

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/343140394>

Determinantes del rendimiento académico de los estudiantes de nuevo acceso a la Universidad Complutense de Madrid

Article in *Revista de Educación (Madrid)* · January 2020

DOI: 10.4438/1988-592X-RE-2020-387-433

CITATIONS

21

2 authors:



María Fernández Mellizo-Soto

Complutense University of Madrid

52 PUBLICATIONS 212 CITATIONS

SEE PROFILE

READS

723



I. Alexander Constante Amores

Universidad Camilo José Cela y Universidad Europea de Madrid

8 PUBLICATIONS 80 CITATIONS

SEE PROFILE

Determinantes del rendimiento académico de los estudiantes de nuevo acceso a la Universidad Complutense de Madrid

Factors associated to the academic performance of new entry students at Complutense University of Madrid

María Fernández-Mellizo
Alexander Constante-Amores



Determinantes del rendimiento académico de los estudiantes de nuevo acceso a la Universidad Complutense de Madrid

Factors associated to the academic performance of new entry students at Complutense University of Madrid

DOI: 10.4438/1988-592X-RE-2020-387-433

María Fernández-Mellizo
Alexander Constante-Amores
Universidad Complutense de Madrid

Resumen

En los últimos años existe un gran interés por conocer los factores o determinantes que inciden en el rendimiento académico del alumnado de la educación superior, de gran utilidad a la hora de evaluar la calidad y la equidad de las universidades. Esta investigación tiene como objetivo analizar los factores individuales (demográficos, socioeconómicos y académicos) asociados al rendimiento académico de los nuevos universitarios, que son aquellos con más posibilidades de abandonar sus estudios. Para ello, se ha analizado la cohorte de estudiantes de nuevo acceso a grados de la Universidad Complutense de Madrid en el curso 2017/18, conformada por un total de 10.720 estudiantes. Los datos se han analizado mediante las técnicas de regresión logística binaria y de árboles de clasificación. Como variables independientes se han seleccionado un total de once, de las cuales cuatro (*nota de acceso a la Universidad, tipo de centro en la educación secundaria, sexo y Comunidad Autónoma familiar*) han resultado estadísticamente significativas en el modelo final. Los resultados muestran que cuanto más elevada es la nota de acceso a la Universidad, más probabilidades tiene el estudiante de obtener un buen rendimiento académico. Del mismo modo, si el alumno ha estudiado Bachillerato en un instituto público, es mujer y/o de Madrid, las probabilidades de obtener un buen rendimiento se incrementan. En general, las variables académicas, y en particular la nota de

acceso a la Universidad, son las que más contribuyen a explicar el desempeño universitario, aunque las variables demográficas tienen cierto peso también. En cambio, las variables socioeconómicas no contribuyen sustantivamente a la explicación de este fenómeno, algo que puede reflejar que los estudiantes, en el nivel superior, están muy seleccionados desde el punto de vista social y, por tanto, la fuerza estadística de estas variables es más reducida que en niveles educativos anteriores.

Palabras clave: educación superior, estudiantes universitarios, rendimiento académico, regresión logística, minería de datos

Abstract

The knowledge of the factors that influence the academic performance of students in higher education is getting attention currently, and it is very useful when evaluating the quality and fairness of universities. The aim of this article is the analysis of individual factors (demographic, socioeconomic and academic) associated to the academic performance of a new cohort of students from the Complutense University of Madrid in the 2017/18 academic year, made up of a total of 10,720 students. Freshmen students are those who are the most likely to drop-out their selected degrees. Data were analysed through binary logistic regression techniques and classification trees. Eleven variables were selected as independent ones. From those, only four have a significant influence in the final model: qualification obtained to access to the university studies, the nature of the institution of origin, gender and the Autonomous Community of the family. The results show that the higher the qualification obtained to access to the university, the more likely the student is to obtain a good academic performance. In the same way, if a student is either women or from Madrid or attended previously to a public institution, the chances of getting a better academic performance will increase. In conclusion, academic variables carry more weight to explain the university performance, although the demographic ones are also important. Among the academic variables, the qualification obtained to access to the university studies is the most relevant. Therefore, socioeconomic variables do not contribute to the explanation, possible due to the already high level of social selection of students at university, which lessens their statistical impact in comparison to previous levels of education.

Key words: higher education, university students, academic achievement, logistic regression, data mining.

Introducción

Los estudiantes universitarios cada vez despiertan más interés por parte de la investigación socio-educativa. Al examen de sus perfiles, estrategias y opiniones se han dedicado numerosas investigaciones, tanto nacionales como internacionales, del mismo modo que se ha tratado con especial atención las características de aquellos que acceden a la enseñanza terciaria (Ariño y Sintés, 2016; EUROSTUDENT, 2018; Fundación BBVA, 2010).

Por su parte, el análisis del rendimiento académico de los estudiantes de educación superior, así como de los factores que inciden en el mismo, ha adquirido bastante notoriedad en los últimos años (García, 2014). La relevancia del rendimiento académico radica, según diversos autores, en que es uno de los indicadores más importantes para determinar la calidad y equidad de la enseñanza universitaria (Rodríguez, Fita y Torrado, 2004). Además, su importancia reside, según otros autores, en que un alto o bajo desempeño por parte del alumnado afecta tanto al discente como a la institución académica (Garbanzo, 2014).

Conocer los factores que explican el rendimiento académico es de utilidad a la hora de evaluar la calidad y la equidad de las universidades, así como para identificar las variables que determinan el éxito o fracaso académico (Garbanzo, 2007), y a la postre incrementar el rendimiento académico de los estudiantes (Lamas, 2015). Incluso, por parte de algunos autores, se reivindica un enfoque en el que analizar los factores asociados al rendimiento académico sirva como estudio previo al diseño de políticas universitarias (Turrul y Roca, 2010).

Las investigaciones que analizan los factores asociados al rendimiento académico en la educación superior no estudian en profundidad los factores individuales (Rodríguez, Ariza y Ramos, 2014). Este artículo surge de la necesidad de estudiar a fondo el impacto de las variables individuales (demográficas, socioeconómicas y académicas) en el rendimiento académico de los estudiantes. Asimismo, a la hora de analizar los determinantes asociados con el rendimiento académico es pertinente y necesario el estudio de los perfiles del alumnado, algo que no ha recibido la atención deseada (Jiménez-Caballero, Camúñez, González-Rodríguez y Fuentes, 2015), y que en este trabajo se ha tratado de abordar. Dicha investigación es crucial realizarla, como se ha hecho en este artículo, en los estudiantes de nuevo acceso o primeros cursos,

ya que son los que tienen más posibilidades de abandonar sus estudios universitarios (García, 2014; Tejedor y García-Valcárcel, 2007) y/o de obtener un bajo rendimiento (Turrul y Roca, 2010).

Debido a todo lo indicado, nos planteamos varias preguntas de investigación: ¿cuáles son los factores individuales (demográficos, socioeconómicos y académicos) que inciden en el rendimiento académico?, ¿cuáles de ellos tienen más importancia en dicho fenómeno? Y, en relación con lo anterior, ¿cuál es el perfil del estudiante que obtiene un buen y un mal rendimiento académico? El objetivo de este estudio es, por tanto, examinar los factores individuales asociados al rendimiento académico del estudiante, así como determinar la importancia de cada uno de ellos y los perfiles de los estudiantes en función de su rendimiento académico.

Para contestar estas preguntas de investigación se ha analizado el caso de la Universidad Complutense de Madrid (UCM), la universidad presencial española más grande en volumen de estudiantes (Manzano, Escribá, Iborra y Safón, 2016). Se ha contado con una base de datos de estudiantes de nuevo ingreso en grado en el curso 2017/18 proporcionada por el Observatorio del Estudiante de dicha universidad.

En el primer apartado se revisa la literatura sobre los factores que inciden en el rendimiento académico en la universidad, sobre todo los que tienen que ver con variables individuales. El segundo apartado se centra en cuestiones metodológicas; se describen los datos, las variables y las técnicas de análisis de datos empleadas. El tercer apartado está dedicado a los resultados de la investigación, tanto los que tienen que ver con los factores individuales que inciden en el rendimiento académico del estudiante como con el establecimiento de perfiles asociados a un buen y a un mal estudiante. Por último, el cuarto apartado presenta una discusión de los resultados y las conclusiones de la investigación.

Factores asociados con el rendimiento académico universitario

No existe consenso por parte de la comunidad científica a la hora de definir el rendimiento académico (Lamas, 2015). Hay que destacar que las notas y los créditos aprobados son los indicadores más utilizados del rendimiento académico (Mañé y Miravet, 2010), siendo las notas el más eficaz según algunos autores (Garbanzo, 2013).

Se han establecido numerosas clasificaciones de los factores asociados al rendimiento académico. En la investigación realizada por Garbanzo (2007) son tres los factores que se encuentran: personales, sociales e institucionales. Por su parte, Tejedor y García-Valcárcel (2007) indican la existencia de cinco determinantes: identificación, psicológicas, académicas, pedagógicas y sociofamiliares. Del mismo modo, el estudio realizado por García (2014) clasifica a los factores que influyen en el rendimiento académico en individuales y organizacionales.

En relación a los factores organizacionales, la literatura consultada encuentra que la asistencia de manera regular tanto a las clases como a las tutorías por parte del alumnado (Clerici y Da Re, 2019) o una buena relación entre el docente y el estudiantado (Garbanzo, 2007), se relacionan positivamente con el rendimiento académico. En esta investigación se ha optado por centrarse en los factores individuales del rendimiento académico, menos estudiados que los organizacionales.

Los factores individuales son divididos por García (2014) en tres grupos: demográficos, socioeconómicos y académicos. La presente investigación utiliza esta última clasificación, ya que se caracteriza por su claridad, precisión terminológica y actualidad.

Factores demográficos

Dentro de los factores demográficos, el más estudiado ha sido el sexo, seguido de la edad. El meta-análisis realizado por Richardson, Abraham y Bond (2012), cuyo objetivo fue determinar los factores psicológicos, demográficos y socioeconómicos asociados con el rendimiento académico en la educación terciaria, incluye un total de 7.167 estudios entre 1997 y 2010, y llega a la conclusión de que el sexo y la edad tienen, no obstante, un tamaño del efecto pequeño.

En cuanto al primero de ellos, las investigaciones realizadas no resultan concluyentes. Algunos estudios señalan que son las mujeres las que demuestran un mejor desempeño académico. La investigación realizada por Pérez-Cárceles, Gómez-Gallego, Gómez-Gallego, Palazón-Pérez y Gómez-García (2014) muestra que las alumnas de la Universidad Católica de San Antonio (Murcia) obtienen un mejor rendimiento académico. De igual manera, el estudio realizado por Duart, Gil, Puyol y Castaño (2008) refiere un rendimiento superior de las mujeres en diversas universidades

de la Comunidad Autónoma de Cataluña. Asimismo, McNabb, Pal y Sloane (2002) hallan que las mujeres universitarias que estudian en Inglaterra y Gales obtienen un mayor desempeño. Sin embargo, otros trabajos muestran que los varones tienen mejor rendimiento académico. La investigación realizada por Santos y Vallelado (2013) revela que los varones pertenecientes a la Licenciatura de Administración y Dirección de Empresas en la Universidad de Valladolid obtienen un mayor rendimiento académico. En la misma línea, Nyikahadzoi, Matamnade, Taderera y Mandimika (2013) señalan que los alumnos (varones) de Contabilidad en la Universidad de Zimbabwe obtienen mejores calificaciones. Además, otros estudios, como el de Danilowicz-Gösele, Lerche, Meya y Schwager (2017) con estudiantes de Göttingen (Alemania), indican que el sexo no es muy importante a la hora de explicar el rendimiento académico.

En lo que concierne a la edad, los estudios tampoco son concluyentes. Por un lado, Jansen (2004) encuentra una relación negativa entre el desempeño y la edad en el alumnado de la Universidad de Groningen (Países Bajos). Por otro lado, varios estudios no encuentran relación entre la edad y el desempeño académico. De Sousa, Taboada, Rivas, Iglesias y López (2017) no encontraron diferencias estadísticamente significativas en el rendimiento académico de los estudiantes universitarios de la República de Angola en función de su edad. Por su parte, Abera (2018), con estudiantes de la universidad de Haramaya (Etiopía), y Bunce, Baird y Jones (2017), haciendo referencia a estudiantes ingleses, tampoco las hallaron.

Otros factores demográficos, menos estudiados, son la nacionalidad (Mañé y Miravet, 2010) y el estado civil. Los estudiantes autóctonos de las universidades nacionales de Argentina (García, 2014) y los estudiantes solteros de la Universidad Católica de Santiago de Guayaquil en Ecuador (Cortez, Tutiven y Villavicencio, 2017) obtienen un mayor desempeño en comparación a los estudiantes inmigrantes y casados, respectivamente.

