

BORDÓN

Revista de Pedagogía



Volumen 70
Número, 1
2018

SOCIEDAD ESPAÑOLA DE PEDAGOGÍA

VALIDEZ PREDICTIVA DEL U-RANKING EN LAS TITULACIONES UNIVERSITARIAS DE CIENCIAS DE LA SALUD

The U-Ranking's predictive validity in university Health Studies

ESTHER LÓPEZ-MARTÍN, ALEXIS MORENO-PULIDO Y EVA EXPÓSITO-CASAS
Universidad Nacional de Educación a Distancia

DOI: 10.13042/Bordon.2017.53478

Fecha de recepción: 20/11/2016 • Fecha de aceptación: 28/05/2017

Autora de contacto / Corresponding Author: Esther López-Martín. E-mail: estherlopez@edu.uned.es

Fecha de publicación *online*: 30/08/2017

INTRODUCCIÓN. Los *rankings* de universidades han alcanzado un enorme protagonismo mediático en los últimos años. La información que publican acerca de la calidad de las actividades docentes, de investigación y de innovación y desarrollo tecnológico de las universidades tiene una gran repercusión social y condiciona la reputación y el prestigio de las mismas, influyendo en la atracción de estudiantes. El presente trabajo estudia la validez predictiva del U-Ranking en egresados de Medicina, Farmacia, Enfermería y Psicología en España. **MÉTODO.** En este trabajo se analiza en qué medida las puntuaciones que obtienen las diferentes universidades españolas de titularidad pública y privada presentes en esta clasificación, permiten predecir el nivel de rendimiento alcanzado por sus egresados en las pruebas selectivas nacionales de formación sanitaria especializada (utilizando datos procedentes de cuatro convocatorias 2012/13, 2013/14, 2014/15 y 2015/16). **RESULTADOS.** Los resultados ponen de manifiesto la ausencia de relación entre las puntuaciones que obtienen las universidades en el U-Ranking y el porcentaje de éxito que alcanzan los estudiantes de dichas instituciones en las pruebas de selección para los aspirantes a las plazas de formación sanitaria especializada en las titulaciones mencionadas. **DISCUSIÓN.** La tendencia marcada en estos resultados evidencia un problema de validez inherente al U-Ranking, asociado al error sistemático en la predicción de un criterio. Asimismo, se destaca la necesidad de replantear el modo en que se construyen los *rankings* de universidades, partiendo de una definición precisa del constructo a evaluar, a fin de mejorar la calidad de la medida asegurando su idoneidad.

Palabras clave: *Rankings universitarios, U-Ranking, Validez predictiva, Impacto de la formación, Ciencias de la Salud, Pruebas selectivas de personal sanitario.*

Introducción

Los *rankings* de universidades tienen como objetivo esencial comparar la actividad de las instituciones de educación superior (IES) en sus tres dimensiones principales: la investigación, la docencia y la innovación (Tomàs-Folch, Feixas-Condom, Ruiz-Ruiz y Bernabeu-Tamayo, 2015). En la construcción de estos *rankings*, se establece una primera fase en la que se definen los indicadores observables en los que se concreta la dimensión evaluada y, posteriormente, se tiende a aglutinar estos indicadores, formando un índice sintético para cada dimensión, según el cual se jerarquizan las IES (Daraio y Bonaccorsi, 2016).

El interés por los *rankings* no ha parado de crecer desde la aparición en el año 2003 del Academic Ranking of World Universities-ARWU o Ranking de Shanghái, al que siguieron el National Taiwan University Ranking-NTU (Ranking de Taiwán), el Quacquarelli Symonds World University Ranking (Ranking QS), el Times Higher Education Ranking (Ranking THE), escindido del anterior, el CWTS Leiden Ranking, el Scimago Institutions Ranking o el Webometrics Ranking of World Universities, por citar algunos de los más destacados (Rauhvargers, 2011; 2013). Asimismo, desde el año 2014, la Unión Europea financia la construcción de una clasificación multidimensional de las IES de todo el mundo (U-Multirank¹). Como puede observarse, los *rankings* de universidades se han extendido por un gran número de países, tanto desarrollados como en vías de desarrollo (Orduña-Malea, 2012).

