in the final exam. Prediction models for dropout and performance only differ in average quiz scores, which are higher for the latter. This is coherent with the student classification considered in our study.

D. Simplicity of dropout and performance prediction models facilitates classification procedure

The simplicity of the dropout and performance models reduces the volume of information to, basically, quiz performance. This in turn leads to two beneficial consequences: (a) time spent gathering and processing the remaining attributes is eliminated; (b) the practitioner can collect the required data directly from the Moodle activity log. Nonetheless, practitioners should be aware that the models are not perfect, due to the medium values of the classification performance metrics.

E. Continuous assessment test grades are associated with final exam performance

The average of the three quiz grades scheduled prior to the third continuous assessment activity (CAA_3) appears in the predictive performance model Model.Failure.T3. The grades of the first three quizzes, included in the CAA_1, CAA_2 and CAA_3's grade, are therefore associated with the response variable failure, and also with the student's final exam performance. This is in line with a similar association found between formative assessment grades and final exam performance in an online calculus course [11], and in a blended information technology course [10].

5.6 Conclusion and further research

In an online statistics course using quizzes as formative assessment instruments, we found that these quizzes played an important role in the early prediction of at-risk students.

The main contribution of the present study is to offer a simple and interpretable procedure to identify dropout-prone and fail-prone students before the halfway point of the semester, by means of tree-based classification models. The procedure is straightforward since students are classified according to a single attribute which is related to student performance in low-stake assessment assignments, such as quizzes set by teachers, and not related to use of elements of the virtual learning environment, such as the forum. In addition, because the information required is easily accessed by the teacher and needs no processing, teachers can control the procedure by themselves. The models interpret that the main factor contributing to final exam performance is continued learning acquired during at least the first half of the course. It shows that, in an always difficult subject such as mathematics or statistics, even more difficult in an online environment, regular engagement in learning activities is paramount. Relying exclusively on the final exam is not a good strategy, at least for the majority of students.

With respect to the simplicity of the procedure, in a similar fully online environment, [8] differs from our study in that they use more complex multiple predictive dropout models. Also working with an online environment, [29]'s models are easy to interpret because they are tree-based, as in our study. But in contrast to ours, [29] does not explicitly show or detail which attributes are most significant. Our performance classification metrics are higher than those of [29], though at the lower edge of those reported by [8] and [23]. Although we acknowledge that our classification performance assessment measures have room to improve, we consider them sufficient to support intervention with students, as does [37], despite reporting even poorer accuracy (between 65% and 70%).

The first limitation worthy of mention is the low classification performance evaluation values, of which teachers need to be aware. The second limitation is associated with the validation set. According to the methodology selected, the students belonging to the training set, to whom the classification models were adjusted, and the students belonging to the validation set, whose performance had been evaluated, were all enrolled in the same academic year. Both these limitations indicate a need for further research. Firstly, more variables should be included in the attribute sets to increase prediction accuracy. Variables related to self-regulated learning strategies, as in [28], and motivation, could be incorporated, for instance. Secondly, a test set using students from a different academic year from the one used to constitute the training set should be considered, as in [25] and [26]. Furthermore, since our research has devised a predictive model capable of providing sufficient actionable information, further research will concentrate on intervention with at-risk students and will include analysis of the efficacy of action taken in terms of academic outcomes and motivation.

Acknowledgment

This research was partially supported by a Fundació IBADA grant and by the Catalan Government Project 2017SGR1619. We would like to thank Mr Paul Garbutt and Dr Laura Calvet for their valuable contributions in helping to improve this study.

References

[1] J. Figueroa-Cañas and T. Sancho-Vinuesa, "Predicting early dropout student is a matter of checking completed quizzes: the case of an online statistics module", *CEUR Workshop Proc.*, vol. 2415, pp. 100–111, 2019.

[2] Y. Lee and J. Choi, "A review of online course dropout research: Implications for practice and future research," *Educ. Technol. Res. Dev.*, vol. 59, no. 5, pp. 593–618, 2011.

[3] H. Vossensteyn et al., Drop-Out and Completion in Higher Education in Europe - Literature Review. European Union, 2015.

[4] LAK11. Description of the 1st International Conference on Learning Analytics and Knowledge 2011 (LAK11). Banff, Alberta. Retrieved from https://tekri.athabascau.ca/analytics/

[5] W. Xing, R. Guo, E. Petakovic, and S. Goggins, "Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory," *Comput. Human Behav.*, vol. 47, pp. 168–181, 2015.

[6] A. Larrabee Sønderlund, E. Hughes, and J. Smith, "The efficacy of learning analytics interventions in higher education: A systematic review," *Br. J. Educ. Technol.*, vol. 50, no. 5, pp. 2594–2618, 2019.

[7] P. Sander and I. Services, "Using Learning Analytics to Predict Academic Outcomes of First-year Students in Higher Education," *CAPSTONE Rep. Pete Sander Manag. Inf. Serv. Oregon State Univ. Univ. Oregon Appl. Inf. Manag. Progr. Spring*, vol. 1277, no. 800, pp. 2–41, 2016.

[8] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," Comput. Electr. Eng., vol. 66, pp. 541–556, 2018.

[9] A. Cohen, "Analysis of student activity in web-supported courses as a tool for predicting dropout," Educ. Technol. Res. Dev., vol. 65, no. 5, pp. 1285–1304, 2017.

[10] M. Ćukušić, Ž. Garača, and M. Jadrić, "Online self-assessment and students' success in higher education institutions," *Comput. Educ.*, vol. 72, pp. 100–109, 2014.

[11] J.Figueroa-Cañas and T. Sancho-Vinuesa, "Investigating the relationship between optional quizzes and final exam performance in a fully asynchronous online calculus module," Interact. Learn. Environ., 2018.

[12] Y. Cui, F. Chen, A. Shiri, and Y. Fan, "Predictive analytic models of student success in higher education: A review of methodology," Inf. Learn. Sci., vol. 120, no. 3–4, pp. 208–227, 2019.

[13] C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," Comput. Educ., vol. 131, no. July 2018, pp. 22–32, 2019.

[14] X. Xu, J. Wang, H. Peng, and R. Wu, "Prediction of academic performance associated with internet usage behaviors using machine learning algorithms," *Comput. Human Behav.*, vol. 98, no. April, pp. 166–173, 2019.

[15] E. Wakelam, A. Jefferies, N. Davey, and Y. Sun, "The potential for student performance prediction in small cohorts with minimal available attributes," *Br. J. Educ. Technol.*, vol. 0, no. 0, 2019.

[16] H. Hirose, "Success/Failure Prediction for Final Examination Using the Trend of Weekly Online Testing," *Proc. - 2018 7th Int. Congr. Adv. Appl. Informatics, IIAI-AAI 2018*, no. 17, pp. 139–145, 2018.

[17] R. Umer, T. Susnjak, A. Mathrani, and S. Suriadi, "A learning analytics approach: Using online weekly student engagement data to make predictions on student performance," *2018 Int. Conf. Comput. Electron. Electr. Eng. ICE Cube 2018*, pp. 1–5, 2019.

[18] C. Lacave, A. I. Molina, and J. A. Cruz-Lemus, "Learning Analytics to identify dropout factors of Computer Science studies through Bayesian networks," *Behav. Inf. Technol.*, vol. 37, no. 10–11, pp. 993–1007, 2018.

[19] A. Ortigosa, R. M. Carro, J. Bravo-Agapito, D. Lizcano, J. J. Alcolea, and Ó. Blanco, "From Lab to Production: Lessons Learnt and Real-Life Challenges of an Early Student-Dropout Prevention System," *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 264–277, 2019.

[20] S. Liu, J. Gomez, and C.-J. Yen, "Community College Online Course Retention and Final Grade: Predictability of Social Presence," J. Interact. Online Learn., vol. 8, no. 2, pp. 165–182, 2009.

[21] Y. Levy, "Comparing dropouts and persistence in e-learning courses," Comput. Educ., vol. 48, no. 2, pp. 185–204, 2007.

[22] P. A. Dupin-bryant, "Pre-Entry Variables Related to Retention in Online Distance Education," Am. J. Distance Educ., vol. 18, no. 4, pp. 199–206, 2011.

[23] E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Araújo, and J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses," Comput. Human Behav., vol. 73, pp. 247–256, 2017.

[24] M. A. Santana, E. B. Costa, B. F. S. Neto, I. C. L. Silva, and J. B. A. Rego, "A predictive model for identifying students with dropout profiles in online courses," in Proceeding of the 8th international conference on educational data mining, EDM workshops, 2015.

[25] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, "Dropout prediction in e-learning courses through the combination of machine learning techniques," Comput. Educ., vol. 53, no. 3, pp. 950–965, 2009.

[26] J. A. Lara, D. Lizcano, M. A. Martínez, J. Pazos, and T. Riera, "A system for knowledge discovery in e-learning environments within the European Higher Education Area - Application to student data from Open University of Madrid, UDIMA," Comput. Educ., vol. 72, pp. 23–36, 2014.

[27] S. B. Kotsiantis, C. J. Pierrakeas, and P. E. Pintelas, "Preventing Student Dropout in Distance Learning Using Machine Learning Techniques," in Proceeding of the 7th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, KES 2003., 2003, no. September 2003, pp. 267–274.

[28] M. Hlosta, Z. Zdrahal, and J. Zendulka, "Ouroboros: Early identification of at-risk students without models based on legacy data," *ACM Int. Conf. Proceeding Ser.*, pp. 6–15, 2017.

[29] A. Wolff, Z. Zdrahal, A. Nikolov, and M. Pantucek, "Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment," in *Third Conference on Learning Analytics and Knowledge (LAK13)*, 2013, no. April 2013.

[30] A. Y. Q. Huang, O. H. T. Lu, J. C. H. Huang, C. J. Yin, and S. J. H. Yang, "Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs," *Interact. Learn. Environ.*, vol. 0, no. 0, pp. 1–25, 2019.

[31] J. A. Ruiperez-Valiente, P. J. Muñoz-Merino, A. Andújar, and C. Delgado-Kloos, "Early Prediction and Variable Importance of Certificate Accomplishment in a MOOC," in Proceedings of the European Conference on Massive Open Online Courses, 2017, no. May, pp. 263–272.

[32] K. Sharma, L. Kidzinski, P. Jermann, and P. Dillenbourg, "Towards Predicting Success in MOOCs: Programming Assignments," Proc. Eur. Stakehold. SUMMIT Exp. best Pract. around MOOCs (EMOOCS 2016), pp. 135–148, 2016.

[33] D. Yang, T. Sinha, D. Adamson, and C. Penstein Rose, "'Turn on, Tune in, Drop out': Anticipating Student Dropouts in Massive Open Online Courses," in Proceedings of the 2013 NIPS Data-driven education workshop, 2013, pp. 1–8.

[34] C. J. Villagrá-Arnedo, F. J. Gallego-Durán, F. Llorens-Largo, P. Compañ-Rosique, R. Satorre-Cuerda, and R. Molina-Carmona, "Improving the expressiveness of black-box models for predicting student performance," *Comput. Human Behav.*, vol. 72, pp. 621–631, 2017.

[35] C. Romero, M. I. López, J. M. Luna, and S. Ventura, "Predicting students' final performance from participation in on-line discussion forums," Comput. Educ., vol. 68, pp. 458–472, 2013.

[36] L. M. Abu Zohair, "Prediction of Student's performance by modelling small dataset size," *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, pp. 1–18, 2019.

[37] R. S. Baker, D. Lindrum, M. J. Lindrum, and D. Perkowski, "Analyzing Early At-Risk Factors in Higher Education e- Learning Courses," *Proc. 8th Int. Conf. Educ. Data Min.*, pp. 150–155, 2015.

[38] M. Yorke, "Formativeassessment in higher education: moves towards theory and the enhancement of pedagogic practice," *Higher Education*, vol. 45, nº. 4, pp. 477–501, 2003.

[39] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, vol. 103. 2013.

[40] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: A review of classification and combining techniques," Artif. Intell. Rev., vol. 26, no. 3, pp. 159–190, 2006.

[41] T. Hothorn, K. Hornik, and A. Zeileis, "Unbiased recursive partitioning: A conditional inference framework. Research Report Series 8, Department of Statistics and Mathematics, WU Wien, 2004," J. Comput. Graph. Stat., vol. 15, no. 3, pp. 651–674, 2006.

[42] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.

[43] R. Blagus and L. Lusa, "Joint use of over-and under-sampling techniques and cross-validation for the development and assessment of prediction models," *BMC Bioinformatics*, vol. 16, no. 1, pp. 1–10, 2015.

[44] R. Pelánek, "A brief overview of metrics for evaluation of student models", *CEUR Workshop Proc.*, vol. 1183, no. 2, pp. 151–152, 2014.

[DS1] https://figshare.com/s/e6efef99ef8deb86c959

[DS2] https://figshare.com/s/da3105c22454d6a9e72f

CAPÍTOL 6

CHANGING THE RECENT PAST TO REDUCE ONGOING DROPOUT: AN EARLY LEARNING ANALYTICS INTERVENTION FOR AN ONLINE STATISTICS COURSE[10]

Practitioners of the statistics course embedded in a computer science programme at a fully online university were concerned with the high dropout rate. In the academic year 2018-19, they decided to carry out a two-phase project in order to address this issue. In the first phase, an early classifier to identify students at risk of dropping out of the 2018-19 statistics course was determined. The second phase was planned to design and implement an early intervention based on the results of the first phase. This article presents the analysis of this intervention. In the 2019-20 online statistics course, 35 students did not submit the first few quizzes before the respective deadlines. They were the target of this intervention, which gave them the chance to change their recent past by submitting the unsubmitted quizzes. Students were encouraged to change by means of an email message. The aim of the intervention was to reduce the dropout rate among the targeted students. Our analysis also includes the students' perception of the intervention collected via a set of interviews. The results show that intervention clearly had a beneficial effect for some students.

Keywords: learning analytics intervention, online statistics, online quizzes, higher education, quiz submission.

[10] En aquest capítol, que **conté el quart article del conjunt que conformen la tesi**, les referències numèriques a figures i taules corresponen a referències exclusivament d'aquest capítol. Les referències bibliogràfiques d'aquest capítol es mostren dins d'aquest mateix capítol.

6.1 Introduction

The figures relating to dropout and retention are still causing major concern in higher education (Delnoij et al. 2020; Lonn et al., 2015), and specially in online environments (Cambruzzi et al., 2015; Figueroa-Cañas and Sancho-Vinuesa, 2019). Previous academic years of a one semester online statistics course in a higher education programme have shown dropout rates of more than one third. Motivated by reducing these worrying rates, practitioners of the 2018-19 statistics course decided to carry out a two-phase project. Taking the Learning Analytics (LA) framework and more specifically the LA cycle (Clow, 2012), the first phase consisted of predicting early on which students were at risk of dropping out of the course. In 2019, Figueroa-Cañas and Sancho-Vinuesa completed this first phase -the three links learner-data-metrics/analytics (Fig. 1) of the LA cycle- resulting in a classifier that predicted that students who did not submit the first few quizzes of the course were at risk of dropping out. The second phase, based on the previous classifier, had to plan and implement an early intervention for targeting the students at risk of dropping out. This article reports on this pilot intervention. To be more precise, the study extends the loop in (i) planning a pilot intervention based on the previous classifier, (ii) implementing the intervention with a cohort of students similar to that in which the classifier was obtained (iii) collecting data of intervened students, and (iv) analysing this data. That is, we complete the first loop (learner-data-metrics/analytics) with *intervention* and move the second loop forward, completing the links learner2-data2-metrics/analytics2 (Fig. 1).

Depending on whether or not the interventions fit within the LA framework, they are identified as LA interventions or non-LA interventions, respectively. Recent examples of systematic reviews have been published on both sorts of interventions: (a) Foster and Francis (2020), Sønderlund et al. (2019) and Wong and Li (2020) on LA interventions, and (b) Delnoij et al. (2020) on non-LA interventions. These four systematic reviews total more than sixty interventions, a sure sign of the current interest in the research area of interventions. Overall, the interventions reviewed in the four previous studies are found to be effective. Nevertheless, both Foster and Francis (2020) and Delnoij et al.(2020) point out the lack of evidence between the interventions and improvement. In addition, Sønderlund et al. (2019) conclude that the best intervention for targeting at-risk students has still not been found.