Factores socioeconómicos

Según varias investigaciones, el nivel socioeconómico de la familia es relevante en el éxito académico universitario (Abera, 2018; Chong, 2017; Dennis, Phinney y Chuateco, 2005), aunque no tanto como lo es en etapas educativas anteriores (Garbanzo, 2013). La mayor parte de los trabajos

establecen una relación positiva entre el origen socioeconómico del estudiante y su rendimiento académico. Por ejemplo, el estudio realizado por Cortez et ál. (2017) con estudiantes de la Universidad Católica de Santiago de Guayaquil (Ecuador) revela que a mayor cualificación del trabajo y nivel educativo de los padres más posibilidad tiene el alumnado de obtener buenas calificaciones. Del mismo modo, el trabajo realizado por Marcenaro y Navarro (2007) pone de manifiesto la influencia de los estudios universitarios del padre en el buen rendimiento académico del alumnado de la Universidad de Málaga. Sin embargo, la investigación realizada por Thiele, Pope, Singleton, Snape y Stanistreet (2017) revela que el estatus socioeconómico familiar repercute negativamente en el desempeño de los estudiantes del Reino Unido únicamente en el caso de clases sociales muy bajas.

Hay pocas investigaciones que aborden la relación entre el trabajo laboral realizado por el estudiante y su rendimiento académico (Simón, Casado y Castejón, 2017), y el desacuerdo entre las mismas es amplio. Según la investigación realizada por Ruesga, Silva y Monsueto (2014), los estudiantes de nuevo acceso a universidades públicas y privadas de España que trabajan obtienen un rendimiento académico más bajo que los que no trabajan. Por el contrario, Simón, et ál. (2017) hallan que el trabajo realizado por los estudiantes de la Universidad de Alicante no repercute en el desempeño académico. Otra variable de naturaleza socioeconómica explorada por alguna literatura es la financiación de los estudios; los estudiantes de la Universidad Privada del Norte del Perú con beca obtienen una mayor nota media (Salinas, Hernández y Barboza-Palomino, 2017).

Factores académicos

Existe una amplia literatura (García, 2014; Jiménez-Caballero et ál., 2015; Vélez y Roa, 2005) que pone de manifiesto que la nota de acceso a la Universidad no sólo está relacionada positivamente con el rendimiento académico de los estudiantes universitarios, sino que es la variable que más contribuye a su explicación. Sin embargo, un trabajo con estudiantes de la Universidad de Alicante muestra como dicho predictor repercute en la probabilidad de finalizar la carrera, pero no en la nota media (Simón, et ál, 2017).

La literatura también encuentra relación entre el orden de elección de los estudios universitarios y el desempeño académico. Los estudiantes que han elegido los estudios en primera opción obtienen mayor rendimiento (Jiménez-Caballero et ál., 2015; Rodríguez, et ál. 2004). Otro factor que se ha estudiado es la vía de acceso; según el estudio de Mañé y Miravet (2010), los estudiantes catalanes que proceden de bachillerato rinden más que los acceden por otras vías.

El tipo de centro también parece encontrarse asociado con el desempeño estudiantil, aunque no hay un consenso por parte de la comunidad científica a la hora de indicar quién rinde más en la universidad. Por un lado, Vinacur (2016) encuentra que los estudiantes de primer año de la Universidad de Buenos Aires que han asistido a centros privados a lo largo de la educación secundaria obtienen un mayor rendimiento que el alumnado que ha ido a un centro público. Por otro lado, Smith y Naylor (2001) hallan que los estudiantes de diversas universidades de Reino Unido que han asistido a centros públicos son los que obtienen un mayor rendimiento académico. En la misma línea, Cerdeira, Catela, Balcão y Seabra (2018) muestran que los estudiantes graduados de Portugal que han acudido previamente a centros privados tienen una ligera desventaja.

La literatura sobre factores individuales asociados al rendimiento académico, por tanto, es bastante concluyente en cuanto al impacto positivo de la nota de acceso y del nivel socioeconómico de los padres del estudiante en su rendimiento académico. La primera hipótesis es formulada en consecuencia:

H1: A mayor nota de acceso de estudiante, mayor rendimiento académico; cuanto más elevado sea el nivel socioeconómico de la familia del estudiante, mejor desempeño en la universidad.

Además, también hay bastante consenso en relación al papel preponderante de la nota de acceso del estudiante en la explicación de su rendimiento académico, y sería lógico pensar que las variables académicas, en general, estén más asociadas al mismo. Por ello, la segunda y tercera hipótesis establecen que:

H2: La nota de acceso a la universidad es la variable que más impacto tiene en el rendimiento académico del estudiante; en general, las variables académicas tienen mayor capacidad explicativa que el resto de variables individuales.

H3: Las variables académicas, y en particular la nota de acceso a la universidad, definen el perfil del estudiante en función de su rendimiento académico.

Como se ha visto, la literatura no es concluyente respecto al impacto de variables como el sexo o la titularidad del centro en el que se ha cursado la educación secundaria, y es escasa respecto a otras variables como la nacionalidad o el orden de elección de estudios, así que no se establecen hipótesis de partida en este sentido.

Metodología

Datos

La presente investigación ha utilizado la base de datos proporcionada por el Observatorio del Estudiante de la UCM. Esta información ha sido obtenida del proceso de preinscripción y de formalización de la matrícula, con la ayuda del Centro de Inteligencia Institucional de dicha universidad. Por consiguiente, los datos analizados incluyen a la población total, son censales.

De los 10,720 estudiantes de nuevo acceso en grado que formalizaron la matrícula en la UCM en el curso 2017/18 se han analizado 8,660 (véase la Tabla I). Esta reducción drástica de sujetos se ha producido debido a que sólo se disponen de las notas de los estudiantes que han accedido por bachillerato (9.468 estudiantes, el 88,3% del total) de los cuales 808 abandonan sus estudios universitarios antes de completar el primer cuatrimestre. No se han tenido en cuenta ni los centros “adscritos” a la UCM, dada la diferente naturaleza de los mismos, ni los dobles grados, puesto que complicaba la clasificación en áreas de conocimiento.

TABLA I. Número de estudiantes por áreas de conocimiento y su correspondiente nota media del primer cuatrimestre

ÁREA	Nº CASOS	NOTA MEDIA DEL PRIMER CUATRIMESTRE*
Ciencias Sociales y Jurídicas	4.919	5,79 (3.828)
Ciencias	1.211	6,23 (1.076)
Artes y Humanidades	1.889	6,70 (1.500)
Ingeniería y Arquitectura	577	5,41 (469)
Ciencias de la Salud	2.124	6,52 (1.787)
Total	10.720	6,13 (8.660)

Fuente: Elaboración propia a partir de la base de datos del Observatorio del Estudiante de la UCM.

Nota: *Entre paréntesis el número de casos que tienen la información de la nota media del primer cuatrimestre.

Variables

La variable dependiente del estudio es el rendimiento académico, medida a través de la nota media obtenida por los estudiantes en el primer cuatrimestre. Se ha optado por dicotomizarla (0 = mal rendimiento académico y 1= buen rendimiento académico)¹. Para su elaboración, lo primero que se realizó fue la agrupación de cada uno de los grados con su área de conocimiento, y se calculó la nota media correspondiente a cada área de conocimiento (Tabla I). Una vez realizado esto, se llevó a cabo la siguiente ecuación con el propósito de dicotomizar la variable dependiente y, de esta manera, diferenciar a los sujetos que obtuvieron un buen y un mal rendimiento académico en el primer cuatrimestre.

$$\text{Rendimiento relativo} = \frac{\text{Nota media del sujeto}}{\text{Nota media del área}} \text{ ecuación (1)}$$

Concretamente, aquellos estudiantes que obtuvieron un valor superior o igual a cero se consideró que alcanzaron un buen rendimiento académico ($Y = 1$), puesto que su nota media es igual o superior a la media de su área de conocimiento. Los valores inferiores a cero, por

⁽¹⁾ No obstante, se han realizado análisis con la variable dependiente continua, obteniendo resultados similares a los mostrados aquí.

su parte, denotan un bajo rendimiento académico ($Y = 0$), al no llegar a la nota media del área de conocimiento. Este procedimiento permite analizar a todos los sujetos de los diferentes grados de la UCM, debido a la disparidad de notas medias entre áreas de conocimiento. En la Tabla II vemos la distribución de los casos en función de la variable dependiente.

TABLA II. Número de estudiantes que obtienen un buen y mal rendimiento académico

Nivel	N	%
Mal rendimiento académico	4032	46,6
Buen rendimiento académico	4628	53,4
Total	8660	100,0

Fuente: Elaboración propia a partir de la base de datos del Observatorio del Estudiante de la UCM.

Se han elegido un total de once variables independientes de la base de datos disponible (Tabla III), basándose en un doble criterio². En primer lugar, se han escogido aquellas tratadas por la literatura. Después, se han incluido en el modelo aquellos predictores (Comunidad Autónoma familiar y convocatoria de admisión) que, aunque no existan estudios sobre su relación con el desempeño, se espera que tengan una influencia en la variable criterio. Dado que la EVAU es diferente en cada Comunidad Autónoma (CCAA), se podría pensar que los estudiantes que provienen de Madrid pueden tener un rendimiento diferente a los estudiantes que provienen de otras comunidades. Además, aquellos estudiantes que acceden en la primera convocatoria, la de junio, podrían tener un rendimiento más elevado que aquellos que acceden en la segunda convocatoria, la de septiembre. Las variables se han dicotomizado para facilitar la interpretación del análisis (Pardo y Ruiz, 2013), aunque se ha comprobado que dicha especificación del modelo no condiciona los resultados³ (véase Anexo, tablas I y II, para los descriptivos de las variables).

⁽²⁾ La base de datos utilizada tiene otra información disponible que no se ha explotado: municipio familiar, provincia familiar, país familiar, ECTS presentados, ECTS matriculados, ocupación del padre, ocupación de la madre y situación laboral del estudiante en el año anterior.

⁽³⁾ Se ha comprobado que especificaciones diferentes de las variables de estudios alcanzados por la madre y el padre, y orden de elección de estudios, no alteran los resultados.

TABLA III. Variables independientes

Factores individuales	Variables independientes	Valores recodificados
Demográficas	Sexo	0 = Hombre 1 = Mujer
	Edad	0 = No idónea (20 o más) 1 = Idónea (17-19)
	Nacionalidad	0 = Otra 1 = Española (autóctono)
	CCAA familiar	0 = Otra CCAA 1 = Madrid
Socioeconómicas	Ocupación del estudiante	0 = Trabaja 1 = No trabaja
	Estudios alcanzados por el padre Estudios alcanzados por la madre	0 = Menos de estudios superiores 1 = Estudios superiores
Académicas	Nota de acceso	5-14
	Tipo de centro en la educación secundaria	0 = Privado 1 = Público
	Convocatoria de admisión	0 = Septiembre 1 = Junio
	Orden de elección de estudios	0 = Segunda a duodécima opción 1 = Primera opción

Fuente: Elaboración propia

Técnicas de análisis de datos

Se han realizado tres tipos de análisis. El primero utiliza la técnica de regresión logística binaria para analizar la contribución relativa de cada variable o conjunto de variables (demográficas, socioeconómicas y académicas) en la probabilidad de obtener un buen rendimiento académico. Para ello, se crearon cuatro modelos. El modelo 1 se encuentra conformado por la nota de acceso a la Universidad, que es la variable que más peso tiene, según la literatura, en el rendimiento⁴, y el modelo 2 incluye además el resto de variables académicas. El modelo 3 introduce

⁽⁴⁾ Introducir la nota de acceso en primer lugar permite examinar cuánta parte del rendimiento académico del estudiante es explicado por el rendimiento previo a la universidad (incluida la EVAU) y, por consiguiente, qué otros factores tienen influencia más allá del rendimiento previo y de la prueba de acceso.

las variables demográficas y, por último, el modelo 4 incluye las variables socioeconómicas.

El segundo análisis se realizó con la técnica de minería de datos de árboles de decisión (Berlanga, Rubio y Vilà, 2013), con el propósito de seleccionar las variables independientes para un modelo de regresión logística mejor especificado, así como más parsimonioso y robusto. De todos los procedimientos disponibles se ha utilizado el algoritmo CART porque es el único que ordena a los predictores según su importancia normalizada. Se ha seguido en este sentido la metodología empleada por Arroyo, Constante-Amores y Asencio (2019). Se han interpretado los coeficientes de regresión en términos de probabilidades, y se han calculado los pronósticos del modelo utilizando una escala probabilística, dada la dificultad de interpretar el logit (Pardo y Ruiz, 2013).

En el tercer análisis se utilizó, de nuevo, la técnica de árboles de clasificación, con el objetivo de identificar los perfiles de los sujetos que obtuvieron un buen y un mal rendimiento académico. Esta técnica de minería de datos se llevó a cabo a través del procedimiento CHAID, ya que, según Berlanga, et ál. (2013), es el algoritmo más adecuado a la hora de realizar perfiles. Los análisis han sido realizados a través del paquete estadístico SPSS versión 25 y STATA versión 15.

Resultados

Variables individuales asociadas al rendimiento académico

La Tabla IV muestra los resultados del primer análisis de regresión logística. En el modelo 1 se aprecia que la nota de acceso se encuentra relacionada positiva y significativamente con el rendimiento académico. El modelo 2 incluye el resto de variables académicas y todas resultaron significativas; haber cursado la secundaria en un centro público, accedido a la universidad en la convocatoria de junio y elegir los estudios cursados en primera opción aumentan la probabilidad de ser buen estudiante. En el modelo 3, en el que se añaden las variables demográficas, no resultaron significativas ni la nacionalidad del estudiante ni la edad. Por su parte, tanto ser mujer como ser de Madrid aumenta la probabilidad de obtener

un buen desempeño. Al introducir las variables socioeconómicas⁵ en el modelo 4 no sólo no resultan significativas, sino que, además, dejan de serlo tanto la convocatoria de admisión como el orden de elección de los estudios⁶.

