

Análisis de la mejora de los resultados de una asignatura inicial de programación tras cambiar la planificación de actividades

Alberto Gómez¹, Maria-Jesús Marco-Galindo², Julià Minguillón², Jordi Escayola Mansilla²

¹Universidad de Extremadura, ²Universitat Oberta de Catalunya
Cáceres Barcelona

agomez@unex.es, mmarcog@uoc.edu, jminguillona@uoc.edu, jescayolam@uoc.edu

Resumen

El objetivo de este trabajo es comprobar que los cambios introducidos en la planificación de actividades de evaluación continua de una asignatura de introducción a la programación han sido el factor determinante en la mejora de los resultados de los estudiantes. La evaluación continua pasó de tres entregas en el semestre a distribuir semanalmente los mismos contenidos y tareas en diez entregas. Aunque un análisis descriptivo de los datos muestra una mejora importante en la tasa de rendimiento tras el cambio, los estudiantes de cada semestre pueden tener perfiles distintos, circunstancia que puede influir también en los resultados. Así, se ha aplicado la técnica de análisis cuasi experimental *Propensity Score Matching* para poder comparar los resultados en estas poblaciones potencialmente distintas y comprobar si la mejora se debe efectivamente a la nueva planificación de actividades. Además, se han analizado varios modelos de regresión logística para intentar explicar la influencia del cambio de la planificación en la mejora de los resultados de la asignatura. A la vista del estudio realizado, se concluye que la temporalización semanal de las actividades ha supuesto un cambio positivo en los resultados de la asignatura.

Abstract

The aim of this project is to prove if the planification changes done in the continuous evaluation for the programming course is the key factor that explains the increase in the student's results. The continuous evaluation of this introductory programming course went from having three deliveries in the semester in the past to distribute the same contents and tasks in ten weekly deliveries. Although the descriptive analysis of the data shows a significant improvement in the performance rate after the change, the students in each semester could have different profiles that also influenced the results. The quasi-experimental analy-

sis technique *Propensity Score Matching* has been applied to be able to compare the results in these potentially different populations and check if the improvement is due to the new planning of activities. In addition, some logistic regression models have been analysed to explain the influence in the planification changes and its relationship with the results improvement. According to the obtained results, it is concluded that the weekly timing of the activities has led to a positive result in the students' performance.

Palabras clave

Propensity Score Matching, learning analytics, planificación de actividades, online education.

1. Introducción

El principal objetivo de este trabajo es comprobar si los cambios introducidos en la planificación de las actividades de evaluación de una asignatura de introducción a la programación en una universidad en línea han influido en la mejora observada en el rendimiento de los estudiantes.

El análisis de las tasas de rendimiento de los dos semestres anteriores y los dos posteriores a la intervención realizada muestra un aumento del porcentaje de estudiantes que supera la asignatura y un acusado descenso del porcentaje de no presentados en la evaluación continua.

Sin embargo, en el análisis de los datos hay que valorar el hecho de que los grupos de estudiantes de cada semestre tengan perfiles distintos puede influir también en la mejora observada. La población de estudiantes de cada semestre es diferente y, por tanto, no se deberían comparar directamente los resultados obtenidos en distintos semestres. Habrá que aplicar técnicas de análisis de datos que permitan comparar poblaciones potencialmente distintas. En nuestro caso se ha usado la técnica de análisis cuasi experimental denominada *Propensity Score Matching* (PSM).

Los diseños cuasi experimentales son una herramienta muy útil para poder comparar resultados cuando se hacen cambios en el diseño y planificación de asignaturas y titulaciones. De esta manera se puede evaluar el alcance de los cambios introducidos cuando los grupos son distintos.

Este trabajo se organiza de la siguiente forma: el apartado 2 revisa trabajos previos relacionados con la enseñanza de la programación y las técnicas de diseño cuasi experimentales. El apartado 3 describe el contexto de aplicación de esta intervención. En el apartado 4 se presenta un breve resumen del análisis de datos realizado. A continuación, se explica el emparejamiento de datos realizado (apartado 5) y los modelos explicativos construidos para determinar la influencia de la intervención realizada (apartado 6). Por último, en el apartado 7, se resumen las conclusiones principales del trabajo y se presentan líneas de trabajo futuras.