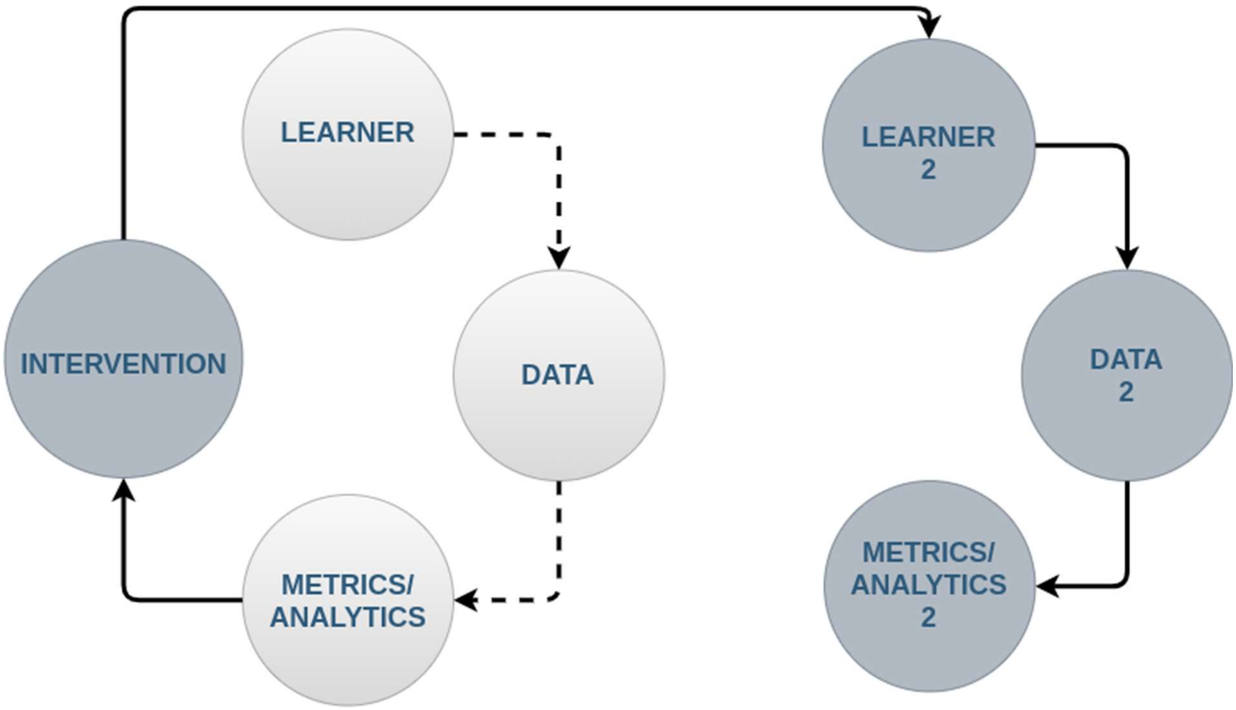


Fig.1 Two LA loops, based on Clow's (2012) LA cycle

This research is contextualized in the field of LA interventions that use predictive analysis to act in time upon students who are at risk of not succeeding in tertiary courses. The studies from Burgos et al. (2018), Cambruzzi et al. (2015), Cameron and Siameja (2019), Corrigan et al. (2015), Jayaprakash et al. (2014), Marcal (2019) and Pardo et al. (2019) present LA interventions that warn students about what may happen and suggest future actions regarding learning activities so as to avoid undesirable outcomes. To the best of our knowledge, no study that gives students at risk of dropping out the chance to improve the performance of past learning activities has been published. The main contribution of the current study is to present a pilot intervention that allows at-risk students to modify their academic results obtained in the recent past. Furthermore, this intervention is quite simple and easy to implement, which enables it to be replicated in very different settings.

The main purpose of the current study is to assess this LA pilot intervention, formatively so as to improve it.

6.2 Literature review

In the following four paragraphs we respond to four questions by means of a literature review: (1) what are interventions based on?, (2) who are the targets of the interventions?, (3) what are interventions aimed at?, and (4) how are the interventions carried out?

The LA interventions, which are used in a great number of educational interventions, share the common feature that they are mainly driven by data (e.g. Arnold & Pistilli, 2012; Burgos et al., 2018; Cameron & Siameja, 2017; Corrigan et al., 2015; Herodotou et al., 2017; Jayaprakash et al., 2014; Khan & Pardo, 2016; Pardo et al., 2019; Pinchbeck & Heaney, 2017; Rahal & Zainuba, 2016). However, we find some theory-guided differences. Thus, there are studies that: (a) also include references to a grounded theory and/or principle (e.g. Jayaprakash et al., 2014; Rahal & Zainuba, 2016), (b) are based on a for-the-study hypothesis (e.g. Herodotou et al., 2017, Pardo et al., 2019), (c) are supported by non-grounded principles (e.g. Cameron & Siameja, 2017; Corrigan et al., 2015; Khan & Pardo, 2016; Pinchbeck & Heaney, 2017), and (d) explicitly do not mention any theories and/or principle (e.g. Burgos et al., 2018; Cambruzzi et al., 2015; Marcal, 2019).

Another aspect that separates the educational interventions is the selfsame learners at whom they are targeted. In 2009, Wilson classified the course-level interventions using a four-level model which depends on which students are targeted. The interventions that belong to the first level are aimed at all students (e.g. Inkelaar & Simpson, 2015; Jayaprakash et al. 2014; Pardo et al. 2019). The second level interventions target specific groups of students. The third level contains interventions focused on at-risk students (e.g. Burgos et al., 2018; Cambruzzi et al. 2015). And finally, interventions targeting students who fail comprise the fourth level (e.g. Atree et al., 2017; Pinchbeck & Heaney, 2017).

The objectives of the interventions are also elements that help to classify them. A great number of studies aim to improve academic outcomes, such as performance (e.g. Arnold and Pistilli, 2012; Cameron and Siameja, 2019; Corrigan et al., 2015; Marcal, 2019) or retention (e.g. Burgos et al., 2018, Cambruzzi et al., 2015; Herodotou et al., 2017, Inkelaar and Simpson, 2015). Davis and Abbit (2013) aim at reducing academic procrastination and Baker et al. (2019), at improving time management.

Others seek to enhance motivation (e.g. Hulleman et al., 2017, Lonn et al., 2015). Engagement, in terms of both self-reflection on learning activities (e.g Kitto et al., 2016) and participation in online discussion forums (e.g. Wise et al., 2014) are also striven for.

Overall, two non-mutually exclusive types of intervention respond to the question of how LA interventions are carried out: supporting students and reporting information to them. Some studies deal with interventions that support students using different modalities: (a) by explicitly mentioning support elements (e.g. Jayaprakash et al., 2014), (b) by means of a tutorial plan (e.g. Burgos et al., 2018), (c) by peer tutoring (e.g. Lunsford et al., 2018), and (d) by containing guidance elements (e.g. Cambruzzi et al., 2015). There are published interventions that report information via email (e.g. Cambruzzi et al., 2015; Corrigan et al., 2015; Herodotou et al., 2017; Jayaprakash et al., 2014; Pardo et al., 2019) or dashboards (e.g. Arnold & Pistilli, 2012; Broos et al., 2020; Khan & Pardo, 2016; Kitto et al., 2016; Lonn et al., 2015). A more detailed look at interventions that provide students with information via email, allows us to separate those interventions that intend to make students aware of their situation within the course (e.g. Cambruzzi et al., 2015; Corrigan et al., 2015; Herodotou et al., 2017; Jayaprakash et al., 2014) from those that do not (e.g. Inkelaar & Simpson, 2015; Pardo et al., 2019). The last two studies are effective in improving retention and exam performance, respectively. Jivet et al. (2017) make the point that giving information to raise awareness is not enough for the intervention to be effective. In line with that, Herodotou et al. (2017) report no differences in final exam attendance among students who were reminded about not having chosen their final exam date from those who were not. Besides, the information released is not based on prior predictive analysis. On the contrary, Cambruzzi et al. (2015), Corrigan et al. (2015) and Jayaprakash et al. (2014) do inform by using predictive analysis. Cambruzzi et al.'s (2015) intervention is also seen to be effective in retention, and the other two, in exam performance.

So far the studies have been reviewed with regard to the features of the interventions. In the final paragraphs of this section we deal with how interventions are assessed. In 2016, Rienties et al. presented the Analytics for an Action Evaluation Framework (A4AEF), which they reported was being used at the UK Open University to assess interventions carried out there. The A4AEF requires the intervention impact to be analysed on learning activities, and/or learning processes and/or learning outcomes. The A4AEF allows for several types of protocol to implement the intervention strategies, among them quasi-experimental designs (e.g. Arnold & Pistilli, 2012; Burgos et al., 2018; Cambruzzi et al., 2015; Corrigan et al., 2015; Jayaprakash et al., 2014; Pardo et al., 2019)) and randomised control trials (e.g. Baker et al., 2019; Heterodotou et al., 2017; Cameron & Siameja, 2017). Rienties et al. (2016) require quasi-experiments to verify group similarity. Arnold and Pisilli (2012), Pardo et al. (2019) and Pinchbeck and Heaney (2017) compare cohorts without reporting on statistical analysis to assess the similarity, whereas Burgos et al. (2018) also compare cohorts and carry out an ANCOVA test to determine the similarity. Corrigan et al. (2015) and Attree et al. (2017) present a comparison between opt-in and opt-out groups with the possibility of a self-selection bias. Marcal (2019, p. 5) compares cohorts from two courses and remarks that both courses are "designed to be identical" (same instructor, textbook, and assessment instruments). In addition, Scheffel (2017) created and validated the Evaluation Framework for Learning Analytics questionnaire (EFLA) to assess LA tools that serve as instruments to implement interventions. The EFLA questionnaire is comprised of three dimensions: data, awareness and reflection, and impact, which are in line with the four stages that Verbert et al. (2013) demand of any LA application/tool: awareness, reflection, sensemaking and impact.

When assessing interventions, not only does their effectiveness in relation to the aims have to be determined, but also changes in the variables affected by the intervention have to be checked. Rienties et al. (2016) point out the necessity to designate which variables are supposed to be influenced by the intervention. What is more, Verbert et al. (2013, p. 1502) state that "in the end, the goal is to induce new meaning or change behaviour". Nevertheless, several studies do not publish variables that have changed due to the intervention, as is the case of: (i) Inkelaar and Simpson (2015), with motivation measure, (ii) Corrigan et al. (2015), with engagement with the Virtual Learning Environment, (iii) Cambruzzi et al. (2015), with the number of students that interact with teachers, (iv) Rahal and Zainuba (2016), with how many students have used the dashboard, and (v) Jayaprakash et al. (2014), with how many students have sought instructor support or accessed external resources.

6.3 Methodology

6.3.1 Learning context

The statistics course assessment instruments are: (a) a compulsory in-person final exam, and (b) non compulsory online continuous assessment throughout the semester. The final mark depends mainly on the final exam. Six continuous assessment activities (CAA), each of which are two tests (a quiz and an R task), form the basis of the continuous assessment. Students can attempt to solve each quiz twice, their mark being the higher of the two attempts. R tasks are questions to be solved using the statistical programme R. The Universitat Oberta de Catalunya (UOC) provides students with a virtual classroom for the statistics course, where the interactions between students and resources take place. The teaching plan, which contains detailed information about the assessment system, is accessible through the virtual classroom. That also includes a discussion forum, a noticeboard and a Moodle site. Students submit the continuous assessment activities through this Moodle classroom, where they receive feedback on their performance. Their participation and performance in solving quizzes and R tasks are recorded by the UOC's information system, as well as their access to the virtual classroom and its resources.

6.3.2 Planning and description of the educational intervention

In 2019, Figueroa-Cañas and Sancho-Vinuesa obtained a classifier to predict early on which students were at risk of dropping out of the course or, more precisely, at risk of not taking the compulsory final exam. The data to generate this classifier came from the cohort of students enrolled on the statistics course in the first semester 2018-19. This course was developed with similar pedagogical factors to that of the first semester 2019-20. These factors were: (a) an identical learning context, (b) four classrooms with the same teachers, and (c) a similar number of enrolled students (197 and 225 in 2018-19 and 2019-20, respectively). Figueroa-Cañas and Sancho-Vinuesa (2019)'s classifier predicted that students who did not submit all three quizzes (Q1, Q2 and Q3) corresponding to the first three continuous assessment activities (CAA-1, CAA-2 and CAA-3) would not take the final exam. That information was available for teachers just before mid term. Due to the similarity between the 2018-19 and 2019-20 courses, the teaching team decided that the classifier obtained with data from the 2018-19 cohort was a useful starting point. Nevertheless, they thought it necessary to advance the intervention to just before the end of the CAA-2 in order for students to have enough time to change their practice. Besides, the teaching team took into account Jayaprakash et al.'s (2014) remark that the intervention may be ineffective if students become aware of being at risk of failing after having done a high number of tests. At the end, the teaching team adapted the original classifier, so that students considered to be at risk of dropping out were those who had not submitted the first two quizzes (Q1 and Q2). These students were the ones targeted by the intervention, which was carried out in a realistic learning setting in which: (a) the learning objectives agreed in the enrolment, (b) the course duration, (c) the contractual relationship between the university and teachers, and (d) the specific features of the UOC, such as an online and asynchronous environment, have been maintained. This pilot intervention, which is driven by the data analyzed by Figueroa-Cañas and Sancho-Vinuesa (2019) and is embedded in the LA cycle (Fig.1), can therefore be considered to be an LA intervention. Since it targets at-risk students, the intervention belongs to that of Wilson's (2009) third level.

Mainly, two elements comprised the intervention: an early email message, and a modification of the teaching plan. The latter consisted of extending the deadline for either of the unsubmitted first two quizzes. Regarding the email message, each teacher sent, within two days after the end of week 4, a single, identical message to all predicted at-risk students. The message wording (see Appendix.A) aimed to provide students with information about their predicted risk of not taking the final exam as well as encouraging them to review didactical resources, ask teachers for extra support and submit the unsubmitted quizzes. It was also intended that students should not think of the prediction as unavoidable and, consequently not endeavour to reverse it, which is an actual risk according to Kitto et al. (2016).

The intervention impact was analyzed at the learning outcome level, which is one of the requirements of the A4AEF (Rientes et al., 2016). To be precise, the learning outcome was retention. The aim of the intervention was to lower the number of students who drop out of the course, that is who do not take the final exam. Based on the predictive analysis of Figueroa-Cañas and Sancho-Vinuesa (2019), the rationale of the core of the intervention was that if the number of students who do not take the first two quizzes is reduced, so is the number of students who do not take the final exam. Consequently, the change sought by the intervention was the submission of the quizzes not submitted during the original time period. Not only that, the intervention also tried to keep students involved in doing regular practice tests (quizzes), which the pedagogical model of statistics and other mathematics courses is based on. That approach is in accordance with Dunlosky et al.'s (2013) point of view that practice testing is an effective learning technique, and also with Roediger et al.'s (2011, p.26) remark that "frequent testing encourages students to study".

6.3.3 Data and analysis of the educational intervention

We have defined three different periods of time within the term: *pre-intervention*, *intervention* and *post-intervention* (Fig. 2). The pre-intervention period, which lasted four weeks, refers to the time between the beginning of the course and the end of the second continuous assessment activity (CAA-2). The intervention period is the two weeks in which intervention took place from the end of the CAA-2 to the beginning of the CAA-3. The post-intervention period corresponds to the time elapsed between the beginning of the CAA-3 and the final exam date.

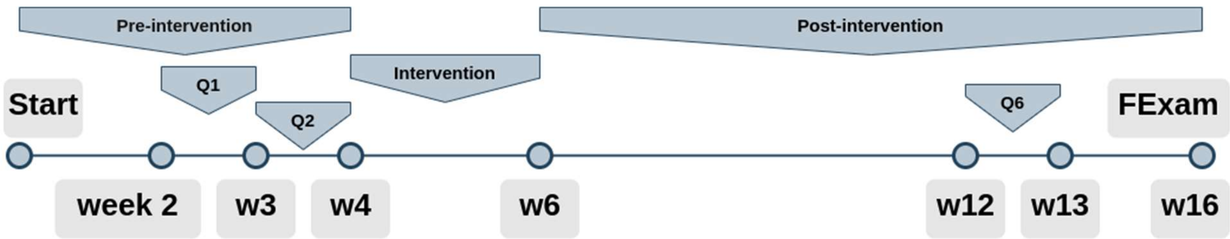


Fig.2 Timeline containing the three periods: pre-intervention, intervention and post-intervention. Quiz 1 (Q1) is open between weeks 2 and 3.

Data collection and analysis was done hierarchically: for the two cohorts, for intervened and potentially intervened students, and for the interviewed students.

The two cohorts

For both cohorts, the 225 students enrolled on the 2019-20 course and the 197 on the 2018-19 one, the participation data in quizzes 1 and 2 was collected. With that, the list of students subject to intervention in the 2019-20 edition was determined, as well as the list of students who would have received intervention in the 2018-19 course if intervention had been implemented then. The latter will be named as potentially intervened students.