TABLA IV. Modelos de regresión logit para explicar el rendimiento académico universitario (Odds Ratio)

	Modelo 1 (nota de acceso)	Modelo 2 (variables académicas)	Modelo 3 (variables demográficas)	Modelo 4 (variables socioeconómicas)
Nota de acceso	1,353***	1,337***	1,352***	1,327***
Centro público		1,531***	1,627***	1,366**
Junio		1,252*	1,288**	1,174
Primera opción		1,166***	1,216***	1,151
Nacionalidad			1,037	0,962
Mujer			1,207***	1,369**
Edad idónea			1,031	1,049
Madrid			1,484***	1,359**
No trabaja				1,104
Estudios superiores de la madre				1,016
Constante (ES)	-2,716***	-3,130***	-3,767***	-3,329***
Pseudo R²	0,061	0,064	0,070	0,059

Fuente: Elaboración propia a partir de la base de datos del Observatorio del Estudiante de la UCM.

Notas: * p < 0,1 ** p < 0,05; *** p < 0,01

⁽⁵⁾ Se ha excluido la variable de estudios alcanzados por el padre al estar muy relacionada con la de estudios de la madre. Se ha optado por conservar esta última debido a que existe una amplia literatura (Garbanzo, 2007; Planck, 2014; Tejedor y García-Valcárcel, 2007) que muestra la importancia que tiene el nivel educativo de la madre en el desempeño universitario.

⁽⁶⁾ Se han realizado estos modelos con la técnica de regresión lineal múltiple (variable dependiente = rendimiento académico relativo -la diferencia entre el rendimiento del estudiante y la media del rendimiento del área de conocimiento-) obteniendo las mismas cuatro variables significativas y con signo positivo en el modelo completo.

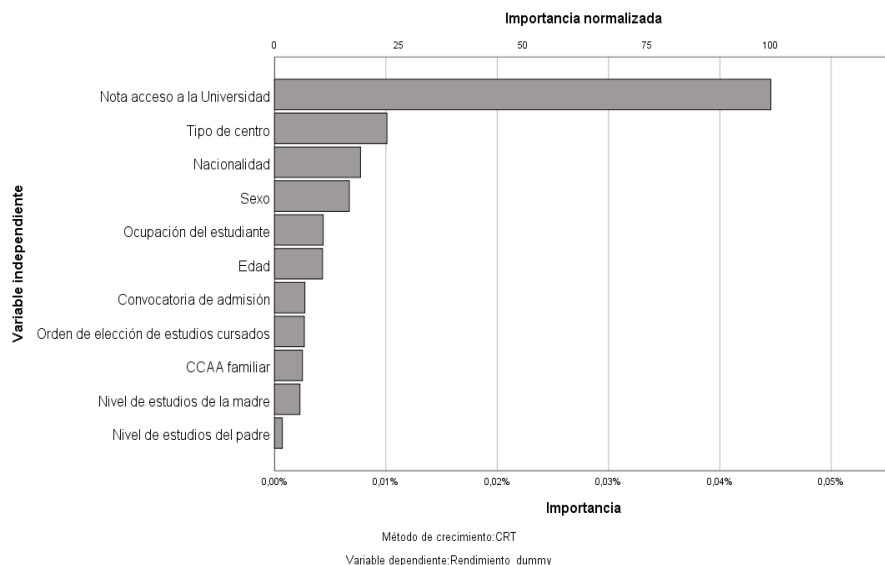
La nota de acceso explica, por si misma, más del 6% de la varianza del rendimiento académico del estudiante (Pseudo R^2), tal y como se observa en el modelo 1. Las variables demográficas añaden un 1% a la explicación (modelo 3). Ni el resto de las variables académicas (modelo 2) ni las variables socioeconómicas (modelo 4) mejoran la explicación (de hecho, en el último modelo la Pseudo R^2 disminuye, empeorando el ajuste del modelo).

A continuación, se utilizó la técnica de árboles de clasificación a través del procedimiento CART con el propósito de reducir el error de especificación y ordenar a los predictores según su importancia normalizada (Gráfico I). Se puede llevar a cabo esta técnica de segmentación, puesto que la base de datos utilizada está conformada por más de 1.000 sujetos (Berlanga et ál., 2013). La variable más relacionada con el rendimiento académico es la nota de acceso a la Universidad, y las que menos el nivel estudios de ambos padres. El tipo de centro al que ha asistido el estudiante en secundaria, también de naturaleza académica, ocupa el segundo lugar en la importancia normalizada. La ocupación del estudiante, así como todas las variables del Gráfico I que están por detrás, tienen un peso menor al 15% de la variable criterio, por lo que se podría hablar de micro-impactos a la hora de influir en el rendimiento académico (Tuero, Cervero, Esteban y Bernardo, 2018).

Antes de realizar el modelo final de regresión se correlacionaron las once variables independientes para eliminar aquellas con una alta correlación. Se observó que los estudios alcanzados por la madre tienen una alta relación con los del padre (lo que se conoce como homogamia educativa), por lo que se decidió eliminar este último predictor, ya que es el que tiene un menor impacto en el rendimiento académico, según el Gráfico I. En cuanto a los supuestos, el modelo de regresión logística cumple tanto el supuesto de no colinealidad, ya que el valor de tolerancia tiene valores por encima de 0,10 y los valores FIV no son mayores de 10, como el de independencia de errores (Durbin-Watson = 1,7). También se comprobaron los casos atípicos, y no se eliminó ninguno ya que la distancia de Cook era inferior a 1. En el modelo final se eliminaron aquellos predictores que no resultaron estadísticamente significativos ($p < 0,05$).

El modelo final está constituido por un total de cuatro variables independientes (véase la Tabla V). Todos los predictores están relacionadas positivamente con el rendimiento académico y explican casi el 7% de la varianza de esta variable.

GRÁFICO I. Importancia normalizada de las variables independientes que explican el rendimiento académico a través del algoritmo CART



Fuente: Elaboración propia a partir de la base de datos del Observatorio del Estudiante de la UCM.

TABLA V. Modelo final de regresión logística para predecir el rendimiento académico

	B (ES)	Exp (B)
Nota de acceso a la Universidad	0,313 (0,13)	1,367***
Centro de público	0,493 (0,51)	1,637***
Mujer	0,190 (0,049)	1,209***
Madrid	0,366 (0,054)	1,442***
Constante	-3,448 (0,147)	0,032***
Pseudo R ²	0,068	

Fuente: Elaboración propia a partir de la base de datos del Observatorio del Estudiante de la UCM.

Nota: ** p < 0,05; *** p < 0,01

En términos de probabilidad, por cada punto más en la nota de acceso, el estudiante tiene un 58% más de probabilidad de obtener un buen rendimiento académico; la probabilidad de obtener un buen rendimiento académico de los que han asistido a un centro público de educación secundaria es un 63% superior a los que han asistido a un centro privado; las mujeres tienen un 20% más de probabilidad de obtener un buen rendimiento académico que los hombres; los estudiantes cuya CCAA familiar se sitúa en Madrid tienen una probabilidad 44% mayor de obtener un buen rendimiento académico que los estudiantes de otras CCAA.

En cuanto a los pronósticos, la probabilidad más alta de obtener un buen rendimiento académico (88%) corresponde a una mujer (sexo = 1) que ha obtenido una nota de acceso igual a 13,96, que ha estudiado en un centro público (centro = 1) y que su CCAA familiar se situó en Madrid (CCAA familiar = 1). Por su parte, la probabilidad más baja de obtener un buen desempeño en la universidad (13%) corresponde a un hombre (sexo = 0), que tiene un 5 en la nota de acceso, que ha estudiado en un centro privado en la educación secundaria (centro = 0) y que su CCAA familiar no se encuentre en Madrid (CCAA familiar = 0).

Perfil de un estudiante con buen y con mal rendimiento académico

Según el procedimiento CHAID, la nota de acceso (nodo 0) es el mejor predictor de la variable dependiente (véanse los Gráficos II y III). La probabilidad más alta de obtener un mal rendimiento académico (70,7%) se da en estudiantes que han obtenido una nota de acceso a la Universidad (sobre 14) menor o igual a 7,64 y que han estudiado en un centro privado en la educación secundaria (nodo 7) (Gráfico II). La probabilidad más alta de obtener un buen rendimiento académico (82,1%) se da en estudiantes que tienen una nota de acceso mayor de 12,29 y que su CCAA familiar se encuentre en Madrid (nodo 19) (Gráfico III). Como este perfil es muy restrictivo, se procede a indicar la segunda y tercera probabilidad más alta de obtener un buen rendimiento académico. La segunda probabilidad más alta (76,9%) se da en estudiantes que han obtenido una nota de acceso entre 10,76 a 12,29, que han estudiado en un centro público en la educación secundaria y que su comunidad familiar sea Madrid (nodo 33). Por último, aquellos estudiantes que obtienen más de un 12,29, cuya residencia familiar no se encuentre en Madrid y que han estudiado en

un centro público de educación secundaria (nodo 35) tienen la tercera probabilidad más alta de obtener un buen rendimiento (75,3%).

GRÁFICO II. Perfil del estudiante que obtienen un mal rendimiento académico

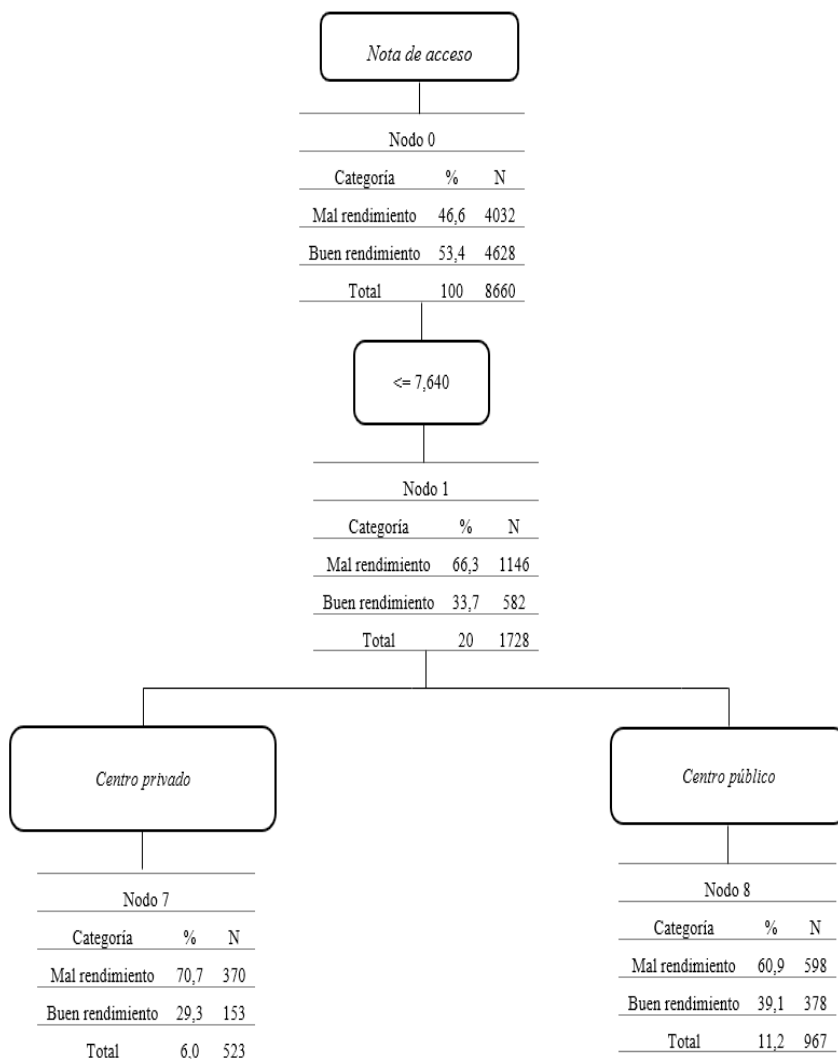
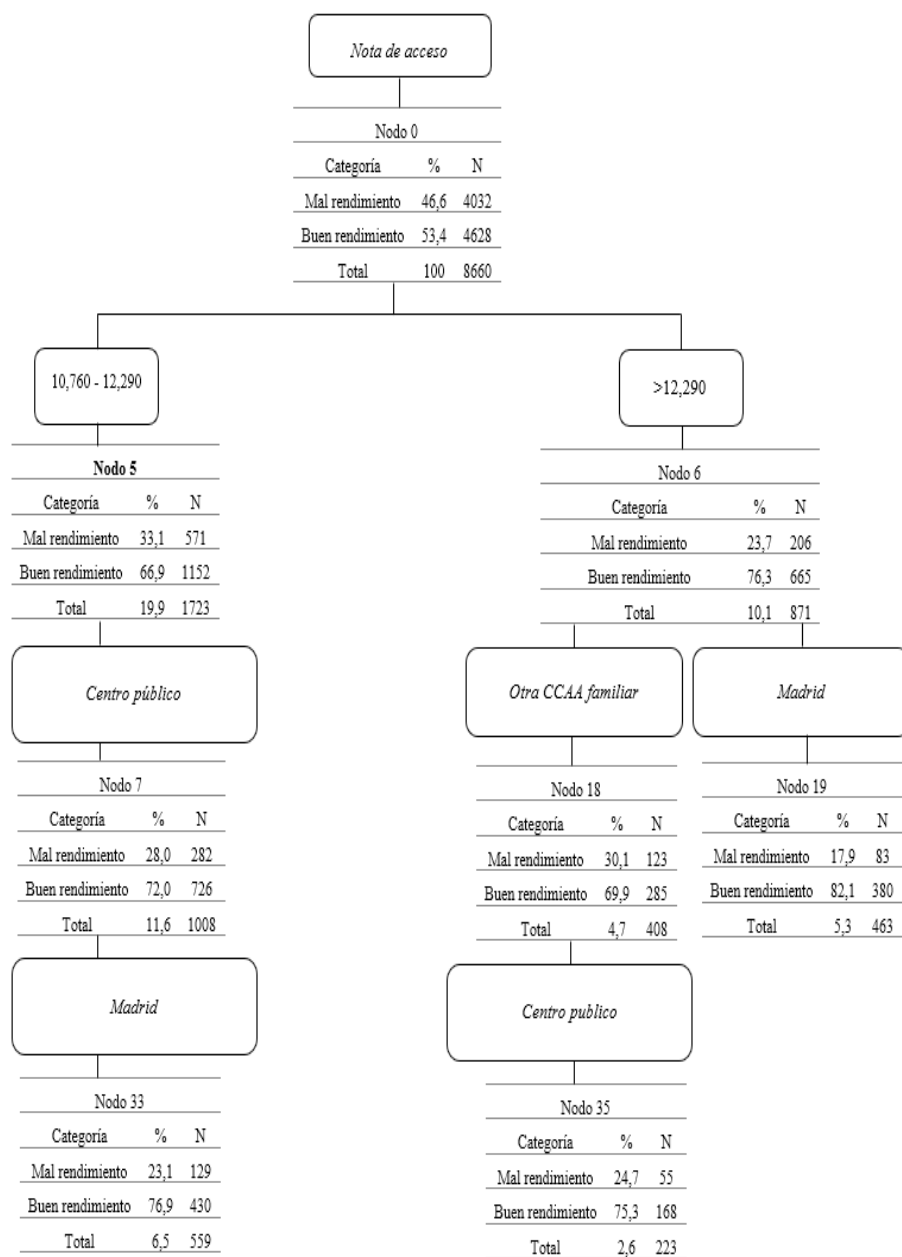