2. Estado de la cuestión

Las asignaturas de introducción a la programación suelen tener un alto nivel de abandonos y suspensos. Hay muchos trabajos que tratan de explicar los motivos del abandono, desarrollar estrategias para minimizarlo y, en general, mejorar el aprendizaje de esta materia básica para los cursos posteriores [15]. La motivación (ya sea intrínseca o extrínseca) se presenta como un aspecto fundamental para conseguir buenos resultados, especialmente si hablamos de estudios no presenciales. Como explica Bueno, “para motivarnos o para conseguir que los demás se motiven, hay que tener un objetivo concreto que sea tangible” [2]. Objetivos concretos, a corto plazo, mantienen y favorecen la motivación. Biggs y Tang indican que las actividades de evaluación a lo largo del curso mantienen la asistencia a clase y el trabajo con los materiales de la asignatura, además de servir para la evaluación formativa y la calificación [1].

Este hecho se puede extrapolar a la enseñanza virtual, donde una adecuada planificación de actividades es imprescindible para que el estudiante se mantenga motivado y participativo. Así, la evaluación formativa continua y el *feedback* que se proporciona a los estudiantes de las actividades que realizan es muy importante e influye en la mejora del rendimiento. Una mala planificación de las actividades de evaluación puede tener consecuencias negativas en estudiantes y profesorado [17]. Dar *feedback* se considera uno de los mejores métodos para incrementar el aprendizaje de los estudiantes [13,18].

La minería de datos educativa (*educational data mining*) es una disciplina que investiga la aplicación de algoritmos de minería de datos sobre distintos tipos de datos educativos para comprender mejor cómo es el proceso de aprendizaje del estudiante [16].

El uso de plataformas de aprendizaje y los estudios a través de Internet han incrementado mucho la cantidad de datos disponibles. Empleando distintas técnicas de análisis de datos, el comportamiento de los estudiantes a lo largo del curso dentro de la plataforma virtual puede dar información importante sobre qué actividades son más importantes, qué factores influyen más en los resultados o cómo dar mejor *feedback* [16]. La predicción del abandono o de los resultados es también un área de interés en la analítica del aprendizaje (*learning analytics*) [4,11]. El resultado del análisis de los datos se utiliza para proporcionar información relevante a los profesores, visualizar datos de estudiantes, construir sistemas de recomendación, predecir su desempeño, categorizarlos, detectar comportamientos no deseables (fraude), planificar actividades, etc. [16].

2.1. Diseño experimental y PSM

Para determinar la influencia de una intervención en un sistema se debe diseñar un experimento donde se establezcan grupos similares que se diferencien solo en ese punto que se está estudiando. Así se podrá analizar posteriormente la influencia de dicha intervención en los resultados obtenidos. En campos como la medicina, el diseño experimental debe ser siempre aleatorio, donde el grupo de control y el grupo sujeto del estudio se formen siguiendo una metodología estricta que asegure que no influyen otros factores y que el estudio no pierda validez [3]. En investigación en educación es difícil hacer diseños aleatorios de experimentos con grupos de control y selección aleatoria. Las normativas universitarias no permiten tratar a los estudiantes del mismo grupo de forma distinta. Además, no sería ético establecer grupos de control o realizar acciones que puedan perjudicar el aprendizaje de algunos estudiantes [6].

En muchos estudios con experiencias de innovación docente se comparan los resultados del grupo en el que se ha aplicado la intervención con los resultados de un grupo del curso anterior, para intentar demostrar que ha habido una mejora. En general, este análisis no es completamente válido porque los grupos no se han formado aleatoriamente y las poblaciones pueden ser distintas. En estos casos estamos ante cuasi experimentos donde no se puede seleccionar a los individuos que están en cada categoría analizada [6]. Por tanto, hay que utilizar las técnicas adecuadas para estudiar estas situaciones.

En los diseños cuasi experimentales, los métodos de emparejamiento por índice de propensión (*Propensity Score Matching*, PSM) son muy utilizados [7,14]. Con estos métodos se pretende reducir el posible sesgo provocado por las variables que intervienen en la formación de los grupos y así poder analizar los datos posteriormente como si provinieran de un ensayo aleatorio. El índice de propensión (*propensity*

score, *PS*) de una observación del conjunto de datos es la probabilidad o propensión de formar parte o no del grupo que ha sido sometido a la intervención, calculada a partir de las variables explicativas con las que se cuenta. Generalmente, los *PS* se calculan mediante un modelo de regresión logística con la participación o no en la intervención como resultado y algunas variables del modelo como predictoras.

Posteriormente, se emplea un algoritmo de emparejamiento de los datos con *PS* similares (*Propensity Score Matching*) a partir de los conjuntos de datos iniciales, rechazando los elementos que no se puedan emparejar. Con los elementos emparejados ya se puede analizar el impacto de la intervención como si los datos se hubieran conseguido en un experimento aleatorio. El emparejamiento entre los dos conjuntos puede hacerse atendiendo a distintos criterios de similitud que dará lugar a distintos conjuntos de datos emparejados. Además, se debe comprobar que no se han perdido demasiadas observaciones en el emparejamiento y que las características de los conjuntos que reciben o no la intervención están equilibradas [8]. Harris presenta una amplia revisión de la literatura donde se usa *PSM* en educación superior [7]. En [10] se muestra un caso donde se usa *PSM* para eliminar el sesgo producido al no haber podido formar el grupo de control de manera aleatoria.