The intervened and potentially intervened students

For the students who received intervention on the 2019-20 course we also collected their performance data in quizzes 1 and 2 during the pre-intervention period, their new participation and performance in quizzes 1 and 2 during the intervention period and also their participation and performance in quizzes 3, 4, 5 and 6. Furthermore, we looked at their access to the virtual classroom and the teaching plan for the three periods.

During the aforementioned three periods, we collected data referring to the contact between intervened students and their teachers. The aim was twofold. First, to know whether students had requested curricular support from their teachers, and second, if students had replied to the email during the intervention period.

We considered the intervention to have had an impact on students if they had sent an email reply or had submitted any of the non-submitted quizzes during the intervention period. The reason for that decision was based on the fact that replying to the email message indicated that the message had been read and therefore the student had received the information and encouragement. With the information provided by teachers regarding the response to the intervention, three a priori intervention response categories were defined: (a) *no_evident_impact*, corresponding to the students who did not send an email reply or who did not make any new submissions, (b) *only_email_reply*, formed by the students who did send an email reply, but who did not submit any quizzes; and (c) *new_submission*, corresponding to the students who did submit one or two of the previously non-submitted quizzes. All the information gathered during the pre-intervention period enabled us to put the students into posteriori pre-intervention categories. The relationship between pre-intervention and intervention response categories was analysed and that resulted in a selection of possible factors

that may have influenced the students' decision to complete and submit any previously non-submitted quizzes. In addition, we selected the data related to final exam attendance for the intervened students. That information was also gathered for those potentially intervened students in the 2018-19 course.
An analysis of the relationship between the three response intervention categories and the participation in the final exam was done. Moreover, we compared the two groups of intervened and potentially intervened students regarding their participation in the final exam by means of a quasi-experimental protocol, one that is accepted by the A4AEF (Rienties et al., 2016). This protocol is the most suitable one for our study because the two courses share similar pedagogical factors. Our case is assimilable to that of Marcal (2019). Its study also analysed an intervention conducted on one course that targeted at-risk students, in relation to the same course in a different academic year but without intervention, stressing that both courses kept the pedagogical factors unchanged.

The interviewed students

The researchers sent an email to all 12 students who had shown any kind of response to the intervention inviting them to participate in a semi-structured interview.

Three main topics were covered by the interview (some guidance questions are shown in Appendix.B). The first topic had to do with the external conditions, such as work and family, at the time when students had to decide whether to take the first two quizzes or not. Students, as agents of their own learning, choose to participate in the learning activities or not. These choices are influenced by external conditions according to Gašević et al. (2015). The second topic referred to the EFLA questionnaire about Learning Analytics tools (Scheffel, 2017). In our case, the tool assessed was the email message. The purpose of our questions was to find out whether the message had made the student: (a) aware of the situation regarding the first two continuous assessment activities, (b) aware of their possible future, (c) reflect on their participation in the pre-intervention learning activities, and (d) reflect on how to perform in future learning activities. And the third topic, the cognitive load, was determined by using the analysis carried out to find factors that had influenced the decision to solve quizzes or not. In addition, to close the interview, students were asked about their overall satisfaction with the initiative of allowing them to submit previously unsubmitted quizzes.

Among the 12 students who were asked to be interviewed, 5 accepted: 3 had succeeded in a previous quiz and submitted a new one, 1 had not succeeded in a previous quiz and submitted a new one, and 1 had not succeeded in a previous quiz and only sent an email reply. The interviews, which lasted between 10-18 minutes (an average of 14 minutes), were conducted by means of video meetings and were recorded at the same time. At a later stage, the audio was transcribed by the authors resulting in a naturalized transcription, that is, "keeping the transcripts in their original form" (Da Silva Nascimento & Kalil Steinbruch, 2019, p. 420). A content analysis was adopted, whose text coding has carried out manually because of the low number of interviewees.

6.4 Results

At the end of the pre-intervention period, 35 students in the 2019-20 cohort and 36 in the corresponding 2018-19 one had not submitted both of the first two quizzes. The former were the intervention target students (intervened students) and the latter, the potentially intervened ones.

With the data collected in the pre-intervention period, we defined three pre-intervention categories: (a) *virtual_classroom_not_accessed*, formed by students who had not accessed the virtual classroom during either of the two periods of time in which the first two quizzes were open to be submitted, (b) *no_quiz_success*, corresponding to the students who had accessed the virtual classroom during at least one of the two periods when Q1 and Q2 were open, but had not been awarded a minimum pass score; and (c) *one_quiz_success*, for the students who had passed the only submitted quiz, the Q1 or Q2. Students were distributed as follows (see final column in Table 1): 14 students had not accessed the virtual classroom when the first two quizzes were available (virtual_classroom_not_accessed students), 12 students had not been awarded a minimum pass score, although they had accessed the virtual classroom when the first two quizzes were available (no_quiz_success students); and 9 students had passed the only submitted quiz (one_quiz_success students). As a result, nearly three quarters of the intervened students (26 out of 35) had not had any successful experiences in solving quizzes prior to the intervention.

Second chance and final examination

According to the intervention response (see last row in Table 1), the count is: 23 students did not send an email reply or did not make any new submission (no_evident_impact students), 5 students did send an email reply, but did not submit any previously unsubmitted quiz (only_email_reply students); and 7 students did submit at least one previously unsubmitted quiz (new_submission students). Hence 12 out of 35 intervened students showed some interaction impact, and what is more remarkable, 7 out of the 12 who showed some interaction impact actually did submit a previously unsubmitted quiz. The correspondence between pre-intervention and intervention response categories is shown in Table 1.

Table 1. Cross reference table between pre-intervention and intervention response categories.

	No_evident_ impact	**Only_email_ reply**	**New_submission**	**Total**
Virtual_classroom_ not_accessed	13	1	0	14
No_quiz_success	9	2	1	12
One_quiz_success	1	2	6	9
Total (students)	23	5	7	35

The most salient point to make, which is the first key result, is that 6 students who had succeeded in one submitted quiz, used the chance given by the intervention to go back and submit the other previously unsubmitted quiz (new_submission students). Apart from these 6 students, one student, who had not passed any quiz and had accessed the virtual classroom when the first two quizzes were available (no_quiz_success student), submitted the two unsubmitted quizzes (new_submission student). Only these 7 students submitted any previously unsubmitted quiz.

Therefore, all students who submitted a previously unsubmitted quiz had not had any earlier unsuccessful experiences with quizzes. It seems reasonable to think that one likely factor that positively influences the decision to submit a previously unsubmitted quiz is having obtained good learning outcomes, in our case a positive quiz performance prior to the intervention.

In 2008, Van Gog and Pass pointed out that learning outcomes should include the mental effort made when completing a task or any other measurement of the cognitive load. The cognitive load determines the working memory resources required for performing a task (Kalyuga, 2012). When the cognitive load exceeds the working memory capacity, students can lose motivation to allocate working memory resources, *germane cognitive resources* (Kalyuga, 2012), and do not do learning tasks. In 2018, Costley and Lange found an association between motivation, specifically the intrinsic goal orientation dimension of motivation, and germane cognitive load. We can ask ourselves whether the cognitive load played a role in the students' decision to submit a previously unsubmitted quiz. This is why the interview included items related to that topic. In 2013, Leppink et al. presented a questionnaire to measure the three widely accepted cognitive load types separately: intrinsic, extraneous and germane. In the present study, we chose a question of each kind and took Leppink et al.'s (2013) questionnaire as a reference (i.e related to the intrinsic load: "were the quiz questions complex?").

Let us now focus on the relationship between the students' intervention response and their final exam behaviour, bearing in mind that the main aim of the intervention was to reduce the number of students who drop out of the course, in our case operationalised as the number of students who do not take the final exam. Table 2 shows the distribution of the 2019-20 students between the ones who took it and the ones who did not in relation to their intervention response categories.

Table 2. Distribution of the 2019-20 cohort according to both their final exam behaviour and intervention response category.

	No_evident_ impact	***Some_impact (Only_email_reply + New_submission)***	**Total (students)**
Took final exam	5	5	10
Did not take Final exam	18	7	25
Total	23	12	35

By observing Table 2, regarding final exam participation, there appears to be a difference between students who had not shown any evident intervention impact, that is students who had neither sent an email reply nor made any new submission (no_evident_impact students), and students who had shown some intervention impact (only_email_reply plus new_submission students). Whereas 18 of the 23 students with no sign of intervention impact did not take the final exam, 7 of the 12 students with evident intervention impact did take the final exam. These numbers indicate that the intervention improved the chances of a student taking the final exam, which is the second key result. Nevertheless, by carrying out a Fisher's exact test, which was chosen because of the smallness of the numbers involved, we found a p-value of 0.2554. This means that in fact there is no statistically significant difference between students who had shown and who had not shown intervention impact regarding taking the final exam. Despite acknowledging the lack of statistical significance, the Fisher test result does not negate the importance of the effect the intervention had on the students.

Now we are going to compare the final examination of two cohorts with and without the intervention. In the 2018-19 cohort without the intervention, among the 36 potentially intervened students 4 eventually took the final exam, which corresponds to 11.1%. As can be seen in Table 2, in the 2019-20 cohort with the intervention, 10 out of 35 students took the final exam (28.6%). So the dropout rate between the intervened students in the 2019-20 course and the potentially intervened students in the 2018-19 course fell by 17.5%. Burgos et al. (2018) reported an improvement of 57.1%, whereas Cambruzzi et al. (2015) presented quite contrasting figures, 23% and 7% for two different courses. Our result is clearly below that of Burgos et al. (2018), while in the range of the two courses reported by Cambruzzi et al. (2015).

Student's perspective

When students were asked whether the external conditions were appropriate during the pre-intervention period, 4 interviewees responded negatively and 1 affirmatively. The latter stated that he had not submitted the quiz due to a personal organization problem. The former gave several different reasons: 2 students because of work, 1 for being away from home and 1 for being a foreign student in a new educational system. With regard to the email message: (a) 2 students responded that the message had made them aware of their situation related to the first two continuous assessment activities and 2 students said the opposite due to their knowing about the activities before receiving the message, (b) 1 student said the message made her aware of her possible academic future whereas 2 students said the opposite, (c) 2 students stated that the message had stimulated them to reflect on their participation in the pre-intervention period tasks while 1 student said the opposite because he knew the reason beforehand; and (d) all 5 students said the message encouraged them to reflect on their future behaviour.

Two students did not submit either of the first two quizzes. The remaining three interviewees, who had all passed the one submitted quiz (one_quiz_success students), considered the questions from the first quiz they submitted as neither complex nor difficult to understand. They also stated that their first submitted quiz helped them to understand the course.

The only student who did not submit a previously unsubmitted quiz, argued that she did not do so because "I had to start from scratch". She did not take the final exam either because she thought she was not well prepared enough. One student who did submit a previously unsubmitted quiz eventually did not take the final exam either because "I had disconnected from the course due to my having other things to deal with" after the third continuous assessment activity. The three remaining students took the final exam and passed it. In the final comment of the interview, all the interviewees expressed their overall satisfaction (e.g. "I am thankful for the help received", "It is positive that teachers make an effort so that students can reconnect").

To sum up the students' experience, it is worthy of note that the email message encouraged the interviewees to reflect on their future behaviour and that all the interviewees were satisfied with the intervention.

6.5 Discussion

The first part of the intervention consisted of the teachers emailing a message. From the previous results, we only know for sure that twelve students read it. Reading a message and receiving the information it contains is essentially a *passive* activity. Taking the step forward from reading to submitting means moving from passive to active engagement, which is not always straightforward.

Five students did not progress from the passive reading of the message to the active submission of quizzes. One interviewee pointed to her workload as the reason for not taking action.

Giving information by means of emails or dashboards is not enough to achieve the necessary changes (Jivet et al. 2017; Kitto et al., 2016). In our case, the interviewees stated that the information about their past behaviour contained in the email message did not encourage them to change, but the explicit chance to submit previously unsubmitted quizzes also contained in the message did. Half of the interviewees, who answered the question clearly, said that the email message did not make them aware of their academic situation because they knew about it beforehand. In general, the fact that information generates awareness is taken for granted. The interviewees' responses introduce an element of doubt about this. A primal explanation may emerge, that is to say, students who are aware beforehand of not having submitted the first few quizzes are more prone to submit them, given the chance.

When interviewees were asked whether the email message had both made them aware of their possible future and also caused them to reflect on their participation in the pre-intervention activities, two of them responded in an unclear way. This leads us to deduce that neither question was understood and both must be reformulated in the future. What is beyond doubt is that all five interviewees found the message made them reflect on how to perform in future learning activities, which is not only remarkable but also encouraging because without reflection it is difficult to change anything.

There is another element to be considered, which is the emotional support provided by the email message. Two interviewees stated that the email message had supported them emotionally. This emotional support may have increased their germane load and positively influenced the decision to submit previously unsubmitted quizzes. In 2018, Costley and Lange showed a positive association between motivation and germane load.

The intervention was essentially designed so that students could submit the previously unsubmitted first two quizzes. The rationale of the intervention design is that the submissions lead to a fall in dropout rates, in our case the number of students who do not take the final exam. We considered the submission of previously unsubmitted quizzes as the only observable evidence that the intervention had produced a real change in student behaviour. In consequence, the seven students who actually submitted (20% of the intervened students) showed us that the intervention had produced a noticeable effect.

It is worthy of comment that 20% of the intervened students engaged with the intervention, therefore it can be said that they have been *active*. Burgos et al. (2018), in an intervention conducted via tutorial plan, reported that 100% of their intervened students had engaged because all of them had received support from the teacher. Khan and Pardo (2016), in an intervention consisting of visualizing data on a dashboard, reported that 77% had accessed the dashboard. Pinchbeck and Heaney (2017), in an intervention consisting of attending a synchronous online tutorial and participating in a forum, reported an engagement of 20% on one course and 58% on another. Only the last study implies that students are active by participating in a forum. Its results, on one course, are identical to ours.

Regarding behaviour and performance prior to the intervention, we distinguish four groups of students. First, some students were disconnected from the virtual classroom for the whole period of time in which the first two continuous assessment activities (CAAs) were open. Those students did not show any observable change. Second, we found students who, despite accessing the virtual classroom when one of the first two CAAs was open, did not submit either of the first two quizzes. Of those students, only one submitted previously unsubmitted quizzes. The workload involved in submitting two quizzes in two weeks may be considered to be extremely heavy, as one interviewee said, even though the email message encouraged her. The third group of students submitted one quiz before the intervention, but failed it. None of them submitted any of the previously unsubmitted quizzes. The unsuccessful experience of solving quizzes may be a relevant factor in deciding not to submit them. None of the interviewees belonged to that group. And finally, the fourth group was made up of students with one successful quiz experience. This may be attributable partly to the fact that the questions were neither complex nor difficult to understand, as all three interviewees from that group stated. For the fourth group of students, submission of previously unsubmitted quizzes was an opportunity six students took advantage of, which prevented them from being left behind by their classmates who had submitted the first two quizzes on time. This is one of the key findings of the present study. Other fourth-group students opted not to submit previously unsubmitted quizzes, maybe due to their cognitive load being too high. This is only a hypothesis because no students from that subgroup were interviewed. To sum up, the intervention led to changes in the behaviour of students who were connected to the virtual classroom and had not had any unsuccessful experiences in the first two quizzes. As one interviewee said, the intervention also produced other undetected behavioural changes such as submitting other quizzes different from the first two, or intended changes like preparing for the final exam and eventually not sitting it.

Seven students submitted quizzes thanks to the opportunity given by the intervention, that is, they showed changes in their behaviour. Nevertheless, four took the final exam and the other three did not, hence three dropped out of the course and four completed it. As we mentioned in the results section, the intervention improves the chances of a student taking the final exam, and especially so for those who submitted previously unsubmitted quizzes. The common factor for the four non-dropout students was their willingness to keep doing the four quizzes following the intervention. A lot of time elapsed, as much as eight weeks, between the implementation of the intervention and the final exam. It is hard to believe that the intervention was the only cause for taking or not taking the final exam. During those eight weeks, the causes that led students not to solve the quizzes or not access the virtual classroom could reappear, as one interviewee expressed about his life experience. In any case, solving quizzes was a relevant factor for final exam participation for two interviewees.

6.6 Conclusion and further research

This study is aimed at analysing a pilot intervention, which intended to reduce the dropout rate in a one-semester online statistics course.