GRÁFICO III. Perfil del estudiante que obtienen un buen rendimiento académico



Discusión y conclusiones

El objetivo principal de este artículo ha sido investigar los factores asociados al rendimiento académico de los estudiantes de nuevo acceso. Se han analizado los datos de la Universidad Complutense de Madrid para esclarecer cuáles son estos determinantes y qué importancia tiene cada uno de ellos. El trabajo se ha articulado en torno a tres preguntas a partir de las cuales, y con la ayuda de la revisión de la literatura, se han formulado sendas hipótesis. Respecto a los factores individuales asociados al rendimiento académico, los resultados del análisis muestran que la nota de acceso a la Universidad tiene un impacto positivo en el rendimiento del alumnado (confirmando, en parte, la primera hipótesis), pero que, sin embargo, el origen socioeconómico del estudiante no incide en el mismo (rechazando, en parte, H1). Además de la nota de acceso, otras variables como el tipo de centro cursado en la educación secundaria (público), el sexo (mujer) y la CCAA familiar (Madrid) también tienen incidencia en el desempeño. Además de algunas variables académicas, por tanto, hay variables de naturaleza demográfica asociadas al rendimiento académico.

La nota de acceso es el predictor más importante del rendimiento académico del estudiante (confirmando, en parte, la segunda hipótesis), tal y como se ha mostrado de forma muy robusta con varios métodos de estimación (regresión logística y árboles de decisión –CART y CHAID). El segundo predictor más importante según el algoritmo CART es también de naturaleza académica, la titularidad del centro al que se ha asistido en secundaria, si bien otras variables de naturaleza académica quedan en posiciones menos relevantes. Además, atendiendo al Pseudo R^2 , la mejora en el ajuste por parte de las variables académica se debe exclusivamente a la nota de acceso. En cambio, las variables demográficas, aunque poco, contribuyen a un mejor ajuste del modelo. Por tanto, si bien H2 se cumple por el efecto de la nota de acceso, hay más dudas sobre que otras variables de naturaleza académica sean tan cruciales.

La nota de acceso es, además, el predictor que mejor discrimina entre ser un buen y mal estudiante en función del desempeño académico. Junto con ella, tanto la titularidad del centro, de naturaleza académica, como la CCAA familiar, de carácter demográfico, son factores muy importantes en el establecimiento de perfiles académicos. La tercera hipótesis queda confirmada, si bien hay que tener en cuenta la relevancia de alguna variable de naturaleza demográfica.

La investigación realizada con los estudiantes de nuevo acceso a grado de la UCM ha revelado varios hallazgos, que abren puertas para la investigación futura. En primer lugar, que los factores individuales juegan un papel en la explicación del rendimiento académico, pero queda una parte importante de este fenómeno que no se explica por las variables consideradas en la investigación (un 93% de la varianza). Habría que expandir el análisis para incluir otro tipo de factores, como los organizacionales.

En segundo lugar, el análisis ha mostrado la relevancia que tiene la nota de acceso a la universidad en el rendimiento académico posterior; a mayor nota de acceso con la que el estudiante accede a la universidad, mejor desempeño académico. Además de este predictor, también resulta importante la titularidad del centro en el que ha cursado la educación secundaria, de naturaleza también académica. El mejor desempeño de los estudiantes que han asistido a un centro público puede deberse a tres posibles explicaciones, no contrastadas todavía: la calidad académica de los centros públicos podría ser mejor; las notas de los centros públicos podrían reflejar con mayor nivel de precisión que las de los centros privados el nivel de desempeño de los estudiantes; por último, la forma de enseñanza en los centros públicos podría ser más similar a la de la universidad pública que en el caso de los centros privados, por lo que los estudiantes provenientes de centros públicos no se expondrían a un cambio tan abrupto como en el caso de los que provienen de los centros privados. Otras variables de naturaleza académica han resultado irrelevantes para el rendimiento: los estudiantes que aprueban la selectividad en septiembre, o que estudian algo que no eligieron en primer lugar, tienen las mismas posibilidades de obtener un buen rendimiento académico.

En tercer lugar, de las variables demográficas y socioeconómicas sólo el sexo del estudiante y la CCAA familiar tienen un impacto en el rendimiento. Si el estudiante es mujer y su familia vive en Madrid, la probabilidad de tener un buen rendimiento académico mejora. Una posible línea de investigación futura podría ser explorar más a fondo el efecto de la CCAA familiar. Los estudiantes provenientes de Madrid en su conjunto tienen la peor nota media de acceso a la Universidad, y, sin embargo, ser de Madrid incide positivamente en el rendimiento una vez dentro de la educación terciaria, algo que podría estar indicando que la prueba de acceso a la universidad (la EVAU) es más difícil en Madrid

que en el resto de CCAA. Una posible explicación de que la nacionalidad del estudiante o el nivel socioeconómico resulten irrelevantes para el rendimiento es que, en el nivel universitario, los estudiantes están ya muy seleccionados desde el punto de vista del origen social, y por tanto la fuerza estadística de estas variables es más reducida que en niveles educativos anteriores. Llama, sin embargo, más la atención que el trabajo laboral realizado por el estudiante no tenga repercusión en su rendimiento académico. Asimismo, tener la edad idónea para acceder a la universidad (tener entre 17 y 19 años) no está relacionado con el rendimiento académico.

En cuarto lugar, desde el punto de vista metodológico, se ha creado una variable dependiente, el rendimiento académico relativo, que permite analizar conjuntamente a los estudiantes de las distintas ramas de estudio, con notas medias dispares entre sí. En un trabajo futuro se puede avanzar en la construcción de una variable que permita controlar la variabilidad de las notas medias de las distintas titulaciones, añadiendo precisión al análisis. Asimismo, los resultados presentados no aportan un modelo explicativo, por lo que una limitación de este trabajo es que no se pueden identificar relaciones de causalidad entre las variables analizadas. Una posible línea de investigación que se abre a partir de este estudio es el análisis de las relaciones de causalidad entre los factores (individuales y organizacionales) y el rendimiento académico a través de modelos de ecuaciones estructurales. Del mismo modo, adicionalmente se podría profundizar en el perfil de los estudiantes universitarios por áreas de conocimiento a través de la técnica estadística de segmentación.

Investigaciones como esta pueden arrojar luz sobre los factores que determinan el rendimiento académico universitario y, a la postre, servir para el diseño de políticas universitarias que mejoren la calidad, y la equidad, del sistema universitario.

Referencias

Abera, G. (2018). Family Socio-Economic Status Effect on Students' Academic Achievement at College of Education and Behavioral

- Sciences, Haramaya University, Eastern Ethiopia. *Journal of Teacher Education and Educators*, 7(3), 207-222.
- Ariño, A., & Sintés, E. (2016). *Via Universitària: Ser estudiant universitari avui*. Barcelona: Fundació Jaume Bofill.
- Arroyo, D., Constante-Amores, I. A., & Asensio, I. (2019). La repetición de curso a debate: un estudio empírico a partir de PISA 2015. *Educación XX1*, 22(2), 69-92, doi: 10.5944/educXX1.22479
- Berlanga, V., Rubio, M. J., & Vilà, R. (2013). Cómo aplicar árboles de decisión en SPSS. *Revista d'Innovació i Recerca en Educació*, 6(1), 65-79. DOI:10.1344/reire2013.6.1615
- Bunce, L., Baird, A., & Jones, S. E. (2017). The student-as-consumer approach in higher education and its effects on academic performance. *Studies in Higher Education*, 42(11), 1958-1978.
- Cerdeira, J. M., Catela, L. C., Balcão A., & Seabra, C. (2018). Predictors of student success in Higher Education: Secondary school internal scores versus national exams. *Higher Education Quarterly*, 72(4), 304-313.
- Chong, E. (2017). Factores que inciden en el rendimiento académico de los estudiantes de la Universidad Politécnica del Valle de Toluca. *Revista Latinoamericana de Estudios Educativos*, 47(1), 91-108.
- Clerici, R., & Da Re, L. (2019). Evaluación de la eficacia de un programa de tutoría formativa. *Revista de Investigación Educativa*, 37(1), 39-56.
- Cortez, F. M., Tutiven, J. L., & Villavicencio, M. N. (2017). Determinantes del Rendimiento Académico Universitario. *Revista Publicando*, 4(10), 284-296.
- Danilowicz-Gösele, K., Lerche, K., Meya, J., & Schwager, R. (2017). Determinants of students' success at university. *Education Economics*, 25(5), 513-532.
- De Sousa, L., Taboada, E., Rivas, R., Iglesias, P., & López, S. (2017). Rendimiento académico y variables sociodemográficas y académicas en la Educación Superior en Angola. *Revista de Estudios e Investigación en Psicología y Educación, Extr.*(14), 112-116.
- Dennis, J. M., Phinney, J. S., & Chuateco, L. I. (2005). The role of motivation, parental support, and peer support in the academic success of ethnic minority first-generation college students. *Journal of College Student Development*, 46, 223-236.
- Duart, J. M., Gil, M., Puyol, M., & Castaño, J. (2008). *La universidad en la sociedad Red*. Barcelona: Ariel.

- EUROSTUDENT. (2018). *Social and economic conditions of student life in Europe. Eurostudent VI 2016-2018. Synopsis of Indicators*. Bielefeld: German Centre for Higher Education Research and Science Studies (DZHW).
- Fundación BBVA. (2010). *Estudio internacional Fundación BBVA sobre estudiantes universitarios de seis países europeos*. Madrid: Dpto. de Estudios Sociales y Opinión Pública.
- Garbanzo, G. M. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Revista de Educación*, 31(1), 43-63.
- Garbanzo, G. M. (2013). Factores asociados al rendimiento académico en estudiantes universitarios desde el nivel socioeconómico: Un estudio en la Universidad de Costa Rica. *Revista Electrónica Educare*, 17(3), 57-87.
- Garbanzo, G. (2014). Factores asociados al rendimiento académico tomando en cuenta el nivel socioeconómico: Estudio de regresión múltiple en estudiantes universitarios. *Revista Electrónica Educare*, 18(1), 119-154.
- García, A. M. (2014). Rendimiento académico y abandono universitario: modelos, resultados y alcances de la producción académica en la Argentina. *Revista Argentina de Educación Superior*, (8), 9-38.
- Jansen, E. P. (2004). The influence of the curriculum organization on study progress in higher education. *Higher Education*, 47(4), 411-435.
- Jiménez-Caballero, J. L., Camúñez, J. A., González-Rodríguez, M. R., & Fuentes, P. (2015). Factores determinantes del rendimiento académico universitario en el espacio europeo de educación superior. *Revista De Ciencias Administrativas y Sociales*, 25(58), 159-176.
- Lamas, H. A. (2015). Sobre el rendimiento escolar. *Propósitos y representaciones*, 3(1), 313-386.
- Manzano, J. A., Escribá, A., Iborra, M., & Safón, V. (2016). *La Universidad Española. Grupos Estratégicos y Desempeño*. Bilbao: Biblioteca Nueva
- Mañé, F., & Miravet, A. (2010). *Rendimiento académico de los estudiantes de primer año en la universidad*. Barcelona: AQU Catalunya. Recuperado de http://www.aqu.cat/doc/doc_81940174_1.pdf
- Marcenaro, O. D., & Navarro, M. L. (2007). El éxito en la universidad: una aproximación cuantílica. *Revista de Economía Aplicada*, 44(15), 5-39.
- McNabb, R., Pal, S., & Sloane, P. (2002). Gender differences in educational attainment: the case of university students in England y Wales. *Economica*, 69(275), 481-503.