La posición que ocupa el sistema universitario español en los *rankings* internacionales ha sido objeto de varios estudios (Alva-Chávez, Díaz-Varela, Franco-Romo y Ballesteros-Carrasco, 2014; Docampo, 2015; Grau, 2013; 2015), que constatan que solo un reducido número de universidades tienen presencia en los mismos. Así, para obtener comparaciones de la actividad de las universidades españolas en su conjunto es

necesario recurrir a *rankings* nacionales. A este respecto, en España se han construido diversos *rankings* de universidades como, por ejemplo, el *ranking* I-UGR², desarrollado por el grupo EC3, o el elaborado por Buela-Casal, Quevedo-Blasco y Guillén-Riquelme (2015), que se centran únicamente en la dimensión investigadora. Otros *rankings*, como los llevados a cabo por la Fundación Conocimiento y Desarrollo³, por el Observatorio IUNE⁴ o por el Instituto Valenciano de Investigaciones Económicas (Ivie) y la Fundación BBVA (U-Ranking)⁵, junto con la dimensión investigadora, también analizan las dimensiones de docencia, innovación y transferencia.

En los últimos años, estos *rankings* han alcanzado un protagonismo mediático considerable. Parte de este éxito puede deberse a que, tal y como señalan González-Riaño, Repiso y Delgado-López-Cózar (2014), son herramientas sencillas que despiertan la atención del público y, en consecuencia, la información que publican tiene una enorme repercusión social y condiciona la reputación y el prestigio de las universidades, influyendo en la atracción de estudiantes. Además, los *rankings* tratan de dar respuesta a un conjunto de necesidades sociales como la de orientar a los estudiantes en la elección de la IES donde cursar estudios, la de informar a los gestores universitarios sobre sus puntos fuertes y débiles y la de facilitar la toma de decisiones en el ámbito de la política educativa (Orduña-Malea, 2012).

A pesar de su posible interés, los *rankings* han sido objeto de un gran número de críticas que se centran, fundamentalmente, en su debilidad metodológica (Delgado-López-Cózar, 2012; Van Raan, 2005). Daraio y Simar (2015) realizan una exhaustiva revisión de estas críticas, agrupándolas en cuatro categorías: (a) *Unidimensionalidad*, que consiste en reducir todas las actividades de la universidad a la dimensión investigadora, en detrimento de la docencia, cuyo análisis se reduce a unos pocos indicadores; (b) *Robustez estadística*, que engloba las

críticas que se centran en el proceso seguido en la construcción de los *rankings* (por ejemplo, en la asignación de pesos a los indicadores o en la norma empleada para contraer varios indicadores en una sola medida), así como en los posibles problemas de falta de validez, fiabilidad y comparabilidad de las medidas; (c) *Dependencia del tamaño de la universidad y mezcla de disciplinas*. Las universidades de mayor tamaño salen favorecidas en la foto que ofrecen los *rankings*. Además, la mezcla de áreas de conocimiento favorece a las universidades en las que dominan las disciplinas científicas, técnicas y médicas, caracterizadas por un elevado ritmo de producción, y perjudica a las que tienen un componente mayor de ciencias sociales y humanidades; y (d) *Falta de consideración de la estructura input-output*. Los *rankings* no relacionan los resultados obtenidos con los recursos utilizados para obtenerlos, es decir, no tienen en cuenta la eficiencia.

¿Qué evalúan los *rankings*? Evidencias de validez

Como se ha adelantado en el apartado anterior, cuando se analizan las propiedades métricas de los índices a partir de los cuales se construyen los *rankings*, se debe prestar especial atención a las evidencias de validez en torno a los mismos. Dentro del ámbito de la psicometría, pero trasladable a este contexto, la validez se presenta como el grado en el que las evidencias empíricas y teóricas amparan la interpretación de las puntuaciones atendiendo a un propósito determinado (American Educational Research Association [AERA], American Psychological Association [APA] y National Council on Measurement in Education [NCME], 2014).

Los estudios de validez persiguen recabar suficientes evidencias que puedan prestar una base científica a la interpretación de las puntuaciones en un uso concreto (Elosua, 2003), analizando, entre otros: la propia definición del constructo evaluado, así como la relevancia y

representatividad de los indicadores en los que se concreta (*validez de contenido* [Sireci y Faulkner-Bond, 2014]); la relación entre las puntuaciones del instrumento de medida con un criterio externo (*validez de criterio —concurrente o predictiva—* [AERA et al., 2014]); o las evidencias que apoyan el uso que se hace de las puntuaciones (*validez basada en las consecuencias* [Cizek, 2012; Lane, 2014]).