3. Contexto

Los datos de este trabajo provienen de la asignatura "Fundamentos de programación" que se oferta como primera asignatura de programación en varios grados y programas formativos de la Universitat Oberta de Catalunya (UOC) [11].

El perfil de las personas que cursan estudios virtuales es muy variado, y puede cambiar mucho de un semestre a otro. La mayoría compatibiliza sus estudios con su trabajo, así que cursan pocas asignaturas cada semestre. En la UOC, cada estudiante consulta con su tutor la idoneidad de las asignaturas elegidas para evitar problemas de coordinación y ajustarlas a su tiempo de dedicación. Hay un proyecto institucional para coordinar asignaturas y mejorar la experiencia de los estudiantes en sus primeros cursos [5].

La mayoría de los estudiantes provienen de grados en Ingeniería en Informática. El modelo de calificación principal está basado en la evaluación continua. A lo largo del semestre se proponen una serie de actividades de evaluación continua (PAC) que ponen en práctica los principales conceptos que se van presentando en la asignatura. Además, los estudiantes deben entregar dos prácticas de programación más complejas que integran todas las competencias (PR1 y PR2). La realización de la primera práctica es requisito imprescindible para poder aprobar la asignatura. Cada semestre se desarrolla a lo largo de unas 15

semanas. La primera práctica se entrega aproximadamente en la novena semana del curso. La entrega de PR1 es un buen indicador de si el estudiante va a poder superar la asignatura o no. Desafortunadamente, hay un alto porcentaje de estudiantes que no llegan a entregarla.

Los cambios en la planificación implantados por el profesorado de la asignatura han estado encaminados a potenciar el trabajo continuo en la asignatura, con entregas más frecuentes y la obtención de *feedback* más a menudo. De esta forma se ha pretendido reducir el abandono y mejorar los resultados. Antes de la modificación de la planificación, los estudiantes entregaban tres pruebas de evaluación continua (PAC1 a PAC3) a lo largo del semestre, una cada 4 o 5 semanas. La entrega de PAC2 se realizaba poco antes de la entrega de PR1. Tras cada prueba, el estudiante recibía *feedback* general con una solución comentada y *feedback* personalizado con la corrección y calificación de su trabajo.

El cambio en la planificación supuso, principalmente, dividir cada PAC anterior en PAC más cortas, de forma que las entregas son más frecuentes. El tipo de problemas incluidos era el mismo. Simplemente se reorganizaron los mismos ejercicios en tres bloques con múltiples entregas. La PAC1 se dividió en 4 (nuevas PAC1 a PAC4), la PAC2 en 4 (nuevas PAC5 a PAC8) y la PAC3 se partió en 2 (nuevas PAC9 y PAC10). Con esta nueva planificación, las entregas pasan a ser semanales, recibiendo el estudiante *feedback* cada semana con la publicación de la solución. Además, hay que considerar también la propia percepción de progreso del estudiante a medida que avanza en sus tareas. El *feedback* individualizado del profesor y la calificación se han seguido entregando en tres momentos del semestre, después de cada bloque.

4. Análisis exploratorio de datos

Para este trabajo se ha contado con los datos de cuatro semestres: dos previos a la modificación de la planificación de actividades (segundo semestre del curso 2016-17 y primero del curso 2017-18) y dos posteriores a los cambios (segundo semestre del curso 2017-18 y primero del 2018-19). Los datos provienen del *Learning Record Store* institucional [12].

Se ha contado con un total de 1487 observaciones: 721 anteriores al cambio y 766, posteriores.

4.1. Variables

Los dos conjuntos de datos tienen las mismas variables que caracterizan al estudiante y a sus estudios. Solo se diferencian en las variables que hacen referencia a las calificaciones de las 3 PAC (datos previos al cambio, *pre*) o de las 10 PAC (datos posteriores al cambio, *post*).

Cada observación describe el perfil y calificaciones de un estudiante en un semestre. Las principales variables analizadas se pueden agrupar en distintas categorías:

- Perfil socio-demográfico: sexo, grupo de edad.
- Perfil de estudios: grado en el que está matriculado, vía de acceso a la titulación, si proviene de estudios de formación profesional o de estudios universitarios, número de asignaturas matriculadas en el semestre, número de semestres que lleva matriculado en el grado, veces que ha repetido la asignatura.
- Calificaciones obtenidas: calificaciones de cada PAC y de PR1 y PR2, calificación final de prácticas y de evaluación continua.