- Nyikahadzoi, L., Matamande, W., Taderera, E., & Mandimika, E. (2013). Determinants of students' academic performance in four selected accounting courses at University of Zimbabwe. *Research in Higher Education Journal*, 21, 1-9.
- Pardo, M., & Ruiz, M. (2013). *Análisis de Datos en Ciencias Sociales y de la Salud III*. Madrid: Síntesis.
- Pérez-Cárceles, M. C., Gómez-Gallego, M., Gómez-Gallego, J. C., Palazón-Pérez de Los Cobos, A., & Gómez-García, J. (2014). El género como variable moderadora de los Resultados Académicos en la Enseñanza Universitaria. *Regional and Sectoral Economic Studies*, 14(3), 55-64.
- Planck, U. (2014). Factores determinantes del rendimiento académico de los estudiantes de la universidad de atacama. *Estudios Pedagógicos*, 40(1), 25-39.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353-387.
- Rodríguez, G., Ariza, M., & Ramos, J. L. (2014). Calidad institucional y rendimiento académico: El caso de las universidades del Caribe colombiano. *Perfiles educativos*, 36(143), 10-29.
- Rodríguez, S., Fita, E., & Torrado, M. (2004). El rendimiento académico en la transición secundaria-universidad. *Revista de Educación*, 334(1), 391-414.
- Ruesga, S. M., Silva, J., & Monsueto (2014). Estudiantes universitarios, experiencia laboral y desempeño académico en España. *Revista de Educación*, 365, 67-95.
- Salinas, D. A., Hernández, A. E. y Barboza-Palomino, M. (2017). Condición de becario y rendimiento académico en estudiantes de una universidad peruana. *Revista Electrónica de Investigación Educativa*, 19(4), 124-133.
- Santos, M. V., & Vallelado, E. (2013). Algunas dimensiones relacionadas con el rendimiento académico de estudiantes de Administración y Dirección de Empresas. *Universitas Psychologica*, 12(3), 739-752.
- Simón, H., Casado, J., & Castejón, J. (2017). Analysis of university student employment and its impact on academic performance. *Electronic Journal of Research in Educational Psychology*, 15(42), 281-306.
- Smith, J., & Naylor, R. (2001). Determinants of Degree Performance in UK Universities: A Statistical Analysis of the 1993 Student Cohort. *Oxford Bulletin of Economics and Statistics*, 63(1), 29-60.

- Tejedor, F. J., & García-Valcárcel, A. (2007). Causas del bajo rendimiento del estudiante universitario (en opinión de los profesores y alumnos). Propuestas de mejora en el marco del EEES. *Revista de Educación*, 342(1), 443-473.
- Thiele, T., Pope, D., Singleton, A., Snape, D., & Stanistreet, D. (2017). Experience of disadvantage: The influence of identity on engagement in working class students' educational trajectories to an elite university. *British Educational Research Journal*, 43(1), 49-67.
- Tuero, E., Cervero, A., Esteban, M., & Bernardo, A. (2018). ¿Por qué abandonan los alumnos universitarios? Variables de influencia en el planteamiento y consolidación del abandono. *Educación XXI*, 21(2), 131-154.
- Turrul, M., & Roca, B. (2010). *El rendimiento académico del alumnado de primer curso de Derecho en la Universidad de Barcelona*. Barcelona: AQU Catalunya. Recuperado de http://www.aqu.cat/doc/doc_61647381_1.pdf
- Vélez, A., & Roa, C. (2005). Factores asociados al rendimiento académico en estudiantes de Medicina. *Educación Médica*, 8(2), 24-32.
- Vinacur, T. (2016). ¿Los alumnos de las escuelas privadas están mejor preparados para ingresar a la universidad? *Revista Colombiana de Educación*, (70), 175-200.

Anexo

TABLA I. Descriptivos de las variables independientes cualitativas

Variable	Niveles de la variable independiente	N	%
Sexo	Hombre	4.086	38,1
	Mujer	6.634	61,9
Edad	No idónea: 20 o más	2.743	25,6
	Idónea: 19 años o menos	7.977	74,4
Nacionalidad	Otra	890	8,3
	Española	9.830	91,7
CCAA familiar	Otra CCAA familiar	3.171	29,6
	Madrid	7.549	70,4
Ocupación del estudiante	Trabaja	716	33,4
	No trabaja	1.430	66,6
Estudios alcanzados por el padre	Estudios secundarios o menos	5.055	51,5
	Estudios superiores	4.757	48,5
Estudios alcanzados por la madre	Estudios secundarios o menos	4.885	48,4
	Estudios superiores	5.216	51,6
Titularidad del centro en educación secundaria	Privado	2.966	36,4
	Público	5.193	63,6
Convocatoria de admisión	Septiembre	549	5,1
	Junio	10.171	94,9
Orden de elección de estudios	Primera opción	6.568	61,3
	Segunda a duodécima opción	4.152	38,7

Fuente: Elaboración propia a partir de la base de datos del Observatorio del Estudiante de la UCM.

TABLA II. Descriptivos de la variable independiente cuantitativa

Variable	N	Media	Desviación típica	Mínimo	Máximo
Nota de acceso	10.720	9,245	2,053	5,00	13,960

Fuente: Elaboración propia a partir de la base de datos del Observatorio del Estudiante de la UCM.

Dirección de contacto: María Fernández-Mellizo. Universidad Complutense de Madrid, Facultad de Educación, Sociología Aplicada, Instituto TRANSOC. Edificio La Almudena. C/Rector Royo Villanova s/n, Ciudad Universitaria. 28040, Madrid. E-mail: mfmellizosoto@edu.ucm.es

Factors associated to the academic performance of new entry students at Complutense University of Madrid

Determinantes del rendimiento académico de los estudiantes de nuevo acceso a la Universidad Complutense de Madrid

DOI: 10.4438/1988-592X-RE-2020-387-433

María Fernández-Mellizo
Alexander Constante-Amores
Universidad Complutense de Madrid

Abstract

The knowledge of the factors that influence the academic performance of students in higher education is getting attention currently, and it is very useful when evaluating the quality and fairness of universities. The aim of this article is the analysis of individual factors (demographic, socioeconomic and academic) associated to the academic performance of a new cohort of students from the Complutense University of Madrid in the 2017/18 academic year, made up of a total of 10,720 students. Freshmen students are those who are the most likely to drop-out their selected degrees. Data were analysed through binary logistic regression techniques and classification trees. Eleven variables were selected as independent ones. From those, only four have a significant influence in the final model: qualification obtained to access to the university studies, the nature of the institution of origin, gender and the family Autonomous Community. The results show that the higher the qualification obtained to access to the university, the more likely the student is to obtain a good academic performance. In the same way, if a student is either women or from Madrid or attended previously to a public institution, the chances of achieving better academic performance will increase. In conclusion, academic variables carry more weight to explain the university performance, although the demographic ones are also important. Among the academic variables, the qualification obtained to access to the university studies

is the most relevant. Therefore, socioeconomic variables do not contribute to the explanation, possible due to the already high level of social selection of students at university, which lessens their statistical impact in comparison to previous levels of education.

Key words: higher education, university students, academic achievement, logistic regression, data mining.

Resumen

En los últimos años existe un gran interés por conocer los factores o determinantes que inciden en el rendimiento académico del alumnado de la educación superior, de gran utilidad a la hora de evaluar la calidad y la equidad de las universidades. Esta investigación tiene como objetivo analizar los factores individuales (demográficos, socioeconómicos y académicos) asociados al rendimiento académico de los nuevos universitarios, que son aquellos con más posibilidades de abandonar sus estudios. Para ello, se ha analizado la cohorte de estudiantes de nuevo acceso a grados de la Universidad Complutense de Madrid en el curso 2017/18, conformada por un total de 10.720 estudiantes. Los datos se han analizado mediante las técnicas de regresión logística binaria y de árboles de clasificación. Como variables independientes se han seleccionado un total de once, de las cuales cuatro (*nota de acceso a la Universidad, tipo de centro en la educación secundaria, sexo y Comunidad Autónoma familiar*) han resultado estadísticamente significativas en el modelo final. Los resultados muestran que cuanto más elevada es la nota de acceso a la Universidad, más probabilidades tiene el estudiante de obtener un buen rendimiento académico. Del mismo modo, si el alumno ha estudiado Bachillerato en un instituto público, es mujer y/o de Madrid, las probabilidades de obtener un buen rendimiento se incrementan. En general, las variables académicas, y en particular la nota de acceso a la Universidad, son las que más contribuyen a explicar el desempeño universitario, aunque las variables demográficas tienen cierto peso también. En cambio, las variables socioeconómicas no contribuyen sustantivamente a la explicación de este fenómeno, algo que puede reflejar que los estudiantes, en el nivel superior, están muy seleccionados desde el punto de vista social y, por tanto, la fuerza estadística de estas variables es más reducida que en niveles educativos anteriores.

Palabras clave: educación superior, estudiantes universitarios, rendimiento académico, regresión logística, minería de datos.

Introducción

University students are becoming more and more interested in socio-educational research. Numerous investigations, both national and international, have been devoted to the examination of their profiles, strategies and opinions, in the same way that the characteristics of those who access tertiary education have been treated with special attention (Ariño and Sintés, 2016; EUROSTUDENT, 2018; BBVA Foundation, 2010).

Moreover, the analysis of the academic performance of higher education students, as well as of the factors that affect it, has gained considerable notoriety in recent years (García, 2014). According to various authors, the relevance of academic performance lies in the fact that it is one of the most important indicators to determine the quality and equity of university education (Rodríguez, Fita and Torrado, 2004). According to other authors, it is also important due to the fact that a high or low performance by students affects both the student and the academic institution (Garbanzo, 2014).

Understanding the factors that explain academic performance is useful when assessing the quality and equity of universities, as is the identification of the variables that determine academic success or failure (Garbanzo, 2007), and ultimately increase student academic performance (Lamas, 2015). Some authors even claim that focusing on analysing the factors associated with academic performance can serve as a preliminary study for the design of university policies (Turrul and Roca, 2010).

Research that analyses the factors associated with academic performance in higher education does not contain an in-depth study of individual factors (Rodríguez, Ariza and Ramos, 2014). This article arises from the need to conduct an in-depth study of the impact of individual variables (demographic, socio-economic and academic) on the academic performance of students. Furthermore, the study of student profiles is relevant and necessary for analysing the determinants associated with academic performance and has not received the desired attention (Jiménez-Caballero, Camúñez, González-Rodríguez and Fuentes, 2015). This study tries to address this. It is crucial to carry out such research on new students or first courses, since they are the ones that have the best chance of dropping out of their university studies (García, 2014; Tejedor and García-Valcárcel, 2007) and/or obtaining low performance levels (Turrul and Roca, 2010).

Due to the above, several research questions must be asked: What are the individual factors (demographic, socio-economic and academic) that affect academic performance? Which of them are more important in this phenomenon? In relation to the above, what is the profile of the student with high and low academic performance? The objective of this study is, therefore, to examine the individual factors associated with students' performance, as well as to determine the importance of each of them and the profiles of the students based on this performance.

To answer these research questions, we have analysed the case of the Complutense University of Madrid (UCM), the largest Spanish university in terms of number of students (Manzano, Escribá, Iborra and Safón, 2016). A database of new degree students in the 2017/18 academic year has been provided by the Student Observatory of the UCM.

The first section contains a review of the literature on the factors that affect academic performance in the university, especially those that have to do with individual variables. The second section focuses on methodological issues; the data, variables and data analysis techniques used are described. The third section is dedicated to the results of the research, both those that have to do with individual factors that affect students' academic performance and with the establishment of profiles associated with a good and a bad student. Finally, the fourth section presents a discussion of the results and conclusions of the research.

Factors associated with the academic performance of university students

There is no consensus in the scientific community when defining academic performance (Lamas, 2015). It should be noted that grades and credits are the most used indicators (Mañé and Miravet, 2010), with grades being the most effective according to some authors (Garbanzo, 2013).

Numerous classifications of the factors associated with academic performance have been established. In the research carried out by Garbanzo (2007), three factors can be found: personal, social and institutional. Additionally, Tejedor and García-Valcárcel (2007) indicate the existence of five determinants: identification, psychological, academic, pedagogical and socio-familial. Similarly, the study conducted by García (2014) classifies the factors that influence academic performance as individual and organizational.