This intervention, which targeted at-risk students who were predicted as such based on a previous classifier (Figueroa-Cañas & Sancho-Vinuesa, 2019), constitutes one step in the LA cycle. It was planned to begin early (after week four of the course) and to last two weeks. What is clearly different from other interventions targeting predicted at-risk students is that students are allowed to change their past behaviour, instead of only their future behaviour (e.g. Burgos et al., 2018; Cambruzzi et al., 2015; Corrigan et al., 2015; Jayaprakash et al, 2014). This is the main contribution of this study. If the predictive analysis points to not having submitted quizzes as the factor that determines which students are prone to drop out, it sounds reasonable to permit students to submit the quizzes until an extended deadline, so they can change their past behaviour.

One conclusion reached is that this intervention was effective, considering the 17.5% fall in the dropout rate between the group of intervened students in the 2019-20 course and the group of students who would have been targeted by the intervention in the 2018-19 course. In the end, seven students submitted some of their previously unsubmitted quizzes, a fact which must also be highlighted because these students increased the amount of learning activities they carried out. It is also worthy of mention that the dropout rate is the lowest among these students compared to those who did not submit any previously unsubmitted quizzes. Another conclusion is that the students who are more prone to submit previously unsubmitted quizzes are those who have previously submitted and passed one quiz. We also conclude that students who are disconnected from the virtual classroom when initial quizzes are opened up are unwilling to take active steps towards submitting quizzes. It is reasonable to assume that among that group of students, some have taken the firm decision to discontinue, so no intervention is effective for them. Notwithstanding this, two limitations have to be pointed out. First, the low number of students subject to the intervention. And second, the likely interference of other factors in the association between the intervention and the dropout rate, such as the continued effort required in solving quizzes.

In order to improve the intervention, it would be necessary to know for certain which students have firmly decided to discontinue. This could be achieved by the implementation of a supplemental tutorial plan, given the outstanding results of Burgos et al. (2018). Tutors would contact all at-risk students to find out whether they are convinced about dropping out or to persuade them to submit quizzes. Maybe one single email is not enough to trigger the change. Furthermore, the tutors would extend the interaction with their students for the remainder of the course so as to dissuade them from dropping out.

Some possible lines of future research are suggested as follows. First, as the students targeted by this intervention are selected from an adaptation of a classifier, searching for an improved classifier would also enhance the effectiveness of the intervention. And second, the intervention could be used in other courses based on continuous assessment by means of quizzes in order to contrast their results with ours.

Acknowledgment

This research was partially supported by a Fundació IBADA grant. We would like to thank Mr Paul Garbutt for their valuable contributions in helping to improve this study.

References

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. *ACM International Conference Proceeding Series*, (April 2012), 267–270. https://doi.org/10.1145/2330601.2330666

Attree, K., Johnston, A., & Livermore, G. (2017). Just a phone call away : the impact of academic intervention on retention and success for repeat fail students in the distance education environment, (July 2014), 1–5. 2014), Proceedings of the 17th International FYHE Conference 2014 (pp.1–5). Queensland University of Technology.

Baker, R., Evans, B., Li, Q., & Cung, B. (2019). *Does Inducing Students to Schedule Lecture Watching in Online Classes Improve Their Academic Performance? An Experimental Analysis of a Time Management Intervention. Research in Higher Education* (Vol. 60). Springer Netherlands. https://doi.org/10.1007/s11162-018-9521-3

Broos, T., Pinxten, M., Delporte, M., Verbert, K., & De Laet, T. (2020). Learning dashboards at scale: early warning and overall first year experience. *Assessment and Evaluation in Higher Education*, *45*(6), 855–874. https://doi.org/10.1080/02602938.2019.1689546

Burgos, C., Campanario, M. L., Peña, D. de la, Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers and Electrical Engineering*, *66*, 541–556. https://doi.org/10.1016/j.compeleceng.2017.03.005

Cambruzzi, W., Rigo, S. J., & Barbosa, J. L. V. (2015). Dropout prediction and reduction in distance education courses with the learning analytics multitrail approach. *Journal of Universal Computer Science*, *21*(1), 23–47.

Cameron, M. P., & Siameja, S. (2017). An experimental evaluation of a proactive pastoral care initiative within an introductory university course. *Applied Economics*, *49*(18), 1808–1820. https://doi.org/10.1080/00036846.2016.1226492

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. *ACM International Conference Proceeding Series*, 134–138. https://doi.org/10.1145/2330601.2330636

Corrigan, O., Glynn, M., McKenna, A., Smeaton, A., & Sinead, S. (2015). Student Data: Data is knowledge – putting the knowledge back in the students' hands. *Proceedings of the European Conference on E-Learning, ECEL*, 165–172.

Costley, J., & Lange, C. (2018). The moderating effects of group work on the relationship between motivation and cognitive load. *International Review of Research in Open and Distance Learning*, *19*(1), 68–90. https://doi.org/10.19173/irrodl.v19i1.3325

Da Silva Nascimento, L., & Kalil Steinbruch, F. (2019). "The interviews were transcribed", but how? Reflections on management research. *RAUSP Management Journal*, *54*(4), 413–429. https://doi.org/10.1108/RAUSP-05-2019-0092

Davis, D. R., & Abbitt, J. T. (2013). An investigation of the impact of an intervention to reduce academic procrastination using short message service (SMS) technology. *Journal of Interactive Online Learning*, *12*(3), 78–102.

Delnoij, L. E. C., Dirkx, K. J. H., Janssen, J. P. W., & Martens, R. L. (2020). Predicting and resolving non-completion in higher (online) education – A literature review. *Educational Research Review*, *29*(January), 100313. https://doi.org/10.1016/j.edurev.2020.100313

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, Supplement*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Figueroa-Cañas, J. & Sancho-Vinuesa, T. (2019). Predicting early dropout students is a matter of checking completed quizzes: The case of an online statistics module. Proceedings of the Learning Analytics Summer Institute Spain 2019: Learning Analytics in Higher Education, Vigo, Spain, June 27-28, 2019, (pp. 100-111), CEUR-WS.org/Vol-2415/paper09.pdf.

Foster, C., & Francis, P. (2020). A systematic review on the deployment and effectiveness of data analytics in higher education to improve student outcomes. *Assessment and Evaluation in Higher Education*, *45*(6), 822–841. https://doi.org/10.1080/02602938.2019.1696945

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1). https://doi.org/10.1007/s11528-014-0822-x

Herodotou, C., Heiser, S., & Rienties, B. (2017). Implementing randomised control trials in open and distance learning: a feasibility study. *Open Learning*, *32*(2), 147–162. https://doi.org/10.1080/02680513.2017.1316188

Huett, J. B., Kalinowski, K. E., Moller, L., & Huett, K. C. (2008). Improving the motivation and retention of online students through the use of arcs-based e-mails. *American Journal of Distance Education, 22(3).International*https://doi.org/10.1080/08923640802224451

Hulleman, C. S., Kosovich, J. J., Barron, K. E., & Daniel, D. B. (2017). Making connections: Replicating and extending the utility value intervention in the classroom. *Journal of Educational Psychology*, *109*(3), 387–404. https://doi.org/10.1037/edu0000146

Inkelaar, T., & Simpson, O. (2015). Challenging the distance education deficit through motivational emails. *Open Learning: The Journal of Open, Distance and E-Learning*, *30*(2), 152–163.

Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47. https://doi.org/10.18608/jla.2014.11.3

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness Is Not Enough : Pitfalls of Learning Practice. *Data Driven Approaches in Digital Education*, *1*(i), 82–96. https://doi.org/10.1007/978-3-319-66610-5

Kalyuga, S. (2012). Interactive distance education: A cognitive load perspective. *Journal of Computing in Higher Education*, *24*(3), 182–208. https://doi.org/10.1007/s12528-012-9060-4

Khan, I., & Pardo, A. (2016). Data2U: Scalable real time student feedback in active learning environments. *ACM International Conference Proceeding Series*, *25-29-NaN-2016*, 249–253. https://doi.org/10.1145/2883851.2883911

Kitto, K., Lupton, M., Davis, K., & Waters, Z. (2016). Incorporating student-facing learning analytics into pedagogical practice. In S. Barker, S. Dawson, A. Pardo, & C. Colvin (Eds.), Show me the Learning. Proceedings ASCILITE *ASCILITE 2016*, (pp. 338–347).

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior*, *47*, 90–97. https://doi.org/10.1016/j.chb.2014.07.013

Lunsford, M. L., Poplin, P. L., & Pederson, J. E. G. (2018). From Research to Practice: Using Assessment and Early Intervention to Improve Student Success in Introductory Statistics. *Journal of Statistics Education*, *26*(2), 125–134. https://doi.org/10.1080/10691898.2018.1483785

Marcal, L. (2019). Early Alert System Pilot in a Microeconomics Principles Course. *Research in Higher Education Journal*, *37*, 1–14.

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology*, *50*(1), 128–138. https://doi.org/10.1111/bjet.12592

Pinchbeck, J., & Heaney, C. (2017). Case report: the impact of a resubmission intervention on level 1 distance learning students. *Open Learning*, *32*(3), 236–242. https://doi.org/10.1080/02680513.2017.1348290

Rahal, A., & Zainuba, M. (2016). Improving students' performance in quantitative courses: The case of academic motivation and predictive analytics. *International Journal of Management Education*, *14*(1), 8–17. https://doi.org/10.1016/j.ijme.2015.11.003

Rienties, B., Boroowa, A., Cross, S., Kubiak, C., Mayles, K., & Murphy, S. (2016). Analytics4Action Evaluation Framework: A Review of Evidence-Based Learning Analytics Interventions at the Open University UK. *Journal of Interactive Media in Education*, *2016*(1), 1–11. https://doi.org/10.5334/jime.394

Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). *Ten Benefits of Testing and Their Applications to Educational Practice. Psychology of Learning and Motivation, 55, 1-36.* https://doi.org/10.1016/B978-0-12-387691-1.00001-6

Scheffel, M. (2017). *The Evaluation Framework for Learning Analytics*. Open Universiteit and SIKS.

Sønderlund, A. L., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, *50*(5), 2594–2618. https://doi.org/10.1111/bjet.12720

Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16–26. https://doi.org/10.1080/00461520701756248

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning Analytics Dashboard Applications. *American Behavioral Scientist*, *57*(10), 1500–1509. https://doi.org/10.1177/0002764213479363

Wilson, K. (2009). The impact of institutional, programmatic and personal interventions on an effective and sustainable first-year student experience. *First Year in Higher Education Conference, (pp. 1–19). Queensland University of Technology.*

Wise, A., Zhao, Y., & Hausknecht, S. (2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics*, *1*(2), 48–71. https://doi.org/10.18608/jla.2014.12.4

Wong, B. T-M, & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education*, *7*(1), 7–28. https://doi.org/10.1007/s40692-019-00143-7

6.7 Appendices

A. Email message (translated from the original language)

Dear student,

As a result of the analysis carried out on previous statistics courses, it has been seen that students who do not submit the first three quizzes corresponding to the first three Continuous Assessment Activities (CAA) are unlikely to take the final exam. Taking your current situation into account, I encourage you to review the teaching resources of the course, express any doubts with your teacher and submit the quizzes previously unsubmitted. You have one single attempt for each quiz*. I am confident that by doing this you can resume the course. The new deadline ends on October 29th, 2019.

If you have any doubts or comments, please do not hesitate to contact me.

Yours sincerely,
the teacher

* The maximum score is 5 out of 10.

P.S. I would appreciate a reply to this message so as to check if you have received it.

B. Guidance questions for the interview (translated from the original language)

Some examples of the interview questions:

- Were the external conditions appropriate?
- Did the email message make you aware of the situation regarding the first two continuous assessment activities?

- Did the email message make you aware of your possible future?
- Did the email message make you reflect on your participation in the pre-intervention learning activities?
- Did the email message make you reflect on how to perform in future learning activities?
- Were the quiz questions complex?
- Were the quiz questions hard to understand?
- Did the quiz questions help you improve your understanding of the topic covered?
- What is your overall level of satisfaction regarding the proposal?

CAPÍTOL 7

RESULTATS I DISCUSSIÓ

Tal com hem apuntat en el capítol 2 de presentació de les publicacions, quatre són els resultats més destacats d'aquesta tesi. Els dos primers corresponen a resultats obtinguts en l'assignatura d'anàlisi matemàtica, mentre que els dos últims, fan referència a l'assignatura d'estadística. A continuació mostrem els quatre resultats.

Resultat 1 (R1).- Els estudiants que realitzen qüestionaris d'entrenament obtenen millors resultats als qüestionaris de mesura.

Resultat 2 (R2).- La qualificació a l'examen final està relacionada amb el fet de realitzar qüestionaris d'entrenament i amb les qualificacions als qüestionaris de mesura.

Resultat 3 (R3).- No presentar-se a l'examen final o no reeixir-hi és predictible a partir de les qualificacions als primers qüestionaris de mesura.

Resultat 4 (R4).- La intervenció docent[11] augmenta la probabilitat de reduir el nombre d'estudiants que no es presenten a l'examen final.

En els següents quatre apartats d'aquest capítol desenvolupem cadascun dels quatre resultats anteriors, altres resultats trobats en els estudis on han estat publicats, així com la seva discussió. Cloem el capítol amb un apartat en què efectuem una discussió conjunta de tots els resultats.

[11] La secció 7.4 conté la descripció de la intervenció docent

7.1 Qüestionaris de mesura i la seva relació amb qüestionaris d'entrenament

El que segueix en aquesta secció correspon a la síntesi dels resultats més destacats, juntament a la seva discussió, del treball presentat a Figueroa-Cañas i Sancho-Vinuesa (2017).

L'anàlisi realitzat, sobre la relació entre l'ús de qüestionaris d'entrenament i resultats en qüestionaris de mesura a l'interior de l'avaluació continuada, s'ha fet sobre 116 estudiants d'anàlisi matemàtica del curs 2015-16. L'avaluació continuada consisteix en la realització de **7 paquets d'activitats**, cadascun format per un qüestionari de mesura i un o diversos qüestionaris d'entrenament. En aquest estudi hem comparat per cadascun dels paquets, la mitjana de les qualificacions dels qüestionaris de mesura obtingudes pels estudiants que havien realitzat qüestionaris d'entrenament, respecte de la mitjana dels estudiants que no n'havien realitzat cap. Els primers estudiants s'anomenen estudiants *actius*, seguint la terminologia de Huisman i Reedijk (2012), i els últims, estudiants *inactius*. Els estudiants actius han practicat, una o més vegades, amb qüestionaris d'entrenament abans d'enfrontar-se amb els qüestionaris de mesura, mentre que els inactius, es llancen directament a la seva resolució. Els actius mostren un major compromís amb la metodologia docent atès que en realitzar qüestionaris dels dos tipus adopten el comportament recomanat per la pròpia metodologia. D'alguna manera, volem esbrinar, per tant, si són diferents les qualificacions aconseguides als qüestionaris de mesura, entre els estudiants més compromesos amb la metodologia i els menys compromesos.

S'observa resumidament que **els estudiants que practiquen amb qüestionaris d'entrenament obtenen millors resultats als qüestionaris de mesura que els seus companys que no practiquen (Resultat 1)**. Es constata que la mitjana de qualificacions dels estudiants actius és superior a la dels estudiants inactius en els set paquets d'activitats. A més, en quatre dels set, la mitjana dels estudiants actius és significativament superior amb un nivell de confiança del 95%, fet que afegeix més consistència al resultat. En cinc paquets d'activitats, els qüestionaris de mesura contenen preguntes de dues tipologies distintes, de correcció automàtica i d'exposició raonada amb correcció manual. En base a això, comparem també les mitjanes obtingudes en les preguntes de cada tipus. Pel que fa a les de correcció automàtica, els estudiants actius presenten una mitjana superior en tots els cinc paquets d'activitats i en tres d'aquest amb una diferència significativament superior amb un nivell de confiança del 95%. Per contra, en les preguntes d'exposició raonada, si bé en quatre paquets d'activitats la mitjana és superior en els estudiants actius, les diferències no arriben a ser significativament superiors en cap dels quatre paquets. Per altra banda, la diferència en les mitjanes dels estudiants actius és més accentuada en les preguntes de correcció automàtica respecte de les d'exposició raonada en els tres paquets d'activitats que mostren una diferència estadísticament significativa.