Las calificaciones vienen dadas con una escala que, ordenada de mayor a menor, es: *A*, *B*, *Cm*, *Cn*, *D*, *N*. La calificación *N* supone que no se entregó la PAC correspondiente; *Cn* y *D* es suspenso; el resto son valores aprobados.

En el análisis previo de los datos se recodificaron algunas variables para agrupar categorías que contaban con pocas observaciones

Además, se ha añadido la variable *TRATAM* para separar las observaciones de los semestres previos al cambio (con valor 0) y los posteriores (con valor 1).

4.2. Comparación de los resultados

El objetivo principal del estudio es comprobar si los cambios en la planificación de las actividades de evaluación continua han provocado la mejora de los resultados y una menor tasa de abandono.

La entrega de la primera práctica (variable *faPR1*) es uno de los indicadores más claros del rendimiento final del estudiante en la asignatura. De las personas que entregan la práctica 1, aprueban las prácticas el 83.4% y superan la evaluación continua el 93.3%. Solo el 21% de los que no entregan la práctica superan la evaluación continua de la asignatura (aunque no pueden aprobar la asignatura al ser la práctica un requisito). Así, el porcentaje de *PR1* presentadas ha aumentado en los semestres posteriores al cambio en la planificación pasando del 48.13% al 60.70%.

Muchos de los que entregan *PR1* aprueban las prácticas *CPR* y la evaluación continua *CPAC*, sobre todo en los datos posteriores al cambio.

Datos	faPR1 = 1	Y CPR aprobada	Y CPAC	Am bas
pre	347	244	315	234
post	465	441	443	426

Cuadro 1: Relación entre *faPR1* y aprobados

El cambio podría haber mejorado también las calificaciones finales de evaluación continua *CPAC*. El

cuadro 2 muestra las frecuencias relativas para los dos conjuntos de datos.

Datos	A	B	Cm	Cn	D	N
pre	0.16	0.12	0.13	0.13	0.07	0.39
post	0.16	0.24	0.18	0.09	0.19	0.14

Cuadro 2: Distribución de notas *CPAC*

Se puede observar un incremento de los aprobados, especialmente de las calificaciones *B* y *Cm*. Y, sobre todo, hay un importante descenso de no presentados. El cambio en la planificación parece haber mejorado también las calificaciones finales de las prácticas *CPR*, tal como muestra el cuadro 3. Se aprecia el incremento en aprobados y el descenso en suspensos y no presentados.

Datos	A	B	Cm	Cn	D	N
pre	0.09	0.15	0.08	0.02	0.12	0.54
post	0.27	0.23	0.06	0.02	0.03	0.39

Cuadro 3. Distribución de notas *CPR*

Todos los estudiantes que superan la práctica pueden optar a aprobar la asignatura por examen final, aunque no hayan superado *CPAC*, así que el incremento de personas que superan *CPR* es importante.

4.3. Evolución de las calificaciones

Resulta interesante comparar la evolución de las calificaciones de los estudiantes en las pruebas de evaluación continua antes y después de los cambios en la planificación. En las figuras 1 y 2 se puede ver la evolución de las calificaciones hasta el momento de la presentación de la práctica 1 en ambos conjuntos de datos. En verde están representados los estudiantes que entregan la *PR1*, y en marrón, los que no.



Figura 1. Evolución de las calificaciones (pre)

En ambos gráficos se observa que las calificaciones de los estudiantes no varían demasiado a lo largo del semestre. Solo cambian mucho si abandonan la asignatura en alguna entrega; si no, la mayoría se mueve entre grados de calificación contiguos.

Se puede apreciar cómo hay estudiantes que, poco a poco, van abandonando la asignatura. Y son pocos

los que, después de no presentar alguna PAC, se recuperan y consiguen presentar la práctica. Se puede ver también que, tras los cambios en la planificación, el porcentaje de personas que abandonan la asignatura en las primeras entregas se ha reducido.

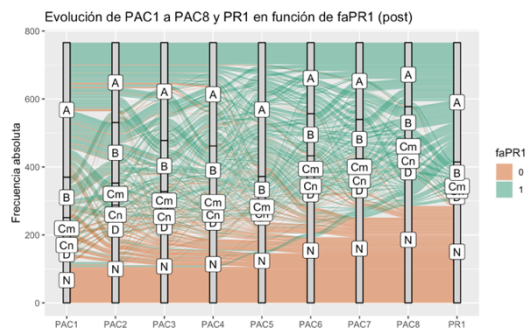


Figura 2. Evolución de las calificaciones (post)

4.4. Análisis exploratorio

También se ha realizado una exploración exhaustiva de las relaciones entre las variables disponibles y no se han encontrado interacciones claras. No se aprecian grandes diferencias en las características generales de los conjuntos de datos anteriores y posteriores a la intervención, más allá de las diferencias en las calificaciones. A partir del análisis exploratorio de los datos parece razonable pensar que el cambio en la planificación de las actividades de evaluación ha mejorado las calificaciones y ha disminuido el abandono de la asignatura. Sin embargo, es necesario preguntarse si esa mejora es significativa, dado que las poblaciones comparadas son distintas.