In relation to organizational factors, the literature consulted finds that regular attendance in both class and tutorials by students (Clerici and Da Re, 2019), or a good relationship between teacher and student (Garbanzo, 2007), are positively related to academic performance. We decided to focus this research on individual factors of academic performance, which have been less studied than organizational ones.

The individual factors are divided by García (2014) into three groups: demographic, socio-economic and academic. This research uses this classification, since it is characterized by its clarity, precise terminology and timeliness.

Demographic factors

Among the demographic factors, gender has been the most studied, followed by age. The meta-analysis conducted by Richardson, Abraham and Bond (2012), whose objective was to determine the psychological, demographic and socio-economic factors associated with academic performance in tertiary education, includes a total of 7,167 studies between 1997 and 2010, but reaches the conclusion that gender and age have little impact.

As for the former, research carried out is non-conclusive. Some studies indicate that it is women who demonstrate better academic performance. The research carried out by Pérez-Cárceles, Gómez-Gallego, Gómez-Gallego, Palazón-Pérez and Gómez-García (2014) shows that students from the Catholic University of San Antonio (Murcia) have better academic performance. Similarly, the study conducted by Duart, Gil, Puyol and Castaño (2008) refers to higher performance by women in various universities of the Autonomous Community of Catalonia. McNabb, Pal and Sloane (2002) also find that female university students studying in England and Wales have better performance. However, other studies show that men have better academic performance. The research carried out by Santos and Vallelado (2013) reveals that men on the Business Administration and Management Degree course at the University of Valladolid have better academic performance. Along the same lines, Nyikahadzoi, Matamnade, Taderera and Mandimika (2013) point out that Accounting male students at the University of Zimbabwe get better grades. Furthermore, other studies, such as Danilowicz-Gösele, Lerche,

Meya and Schwager (2017) with students from Göttingen (Germany), indicate that gender is not very important when explaining academic performance.

Studies are also non-conclusive with regard to age. On the one hand, Jansen (2004) finds a negative relationship between performance and age in students from the University of Groningen (Netherlands). On the other hand, several studies find no relationship between age and academic performance. De Sousa, Taboada, Rivas, Iglesias and López (2017) found no statistically significant differences in the academic performance of university students in the Republic of Angola according to their age. Abera (2018), with students from the University of Haramaya (Ethiopia), and Bunce, Baird and Jones (2017), with English students, did not find any either.

Other less frequently studies demographic factors are nationality (Mañé and Miravet, 2010) and marital status. National students from the universities of Argentina (García, 2014) and single students from the Catholic University of Santiago de Guayaquil in Ecuador (Cortez, Tutiven and Villavicencio, 2017) obtain better performance than immigrant and married students, respectively.

Socio-economic factors

According to several studies, the socio-economic level of the family is relevant for university success (Abera, 2018; Chong, 2017; Dennis, Phinney and Chuateco, 2005), although not as much as in previous educational stages (Garbanzo, 2013). Most of the studies establish a positive relationship between the socio-economic origin of the student and their academic performance. For example, the study conducted by Cortez et al. (2017) with students from the Catholic University of Santiago de Guayaquil (Ecuador) reveals that the higher the qualification of the job and educational level of the parents, the more likely it is for students to obtain good grades. Similarly, the work carried out by Marcenaro and Navarro (2007) shows the influence of the father's university studies on the good academic performance of students from the University of Málaga. However, research conducted by Thiele, Pope, Singleton, Snape and Stanistreet (2017) reveals that a family's socio-economic status has a negative impact on the performance of UK students but only in the case of very low social classes.

There is little research that addresses the relationship between the job performed by the student and their academic performance (Simón, Casado and Castejón, 2017), and there is considerable disagreement between them. According to the research carried out by Ruesga, Silva and Monsueto (2014), new working students admitted to public and private universities in Spain obtain lower grades than those who do not work. On the contrary, Simon, et al. (2017) find that the job performed by students from the University of Alicante does not affect academic performance. Another variable of a socio-economic nature explored by some works is the source of funding for studies; scholarship students from the Universidad Privada del Norte del Perú obtain a higher average grade (Salinas, Hernández and Barboza-Palomino, 2017).

Academic factors

Many academic works (García, 2014; Jiménez-Caballero et al., 2015; Vélez and Roa, 2005) show that the university entry grade is not only positively related to the academic performance of university students, but is also the variable that best explains this. However, a study done with students from the University of Alicante shows how this predictor has an impact on the probability of finishing their higher education, but not on the average grade (Simón, et al, 2017).

The literature also finds a relationship between the order of choice of university studies and academic performance.

Students who have chosen studies in the first option obtain higher performance levels (Jiménez-Caballero et al., 2015; Rodríguez, et al. 2004). Another factor that has been studied is university entrance; according to the study by Mañé and Miravet (2010), Catalan students who come from baccalaureate studies show better academic performance than those entering through other paths.

The type of centre also seems to be associated with student performance, although there is no consensus on the part of the scientific community when it comes to indicating who performs better at university. On the one hand, Vinacur (2016) finds that freshmen from the University of Buenos Aires who have attended private centres throughout their secondary education obtain higher performance levels than students who went to a public school. On the other hand, Smith and Naylor (2001) find

that students from various universities in the United Kingdom who have attended public centres obtain higher levels of academic performance. Along the same lines, Cerdeira, Catela, Balcão and Seabra (2018) show that graduated students from Portugal who have previously attended private centres have a slight disadvantage.

Literature on individual factors associated with academic performance is therefore quite conclusive regarding the positive impact of the entry grade and the socio-economic level of the student's parents on their academic performance. The first scenario is formulated accordingly:

H1: The higher the student entry grade, the higher their academic performance; the higher the socio-economic level of the student's family, the higher their performance levels at the university.

Furthermore, there is also considerable consensus regarding the preponderant role of the student's entry grade when explaining their academic performance, and it would be logical to think that academic variables, in general, are more associated with it. Therefore, the second and third scenarios establish that:

H2: The university entry grade is the variable that has the most impact on the student's academic performance; in general, academic variables have a greater explanatory capacity than the rest of individual ones.

H3: Academic variables, and in particular the university entry grade, define the student's profile based on their academic performance.

As we have seen, literature is non-conclusive regarding the impact of variables such as gender or type of secondary school, and is scarce concerning the influence of other variables such as nationality or order of study choice, so no starting scenario is established in this regard.

Methodology

Data

The database of the UCM Student Observatory was used for this research. This information has been obtained from the pre-registration and registration process, with the help of the university's Institutional Intelligence Centre. Therefore, the analysed data includes the total population included in the census.

8,660 of the 10,720 newly graduated students who enrolled at the UCM in the 2017/18 academic year have been analysed (see Table I). This drastic reduction in numbers is due to the fact that only the grades of students entering through baccalaureate (9,468 students, 88.3% of the total) are available, of which 808 drop out of university studies before completing the first semester. The centres “attached” to the UCM have not been taken into account, given their different nature, or double degrees, since it complicated the classification in areas of knowledge.

TABLE I. Number of students by areas of knowledge and their corresponding average grade for the first semester

AREA	NO. OF CASES	AVERAGE GRADE OF THE FIRST QUARTER*
Social Sciences and Legal Studies	4,919	5.79 (3,828)
Sciences	1,211	6.23 (1,076)
Arts and Humanities	1,889	6.70 (1,500)
Engineering and Architecture	577	5.41 (469)
Health Sciences	2,124	6.52 (1,787)
Total	10,720	6.13 (8,660)

Source: Own elaboration from the database of the UCM Student Observatory.

Note: *In brackets, the number of cases that contain information on the average grade of the first semester.

Variables

The dependent variable of the study is academic performance, measured through the average grade obtained by students in the first semester. We have decided to dichotomize it (0 = low academic performance and 1 = high academic performance)¹. The first thing we did was to group together each degree with their area of knowledge, and the average grade corresponding to each area was calculated (Table I). The following equation was then used to dichotomize the dependent variable

⁽¹⁾ However, analyses have been performed with the continuous dependent variable, obtaining results similar to those shown here.

and distinguish between the students who obtained high and low grades in the first semester.

Relative performance = Average grade of the student - Average grade of the area equation (1)

More specifically, those students who obtained a value greater than or equal to zero were considered to have achieved high academic performance levels ($Y = 1$), since their average grade is equal to or greater than the average of their area of knowledge. However, values below zero denote a low academic performance ($Y = 0$), not reaching the average grade of the area of knowledge. This procedure allows us to analyse all students from different degree courses at the UCM, despite the disparity of average grades in different areas of knowledge. Table II shows the case distribution according to the dependent variable.

TABLE II. Number of students who obtain high and low academic performance levels

Level	N	%
Low academic performance level	4032	46.6
High academic performance level	4628	53.4
Total	8660	100.0

Source: Own elaboration from the database of the UCM Student Observatory.

A total of eleven independent variables have been chosen from the available database (Table III), based on a double standard². First, those discussed in the literature have been chosen. Then, those predictors (Family Autonomous Community and call for admission) that have not previously been included in studies on their relationship with performance, and which are expected to have an influence on the standard variable, have

⁽²⁾ The database used contains other available information that has not been utilised: family municipality, family province, family country, ECTS presented, ECTS enrolled, father's occupation, mother's occupation and student's employment status in the previous year.

been included in the model. Since EVAU (Evaluation for University Access) is different in each Autonomous Community (AC), one might think that students who come from Madrid may have a different performance than those from other communities. Furthermore, those students who enter in the first call in June could have a higher performance than those who enter in the second call in September. The variables have been dichotomized to facilitate the interpretation of the analysis (Pardo and Ruiz, 2013), although it has been verified that this model specification does not condition the results³ (see Annex, tables I and II, for descriptive variables).

TABLE III. Independent variables

Individual factors	Independent variables	Recoded values
Demographic	Gender	0 = Man 1 = Woman
	Age	0 = Unqualified (20 or over) 1 = Qualified (17-19)
	Nationality	0 = Other 1 = Spanish (local)
	Family Autonomous Community	0 = Other Autonomous Community 1 = Madrid
Socio-economic	Student occupation	0 = Working 1 = Not working
	Father's level of education Mother's level of education	0 = Less than higher education 1 = Higher education
Academic	Entry grade	5-14
	Type of secondary school	0 = Private 1 = Public
	Call for admission	0 = September 1 = June
	Order of study choice	0 = According to option twelve 1 = First option

Source: Compiled by the author

⁽³⁾ It has been proven that different specifications of the variables of studies reached by the mother and father, and order of choice of studies, do not alter the results.

Data analysis techniques

Three types of analyses have been performed. The first uses the binary logistic regression technique to analyse the relative contribution of each variable or set of variables (demographic, socio-economic and academic) regarding the probability of obtaining high academic performance levels. Four models were created for this purpose. Model 1 is made up of the university entry grade, which is the variable of greatest impact on academic performance, according to literature⁴, and model 2 also includes other academic variables. Model 3 introduces demographic variables and, finally, model 4 includes socio-economic variables.

The second analysis was carried out using the data mining technique of decision tree (Berlanga, Rubio and Vilà, 2013), with the purpose of selecting the independent variables to achieve a well-specified logistic regression model, which is also more parsimonious and robust. Out of all the available procedures, the CART algorithm has been used because it is the only one that arranges the predictors according to their normalized importance. The methodology used by Arroyo, Constante-Amores and Asensio (2019) has been followed in this regard. Regression coefficients have been interpreted in terms of probabilities, and model forecasts have been calculated using a probabilistic scale, given the difficulty of interpreting the logit (Pardo and Ruiz, 2013).

In the third analysis, the classification tree method was used again in order to identify the profiles of the students who obtained high and low academic performance levels. This data mining technique was carried out through the CHAID procedure, since, according to Berlanga, et al. (2013), it is the most appropriate algorithm when making profiles. The analyses have been carried out using SPSS version 25 and STATA version 15.

Results

Individual variables associated with academic performance

Table IV shows the results of the first logistic regression analysis. Model 1 shows that the entry grade is positively and significantly related to

⁽⁴⁾ By introducing the entry grade, you can first of all examine how much of the student's academic performance is explained by their pre-university performance (including the EVAU) and, consequently, what other factors have an influence beyond previous performance and the admission exam.

academic performance. Model 2 includes other academic variables and all were significant; having attended high school in a public school, entering the university in June and choosing the first option studies increases the probability of being a good student. In model 3, in which demographic variables are added, neither the student's nationality nor age were significant. Moreover, both being a woman and from Madrid increases the probability of obtaining high performance levels. By introducing the socio-economic⁵ variables, not only are they not significant, but the call for admission and the order of study choice also cease to be so⁶.

TABLE IV. Logit regression models to explain university academic performance (Odds Ratio)

	Model 1 (entry grade)	Model 2 (academic variables)	Model 3 (demographic variables)	Model 4 (socio-econom- ic variables)
Entry grade	1.353***	1.337***	1.352***	1.327***
Public school		1.531***	1.627***	1.366**
June		1.252*	1.288**	1.174
First option		1.166***	1.216***	1.151
Nationality			1.037	0.962
Woman			1.207***	1.369**
Ideal age			1.031	1.049
Madrid			1.484***	1.359**
Not working				1.104
Mother's higher education level				1.016
Constant (ES)	-2.716***	-3.130***	-3.767***	-3.329***
Pseudo R²	0.061	0.064	0.070	0.059

Source: Own elaboration from the database of the UCM Student Observatory.