La repercusión social de los *rankings*, así como su influencia sobre el proceso de toma de decisiones (individuales o administrativas), pone de manifiesto la necesidad de analizar aquellas evidencias que garanticen la validez de los datos que sustentan los mismos. En el caso concreto de los *rankings* de docencia, cuyo análisis constituye el núcleo de este trabajo, las posiciones relativas de las universidades son calculadas a partir de sus puntuaciones individuales en un conjunto de indicadores y, en consecuencia, es necesario que tales indicadores sean manifestaciones observables de la realidad que está siendo evaluada: la calidad de la docencia universitaria. Esta calidad docente se concreta en la capacidad, a nivel individual (profesor) o institucional (universidades), de proporcionar una ayuda educativa a los alumnos que repercuta en la construcción de nuevos conocimientos y en el logro de los objetivos educativos (Mauri, Coll y Onrubia, 2008), asociándose, por tanto, con los resultados educativos alcanzados y con las condiciones o el proceso educativo que permiten conseguir dichos resultados (Ramírez Garzón y Montoya Vargas, 2014).

Una vez que se ha garantizado la precisión de la medida pretendida, resulta esencial proporcionar evidencias de su validez predictiva. Tradicionalmente se ha considerado la validez de criterio como aquella que se basa en la relación de las puntuaciones de un instrumento de medición con un criterio externo, siendo este una medida operacionalmente distinta respecto a la medida cuya validez se desea analizar. La fórmula convencional de referirse a este tipo de evidencia de validez establece dos tipologías,

predictiva y concurrente (AERA *et al.*, 2014), cuya diferenciación se basa en el punto temporal en el que se realiza la toma de datos de la variable criterio (en el mismo momento —concurrente— o en distinto momento —predictiva—). Los resultados de la calidad de la docencia de las universidades, difundidos por los *rankings*, pueden ser analizados tomando en consideración las evidencias de validez basadas en la predicción, entendiéndose estas como el grado en el que los datos aportados por una medida en un momento determinado pueden predecir las puntuaciones en un criterio externo obtenidas con posterioridad (AERA *et al.*, 2014). Por ejemplo, la calidad de la docencia de las universidades, medida indirectamente a través de los *rankings*, podría considerarse un predictor relevante de la formación recibida por los estudiantes y, consecuentemente, de su futuro éxito laboral.

Objetivo de investigación

El principal objetivo de este trabajo ha sido analizar la validez predictiva del U-Ranking. Para ello, se ha comprobado en qué medida la puntuación asignada a las distintas universidades en este *ranking*, se relaciona con el porcentaje de egresados que, de dichas instituciones, superan las pruebas selectivas para el acceso a plazas de formación sanitaria especializada para médicos, farmacéuticos, enfermeros y psicólogos.

Teniendo en cuenta que el U-Ranking proporciona información sobre la calidad de las actividades docentes, de investigación y de innovación y desarrollo tecnológico de las universidades, la capacidad predictiva de esta herramienta quedará probada, si son las instituciones que ocupan las primeras posiciones en este *ranking* (fundamentalmente en la dimensión de docencia), las que más tasas de aprobados obtienen en estas pruebas competitivas y, viceversa, los estudiantes de aquellas universidades que puntúan bajo en el U-Ranking, tenderán a superar estas pruebas selectivas en menor medida.

Método

A continuación se describen las variables consideradas, así como las instituciones que han formado parte de este estudio.

Variables e instrumentos utilizados

El U-Ranking forma parte del proyecto Indicadores Sintéticos del Sistema Universitario Español (ISSUE), desarrollado en colaboración por la Fundación BBVA y el Instituto Valenciano de Investigaciones Económicas (Ivie), y se presenta como una herramienta que jerarquiza anualmente la calidad de las actividades docentes, de investigación y de innovación y desarrollo tecnológico de las universidades españolas (Pérez, Aldás, Aragón y Zaera, 2016). La evaluación de cada una de estas dimensiones se establece en torno a cuatro ámbitos (acceso a recursos, producción obtenida, calidad e internacionalización de las actividades) que, a su vez, se concretan en un conjunto de indicadores (en el anexo 1 se incluyen los indicadores relativos a la dimensión de docencia).

Entre los resultados de este proyecto, se encuentra una herramienta web⁶ que permite al usuario construir *rankings* personalizados teniendo en cuenta los siguientes criterios: *titulación*, *comunidad autónoma* y *ponderación* que se desea otorgar a las diferentes actividades evaluadas: *docencia*, *investigación* e *innovación* y *desarrollo tecnológico*. En relación a este último aspecto, en el caso de que no se personalice esta ponderación, la herramienta aplica por defecto los siguientes pesos: 56% para docencia, 34% para investigación y 10% para innovación y desarrollo tecnológico.