5. Emparejamiento de los datos

Como los datos provienen de semestres distintos, es posible que la distribución de las variables que definen los perfiles de los estudiantes sean diferentes y, por tanto, los resultados de la evaluación estén sesgados por las diferencias en los conjuntos de datos y no solo por la influencia de la intervención realizada. La técnica cuasi experimental *Propensity Score Matching* permite conseguir datos comparables previos y posteriores a la intervención, y así poder analizar el impacto de la intervención como si los datos provinieran de un estudio aleatorio.

Existen muchos algoritmos para emparejar las observaciones que tienen PS similares. En este estudio se han probado varios de los algoritmos más usuales para comparar los conjuntos de datos emparejados que se consiguen y seleccionar los que presentaban un grado mayor de equilibrio.

Los algoritmos que se han usado, según el nombre en el paquete *MatchIt* [9] de R con el que se han desarrollado los cálculos, son los siguientes:

- *Exact Matching*: se emparejan las observaciones que tienen exactamente los mismos valores en todas las variables.
- *Subclassification*: se forma un número determinado de clases disjuntas en las que la distribución de covariables es tan similar como sea posible.
- *Nearest Neighbor Matching*: se usa el PS para emparejar cada observación de un conjunto con el que tenga un valor más parecido en el otro. Se hace la asignación en orden, con un algoritmo voraz, que a veces empareja observaciones con valores muy distintos. En estos casos se puede usar una opción adicional, un calibre (*caliper*) que obliga a que la diferencia entre los valores emparejados sea menor que la proporción indicada de la desviación típica de las distancias calculadas.
- *Full Matching*: se forman tantas clases disjuntas como sea necesario. En cada clase hay una observación de uno de los dos conjuntos y tantas del otro conjunto como coincidan con ella.

5.1. Equilibrio tras el emparejamiento

Aunque se han realizado pruebas con todos los algoritmos de emparejamiento, los resultados no son adecuados en algunos de ellos. Por ejemplo, con el algoritmo *Exact Matching*, aunque el equilibrio final es absoluto, se pierde más de la mitad de las observaciones, al emparejar solo las observaciones idénticas de cada conjunto. En todos los métodos se han descartado 8 casos del conjunto de datos *post* porque los valores de PS están fuera de la región de soporte común. Este valor tan bajo de observaciones descartadas ya da una idea de que las poblaciones son bastante similares incluso sin emparejamiento. Los resultados han sido adecuados, con pocas pérdidas de datos y variables finales equilibradas, en el caso de los algoritmos *Nearest Neighbor Matching* con calibres 0.1 y 0.2, y con el algoritmo *Full Matching*, que proporciona los mejores resultados. Solo descarta 8 observaciones y el equilibrio es adecuado en todas las covariables. La distribución de las PS emparejadas y descartadas se puede ver en la figura 3.

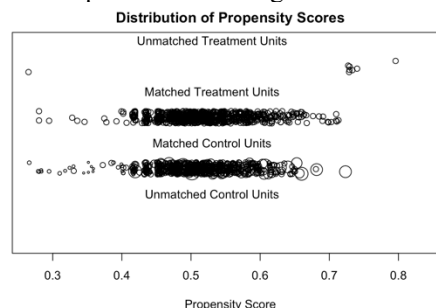


Figura 3. Distribución de PS

Todas las covariables están equilibradas tras el emparejamiento, como puede apreciarse en la figura 4.

Con los datos resultantes tras este emparejamiento equilibrado ya se puede realizar el estudio comparativo que se desee, considerando que los datos que tenemos provienen, a efectos prácticos, de una población homogénea.

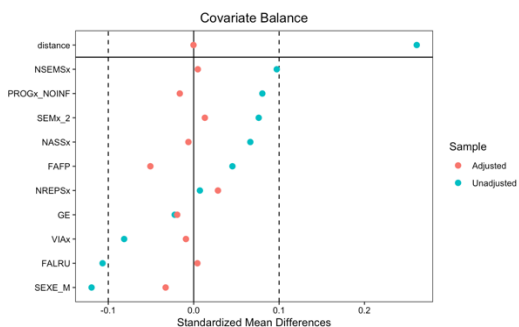


Figura 4. Equilibrio tras el emparejamiento

6. Modelos explicativos

Una vez realizado el proceso de emparejamiento, se puede trabajar con estos datos como si procedieran de un experimento aleatorio. En nuestro caso, el grupo de control está formado por los estudiantes que cursaron la asignatura en los dos primeros semestres, antes del cambio en la planificación. El grupo que ha recibido el tratamiento estaría formado por los estudiantes que han cursado la asignatura en los dos semestres posteriores.