Notes: * p < 0.1 ** p < 0.05; *** p < 0.01

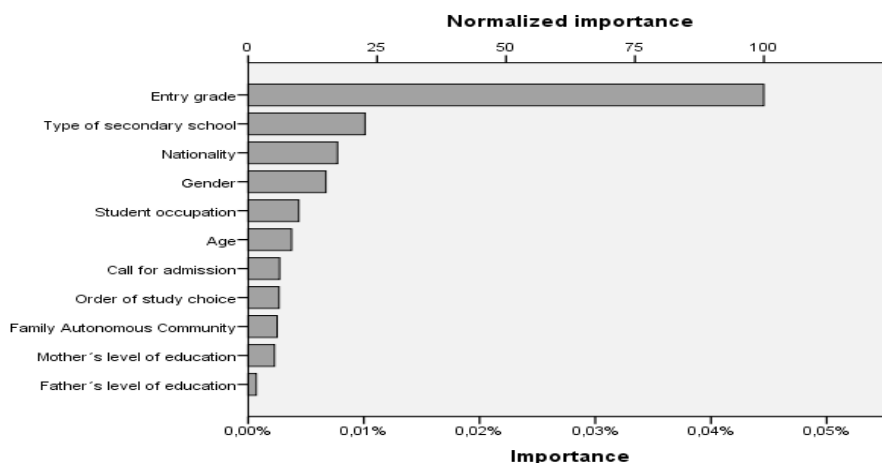
⁽⁵⁾ The variable of studies achieved by the father has been excluded as it is closely related to that of the mother's studies. It has been decided to keep the latter because there is considerable literature (Garbanzo, 2007; Planck, 2014; Tejedor and García-Valcárcel, 2007) that shows the importance of the mother's level of education on university performance.

⁽⁶⁾ These models have been made using the multiple linear regression technique (dependent variable = relative academic performance -the difference between the student's performance and the average performance of the area of knowledge-) obtaining the same four significant and positive signs in the full model.

The entry grade itself explains more than 6% of the variance in the student's academic performance (Pseudo R^2), as seen in model 1. Demographic variables add 1% to the explanation (model 3). Neither the rest of the academic variables (model 2) nor the socio-economic variables (model 4) improve the explanation (in fact, in the last model, the Pseudo R^2 decreases, making the adjustment of the model worse).

Next, the CART classification tree method was used to reduce the specification error and arrange the predictors according to their normalized importance (Graph I). This segmentation technique can be employed since the database has more than 1,000 students (Berlangua et al., 2013). The variable most related to academic performance is the university entry grade, and the lowest is the level of education of both parents. The type of school attended by the student in secondary school, also of an academic nature, occupies second place in normalized importance. The student's occupation, as well as all the variables in Graph I that are behind it, are worth less than 15% of the criterion variable, so we could talk about micro-impacts influencing academic performance (Tuero, Cervero, Esteban and Bernardo, 2018).

GRAPH I. Normalized importance of independent variables that explain academic performance through the CART algorithm



Source: Own elaboration from the database of the UCM Student Observatory.

Before making the final regression model, the eleven independent variables were correlated to eliminate those with a high correlation. The mother's level of education was seen to be closely related to that of the father (which is known as educational homogamy), so we decided to eliminate this last predictor, since it has the least impact on academic performance, according to Graph I. As far as the logistic regression model is concerned, it meets both the assumption of non-collinearity, since tolerance values are above 0.10 and VIF values are no greater than 10, like that of independent errors (Durbin-Watson = 1.7). Atypical cases were also checked, and none were eliminated since Cook's distance was lower than 1. In the final model, those predictors that were not statistically significant ($p < 0.05$) were eliminated.

The final model consists of a total of four independent variables (see Table V). All predictors are positively related to academic performance and explain almost 7% of the variance of this variable.

TABLE V. Final logistics regression model to predict academic performance

	B (ES)	Exp (B)
University entry grade	0.313 (0.13)	1.367***
Public school	0.493 (0.51)	1.637***
Woman	0.190 (0.049)	1.209***
Madrid	0.366 (0.054)	1.442***
Constant	-3.448 (0.147)	0.032***
Pseudo R²	0,068	

Source: Own elaboration from the database of the UCM Student Observatory

Note: ** $p < 0.05$; *** $p < 0.01$

In terms of probability, for each additional point in the entry grade, the student is 58% more likely to obtain high academic performance levels; the probability of obtaining high academic performance levels

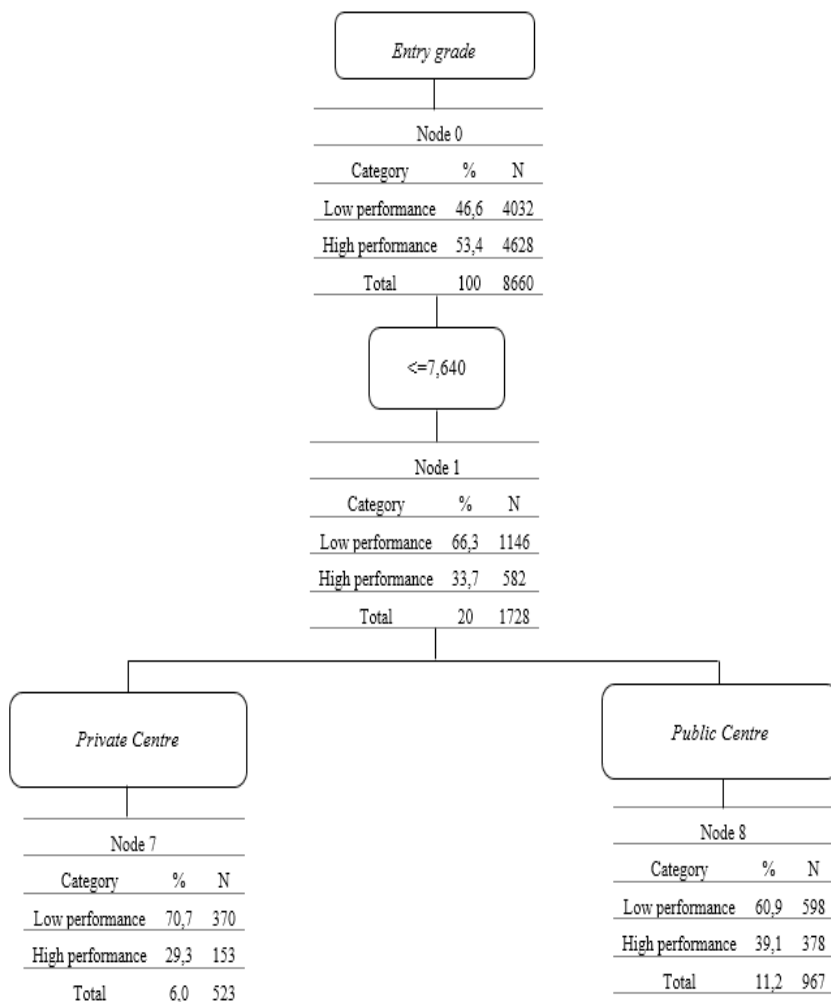
from those who have attended a public secondary school is 63% higher than those who have attended a private school; women are 20% more likely to get good grades than men; students whose family Autonomous Community is located in Madrid are 44% more likely to obtain good grades than students from other Autonomous Communities.

As for forecasts, the highest probability of obtaining high academic performance levels (88%) belongs to a woman (gender = 1) who has obtained an entry grade equal to 13.96, who has studied in a public school (school = 1) and whose family Autonomous Community is located in Madrid (family Autonomous Community = 1). Furthermore, the lowest probability of obtaining low university performance levels (13%) belongs to a man (gender = 0), who has obtained an entry grade of 5, who has studied in a private secondary school (school = 0) and whose family Autonomous Community is not located in Madrid (family Autonomous Community = 0).

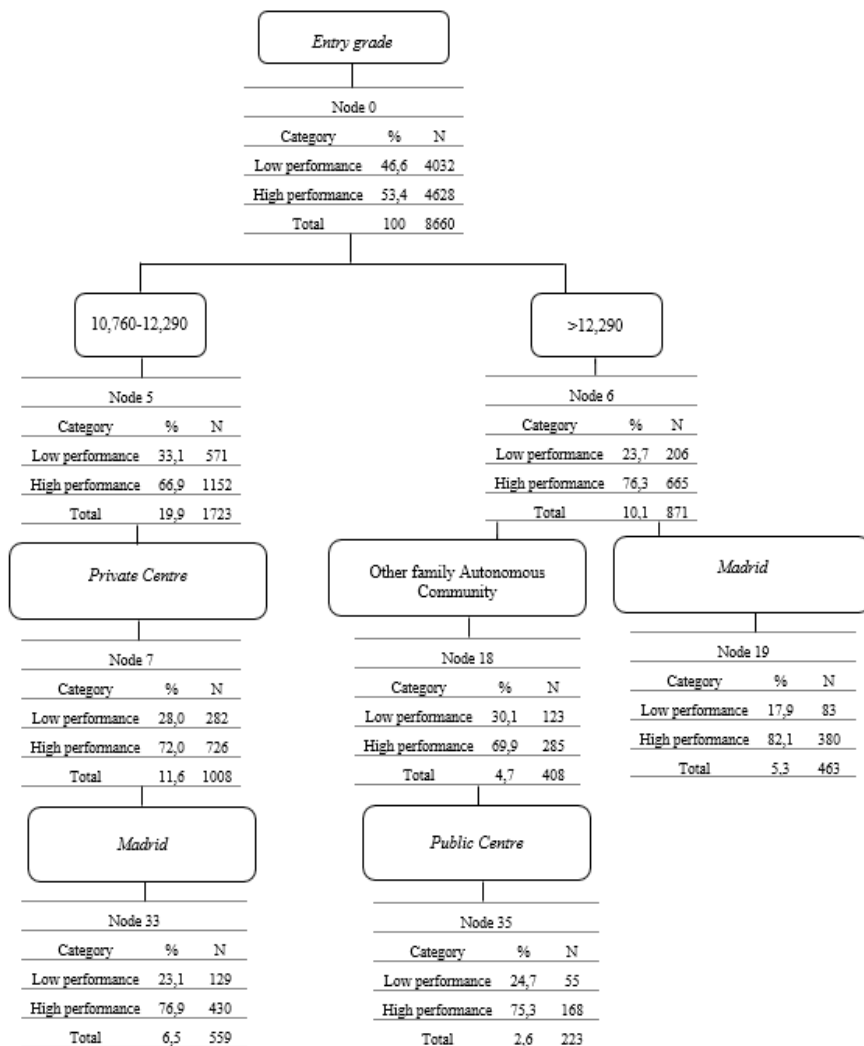
Profile of a student with high and low academic performance

According to the CHAID procedure, the entry grade (node 0) is the best predictor of the dependent variable (see Graphs II and III). The highest probability of obtaining a poor academic performance (70.7%) corresponds to students who have obtained a university entry grade (out of 14) of less than or equal to 7.64 and who have studied in a private secondary school (node 7) (Graph II). The highest probability of obtaining a good academic performance (82.1%) corresponds to students who have an entry grade of over 12.29 and whose family Autonomous Community is in Madrid (node 19) (Graph III). As this profile is very restrictive, the second and third highest probability of obtaining high academic performance levels is indicated. The second highest probability (76.9%) corresponds to students who have obtained an entry grade between 10.76 and 12.29, who have studied in a public secondary school and whose family community is Madrid (node 33). Finally, students who obtain more than 12.29, whose family residence is not in Madrid and who have studied in a public secondary school (node 35) have the third highest probability of obtaining a good performance (75.3%).

GRAPH II. Profile of the student who obtains low academic performance levels



GRAPH III. Profile of the student who obtains high academic performance levels



Discussion and conclusions

The main objective of this article has been to investigate the factors associated with the academic performance of new students. Data from the Complutense University of Madrid has been analysed to clarify what these determinants are and how important each one is. The work has been articulated around three questions from which two scenarios have been formulated with the help of literature reviews. Regarding the individual factors associated with academic performance, the results of the analysis show that the university entry grade has a positive impact on student performance (confirming, in part, the first scenario), although the socio-economic origin of students does not affect it (rejecting, in part, H1). In addition to the entry grade, other variables such as the type of secondary school attended (public), gender (women) and the family Autonomous Community (Madrid) also have an impact on performance. Therefore, in addition to some academic variables, demographic ones are also associated with academic performance.

The entry grade is the most important predictor of the student's academic performance (confirming, in part, the second scenario), as shown in detail using several estimation methods (logistic regression and decision trees –CART and CHAID). The second most important predictor according to the CART algorithm is also of an academic nature, the type of secondary school attended, although other variables of an academic nature remain in less relevant positions. In addition, according to Pseudo R^2 , improved adjustment by academic variables is exclusively due to entry grade. Demographic variables also help the model to adjust better, although only slightly. Therefore, although H2 is fulfilled by the effect of the entry grade, there is growing doubt as to whether other academic variables are so crucial.