Discussió

A la vista del resultat 1, podem pensar que, en general, **la pràctica** fa que els estudiants obtinguin millors resultats en els qüestionaris de mesura. Quan observem els resultats distingint en el tipus de preguntes d'aquests qüestionaris, mentre que en les preguntes de correcció automàtica s'ha detectat una diferència significativa de qualificacions entre estudiants que practiquen i que no practiquen, no ha estat així en les preguntes d'exposició raonada. En conseqüència el tipus de preguntes té incidència en la millora dels resultats. El fet que els qüestionaris d'entrenament només contenen preguntes de correcció automàtica (d'elecció múltiple, d'aparellament, de resposta breu) fa que en realitzar aquest tipus de qüestionari només es practiquin les preguntes de correcció automàtica. Això dona força a considerar la pràctica com un element essencial en l'obtenció de millors resultats. Per altra banda, l'absència de preguntes d'exposició raonada en els qüestionaris d'entrenament fa que els estudiants no rebin feedback de com respondre preguntes d'aquest tipus. És a dir, no reben retroalimentació de com explicar i escriure els passos intermedis seguits en la resolució de les activitats. Així doncs, la provisió de feedback en les preguntes de correcció automàtica constitueix un element diferencial respecte de les d'exposició raonada. Sembla raonable pensar, doncs, que aquest element diferencial, **el feedback**, podria també incidir que els estudiants actius mostrin, de forma estadísticament significativa, millors resultats en les de correcció automàtica que en les altres. Per últim, algunes característiques personals dels estudiants, com ara els coneixements previs d'anàlisi matemàtica o les càrregues familiars i/o professionals, no varien quan aquests s'enfronten, en els qüestionaris de mesura, a una pregunta de correcció automàtica o d'exposició raonada. Això atorgaria major rellevància a l'element que sí varia, és a dir la pràctica d'un tipus de preguntes (de correcció automàtica) i no de les altres. En parlar de característiques personals no incloem la motivació atès que no es pot assegurar que aquesta sigui la mateixa quan un estudiant intenta realitzar un tipus de pregunta o una altra. És més, es podria pensar que en les preguntes d'aparellament o d'elecció múltiple la motivació podria ser més alta que en les de resposta breu o d'exposició raonada.

7.2 Els qüestionaris, d'entrenament i de mesura, en relació a l'examen final

El que segueix en aquesta secció correspon a la síntesi dels resultats més destacats, juntament a la seva discussió, del treball presentat a Figueroa-Cañas i Sancho-Vinuesa (2021a).

L'anàlisi realitzat, sobre la relació entre l'ús de qüestionaris d'entrenament, qualificació dels qüestionaris de mesura i qualificació a l'examen final; s'ha dut a terme sobre 176 estudiants d'anàlisi matemàtica del curs 2016-17. La realització de qüestionaris d'entrenament es pot considerar com una variable relacionada amb el **comportament** de l'estudiant (Rienties et al., 2017), mentre que la qualificació obtinguda als qüestionaris de mesura es correspondria amb una variable relacionada amb el **coneixement** ("cognition") de l'estudiant (Rienties et al., 2017). El treball de Figueroa-Cañas i Sancho-Vinuesa (2021a) es pregunta, doncs, si les qualificacions assolides per l'estudiant a l'examen final estan relacionades amb el seu comportament mostrat al llarg del curs i amb el coneixement demostrat durant el curs.

En base a una regressió lineal múltiple proposada per Angus i Watson (2009), s'ha determinat que **la qualificació a l'examen final està relacionada amb el fet de realitzar qüestionaris d'entrenament i amb les qualificacions als qüestionaris de mesura (Resultat 2)**. Per tant, efectivament, les qualificacions assolides per l'estudiant a l'examen final estan relacionades amb el seu comportament mostrat al llarg del curs i amb el coneixement demostrat durant el curs.

Quan la metodologia emprada es basa en una regressió lineal, cal introduir-hi variables de control per tal que les qualificacions a l'examen final es puguin atribuir a les variables que es pretenen analitzar (realització de qüestionaris d'entrenament i qualificacions als qüestionaris de mesura). La competència matemàtica prèvia ("prior mathematical competence") de l'estudiant és una variable que Angus i Watson (2009), Johnson i Mckenzie (2013) i Palocsay i Stevens (2008) consideren cal incloure com a variable de control en estudis de rendiment acadèmic. L'assignatura d'anàlisi matemàtica del curs 2016-17 incloïa una prova inicial opcional sobre coneixements d'anàlisi matemàtica de nivell d'estudis de secundària. Hem considerat que la qualificació d'aquesta prova inicial constituïa una variable relacionada amb la competència matemàtica prèvia. Atès que es tracta d'una prova opcional, dels 176 estudiants de l'estudi, un total de 115 estudiants l'han realitzada. Pel fet de disposar de les qualificacions de la prova inicial de coneixements previs per a 115 estudiants, hem dividit l'estudi en dues parts. En la primera s'analitza el grup complet de 176 estudiants i en la segona, només el grup reduït de 115 estudiants. En tots dos grups s'han observat dues relacions entre qüestionaris i examen final. La primera és que com més gran és el nombre d'activitats dels qüestionaris d'entrenament realitzades, més alta és la qualificació a l'examen final. La segona, que com més altes són les qualificacions en els qüestionaris de mesura, més alta és la qualificació a l'examen final. Per a aquest grup reduït de 115 estudiants, s'ha observat una tercera relació: com més alta és la qualificació a la prova inicial de coneixements, més alta és la qualificació a l'examen final.

Discussió

Al grup reduït de 115 estudiants, havent tingut en compte la qualificació en la prova inicial de coneixements com a variable de control, fa que les dues relacions trobades entre qüestionaris i examen final no es puguin atribuir a la competència matemàtica prèvia de l'estudiant. Això dona més rellevància a les pròpies relacions.

En tant que les qualificacions quantifiquen aprenentatges assolits, podríem afirmar que els aprenentatges assolits a cadascun dels instruments d'avaluació (qüestionaris de mesura i examen final) estan relacionats. Cal valorar l'afirmació anterior en tant que els aprenentatges assolits es demostren en dos instruments diametralment oposats. Veiem-ne les principals diferències. Per una banda, l'examen final consta d'una única prova, amb vigilància, amb poc temps (2 hores), sobre tot el contingut de l'assignatura i amb preguntes exclusivament d'exposició raonada. Per l'altra, els qüestionaris de mesura consten de diverses proves, sense vigilància, amb molt temps (dies), sobre una part del contingut i amb preguntes minoritàriament d'exposició raonada.

7.3 Els qüestionaris de mesura permeten fer prediccions sobre l'examen final

El que segueix en aquesta secció correspon a la síntesi dels resultats més destacats, juntament a la seva discussió, dels treballs presentats a Figueroa-Cañas i Sancho-Vinuesa (2019) i a Figueroa-Cañas i Sancho-Vinuesa (2020).

En una primera aproximació, el problema de predir de manera primerenca els estudiants en risc que abandonin l'assignatura ha estat abordat sobre 197 estudiants d'estadística del curs 2018-19 (Figueroa-Cañas i Sancho-Vinuesa, 2019). La presentació o no a l'examen final és considerada com una característica definitòria de l'abandonament de l'assignatura ("course dropout") per part de l'estudiant, segons Burgos et al. (2018) i Lara et al. (2014). El nostre resultat més destacat és que **no presentar-se a l'examen final és predictible si detectem que no han estat realitzats els primers qüestionaris de mesura.**

Del treball resulten quatre predictors, cadascun executable en l'instant en què finalitza cadascuna de les quatre primeres proves d'avaluació continuada. Els predictors es generen en forma d'arbre de decisió en què la realització dels qüestionaris de mesura és l'únic atribut amb capacitat predictiva. Concretament, els predictors determinen que aquells estudiants que no realitzen els qüestionaris de mesura anteriors a la data de finalització del lliurament de cadascuna de les quatre primeres proves d'avaluació continuada no es presentaran a l'examen final. Per tant, no presentar-se a l'examen final és predictible a partir del comportament als primers qüestionaris de mesura. Les mesures de rendiment de predicció (exactitud, precisió i sensibilitat) del predictor executable a l'acabament de la tercera prova d'avaluació continuada assoleixen valors molt satisfactoris. En la construcció dels predictors es van incloure inicialment altres atributs, que l'algoritme de generació dels predictors ha desestimat en front de la ja esmentada realització de qüestionaris de mesura. Entre els atributs descartats per l'algoritme trobem atributs relacionats amb indicadors de compromís a l'aula virtual (visualització del pla docent, accessos tant al tauler virtual de l'assignatura com al fòrum de discussió) i atributs de caràcter personal i invariant des de l'inici de la matriculació (si l'estudiant és repetidor/a i el nombre d'assignatures matriculades en el semestre).

En una segona aproximació, més robusta, sobre els mateixos 197 estudiants d'estadística del curs 2018-19 s'ha abordat un altre cop el problema de predir de manera primerenca els estudiants en risc que abandonin l'assignatura i el nou problema de predir els estudiants que no reeixiran a l'examen final (Figueroa-Cañas i Sancho-Vinuesa, 2020). El nostre resultat més destacat és que **no presentar-se a l'examen final o no reeixir-hi és predictible a partir de les qualificacions als primers qüestionaris de mesura (Resultat 3).**

S'observa que, considerant atributs relacionats amb les qualificacions en els qüestionaris de mesura en lloc de amb la realització en els mateixos qüestionaris, s'obtenen predictors amb millors mesures de rendiment de predicció. Els dos predictors amb millors mesures de rendiment s'obtenen quan acaba la tercera prova d'avaluació contínua. Aquests predictors, en forma d'arbre de decisió, tenen com a únic atribut amb capacitat predictiva la mitjana dels tres primers qüestionaris de mesura. Tal com havíem mencionat anteriorment, els predictors també han desestimat altres atributs (relacionats amb indicadors de compromís a l'aula virtual i de caràcter personal i invariant des de l'inici de la matriculació). El primer predictor determina que els estudiants que abandonaran l'assignatura són aquells/es que obtenen una mitjana dels tres primers qüestionaris de mesura no superior a 6,3. El segon predictor estableix que els estudiants que no reeixiran a l'examen final són aquells/es que assoleixen una mitjana dels tres primers qüestionaris de mesura no superior a 7,8**.** Així, s'observa que la diferència entre predir com a estudiant en rics d'abandonar (no presentar-se a l'examen) i de no reeixir a l'examen és d'1,5 punts.

Discussió

Els algoritmes basats en arbres de decisió tenen l'avantatge que produeixen predictors simples i fàcils d'interpretar (James et al., 2013). Els predictors obtinguts en Figueroa-Cañas i Sancho-Vinuesa (2019) i Figueroa-Cañas i Sancho-Vinuesa (2020) resulten ser extremadament simples. Això permet reduir el temps necessari per recollir, processar i executar els predictors. A més, atès que només depèn de dades relacionades amb els qüestionaris de mesura, el mateix professorat de l'assignatura pot recollir la informació directament accedint al registre d'activitat del Moodle de l'aula i, posteriorment, executar els predictors. El fet que disposi d'aquesta informació abans de la meitat del curs, facilita fer una intervenció a temps que reverteixi la situació de risc de no superar l'assignatura.

En el cas del predictors de Figueroa-Cañas i Sancho-Vinuesa (2020), observem que la diferència entre el predictor que determina si un estudiant no es presenta a l'examen o no el supera rau en el valor de la mitjana dels tres primers qüestionaris de mesura. De manera que els estudiants susceptibles de no presentar-se també són susceptibles de no reeixir. Com que la realitat és que els estudiants que no es presenten a l'examen final formen part del grup dels que no reïxen, el predictor fa una predicció que resulta coherent amb la realitat.

7.4 Una intervenció docent ajuda a reduir l'abandonament

El que segueix en aquesta secció correspon a la síntesi dels resultats més destacats, juntament a la seva discussió, del treball presentat a Figueroa-Cañas i Sancho-Vinuesa (2021b).

En l'assignatura d'estadística del primer semestre del curs 2019-20 es va implementar una intervenció docent sobre 35 estudiants que no havien lliurat els dos primers qüestionaris de mesura, fet que els classificava com a estudiants en risc de no presentar-se a l'examen final, i en conseqüència a abandonar l'assignatura. La intervenció consistia en l'enviament d'un correu electrònic informant-los de no haver lliurat tots dos qüestionaris de mesura, de la predicció que probablement no es presentarien a l'examen final i de l'oportunitat que se'ls oferia de realitzar aquell/s qüestionari/s de mesura no lliurats. En l'acabament del missatge, se'ls demanava que retornessin un altre missatge de resposta, de manera anàloga a un acusament de recepció. El resultat més destacat és que **la intervenció docent augmenta la probabilitat de reduir el nombre d'estudiants que no es presenten a l'examen final (Resultat 4).**

Hem considerat que l'estudiant ha mostrat algun signe d'impacte de la intervenció si ha respost el missatge o bé ha lliurat algun qüestionari de mesura dels no realitzats prèviament. S'observa que entre els estudiants que han mostrat algun signe d'impacte el percentatge de presentats a l'examen final és superior al percentatge obtingut entre els estudiants que no han presentat cap signe. El Resultat 4 es veu refermat quan comparem dues cohorts d'estudiants de dues promocions de la mateixa assignatura (primer semestre del curs 2018-19 i primer semestre del curs 2019-20) amb un context d'aprenentatge idèntic i un nombre d'estudiants matriculats similar. S'ha observat una disminució de l'11% en la taxa d'estudiants no presentats a l'examen final entre els estudiants de la cohort del curs 2018-19 susceptibles a haver rebut la intervenció docent, respecte la cohort del curs 2019-20 que sí la van rebre.

A més, cal destacar que 6 estudiants que havien reeixit en l'únic qüestionari lliurat van aprofitar l'oportunitat que se'ls oferia, i van lliurar el qüestionari que tenien pendent. Aquests estudiants representen dues terceres parts dels estudiants que havien aprovat l'únic qüestionari lliurat.

A banda dels resultats anteriors derivats de l'anàlisi quantitativa, des de l'experiència dels estudiants recollida a través de l'entrevista, observem que el missatge va encoratjar a reflexionar sobre el seu comportament futur a tots 5 estudiants entrevistats. Cal també mencionar que tots els entrevistats van expressar la seva satisfacció per l'oportunitat que se'ls havia ofert.

Discussió

No tots els estudiants que van respondre el missatge, fet que demostra que l'havien llegit, van lliurar els qüestionaris no realitzats. La lectura pot ser considerada com un comportament passiu, mentre que la realització dels qüestionaris es tracta d'un comportament actiu. Avançar des d'un comportament passiu cap a un actiu no resulta una tasca directa. Jivet et al. (2017) considera que només subministrar informació no és suficient per garantir que els estudiants facin el pas de modificar el seu comportament. La nostra experiència va en la línia de Jivet et al. (2017).

La resposta activa a la intervenció docent, haver realitzat qüestionaris pendents, ha estat seguida pel 20% dels estudiants que l'han rebuda. Cal posar en valor aquesta dada en referència a les xifres reportades en una intervenció en què la resposta activa era la participació en un fòrum de discussió (Pinchbeck i Heaney, 2017). I més, tenint en compte que probablement resulta més complicat resoldre un qüestionari que participar en un fòrum.

Cap dels estudiants que han lliurat un sol qüestionari i l'han suspès ha tingut una resposta activa a la intervenció. En canvi, dues terceres parts dels estudiants que han lliurat un sol qüestionari i l'han aprovat han tingut una resposta activa. En realitat, cap dels estudiants amb una resposta activa ha tingut una experiència no reeixida amb els qüestionaris. Semblaria raonable que aquest fet tingués alguna incidència en la decisió de respondre activament a la intervenció. És a dir, una mala experiència rebuda podria conduir a no realitzar els qüestionaris pendents, mentre que una bona experiència podria induir al contrari.

La resposta activa no condueix a presentar-se a l'examen final (no tots els estudiants que van completar els qüestionaris pendents s'hi van presentar). Aquells que sí s'hi van presentar van continuar fent els quatre qüestionaris restants programats. Un cop finalitzat el breu període de quinze dies en què es desenvolupa la intervenció, resten vuit setmanes fins al dia de l'examen final. Resulta difícil pensar que la intervenció és l´única causa per a presentar-se o no a l'examen final donada la separació temporal. Sembla més raonable que, finalitzada la intervenció, mantenir-se enganxat a la realització de qüestionaris tingui una influència més gran.

7.5 Discussió conjunta dels resultats obtinguts

Els resultats mencionats anteriorment, resumidament, estableixen que la realització de qüestionaris durant el semestre està associada amb els resultats acadèmics. Els resultats acadèmics de l'estudiant són producte del procés d'ensenyament/aprenentatge, que segons Anderson (2003) se sustenta sobre una xarxa complexa d'interaccions. En aquest sentit, podem pensar que els resultats acadèmics tenen a veure amb aquesta xarxa complexa d'interaccions. Si considerem la realització de qüestionaris com una manifestació de les interaccions entre estudiant i sistema d'avaluació, podem dir que aquestes interaccions tenen incidència sobre els resultats acadèmics. És més, el nostre estudi ens permet afirmar que les interaccions entre estudiant i sistema d'avaluació juguen un paper més destacat que altres interaccions com les produïdes entre estudiant i interfície o estudiant i altres recursos no didàctics, si més no al començament del curs. Prova d'això seria que el predictor primerenc, determinat abans de la meitat del curs en l'assignatura d'estadística, de no presentar-se a l'examen final o de no aprovar-lo, inclou exclusivament la realització de qüestionaris. Aquest predictor, per contra, no incorpora accessos a l'aula virtual, interpretables com a manifestacions d'interaccions estudiant i interfície (Joksimović et al., 2015), ni accessos al pla docent, interpretables com interaccions estudiant i altres recursos no didàctics (Sabry i Baldwin, 2003).