El objetivo ahora es determinar si el cambio en la planificación ha influido positivamente en la superación de la asignatura. Más concretamente, interesa saber si este cambio ha disminuido de forma significativa el abandono de la evaluación continua en las primeras semanas del curso y si ha aumentado el porcentaje de personas que presentan la primera práctica de la asignatura, requisito para poder aprobar la asignatura y un buen indicador del resultado final.

Para ello se han desarrollado y analizado una serie de modelos de regresión logística donde se explorará la influencia de la variable *TRATAM*. Los modelos de regresión logística sirven para predecir el resultado de variables categóricas binarias.

Estos modelos, en función de las variables consideradas, calculan la probabilidad de que un individuo termine en una categoría u otra de la variable objetivo.

6.1. Homogeneización de calificaciones

Para poder desarrollar modelos de regresión con los datos disponibles hay que tener en cuenta que no hay el mismo número de variables *PAC* en los dos conjuntos de datos y que las variables *PAC1* y *PAC2* no hacen referencia a actividades directamente compara-

bles en ambos conjuntos de datos. Se ha creado una nueva variable *NOTAPAC1* para homogeneizar las calificaciones obtenidas durante las primeras semanas del curso. *NOTAPAC1* tiene el valor de la calificación de la *PAC 1* en el conjunto de datos previo y la media aritmética de las cuatro primeras *PAC* en el conjunto de datos posterior. Los contenidos que se califican coinciden. También se ha definido la variable *faNOTAPAC1* como indicador del seguimiento de la evaluación continua durante las primeras cuatro actividades. Será 0 si no se ha presentado nada de la evaluación continua y 1 en caso contrario.

6.2. Modelos analizados

Se han desarrollado distintos modelos de regresión logística con los datos obtenidos tras la fase de PSM con los algoritmos *Full Matching* y *Nearest Neighbor Matching* con calibre 0.2, el más utilizado en la literatura. Los resultados han sido similares.

Los modelos de regresión se han construido para determinar la influencia directa de las variables en la variable binaria elegida como objetivo, así como las posibles interacciones. Aunque eran especialmente interesantes las interacciones de la variable de control *TRATAM* con las calificaciones, también se han explorado otras relaciones con variables como el sexo, programa de estudios, vía de acceso, etc. El análisis de la influencia se suele hacer a partir de los *odds ratio*, la exponencial de los coeficientes del modelo. Para cada incremento de una unidad en una variable cuantitativa, el *odds ratio* se puede interpretar como la oportunidad de encontrar el valor esperado entre los individuos con ese incremento en el valor de la variable, frente a los individuos que no lo muestran. En las variables categóricas, el incremento se refiere al paso desde el valor base a cada uno de los valores de las variables ficticias.

A continuación, se presentan dos de los modelos desarrollados con una capacidad de explicación buena y con indicadores de validez y calidad adecuados.

6.3. Seguimiento de la evaluación continua

En el primer modelo se intenta explicar la variable que indica si se han presentado las actividades de las primeras semanas del curso, *faNOTAPAC1*, a partir de la información del perfil del estudiante. De forma resumida, el pseudoajuste R^2 es muy bajo, lo que representa un mal ajuste del modelo. Las únicas variables significativas en este modelo son *TRATAM* (que indica si pertenece o no a los semestres posteriores al cambio) y *NREPSx* (el número de repeticiones de la asignatura). A medida que aumenta el número de veces que se repite la asignatura parece que disminuye la probabilidad de seguir la evaluación continua. Pertener a los semestres posteriores al cambio influye en una menor probabilidad de haber abando-

nado la asignatura desde las primeras semanas. Ninguna de las interacciones entre variables exploradas en este modelo ha resultado significativa.

Este primer modelo, aunque no demasiado bueno, da una primera idea de que el cambio de planificación de las actividades ha supuesto un aumento de la probabilidad de seguir la evaluación continua y de no abandonar en las primeras semanas.