The entry grade is also the best predictor to discriminate between being a good and a bad student based on academic performance. Along with it, both the type of school, an academic variable, and the family Autonomous Community, of a demographic nature, are very important factors in the establishment of academic profiles. The third scenario is confirmed, although the relevance of one or other demographic variable must be taken into account.

The research carried out with the new students accessing a UCM degree course has made several new discoveries, which open doors for future

research. First, that individual factors play a role in explaining academic performance, but there is still an important part of this phenomenon that is not explained by the variables considered in the research (93% of the variance). The analysis would have to be expanded to include organizational factors, among others.

Secondly, the analysis has shown the relevance of the university entry grade in subsequent academic performance; the higher the university entry grade, the better the academic performance. In addition to this predictor, the type of secondary school attended is also important, another academic factor. The best performance of students who have attended a public school may have three possible explanations, not yet proven: the academic quality of public schools could be better; the grades of public schools could reflect the level of student performance more accurately than those of private schools; finally, the teaching methods used in public schools could be more similar to those of the public university than in the case of private schools, so students from public schools would not be exposed to such an abrupt change as those that come from private schools. Other academic variables have proved irrelevant to performance: students who are selected in September, or who study something they did not choose in the first place, have the same chances of obtaining high academic performance levels.

Thirdly, out of the demographic and socio-economic variables, only the student's gender and the family Autonomous Community have an impact on performance. If the student is a woman and her family lives in Madrid, the probability of having high academic performance levels increases. A possible line of future research could be to further explore the effect of family Autonomous Communities. Students from Madrid as a whole have the worst average University entry grade, and yet, being from Madrid positively affects performance after starting tertiary education, which could be an indication that the entrance exam to the university (the EVAU) is more difficult in Madrid than in the rest of the Autonomous Communities. A possible explanation for the student's nationality or socio-economic status being irrelevant to performance is that, at the university level, preference is already being given to the selection of students according to their social origin, and therefore the statistical significance of these variables is lower than in previous levels of education. However, it attracts more attention than the work done by students has no impact on their academic performance. Likewise, having

the ideal age to enter the university (between 17 and 19 years old) has nothing to do with academic performance.

Fourthly, from the methodological point of view, a dependent variable has been created, the relative academic performance, which allows students from different fields of study to be analysed together, with different average grades from each other. In a future work, progress can be made in the building of a variable to control the variability of the average grades of different degrees, adding precision to the analysis. Likewise, the results presented do not provide an explanatory model, so this work is limited in such a way that causal relationships between the analysed variables cannot be identified. A possible line of research launched from this study is the analysis of causal relationships between factors (individual and organizational) and academic performance through structural equation models. In the same way, the profile of university students could also be delved into by areas of knowledge through the statistical technique of segmentation.

Research such as this can shed light on factors that determine university performance and ultimately serve to design university policies that improve the quality and equity of the university system.

References

- Abera, G. (2018). Family Socio-Economic Status Effect on Students' Academic Achievement at College of Education and Behavioral Sciences, Haramaya University, Eastern Ethiopia. *Journal of Teacher Education and Educators*, 7(3), 207-222.
- Ariño, A., & Sintés, E. (2016). *Via Universitària: Ser estudiant universitari avui*. Barcelona: Fundació Jaume Bofill.
- Arroyo, D., Constante-Amores, I. A., & Asensio, I. (2019). La repetición de curso a debate: un estudio empírico a partir de PISA 2015. *Educación XX1*, 22(2), 69-92, doi: 10.5944/educXX1.22479
- BBVA Foundation. (2010). *Estudio internacional Fundación BBVA sobre estudiantes universitarios de seis países europeos*. Madrid: Dpto. de Estudios Sociales y Opinión Pública.

- Berlanga, V., Rubio, M. J., & Vilà, R. (2013). Cómo aplicar árboles de decisión en SPSS. *Revista d'Innovació i Recerca en Educació*, 6(1), 65-79. DOI:10.1344/reire2013.6.1615
- Bunce, L., Baird, A., & Jones, S. E. (2017). The student-as-consumer approach in higher education and its effects on academic performance. *Studies in Higher Education*, 42(11), 1958-1978.
- Cerdeira, J. M., Catela, L. C., Balcão A., & Seabra, C. (2018). Predictors of student success in Higher Education: Secondary school internal scores versus national exams. *Higher Education Quarterly*, 72(4), 304-313.
- Chong, E. (2017). Factores que inciden en el rendimiento académico de los estudiantes de la Universidad Politécnica del Valle de Toluca. *Revista Latinoamericana de Estudios Educativos*, 47(1), 91-108.
- Clerici, R., & Da Re, L. (2019). Evaluación de la eficacia de un programa de tutoría formativa. *Revista de Investigación Educativa*, 37(1), 39-56.
- Cortez, F. M., Tutiven, J. L., & Villavicencio, M. N. (2017). Determinantes del Rendimiento Académico Universitario. *Revista Publicando*, 4(10), 284-296.
- Danilowicz-Gösele, K., Lerche, K., Meya, J., & Schwager, R. (2017). Determinants of students' success at university. *Education Economics*, 25(5), 513-532.
- De Sousa, L., Taboada, E., Rivas, R., Iglesias, P., & López, S. (2017). Rendimiento académico y variables sociodemográficas y académicas en la Educación Superior en Angola. *Revista de Estudios e Investigación en Psicología y Educación*, Extr.(14), 112-116.
- Dennis, J. M., Phinney, J. S., & Chuateco, L. I. (2005). The role of motivation, parental support, and peer support in the academic success of ethnic minority first-generation college students. *Journal of College Student Development*, 46, 223-236.
- Duart, J. M., Gil, M., Puyol, M., & Castaño, J. (2008). *La universidad en la sociedad Red*. Barcelona: Ariel.
- EUROSTUDENT. (2018). *Social and economic conditions of student life in Europe. Eurostudent VI 2016-2018. Synopsis of Indicators*. Bielefeld: German Centre for Higher Education Research and Science Studies (DZHW).
- Garbanzo, G. M. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Revista de Educación*, 31(1), 43-63.

- Garbanzo, G. M. (2013). Factores asociados al rendimiento académico en estudiantes universitarios desde el nivel socioeconómico: Un estudio en la Universidad de Costa Rica. *Revista Electrónica Educare*, 17(3), 57-87.
- Garbanzo, G. (2014). Factores asociados al rendimiento académico tomando en cuenta el nivel socioeconómico: Estudio de regresión múltiple en estudiantes universitarios. *Revista Electrónica Educare*, 18(1), 119-154.
- García, A. M. (2014). Rendimiento académico y abandono universitario: modelos, resultados y alcances de la producción académica en la Argentina. *Revista Argentina de Educación Superior*, (8), 9-38.
- Jansen, E. P. (2004). The influence of the curriculum organization on study progress in higher education. *Higher Education*, 47(4), 411-435.
- Jiménez-Caballero, J. L., Camúñez, J. A., González-Rodríguez, M. R., & Fuentes, P. (2015). Factores determinantes del rendimiento académico universitario en el espacio europeo de educación superior. *Revista De Ciencias Administrativas y Sociales*, 25(58), 159-176.
- Lamas, H. A. (2015). Sobre el rendimiento escolar. *Propósitos y representaciones*, 3(1), 313-386.
- Manzano, J. A., Escribá, A., Iborra, M., & Safón, V. (2016). *La Universidad Española. Grupos Estratégicos y Desempeño*. Bilbao: Biblioteca Nueva
- Mañé, F., & Miravet, A. (2010). *Rendimiento académico de los estudiantes de primer año en la universidad*. Barcelona: AQU Catalunya. Retrieved from http://www.aqu.cat/doc/doc_81940174_1.pdf
- Marcenaro, O. D., & Navarro, M. L. (2007). El éxito en la universidad: una aproximación cuantílica. *Revista de Economía Aplicada*, 44(15), 5-39.
- McNabb, R., Pal, S., & Sloane, P. (2002). Gender differences in educational attainment: the case of university students in England y Wales. *Economica*, 69(275), 481-503.
- Nyikahadzoi, L., Matamande, W., Taderera, E., & Mandimika, E. (2013). Determinants of students' academic performance in four selected accounting courses at University of Zimbabwe. *Research in Higher Education Journal*, 21, 1-9.
- Pardo, M., & Ruiz, M. (2013). *Análisis de Datos en Ciencias Sociales y de la Salud III*. Madrid: Síntesis.
- Pérez-Cárceles, M. C., Gómez-Gallego, M., Gómez-Gallego, J. C., Palazón-Pérez de Los Cobos, A., & Gómez-García, J. (2014). El género como variable moderadora de los Resultados Académicos en la Enseñanza Universitaria. *Regional and Sectoral Economic Studies*, 14(3), 55-64.

- Planck, U. (2014). Factores determinantes del rendimiento académico de los estudiantes de la universidad de atacama. *Estudios Pedagógicos*, 40(1), 25-39.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353-387.
- Rodríguez, G., Ariza, M., & Ramos, J. L. (2014). Calidad institucional y rendimiento académico: El caso de las universidades del Caribe colombiano. *Perfiles educativos*, 36(143), 10-29.
- Rodríguez, S., Fita, E., & Torrado, M. (2004). El rendimiento académico en la transición secundaria-universidad. *Revista de Educación*, 334(1), 391-414.
- Ruesga, S. M., Silva, J., & Monsueto (2014). Estudiantes universitarios, experiencia laboral y desempeño académico en España. *Revista de Educación*, 365, 67-95.
- Salinas, D. A., Hernández, A. E. y Barboza-Palomino, M. (2017). Condición de becario y rendimiento académico en estudiantes de una universidad peruana. *Revista Electrónica de Investigación Educativa*, 19(4), 124-133.
- Santos, M. V., & Vallelado, E. (2013). Algunas dimensiones relacionadas con el rendimiento académico de estudiantes de Administración y Dirección de Empresas. *Universitas Psychologica*, 12(3), 739-752.
- Simón, H., Casado, J., & Castejón, J. (2017). Analysis of university student employment and its impact on academic performance. *Electronic Journal of Research in Educational Psychology*, 15(42), 281-306.
- Smith, J., & Naylor, R. (2001). Determinants of Degree Performance in UK Universities: A Statistical Analysis of the 1993 Student Cohort. *Oxford Bulletin of Economics and Statistics*, 63(1), 29-60.
- Tejedor, F. J., & García-Valcárcel, A. (2007). Causas del bajo rendimiento del estudiante universitario (en opinión de los profesores y alumnos). Propuestas de mejora en el marco del EEES. *Revista de Educación*, 342(1), 443-473.
- Thiele, T., Pope, D., Singleton, A., Snape, D., & Stanistreet, D. (2017). Experience of disadvantage: The influence of identity on engagement in working class students' educational trajectories to an elite university. *British Educational Research Journal*, 43(1), 49-67.
- Tuero, E., Cervero, A., Esteban, M., & Bernardo, A. (2018). ¿Por qué abandonan los alumnos universitarios? Variables de influencia en el

planteamiento y consolidación del abandono. *Educación XX1*, 21(2), 131-154.

- Turrul, M., & Roca, B. (2010). *El rendimiento académico del alumnado de primer curso de Derecho en la Universidad de Barcelona*. Barcelona: AQU Catalunya. Retrieved from http://www.aqu.cat/doc/doc_61647381_1.pdf
- Vélez, A., & Roa, C. (2005). Factores asociados al rendimiento académico en estudiantes de Medicina. *Educación Médica*, 8(2), 24-32.
- Vinacur, T. (2016). ¿Los alumnos de las escuelas privadas están mejor preparados para ingresar a la universidad? *Revista Colombiana de Educación*, (70), 175-200.

Annex

TABLE I. Descriptions of independent qualitative variables

Variable	Independent variable levels	N	%
Gender	Man	4,086	38.1
	Woman	6,634	61.9
Age	Not ideal: 20 or over	2,743	25.6
	Ideal: 19 or under	7,977	74.4
Nationality	Other	890	8.3
	Spanish	9,830	91.7
Family Autonomous Community	Another family autonomous community	3,171	29.6
	Madrid	7,549	70.4
Student occupation	Working	716	33.4
	Not working	1,430	66.6
Father's level of education	Secondary education or less	5,055	51.5
	Higher education	4,757	48.5
Mother's level of education	Secondary education or less	4,885	48.4
	Higher education	5,216	51.6
Type of secondary school	Private	2,966	36.4
	Public	5,193	63.6
Call for admission	September	549	5.1
	June	10,171	94.9
Order of study choice	First option	6,568	61.3
	Second to twelfth option	4,152	38.7

Source: Own elaboration from the database of the UCM Student Observatory.

TABLE II. Descriptions of the individual quantitative variable

Variable	N	Mean	Standard deviation	Minimum	Maximum
Entry grade	10,720	9.245	2.053	5.00	13.960

Source: Own elaboration from the database of the UCM Student Observatory.

Contact address: María Fernández-Mellizo. Universidad Complutense de Madrid, Facultad de Educación, Sociología Aplicada, Instituto TRANSOC. Edificio La Almudena. C/Rector Royo Villanova s/n, Ciudad Universitaria. 28040, Madrid. E-mail: mfmellizotoso@edu.ucm.es