A partir de la información proporcionada por esta herramienta, para el caso concreto de este trabajo, se han construido un total de ocho *rankings* personalizados combinando los siguientes criterios:

- Titulación: Medicina, Farmacia, Psicología y Enfermería.
- Comunidad Autónoma: todas.
- Ponderación: a) 100% docencia, 0% investigación y 0% innovación y desarrollo; y b) 56% docencia, 34% investigación y 10% innovación y desarrollo (ponderación por defecto).

El resultado de esta selección ofrece la siguiente información: una primera columna con la posición de cada universidad en el grado elegido y, a continuación, el valor del índice calculado para cada institución. Pérez *et al.* aclaran que “varios grados pueden ocupar la misma posición en el *ranking*, dado que los índices están redondeados a un decimal porque no se considera que mayor precisión refleje, con mayor exactitud, diferencias entre grados” (2016: 34). Asimismo, los autores defienden la robustez del U-Ranking respecto a los cambios en los pesos y afirman que “una universidad no pasa de los primeros lugares a los últimos por sustanciales que sean los cambios en los pesos, pero es cierto que algunas pueden mejorar en el *ranking* algunos puestos si se reconoce mayor importancia a la docencia o a la investigación” (2016: 48).

En definitiva, para cada una de las cuatro titulaciones se han obtenido dos clasificaciones diferentes que ordenan a las universidades que imparten dicha titulación en función de la ponderación asignada a las diferentes actividades (100% docencia y ponderación por defecto). Cabe señalar que no se ha considerado la clasificación de las universidades cuando se evalúa de manera exclusiva la investigación o la innovación y el desarrollo, ya que podría considerarse que estas actividades están más alejadas de la calidad de la formación que reciben los estudiantes en las diferentes instituciones y, consecuentemente, su posible influencia sobre el porcentaje de sujetos que superan las pruebas selectivas para el acceso a plazas de formación sanitaria especializada podría ser cuestionada.

La tabla 1 recoge el número de universidades que se incluyen en el U-Ranking 2016 para las titulaciones de Enfermería, Medicina, Farmacia y Psicología.

TABLA 1. Universidades incluidas en el U-Ranking por titulación

	N total	N públicas	N privadas
Enfermería	48	40	8
Medicina	17	13	4
Farmacia	34	29	5
Psicología	38	29	9

Tomando en consideración la variable criterio de este estudio, dicha variable está constituida por el porcentaje de sujetos de las distintas universidades que superan la prueba de selección para los aspirantes a las plazas de formación sanitaria especializada en Medicina, Farmacia, Enfermería y Psicología. Esta prueba selectiva contempla la realización de un ejercicio de preguntas múltiples, distinto para cada una de las titulaciones que, en la última convocatoria (Orden SSI/1892/2015, de 10 de septiembre), estuvo compuesto por un total de 225 preguntas y cuya duración se estableció en cinco horas. La superación de este ejercicio es requisito *sine qua non* para que los aspirantes participen en los actos de adjudicación de plazas aunque, para la asignación definitiva de las mismas, se valora además el expediente académico de los participantes.

La objetividad de esta prueba, el carácter nacional de la convocatoria y el hecho de que las preguntas aborden contenidos relativos a cada una de las titulaciones son algunos de los aspectos que justifican la inclusión, como variable criterio de este estudio, de los resultados obtenidos por los aspirantes de las diferentes instituciones en esta prueba. Concretamente, se ha considerado el porcentaje⁷ de sujetos de las distintas universidades que han superado la nota de corte sobre el total de candidatos

presentados a examen en las convocatorias 2012-13, 2013-14, 2014-15 y 2015-16, excepto en el caso de Psicología que solo se dispone de datos de las tres últimas convocatorias.

En este punto, es importante señalar que, con la finalidad de incrementar la precisión de la medida, se ha utilizado la información disponible para todo el periodo, en lugar de considerar un año concreto. En cualquier caso, el tener en cuenta los datos de las cuatro últimas convocatorias no afecta a una de las principales condiciones que se debe cumplir al analizar la validez predictiva, como es que la medida del criterio debe ser posterior a la medida de la escala objeto de estudio, ya que, aunque los datos del U-Ranking proceden de la última edición (U-Ranking 2016), en la construcción de los diferentes índices se utiliza información referida a los siguientes periodos: 2008-2015 en el caso de docencia y 2008-2014 para la evaluación de las actividades investigadoras y de innovación y desarrollo tecnológico (Pérez *et al.*, 2016).). A su vez, la tabla 2 permite observar la alta correlación que existe entre este porcentaje global y los resultados de las universidades en

las diferentes convocatorias; lo cual garantiza la estabilidad de la medida.