6.4. Presentación de la PR1

Como la entrega de la primera práctica es un buen indicador del seguimiento de la asignatura, se ha intentado explicar esa variable *faPR1* en función del resto de variables de perfil y *NOTAPAC1*. También se han explorado algunas interacciones entre variables (como estar en una titulación de Ingeniería Informática con la experiencia universitaria previa o con el número de semestres en la titulación), pero la única significativa es la interacción entre *NOTAPAC1* y la variable *TRATAM*. En este modelo, mostrado en el cuadro 4, el ajuste es adecuado.

Predictors	faPR1		
	Odds Ratios	CI	p
(Intercept)	0.20	0.08 – 0.49	<0.001
TRATAM	0.74	0.37 – 1.45	0.377
SEXE [M]	0.70	0.48 – 1.01	0.055
GE	0.75	0.66 – 0.85	<0.001
PROGx [NOINF]	1.13	0.73 – 1.74	0.586
EXP [SIN]	1.08	0.74 – 1.58	0.681
FAFP	0.65	0.42 – 1.00	0.050
FALRU	1.04	0.28 – 3.91	0.958
NASSx	1.11	0.98 – 1.26	0.099
NSEMSx	1.10	0.93 – 1.30	0.290
NREPSx	0.67	0.51 – 0.87	0.003
SEMx	0.88	0.66 – 1.16	0.360
NOTAPAC1	1.59	1.48 – 1.70	<0.001
TRATAM * NOTAPAC1	1.19	1.08 – 1.33	0.001
PROGx [NOINF] * NSEMSx	1.16	0.92 – 1.46	0.221
PROGx [NOINF] * EXP [SIN]	1.06	0.56 – 1.99	0.859
Observations	1479		
R ² Tjur	0.444		

Cuadro 4. Resultado del modelo logístico para *faPR1*

Las variables que resultan significativas son el grupo de edad, si se proviene de estudios de Formación Profesional o no, el número de repeticiones de la asignatura, *NOTAPAC1* y la interacción entre *TRATAM* y *NOTAPAC1*.

Además, hay que tener en cuenta que, mientras las primeras son variables dicotómicas o con pocos

valores, *NOTAPAC1* tiene un rango de valores superior (entre 0 y 10). Por tanto, la interacción con la variable *TRATAM* puede suponer un incremento de la probabilidad de entregar la práctica mucho mayor. Un *odds ratio* superior a uno supone un incremento de la probabilidad de que el valor de *faPR1* sea 1 con un cambio de la variable correspondiente. Un valor inferior a uno indica que la probabilidad disminuye. Por tanto, según este modelo, a medida que aumenta la calificación de *NOTAPAC1* (la *PAC1* de antes del cambio o la media aritmética de las 4 primeras *PAC* tras el cambio), la probabilidad de presentar *PR1* aumenta. Además, esa influencia se ve reforzada en los semestres posteriores al cambio por la interacción positiva entre *TRATAM* y *NOTAPAC1*.

El *odds ratio* de la variable *TRATAM* es menor que 1, lo que parecería indicar que la intervención disminuye la probabilidad de presentar la práctica 1, justo lo contrario de que lo queríamos conseguir. Sin embargo, la interacción de *TRATAM* con *NOTAPAC1* es positiva, así que ese coeficiente negativo de *TRATAM* solo suaviza un poco el valor de la interacción. Además, no resulta significativo en este modelo.

7. Conclusiones y trabajo futuro

En este trabajo se han analizado los datos de cuatro semestres de los estudiantes que han cursado la asignatura "Fundamentos de programación" para determinar si los cambios en la planificación de las actividades han supuesto una mejora en el rendimiento y una disminución del abandono en la asignatura. Aumentar la frecuencia de las entregas ha mejorado las tasas de seguimiento de la evaluación continua y de presentación de la primera práctica de la asignatura, lo que repercute en un mayor índice de superación de la asignatura. Esta mejora está alineada con los resultados que se presentan en la literatura de investigación en educación en informática [1,13,15,18].

El aumento del *feedback* mejora los resultados y disminuye el abandono de los estudiantes. Con las entregas frecuentes de actividades, los estudiantes reciben *feedback* grupal más frecuentemente, además de tener que gestionar mejor su tiempo. Aunque el *feedback* personalizado por parte del profesorado no se ha incrementado, el estudiante obtiene retroalimentación al ir consiguiendo entregar a tiempo las actividades y puede pedir ayuda antes si es necesario.

La comparación directa de los resultados de semestres distintos no sería correcta porque los perfiles de los estudiantes pueden variar de un semestre a otro. Ante la imposibilidad de un diseño experimental aleatorio en esta intervención, se ha usado la técnica cuasi experimental *Propensity Score Matching* para construir un conjunto de datos equilibrado. Las poblaciones de estudiantes en cada semestre son similares, como se deduce del resultado del PSM.