El treball realitzat en relació a la intervenció docent sobre un grup d'estudiants d'estadística ens ha aportat evidències sobre l'experiència, la percepció de l'estudiant. Quan els entrevistats afirmen que el missatge rebut els ha fet reflexionar sobre el comportament futur, evidencien interaccions estudiant-estudiant, és a dir, aquelles que l'estudiant té amb ell mateix (Moore, 1989). Tot i així, aquestes interaccions estudiant-estudiant, produïdes durant el breu període en què ha tingut lloc la intervenció, no han estat capaces d'assegurar que l'estudiant es presenti a l'examen final. El llarg període transcorregut (més de vuit setmanes) entre aquestes interaccions i la decisió de presentar-se o no a l'examen fa que, probablement, la seva influència es vagi esmorteint a mesura que passa el temps.

Els estudiants que han respost activament a la intervenció també han generat noves interaccions estudiant i sistema d'avaluació perquè han realitzat un qüestionari no fet prèviament. Altre cop, són aquestes les interaccions més determinants. Això ho prova el fet que d'entre tots els estudiants que han respost activament a la intervenció, el factor comú dels que finalment s'han presentat a l'examen final ha estat la persistència en produir aquestes interaccions fins a final de curs, realitzant tota la resta de qüestionaris programats. Semblaria, doncs, que mantenir les interaccions estudiant i sistema d'avaluació durant tot el curs, amb la realització continuada de qüestionaris, resulta de gran rellevància per tal de no abandonar l'assignatura.

CAPÍTOL 8

CONCLUSIONS

Aquesta tesi és fruit del nostre treball realitzat al llarg dels darrers sis anys, materialitzat en quatre publicacions. Es va iniciar amb la finalitat d'aprofundir en l'estudi de l'eficàcia d'una metodologia docent per a l'ensenyament/aprenentatge de matemàtiques superiors en línia basada en la realització continuada de qüestionaris amb correcció i feedback automàtics. Estudis previs (Calm et al., 2013; Escudero-Viladoms, 2012; Escudero-Viladoms i Sancho-Vinuesa, 2016; Sancho-Vinuesa i Escudero-Viladoms, 2012; Sancho-Vinuesa et al., 2013; Sancho-Vinuesa et al., 2018) havien apuntat la bondat d'aquesta metodologia a diferents assignatures de matemàtiques dels estudis d'enginyeria de la UOC. El nostre treball pretenia reavaluar la bondat apuntada sota una perspectiva diferent, tot profunditzant en el coneixement de les relacions entre qüestionaris amb correcció i feedback automàtics i resultats acadèmics. Hem adoptat una perspectiva basada en l'evidència, en l'ús de les dades del procés d'ensenyament-aprenentatge: les analítiques d'aprenentatge. Concretament, hem definit una seqüència de cicles encadenats basats en el cicle de les analítiques d'aprenentatge de Clow (2012). Aquesta aproximació ens ha permès, de manera progressiva, comprendre millor com les interaccions que tenen lloc entre l'estudiant i el sistema d'avaluació d'assignatures de matemàtiques del grau d'enginyeria informàtica de la UOC incideixen en el seu aprenentatge. En termes generals, el que hem après en cloure la tesi és que **mantenir el compromís al llarg del curs amb la realització de qüestionaris amb correcció i feedback automàtics ajuda l'estudiant a l'assoliment dels objectius d'aprenentatge.**

La tesi està estructurada sobre quatre articles, que es poden separar en dos grups. El primer està format pels tres primers treballs, cronològicament parlant, en què s'han analitzat, des de l'òptica d'un observador extern a les aules, les relacions entre qüestionaris i resultats acadèmics. El segon grup conté el quart treball, en què a partir de l'anàlisi efectuada en el tercer treball del primer grup, s'ha dut a terme i s'ha analitzat una intervenció docent aplicada sobre estudiants en risc d'abandonar l'assignatura d'estadística. Pel que fa als treballs del primer grup, ens porten a concloure que la realització de qüestionaris, d'entrenament i/o mesura, amb correcció i feedback automàtics estan relacionats amb els resultats acadèmics, com ara qualificacions en els qüestionaris de mesura o qualificacions en l'examen final. Això s'ha vist en dues assignatures diferents, l'anàlisi matemàtica i l'estadística, amb un model pedagògic molt similar basat en la realització regular de qüestionaris amb correcció i feedback automàtics. Anant un pas més enllà, podem concloure que mantenir el compromís amb l'assignatura al llarg del curs amb la realització de qüestionaris amb correcció i feedback automàtics ajuda a l'assoliment dels objectius d'aprenentatge. Aquesta conclusió no significa que la relació entre qüestionaris i resultats acadèmics sigui una relació causa-efecte, en el sentit que la realització dels qüestionaris sigui la causa que els resultats acadèmics siguin millors. Cal considerar això com una limitació atribuïble als dissenys ex-post facto i correlacional, emprats en els treballs sobre l'assignatura d'anàlisi matemàtica, que impedeixen inferir causalitat entre les variables. Quant al treball del segon grup, ens permet concloure que una intervenció docent que ofereix als estudiants la possibilitat de tornar a enganxar-se en la realització regular de qüestionaris, millora les seves opcions de no abandonar l'assignatura, incidint una altra vegada en com de clau resulta **persistir en la realització dels qüestionaris**. La principal limitació cal expressar-la com el reduït nombre d'estudiants que ha rebut la intervenció i, especialment que hi ha respost activament, és a dir, que ha realitzat el qüestionari no lliurat dins del termini previst inicialment.

La millora continua en la pràctica docent implica l'anàlisi de l'experiència i la introducció de certes accions que persegueixin la millora de l'aprenentatge. El fet de basar aquestes accions en evidències, en dades, fa que estiguin justificades i siguin més robustes. Les anàlisis realitzades en aquest treball de recerca ens han portat a pensar que cal fomentar la realització de qüestionaris. El fet de tenir un predictor de "no superació de l'assignatura" ens ha permès definir una intervenció senzilla orientada a fomentar que els estudiants "continuïn fent", això és, estiguin compromesos amb l'assignatura. Atès que la intervenció ha resultat efectiva ens permet concloure que aquest camí basat en el cicle dades-anàlisi-intervenció (Clow, 2012) és un procediment que resulta útil a l'hora de prendre decisions en la pràctica docent.

Aquesta tesi ha estat enfocada cap a l'estudi de l'ús de qüestionaris des del punt de vista del comportament i coneixement dels estudiants. Rienties et al. (2017) consideren que les **actituds** dels estudiants afecten tant el seu comportament com el seu coneixement, de manera que constitueixen factors que incideixen sobre el seu aprenentatge. En consonància amb això, hem pogut veure durant la pandèmia iniciada el 2020 com de rellevant és el paper que juguen els factors afectius en l'aprenentatge de l'estudiant. Suggerim, com a primera línia de futurs treballs, incorporar variables relacionades amb les actituds (motivació i autoconfiança, per exemple) per aprofundir en la comprensió de les relacions entre qüestionaris i resultats acadèmics, així com per poder dissenyar i implementar intervencions docents per a la millora d'aquests resultats. Entenem que fora recomanable avançar cap a la cerca de perfils actitudinals, en relació a la realització de qüestionaris, que permetessin dur a terme diferents intervencions docents adreçades a diferents grups d'estudiants, segons aquests perfils. Proposem, com a segona línia de futurs treballs, estendre l'anàlisi de les relacions entre qüestionaris i resultats acadèmics a altres assignatures de matemàtiques en línia que fan servir els qüestionaris com a eina d'aprenentatge (àlgebra, mètodes numèrics, i probabilitat i processos estocàstics, en són tres exemples), de manera que el rang d'assignatures estudiades permeti contrastar els resultats obtinguts en aquesta tesi. Finalment, com a tercera línia de futurs treballs, suggerim que l'anàlisi de la metodologia docent basada en la realització de qüestionaris inclogui d'una banda la perspectiva de l'estudiant, seguint el camí iniciat en el treball sobre la intervenció docent (Figueroa-Cañas i Sancho-Vinuesa, 2021b), i de l'altra, la perspectiva del professorat.

BIBLIOGRAFIA

Abrami, P. C., Bernard, R. M., Bures, E. M., Borokhovski, E., i Tamim, R. M. (2011). Interaction in distance education and online learning: Using evidence and theory to improve practice. *Journal of Computing in Higher Education*, *23*(2–3), 82–103. https://doi.org/10.1007/s12528-011-9043-x

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., i Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, *31*(1), 542–550. https://doi.org/10.1016/j.chb.2013.05.031

Anderson, T. (2003). Getting the mix right again: An updated and theoretical rationale for interaction. *International Review of Research in Open and Distance Learning*, *4*(2), 126–141. https://doi.org/10.19173/irrodl.v4i2.149

Angus, S. D., i Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. *British Journal of Educational Technology*, *40*(2), 255–272.

Archer, E. (2017). The Assessment Purpose Triangle: Balancing the Purposes of Educational Assessment. *Frontiers in Education*, *2*(August), 1–7. https://doi.org/10.3389/feduc.2017.00041

Baleni, Z. G. (2015). Online formative assessment in higher education: its pros and cons. *The Electronic Journal of e-Learning* , *13*(4), 228-236.

Berge, Z. L. (2002). Active, interactive and reflective eLearning. *Quarterly Review of Distance Education*, *3*(2), 181–190.

Biggs, J., i Tang, C. (2011). *Teaching for quality learning at university. (2nd Edn.).* Open University Press/McGrawHill.

Burgos, C., Campanario, M. L., Peña, D. de la, Lara, J. A., Lizcano, D., i Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers and Electrical Engineering*, *66*, 541–556. https://doi.org/10.1016/j.compeleceng.2017.03.005

Calm, R., Masià, R., Olivé, C., Parés, N., Pozo, F., Ripoll, J., i Sancho-Vinuesa, T. (2013). Wiris Quizzes: un sistema de evaluación continua con feedback automático para el aprendizaje de matemáticas en línea. *Teoría de La Educación. Educación y Sultura En La Sociedad de La Información*, *14*(2), 452–472.

Cambruzzi, W., Rigo, S. J., i Barbosa, J. L. V. (2015). Dropout prediction and reduction in distance education courses with the learning analytics multitrail approach. *Journal of Universal Computer Science*, *21*(1), 23–47.

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. Dins *Proceedings of the Second International Conference on Learning Analytics amd Knowledge (LAK 2012)*: 29 abril-2 maig 2012 (pp. 134-138). New York:ACM. https://doi.org/10.1145/2330601.2330636

Corrigan, O., Glynn, M., McKenna, A., Smeaton, A., i Sinead, S. (2015). Student Data: Data is knowledge – putting the knowledge back in the students' hands. *Proceedings of the European Conference on E-Learning, ECEL*, 165–172.

Duart, J., i Sangrà, A. (2000). Formación universitaria por medio de la web: un modelo integrador para el aprendizaje superior. *Aprender En La Virtualidad*, 7–33.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., i Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, Supplement*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Engelbrecht, J., i Harding, A. (2005). Teaching undergraduate mathematics on the internet. *Educational Studies in Mathematics*, *58*(2), 253–276. https://doi.org/10.1007/s10649-005-6457-2

Escudero-Viladoms, N. (2012). *Feedback, confiança matemàtica i aprenentatge matemàtic en un entorn d'aprenentatge en línia*. Universitat Oberta de Catalunya.

Escudero-Viladoms, N., i Sancho-Vinuesa, T. (2016). *Confidence and learning: affective and cognitive aspects in online mathematics with automatic feedback*. Dins S. Caballé, i R. Clarisó (Coords.), *Formative assessment, learning data analytics and gamification*. (pp. 87-195). Academic Press.

Figueroa-Canas, J., i Sancho-Vinuesa, T. (2017). Exploring the Efficacy of Practicing with Wiris-Quizzes in Online Engineering Mathematics. *Revista Iberoamericana de Tecnologias Del Aprendizaje*, *12*(3), 141–146. https://doi.org/10.1109/RITA.2017.2735499

Figueroa-Cañas, J., i Sancho-Vinuesa, T. (2019). Predicting early dropout students is a matter of checking completed quizzes: The case of an online statistics module. Dins *Proceedings of the Learning Analytics Summer Institute:* 27–28 juny (pp. 100–111). CEUR Workshop Proceedings 2415

Figueroa-Cañas, J., i Sancho-Vinuesa, T. (2020). Early Prediction of Dropout and Final Exam Performance in an Online Statistics Course. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 15(2), 86-94. DOI: 10.1109/RITA.2020.2987727

Figueroa-Cañas, J. i Sancho-Vinuesa, T. (2021a). Investigating the relationship between optional quizzes and final exam performance in a fully asynchronous online calculus module. *Interactive Learning Environments*, 29(1), 33-43. DOI: 10.1080/10494820.2018.155986

Figueroa-Cañas, J. i Sancho-Vinuesa, T. (2021b). Changing the recent past to reduce ongoing drop-out: an early learning analytics intervention for an online statistics course. *Open Learning: The Journal of Open, Distance and e-Learning*. DOI: 10.1080.02680513.2021.1971963.

Gikandi, J. W., Morrow, D., i Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers and Education*, *57*(4), 2333–2351. https://doi.org/10.1016/j.compedu.2011.06.004

Hattie, J., i Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hillman, D. C. A., Willis, D. J., i Gunawardena, C. N. (1994). Learner-interface interaction in distance education: An extension of contemporary models and strategies for practitioners. *American Journal of Distance Education*, *8*(2), 30–42. https://doi.org/10.1080/08923649409526853

Holden, J., i Westfall, P. (2007). *An instructional media selection guide for distance learning (4th Edition)*. United states distance learning association.

Hothorn, T., Hornik, K., i Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. Research Report Series 8, Department of Statistics and Mathematics, WU Wien, 2004. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674. https://doi.org/10.1198/106186006X133933

Huisman, R., i Reedijk, H. E. (2012). The impact of online tests in addition to group assignments on student learning. *ICICTE 2012 Proceedings 654*, 654–667.

James, G., Witten, D., Hastie, T., i Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). https://doi.org/10.1007/978-1-4614-7138-7

Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., Regan, J. R., i Baron, J. D. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47. https://doi.org/10.18608/jla.2014.11.3

Jivet, I., Scheffel, M., Drachsler, H., i Specht, M. (2017). Awareness Is Not Enough : Pitfalls of Learning Practice. *Data Driven Approaches in Digital Education*, *1*(i), 82–96. https://doi.org/10.1007/978-3-319-66610-5

Johnson, J. A., i Mckenzie, R. (2013). The effect on student performance of web-based learning and homework in microeconomics. *Journal of Economics and Economic Education Research*, *14*(2), 115–126.

Joksimović, S., Gašević, D., Loughin, T. M., Kovanović, V., i Hatala, M. (2015). Learning at distance: Effects of interaction traces on academic achievement. *Computers and Education*, *87*, 204–217. https://doi.org/10.1016/j.compedu.2015.07.002

Kalyuga, S. (2012). Interactive distance education: A cognitive load perspective. *Journal of Computing in Higher Education*, *24*(3), 182–208. https://doi.org/10.1007/s12528-012-9060-4

Kanuka, H. (2011). Interaction and the online distance classroom: Do instructional methods effect the quality of interaction? *Journal of Computing in Higher Education*, *23*(2–3), 143–156. https://doi.org/10.1007/s12528-011-9049-4

Karal, H., Kokoc, M., Colak, C., i Yalcin, Y. (2013). Using Pen- Based Technology in Online Mathematics Course: An Evaluation Study. *European Journal of Open, Distance and e- Learning*, *16*(2), 152–164.