Los modelos de regresión logística analizados permiten confirmar la influencia positiva de la nueva planificación de la asignatura en la disminución del abandono y en el aumento del seguimiento de la evaluación continua y de la entrega de la primera práctica, lo que influye en las tasas de éxito.

Aunque los modelos de regresión logística permiten valorar qué variables e interacciones influyen más en la entrega de la práctica, la capacidad predictiva tiene mucho margen de mejora. No se cuenta con todas las variables que influyen en el seguimiento de la asignatura por parte del estudiante. Por ejemplo, no se dispone de información sobre horas de dedicación y hábitos de estudio, lo que sería muy interesante analizar en futuros trabajos.

Por otro lado, los perfiles de los estudiantes no son homogéneos en todos los programas de estudio. Con un análisis de los programas con más estudiantes (Ingeniería Informática, especialmente), se podrían obtener datos más ajustados y mejores modelos.

Ahora que se ha comprobado que este nuevo modelo de planificación ha influido positivamente, se podrían usar los datos para construir modelos de predicción más completos que permitan construir perfiles de estudiantes según su comportamiento, detectar en qué actividades aumenta el riesgo de abandono o en cuáles suspenden más, para así poder modificarlas y prestar más ayuda a cada estudiante.

Referencias

- [1] John Biggs y Catherine Tang. 2003. *Teaching for Quality Learning at University*. SRHE and Open University Press.
- [2] David Bueno. 2017. *Neurociencia para educadores*.
- [3] Glynis Cousin. 2009. *Researching Learning in Higher Education: An Introduction to Contemporary Methods and Approaches*. Taylor & Francis.
- [4] A. Dinesh Kumar, R. Pandi Selvam, y K. Sathesh Kumar. 2018. Review on prediction algorithms in educational data mining. *International Journal of Pure and Applied Mathematics*.
- [5] Loles González, Julià Minguillón, Josep Antoni Martínez-Aceituno, y Julio Meneses. 2018. Institutional support to provide freshmen with flexible learning paths at course and semester level in open higher education. En *Proceedings of the 10th European Distance and E-Learning Network Research Workshop, 2018 Barcelona, 24-26 Octubre, 2018 ISBN 978-615-5511-25-7*.
- [6] Patricia Haden. 2019. Descriptive Statistics. En *The Cambridge Handbook of Computing Education Research*. Cambridge University Press, 102-132.
- [7] Heather D Harris. 2015. Propensity score matching in higher education assessment. *Masters Theses*.
- [8] Heather Harris y S. Jeanne Horst. 2016. A brief guide to decisions at each step of the propensity score matching process. *Practical Assessment, Research and Evaluation* 21, 4.
- [9] Daniel E Ho, Kosuke Imai, Gary King, y Elizabeth A Stuart. 2011. {MatchIt}: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software* 42, 8: 1-28.
- [10] Lisa Angelique Lim, Sheridan Gentili, Abelardo Pardo, Vitomir Kovanović, Alexander Whitelock-Wainwright, Dragan Gašević, y Shane Dawson. 2019. What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course. *Learning and Instruction*.
- [11] Maria-Jesús Marco-Galindo, Julià Minguillón, y Teresa Sancho-Vinuesa. 2020. Análisis de la progresión de los estudiantes en una asignatura introductoria a la programación mediante redes bayesianas. *Actas de las Jornadas sobre Enseñanza Universitaria de la Informática* 5.
- [12] Julià Minguillón, Jordi Conesa, M Rodríguez, y F Santanach. 2018. Learning Analytics in Practice: Providing E-Learning Researchers and Practitioners with Activity Data: Emerging Technologies for Teaching and Learning. En *Lecture Notes in Educational Tech*. 145-167.
- [13] Margaret Price, Karen Handley, y Jill Millar. 2011. Feedback: Focusing attention on engagement. *Studies in Higher Education* 36, 8: 879-896.
- [14] Claudia Ramírez Ovalle. 2015. Sobre la Técnica de Puntajes de Propensión (Propensity Score Matching) y sus usos en investigación en Educación. *Educación y Ciencia* 4, 43:81-89.
- [15] Anthony V. Robins. 2019. Novice Programmers and Introductory Programming. En *The Cambridge Handbook of Computing Education Research*. Cambridge University Press, 327-376.
- [16] Cristóbal Romero y Sebastián Ventura. 2010. Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 40, 601-618.
- [17] Jesús Serrano Guerrero, Francisco Pascual Romero Chicharro, y José Ángel Olivás Varela. 2011. La sobre-evaluación: efectos negativos de una mala planificación de la evaluación. *ReVisión* 4, 2.
- [18] Valerie J. Shute. 2008. Focus on Formative Feedback. *Review of Educational Research* 78, 1: 153-189.