Keppell, M., Au, E., Ma, A., i Chan, C. (2006). Peer learning and learning-oriented assessment in technology-enhanced environments. *Assessment and Evaluation in Higher Education*, *31*(4), 453–464. https://doi.org/10.1080/02602930600679159

Keselman, H. J., Othman, A. R., Wilcox, R. R., i Fradette, K. (2004). The new and improved two-sample T test. *Psychological Science*, *15*(1), 47–51. https://doi.org/10.1111/j.0963-7214.2004.01501008.x

Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., i Riera, T. (2014). A system for knowledge discovery in e-learning environments within the European Higher Education Area - Application to student data from Open University of Madrid, UDIMA. *Computers and Education*, *72*, 23–36. https://doi.org/10.1016/j.compedu.2013.10.009

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., i Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Lowenthal, P. R., Snelson, C., i Dunlap, J. C. (2017). Live synchronous web meetings in asynchronous online courses: Reconceptualizing virtual office hours. *Online Learning Journal*, *21*(4), 177–194. https://doi.org/10.24059/olj.v21i4.1285

Moore, M. G. (1989). Editorial: Three Types of Interaction. *American Journal of Distance Education*, *3*(2), 1–7. https://doi.org/10.1080/08923648909526659

Newton, P. E. (2007). Clarifying the Purposes of Educational Assessment. *Assessment in Education*, *14*(2), 149–170. https://doi.org/10.1080/09695940701478321

Palocsay, S. W., i Stevens, S. P. (2008). A Study of the Effectiveness of Web-Based Homework in Teaching Undergraduate Business Statistics. *Decision Sciences Journal of Innovation Education*, *6*(2), 213–232.

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., i Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology*, *50*(1), 128–138. https://doi.org/10.1111/bjet.12592

Pinchbeck, J., i Heaney, C. (2017). Case report: the impact of a resubmission intervention on level 1 distance learning students. *Open Learning*, *32*(3), 236–242. https://doi.org/10.1080/02680513.2017.1348290

Rienties B., Cross S., i Zdrahal Z. (2017) Implementing a Learning Analytics Intervention and Evaluation Framework: What Works?. Dins Kei Daniel B. (eds) *Big Data and Learning Analytics in Higher Education* (pp. 147-166). Springer, Cham. https://doi.org/10.1007/978-3-319-06520-5_10

Roediger, H. L., Putnam, A. L., i Smith, M. A. (2011). *Ten Benefits of Testing and Their Applications to Educational Practice*. *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 55). https://doi.org/10.1016/B978-0-12-387691-1.00001-6

Sabry, K., i Baldwin, L. (2003). Web-based learning interaction and learning styles. *British Journal of Educational Technology*, *34*(4), 443–454. https://doi.org/10.1111/1467-8535.00341

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education*, *35*(5), 535–550. https://doi.org/10.1080/02602930903541015

Sancho-Vinuesa, T., Escudero-Viladoms, N., i Masià, R. (2013). Continuous activity with immediate feedback: a good strategy to guarantee student engagement with the course. *Open Learning: The*

Journal of Open, Distance and e-Learning, *28*(1), 51–66. https://doi.org/10.1080/02680513.2013.776479

Sancho-Vinuesa, T., i Escudero-Viladoms, N. (2012). ¿ Por qué una propuesta de evaluación formativa con feedback automático en una asignatura de matemáticas en línea ? *Revista de UniversiDad y Sociedad Del Conocimiento (RUSC)*, *9*(2), 59–79.

Sancho-Vinuesa, T., Masià, R., Fuertes-Alpiste, M., i Molas-Castells, N. (2018). Exploring the effectiveness of continuous activity with automatic feedback in online calculus. *Computer Applications in Engineering Education*, *26*(1), 62–74. https://doi.org/10.1002/cae.21861

Scheffel, M. (2017). *The Evaluation Framework for Learning Analytics*. Open Universiteit and SIKS.

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Smith, G., Torres-Ayala, A. T., i Heindel, A. J. (2008). Disciplinary Differences in E-learning Instructional Design: The Case of Mathematics. *The Journal of Distance Education*, *22*(3), 63–88.

Spector, J. M., Ifenthaler, D., Sampson, D., Yang, L. J., Mukama, E., Warusavitarana, A., Dona, K. L., Eichhorn, K., Fluck, A., Huang, R., Bridges, S., Lu, J., Ren, Y., Gui, X., Deneen, C. C., San Diego, J., i Gibson, D. C. (2016). Technology enhanced formative assessment for 21st century learning. *Educational Technology and Society*, *19*(3), 58–71.

Swan, K. (2001). Virtual interaction: Design factors affecting student satisfaction and perceived learning in asynchronous online courses. *Distance Education*, *22*(2), 306–331. https://doi.org/10.1080/0158791010220208

Taras, M. (2005). Assessment - Summative and formative - Some theoretical reflections. *British Journal of Educational Studies*, *53*(4), 466–478. https://doi.org/10.1111/j.1467-8527.2005.00307.x

Tempelaar, D., Rienties, B., Mittelmeier, J., i Nguyen, Q. (2018). Student profiling in a dispositional learning analytics application using formative assessment. *Computers in Human Behavior*, *78*, 408–420. https://doi.org/10.1016/j.chb.2017.08.010

Trenholm, S., Alcock, L., i Robinson, C. (2015). An investigation of assessment and feedback practices in fully asynchronous online undergraduate mathematics courses. *International Journal of Mathematical Education in Science and Technology*, *46*(8), 1197–1221. https://doi.org/10.1080/0020739X.2015.1036946

Van der Kleij, F. M., Feskens, R. C. W., i Eggen, T. J. H. M. (2015). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis. *Review of Educational Research, 85*(4), 475-511.

Wan Niu Voon, B., Wong, L. S., i Marhana, S. (2014). Online Mathematics Learning in Tertiary Education : A Study on Students ' Behavior. *Academic Research International*, *5*(5), 143–147.

Wang, Z., Chen, L., i Anderson, T. (2014). A framework for interaction and cognitive engagement in connectivist learning contexts. *International Review of Research in Open and Distance Learning*, *15*(2), 121–141. https://doi.org/10.19173/irrodl.v15i2.1709

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, *37*(1), 3–14. https://doi.org/10.1016/j.stueduc.2011.03.001

Yorke, M. (2003). Formative Assessment in Higher Education : Moves Towards Theory and the Enhancement of Pedagogic Practice. *Higher Education*, *45*(4), 477–501.

ANNEX

PREDICTING EARLY DROPOUT STUDENTS IS A MATTER OF CHECKING COMPLETED QUIZZES: THE CASE OF AN ONLINE STATISTICS MODULE[12]

Higher education students who either do not complete the subjects they enrolled in or interrupt indefinitely their studies without certification, the so-called college dropout problem, still continues to be a major concern for practitioners and researchers. Within the subjects, an early prediction of dropout students has aided teachers to focus their intervention in order to reduce dropout rates. Several machine-learning techniques have been used to classify/predict dropout students, including the tree-based methods which are not the best performers, but in their favour, are easily interpretable. This study presents a procedure to identify dropout-prone students at an early stage in an online statistics module, based on decision tree models. Although the attributes initially considered in the creation of the trees were mainly related to quiz completion, participation in the forum and access to the bulletin board, the final models show that the former is the only attribute with significant discriminatory power. We have evaluated the classification performance by means of a validation set. The performance measure of accuracy shows values above 90%, whereas that of recall and precision slightly under 90%.

Keywords: Dropout prediction, decision trees, quiz completion, online education.

Introduction

Among education practitioners and researchers, students who do not complete a single module/subject or indefinitely interrupt their studies without having achieved the certificate have been a matter of considerable concern for a long time. These students are usually called dropout students. In online courses, the high dropout rates of students justify the abundant research on this particular topic, as shown

[12] En aquest annex, les referències numèriques a taules corresponen a referències exclusivament d'aquest annex. Les referències bibliogràfiques d'aquest annex es mostren dins d'aquest mateix annex.

in the extensive review of [1], where 159 studies published between 1999 and 2009 were analysed. More recently, in the European framework, reducing the dropout student rate in higher education is considered a key strategy to attain the ambitious objective of not less than 40% of people in their thirties who have completed higher education studies by 2020 [2]. Concerned as teachers and guided by European strategy, the authors have decided to carry out research on dropout students in the statistics module at the Universitat Oberta de Catalunya

In a higher education context, two levels of dropout can be differentiated: (a) the micro-level dropout, and (b) the macro-level one. In the former, the fact of dropout takes place inside the module or subject [3], where teachers can intervene in case they have convenient information at an early stage in order to reduce it. In line with that, Burgos [4] shows a reduction of 14% in dropout rates of students by means of a tutoring plan action after the dropout-prone students have been identified early. In macro-level dropout, withdrawal from studies occurs, in general, outside the subjects so that the interventions are the responsibility of other staff different from the teachers of the subject.

The main purpose of the present study is to design a procedure to identify as many dropout-prone students as possible in an online statistics module, as soon as possible. This procedure is based on the prediction/classification provided by binary conditional decision trees generated in several instants of time throughout the module duration, from the data related, mainly to test completion and participation in both the online forum and the bulletin board.

Literature review

According to [1], there is an absence of consensus on the definition of both the micro-level dropout and the macro-level one. With regard to the latter, even online and face-to-face universities do not share the dropout definition [5]. Grau-Valldosera [5] claims the time accepted without any enrolled subjects in an online university has to be extended compared with that in a face-to-face university because of the students' characteristics.

As illustrations of the micro-level dropout definitions, we have chosen the three that follow. First, Liu [6] straightforwardly associates subject dropout with subject failure. Dropout students are those who do not attain A, B, or C, that is, those who fail the subject. Second, Levy [7] defines dropout students as those who do not complete the subject and their tuition fees have not been refunded. And third, Dupin [8] considers dropout students as those who are non-completers, understood in a broad sense.

The studies about dropout students by Cohen [3], Burgos [4], Costa [9], Santana [10], Lykourentzou [11], Lara [12] and Kotsiantis [13] are focused on the micro level (university subjects), all in an online but [3] blended environment. In addition, all of them are concerned with early prediction and show considerable high values of several evaluation measures of classification performance, such as accuracy, recall, precision or F1-measure. Cohen [3] reports a maximum precision of 80%, Burgos [4] a recall of 96.73%, Costa [9] a maximum F1-measure of 82%, Santana [10] a maximum accuracy of 86%, Lykourentzou [11] a maximum recall of 95% and Kotsiantis [13] a maximum accuracy of 83.89%. Lara [12] found an accuracy above 90%, a figure that is "a very acceptable percentage for the problem domain" [12, pp. 31]. In the following four paragraphs, we present a comparative review between [3-4, 9-14] regarding dropout definition, single/multiple predicting instants of time, attributes selected as predictors and classification method to carry out the prediction.

The dropout definition from the failure perspective [6] is the one used in the studies of Cohen [3], Costa [9] and Santana [10]. The definition of Levy [7] is explicitly mentioned in Lykourentzou [11], who adds another requirement: that the dropout student has to access the e-learning platform at least once throughout the subject duration. That means the student has to leave a trace in the information system before leaving the subject in order to be considered a dropout student. For Burgos [4] and Lara [12] students who do not sit the final exam are those defined as dropout students. And finally, Kotsiantis [13] does not precisely define the non-completer students.

Predicting in a single instant of time is the option chosen by Santana [10] and Kotsiantis [13]. The latter argues that prediction has to be released before the subject is half over because otherwise it would not be useful for the teachers to intervene in time. Santana [10] predicts dropouts after the first exam, which also coincides with half of the subject duration. In contrast, multiple instants of time, albeit not the same ones, are contained in the proposals of [3-4, 9, 11-12]. Lykourentzou [11] released predictions into each of the 7 sections that the subject is divided into. Similarly, Burgos [4] predicts in each of the 12 assessment activities. The proposals of [3,9,12], based mainly on regular time intervals, are slightly different: Cohen [3] predicts dropouts monthly, in a one semester course, Lara [9] weekly in 15-20 week courses, and finally Costa [9] also weekly in a 10-week course and after releasing the mid-course exam marks.

All the attributes employed in [3-4, 9-13] can be grouped into three main categories: demographics, usage of educational tools, and assessment activities or exam performance. The first category is formed by time-invariant data available at the beginning of the course, whereas the other two categories include time-varying data which are incrementally collected throughout the course. Demographic attributes such as gender and professional information are used by [9-11, 13] alike. Some studies also consider other specific demographic attributes, like English language literacy [13]. The usage of educational tools in general, and particularly participation in the forum is included in the set of attributes that form the models of Cohen [3], Costa [9], Santana [10], Lykourentzou [11] and Lara [12]. Finally, the marks attained in assessment activities or exams are analysed in the studies of Burgos [4], Costa [9], Santana [10], Lykourentzou [11] and Kotsiantis [13].

Regarding classification methods, apart from Cohen [3] who uses a unique method based on comparing changes in attribute values of a student with respect to the mean of attribute values of the whole group of students, the studies of [4, 9-13] use a great variety of machine-learning techniques. Algorithms based on neural networks and support vector machines are common to [4, 9-13], whereas naive Bayes and decision tree classifiers are only employed by Costa [9], Santana [10] and Kotsiantis [13]. Finally, logistic regression is also included in the set of classifiers of Burgos [4], Lara [12] and Kotsiantis [13].

Although the study of Romero [14] does not explicitly mention the dropout problem, as it aims to predict the final performance of students by classing them as passed or failed, it could be deemed as a dropout problem according to Liu's definition [6]. Moreover, like some of the references previously reviewed, an early prediction is released and the usage of the forum is the source of information to feed the attributes. The study stands out for the comparative performance of 14 classification algorithms and reaches the conclusion that the sequential minimal optimization (SMO) algorithm, related with support vector machines, is the better performer. It is worth recalling that the studies of [3-4, 9-13] all included that machine-learning technique.

The high dropout rates are also a major source of concern in Massive Open Online Courses [15] and, in order to reduce them, several studies have dealt with their early prediction [15-17]. These studies differ both in the type of dependent variables and the machine learning methods used in their models. First, whereas the studies by Ruiperez-Valiente [15] and Sharma [16] include the scores awarded after assignment submission, Yang [17]'s only takes into account the behaviour in the discussion forum. And second, prediction algorithms based on artificial neural networks are the ones chosen by Sharma [16], while Ruiperez-Valiente [15] implemented random forests, generalised boosted regression modelling, K-nearest neighbours and a logistic regression, and Yang [17] used a survival model. Sharma [16] finds a relationship between students failing in assignments and dropping out of the course.

M e t h o d o l o g y

Participants and learning context

The participants in this study were the 197 students enrolled in the first semester of the 2018/19 fully asynchronous online one-semester statistics module, which formed part of the Computer Engineering degree at the Universitat Oberta de Catalunya.

The teaching plan for this statistics module allowed students to complete optional quizzes (Quizzes) and constructed-response questions (R.Questions) that had to be solved by using the statistical program R. Six different pairs (Quiz, R.Question), named continuous assessment tests, were scheduled throughout the semester. Quizzes were corrected and marked immediately, providing automated feedback. R.Questions required manual teacher correction and feedback was delayed. The scores attained, which formed part of the continuous assessment mark, could be included in the final mark. The module included two assessment instruments: (a) a compulsory in-person final exam, and (b) non-compulsory online continuous assessment throughout the semester. The final mark for the module was mainly based on the final exam mark, which could be modified slightly by the continuous assessment mark. In addition, during the first week teachers assigned an initial test to ascertain students' prior knowledge of secondary-education statistics. In order to encourage participation, students who voluntarily completed and submitted the test obtained a bonus, which also formed part of the continuous assessment mark.

An e-learning platform provides students enrolled in the statistics module of the Universitat Oberta de Catalunya with a communication tool: a forum, and an information tool: a bulletin board. The latter was used by teachers to upload course information which was mostly only accessible by students via that bulletin board. The former allowed students and teachers to interact with each other, in general, asynchronously. The e-learning platform also included direct access to view the teaching plan, which contained precise information about the assessment system. All reading access to the bulletin board, forum and teaching plan, as well as writing acces to the forum were recorded by the information system of the Universitat Oberta de Catalunya.

Measure and data collection

The data has been collected in four instants of time, which coincide with the first four continuous assessment test submission deadlines, the only ones in the first half of the course. The separation between submission deadlines is variable, ranging from 1 to 3 weeks. We define four periods of time (Period.1, ..., Period.4) from the previous submission deadline as follows: Period.1 is the interval of time between the first day of the semester and the first submission deadline, Period.2 is the interval of time between the first and second submission deadlines, and so on for Period.3 and Period.4.

During the first period (Period.1), we gathered students' register data such as the number of courses enrolled on in the semester and whether they were repeater students or not. This data, contained in the information system of the Universitat Oberta de Catalunya and anonymously delivered to us, filled the instances of the attributes Repeating and Enrolled_Courses (see Table.1). The Moodle activity log was the source of information to determine whether the student had submitted the initial test or not, and likewise the first continuous assessment test. With that data, the instances of the attributes Initial_Test, Quiz_Till_Period.1 and R.Question_Till_Period.1 were filled (see Table.1). The e-learning platform activity log provided the date and time of all access to the platform which, after being pre-processed, filled the instances of the attributes BBoard_Till_Period.1, Forum_Wr_Till_Period.1, Forum_Re_Till_Period.1 and Teaching_Plan_Viewed_Till_Period.1 (see Table.1). All the previous data, transferred to the second period (Period.2) and incremented with the specific information collected in Period.2, filled the attributes ending in _Till_Period.2. This procedure was repeated for Period.3 and Period.4 (see Table.1).

The attribute selection of our study is based on the references in section Literature review [4,9]. Nevertheless, we have not considered the scores of assessment activities as attributes as [4] does. Instead, we have opted for the completion or non-completion of Quizzes and R.Questions. There are two main reasons for this decision. The first reason is Quizzes and R.Questions submission data are available faster than definite marks since both R.Questions are marked manually (as mentioned in section 3.1), and students may apply for marking reviews. The second reason is the likely high correlation between completion of assessment activity and its mark, as is shown in a calculus module of the same degree and in a very similar educational context [18].

In the present study, we have defined, based on [7], a dropout student as the student who attains a final mark of “Not Completed”, which means the student has not taken the compulsory final exam.

That approach is in line with that of the [4, 12]. The boolean response variable Y=Dropout indicates whether the student complies (I.Dropout) or not (I.Completer) with the previous definition, that is, whether they belong to the dropout student or to the completer student class.

Table.1 Attributes for the Period.i, with i=1,..., 4

Name	Description	Types and Values
Repeating	Indicates whether the student is repeating the subject or not	Type: Boolean. Values: I.RP, N.RP
Enrolled _Courses	Indicates the total of courses enrolled on in the semester.	Type: Integer Values: {1, ...}
Initial_Test	Indicates whether the student has or has not completed and submitted the initial test.	Type: Boolean. Values: H.IT, N.IT
Teaching _Plan_ Viewed	Indicates whether the student has or has not viewed the teaching plan until the last day of the Period.i	Type: Boolean. Values: H.TPV, N.TPV
Quiz_Till_ Period.i	Indicates the number of quizzes completed and submitted until the last day of the Period.i	Type: Integer Values: {0, 1, ..., i}
R.Question_ Till_Period.i	Indicates the number of quizzes completed and submitted until the last day of the Period.i	Type: Integer Values: {0, 1, ..., i}
BBoard_ Till_Period.i	Indicates the number of periods in which the student has accessed the board until the last day of the Period.i	Type: Integer Values: {0, 1, ..., i}
Forum_Wr_ Till_Period.i	Indicates the number of periods in which the student has written messages on the forum until the last day of the Period.i	Type: Integer Values: {0, 1, ..., i}
Forum_Re_ Till_Period.i	Indicates the number of periods in which the student has read messages on the forum until the last day of the Period.i	Type: Integer Values: {0, 1, ..., i}

To fill the instances of that variable, the information system of the Universitat Oberta de Catalunya has anonymously delivered the final marks to us. By combining attributes, like predictors, and response variable Y, four sets of data are available, and each of them contains the instances of attributes of each period and the instances of the Dropout variable.

Classification method

We pose a classification problem, the result of which will be a binary classification model or binary classifier in order to predict whether a student will be classed as a dropout student or completer student at the end of the semester. In addition, we require the classifier to be easily interpretable, although at the expense of it not being the best performer in terms of the usual evaluation measures of classification performance like accuracy, precision or recall. Due to "tree-based methods being simple and useful for interpretation " [19, pp303], we have decided to use those methods of classification in our study. Basically, a binary decision tree is an oriented graph that starts in a node called root, follows through arcs called branches, and ends in the terminal nodes called leaves. Each non terminal node, including the root, represents an attribute (a test on the attribute), and each leaf represents one of the two classes (dropout student or completer student) or the proportion of students that belong to each class. The branches that come out of a node represent the values of the attribute associated with the node (the answer to the test on the attribute) [20].

Given our attribute selection (see Table.1), we observe that not all the attributes have the same number of possible values. A widely identified issue detected in studies using decision tree models is the bias, in creating the nodes, to attributes with a large number of possible values [21]. Conditional tree models mitigate that bias [21], and for that reason those models are the classification methods we have chosen. To grow our conditional trees we have used the ctree() function provided by the statistical program R.

For each of the four data sets a classification model has been built (Model.1, Model.2, Model.3, Model.4). In order to evaluate the performance of the models, firstly the whole data set can be split into two mutually exclusive sets: the training set and the validation one. Secondly, with the training set the classification model is fitted. And finally, the evaluation of the performance is carried out using the validation set [19]. In our study, we have conducted a random stratified split into a training set (80% of the whole set) and a validation set (20%), keeping the same class distribution of the whole set in each subset.

Taking into account that our main purpose was to identify dropout-prone students, we have considered students predicted as dropouts, that is, those whose predicted class is I.Dropout, as “Positive” cases, and the others, those whose predicted class is I.Completer, as “Negative” cases. Moreover, as usual, we differentiate between “True” or “False” depending on whether the predicted class coincides with the observed class or not, respectively. Table.2 depicts the four possible pairs when applying the validation set to the model fitted with the training set.

Table.2 Possible pairs in terms of predicted and observed classes

	Predicted class I.Dropout	**Predicted class I.Completer**
Observed class I.Dropout	True Positive (TP)	False Negative (FN)
Observed class I.Completer	False Positive (FP)	True Negative (TN)

In our study we have decided to use three evaluation measures of the classification performance: Accuracy (1), Precision (2) and Recall (3), according to the following definitions [19]:

$$Accuracy = \frac{TP+T}{TP+TN+FN+FP} \quad (1)$$

$$Precision = \frac{TP}{TP+F} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Results

The four classifiers created from the training set are extremely simple, each of them contains one single node. Table.3 depicts each model as a decision rule. The only attribute shown in the models, a result that reveals that it is the one with the strongest association with the response Dropout [20], is the completion of Quizzes (Quiz_Till_Period.i). Using the first three models (Model.1, Model.2, Model.3), students are classified/predicted as dropout students (class I.Dropout) if they have not completed all the Quizzes scheduled until the end of the period associated with the model, in other words, if they have not completed one or more of those Quizzes. As an example, at the end of Period.3, students that has not completed all three Quizzes corresponding to the first three continuous assessment tests are classified/predicted as dropout students. Only students who have completed all three Quizzes are classified/predicted as completer students (class I.Completer). In Model.4, the last condition is softened, so that students are classified/predicted as completer students even if they have not completed all four Quizzes. They can have decided to skip one Quiz, at the most.

Table.3 Decision rules for the four models: Model.1, Model.2, Model.3, Model.4

Model.1*: **IF** Quiz_Till_Period.1=1 **THEN** I.Completer **ELSE** I.Dropout	**Model.2***: **IF** Quiz_Till_Period.2=2 **THEN** I.Completer **ELSE** I.Dropout
Model.3*: **IF** Quiz_Till_Period.3=3 **THEN** I.Completer **ELSE** I.Dropout	**Model.4***: **IF** Quiz_Till_Period.4 >2 **THEN** I.Completer **ELSE** I.Dropout

* p-value < 0.001. H_0: D(Dropout | Quiz_Till_Period.i) = D(Dropout), that is, H_0: The response Dropout is independent of the predictor Quiz_Till_Period.i

Using the validation test, three results stand out in the evaluation measures of the classification performance (see Table.4). First, Accuracy shows a gradual increase from the first model and, in Model.3 passes the figure of 90%, which is considered acceptable by [12]. Second, Precision also grows from the second model and reaches the value 100% in Model.4. And third, Recall also rises from the first model and reaches its highest value in Model.3, just when Accuracy attains the level of "acceptable".

Table.4 Performance measures

	Recall	**Precision**	**Accuracy**
Model.1	53.8%	87.5%	82.9%
Model.2	61.5%	80.0%	82.9%
Model.3	84.6%	84.6%	90.2%
Model.4	84.6%	100%	95.1%

Discussion

From the very beginning we aimed to find a classification model that was easily interpretable, even at the expense of not finding the best performer classifier. The four classification models (see Table.3) entirely comply with the previous requirement. In the rest of the section, we discuss the following three statements: (a) completing evaluative quizzes is the only attribute that determines the classification process, (b) the simplicity of the models eases the creation of an overall classification procedure that includes all the models, and (c) applying the models separately, Model.3 is the best.

Above all, it is worth noticing that only one attribute, the Quiz_Till_Period.i, intervenes in the classification process as the four models show (see Table.3). The Quiz_Till_Period.i attribute, directly related with the completion of Quizzes, has basically an evaluative character, which sets it apart from the attributes related to the usage of the e-learning platform, such as the forum. The dominance of evaluative attributes is in line with the study of Costa [9], who found that the most important attribute was the midterm marks. On the other hand, completion of R.Questions likewise has evaluative character, but nonetheless does not intervene in the final models. The main difference between Quizzes and R.Questions lies in that the latter require students to apply higher level skills than the former. Consequently, it seems reasonable to argue that students who do not even complete the least-demanding assessment assignments, such as Quizzes, are the most prone to becoming dropout students. And last but not least, the three first models separate dropout and completer students depending on whether they have or have not completed all the Quizzes scheduled until the moment the model is applied. Therefore, we can interpret that the continued "doing" of Quizzes is the relevant aspect in differentiating those who complete the course from those who do not.

The simplicity of the model reduces the volume of information actually being used to only that related to completion of Quizzes, which in turn entails two beneficial consequences: (a) the obvious elimination of time spent gathering and processing the rest of attributes, (b) the teacher himself/herself can collect the required data directly from the Moodle activity log. Using the three first models in cascade, at the end of the first continuous assessment test submission deadline, the teacher can create a list of dropout-prone students by selecting those who have not completed the first Quiz. After the second submission deadline, the teacher can add new dropout-prone students to the previous list by selecting those who have not completed the second Quiz, and likewise regarding those who have not complete the third one. So, by following that simple procedure the teacher step by step adds to the list of dropout-prone students, which can be useful when deciding possible measures in order to change the unsuccessful predicted result.

The performance measures (see Table.4) indicate that, for Model.1, Precision is quite high, but Recall is not, which can be interpreted as follows: that a limited amount of students classed as a I.Dropout will eventually become completers, whereas a significant number of students classed as I.Completer will finally become dropouts. As a result, a limited number of students can be the target of unnecessary teacher intervention, but what is worse, a significant number of students will be outside the scope of teacher intervention, which would have been useful if they had been correctly classified. Due to the fact that our purpose is to identify as many dropout-prone students as possible, Recall prevails over Precision. As a consequence, Model.1 turns out to have a low degree of satisfaction. Model.2 is slightly more satisfactory than the Model.1 because of its higher Recall, but Model.3 is the best option owing to its reasonably high values of Precision, Recall, and also Accuracy (90.2%, which is therefore acceptable according to [12]). Moreover, Model.3 can be applied after the seventh week of the course, someway before the halfway point of the semester. And finally, because Model.4's Recall does not improve that of the Model.3, and given that our purpose included identification "as soon as possible", we can state that Model.3 is better than Model.4.

Conclusion and further research

The main contribution of the present study is to provide a simple and easy-to-use procedure, by means of several classification conditional tree-based models, to identify dropout-prone students before the halfway point of the semester. Firstly, it is simple since there is only a single attribute that contributes to classifying students. That attribute is related to students' behaviour with respect to the completion of low-stake assessment assignments such as quizzes posed by teachers and not related to the usage of the e-learning platform, like forum participation. And secondly, it is easy to use because simply by knowing every time a student has not completed one of the first three posed quizzes is enough to identify him/her directly as a dropout-prone student. Furthermore, because the information required is not only easily accessible by the teacher, but also does not need to be processed, teachers can control the procedure by themselves and implement it once the first quiz is submitted. If the performance measures entail a serious concern for the teacher, the previous procedure has to be modified in some way, although it remains simple and easy-to-use. The procedure consists of checking whether students have completed all of the first three quizzes. If the answer is no, the student is identified as a dropout student.

According to the methodology selected, the students that belong to the training set, with whom the classification models have been fitted, and the students of the validation set, whose performance has been evaluated, are enrolled all together in the same academic year. This limitation could lead to further research. The studies of Lykourentzou [11], Lara [12] and Kotsiantis [13], which create the training set in one academic period and the test set in a different one, are references that it would be useful to bear in mind. A second aspect that could be included in further research is the extension of the identification procedure to the fail-prone students [6], so that a richer approach to the dropout prediction problem could be achieved.

Acknowledgements

This paper has been partially supported by a Fundació IBADA grant. We would like to thank Dr Laura Calvet and Mr Paul Garbutt for their valuable contributions in helping to improve this study.

References

1. Lee, Y., Choi, J.: A review of online course dropout research: Implications for practice and future research. Educ. Technol. Res. Dev. 59, 593–618 (2011).
2. Vossensteyn, H., Kottmann, A., Jongbloed, B., Kaiser, F., Cremonini, L., Stensaker, B., Hovdhaugen, E., Wollscheid, S.: Drop-Out and Completion in Higher Education in Europe - Literature Review. (2015).
3. Cohen, A.: Analysis of student activity in web-supported courses as a tool for predicting dropout. Educ. Technol. Res. Dev. 65, 1285–1304 (2017).
4. Burgos, C., Campanario, M.L., Peña, D. de la, Lara, J.A., Lizcano, D., Martínez, M.A.: Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. Comput. Electr. Eng. 66, 541–556 (2018).
5. Grau-Valldosera, J., Minguillón, J.: Rethinking dropout in online higher education: The case of the universitat oberta de catalunya. Int. Rev. Res. Open Distance Learn. 15, 290–308 (2014).
6. Liu, S., Gomez, J., Yen, C.-J.: Community College Online Course Retention and Final Grade: Predictability of Social Presence. J. Interact. Online Learn. 8, 165–182 (2009).
7. Levy, Y.: Comparing dropouts and persistence in e-learning courses. Comput. Educ. 48, 185–204 (2007).
8. Dupin-bryant, P.A.: Pre-Entry Variables Related to Retention in Online Distance Education. Am. J. Distance Educ. 18, 199–206 (2011).
9. Costa, E.B., Fonseca, B., Santana, M.A., de Araújo, F.F., Rego, J.: Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Comput. Human Behav. 73, 247–256 (2017).
10. Santana, M.A., Costa, E.B., Neto, B.F.S., Silva, I.C.L., Rego, J.B.A.: A predictive model for identifying students with dropout profiles in online courses. In: Workshop Proceedings of the EDM 2015 International Conference on Educational Data Mining Vol 1446.
11. Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., Loumos, V.: Dropout prediction in e-learning courses through the combination of machine learning techniques. Comput. Educ. 53, 950–965 (2009).
12. Lara, J.A., Lizcano, D., Martínez, M.A., Pazos, J., Riera, T.: A system for knowledge discovery in e-learning environments within the European Higher Education Area - Application to student data from Open University of Madrid, UDIMA. Comput. Educ. 72, 23–36 (2014).

13.Kotsiantis, S.B., Pierrakeas, C.J., Pintelas, P.E.: Preventing Student Dropout in Distance Learning Using Machine Learning Techniques. In: Proceeding of the 7th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, KES 2003, pp. 267–274, Oxford, UK (2003).

14.Romero, C., López, M.I., Luna, J.M., Ventura, S.: Predicting students' final performance from participation in on-line discussion forums. Comput. Educ. 68, 458–472 (2013).

15.Ruiperez-Valiente JA, Muñoz-Merino PJ, Andújar A, Delgado-Kloos C.: Early Prediction and Variable Importance of Certificate Accomplishment in a MOOC. Proceedings of the European Conference on Massive Open Online Courses, 263-272 (2017).

16.Sharma K, Kidzinski L, Jermann P, Dillenbourg P.: Towards Predicting Success in MOOCs: Programming Assignments. Proceedings of the European Stakehold SUMMIT on Experiences and Best Practices Around MOOCs (EMOOCS), 135–148 (2016).

17.Yang D, Sinha T, Adamson D, Penstein Rose C.: "Turn on, Tune in, Drop out": Anticipating Student Dropouts in Massive Open Online Courses. Proceedings of the 2013 NIPS Data-driven education workshop, 1-8 (2013).

18.Figueroa-Cañas J, Sancho-Vinuesa T.: Investigating the relationship between optional quizzes and final exam performance in a fully asynchronous online calculus module. Interact Learn Environ. (2018).

19. James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning. (2013).

20. Kotsiantis SB, Zaharakis ID, Pintelas PE (2006) Machine learning: A review of classification and combining techniques. Artif Intell Rev 26:159–190.

21.Hothorn, T., Hornik, K., Zeileis, A.: Unbiased recursive partitioning: A conditional inference framework. Research Report Series 8, Department of Statistics and Mathematics, WU Wien, 2004. J. Comput. Graph. Stat. 15, 651–674 (2006).