Análisis de datos

El procedimiento utilizado con mayor frecuencia a la hora de analizar evidencias de validez basadas en la dimensión predictiva es el cálculo de la correlación entre la medida de interés y algún criterio relevante (Abad, Olea, Ponsoda y García, 2011). A la correlación entre ambas variables, se le denomina coeficiente de validez y refleja la magnitud y la intensidad de la relación entre ambas variables. Dicha correlación pone de manifiesto en qué grado la variable predictora (X) y la variable criterio (Y) miden el mismo constructo, es decir, informa sobre la validez de las mismas en torno a su objeto de interés.

En el contexto de este trabajo, el coeficiente de validez indicará en qué medida las puntuaciones que obtienen las universidades en el U-Ranking permiten pronosticar el porcentaje de éxito que alcanzan los estudiantes de dichas

TABLA 2. Correlación entre el porcentaje de aprobados global y el porcentaje de aprobados en cada una de las convocatorias

		Convocatoria 2012-13	Convocatoria 2013-14	Convocatoria 2014-15	Convocatoria 2015-16
Enfermería	Correlación	0,932**	0,915**	0,954**	0,837**
	N	51	51	56	55
Farmacia	Correlación	0,826**	0,836**	0,612**	0,700**
	N	15	16	18	19
Medicina	Correlación	0,793**	0,932**	0,938**	0,905**
	N	28	29	34	34
Psicología	Correlación		0,752**	0,637**	0,801**
	N		37	36	40

** Correlación significativa a un nivel de confianza del 99% ($\alpha = 0,01$).

Nota: Se estima el coeficiente de correlación de Spearman ya que, en una parte importante de los casos, las variables no se distribuyen normalmente o se trabaja con muestras pequeñas ($n < 30$) (anexo II).

instituciones en las prueba de selección para los aspirantes a las plazas de formación sanitaria especializada en Medicina, Farmacia, Enfermería y Psicología.

En la estimación de dicha relación, generalmente se utiliza el coeficiente de Correlación Producto-Momento de Pearson cuando las variables objeto de estudio son continuas; sin embargo, si las variables se expresan de otra manera, suelen calcularse otro tipo de coeficientes de correlación (Anastasi y Urbina, 1998). En el presente estudio, aunque las variables objeto de interés son continuas, en ocasiones no se distribuyen normalmente (anexo II) o se trata de muestras pequeñas (< 30) y, por tanto, resulta más apropiado utilizar el coeficiente de correlación de Spearman, que es igual a:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Donde, n es el número de parejas de datos y d_i es la diferencia entre el i -ésimo par de rangos $[R(X_i)-R(Y_i)]$.

Resultados

Los resultados presentados en la tabla 3 muestran cómo en el caso de las cuatro titulaciones consideradas (Enfermería, Farmacia, Medicina y Psicología), la calidad docente establecida por el U-Ranking no se relaciona con el porcentaje de estudiantes que, de estas universidades, superan las pruebas selectivas de personal sanitario, es decir, el índice de docencia no permite predecir el éxito de los estudiantes en dichas pruebas.

En relación al índice establecido por defecto por el U-Ranking que, junto con la dimensión docente, considera las dimensiones de investigación y de innovación y desarrollo tecnológico de las IES, los resultados son similares, es decir,

TABLA 3. Correlación entre el porcentaje de aprobados en cada una de las titulaciones y los índices (Docencia y Ponderación por defecto) proporcionados por el U-Ranking

		Índice de docencia	Índice de ponderación por defecto
Enfermería	Correlación	0,171	0,242
	N	46	46
Farmacia	Correlación	-0,246	-0,247
	N	15	15
Medicina	Correlación	-0,013	0,113
	N	33	33
Psicología	Correlación	0,230	0,307
	N	31	31

Nota: en el caso de Psicología, no se han considerado las puntuaciones de tres universidades de las cuales se han presentado un número de egresados igual o inferior a tres en el periodo de estudio. Siguiendo el mismo criterio, se ha excluido una IES de Farmacia.

la correlación entre este índice y el porcentaje de egresados de las distintas universidades que superan dichas pruebas selectivas no resulta significativa.

Discusión

Los resultados presentados en el apartado anterior ponen de manifiesto la falta de relación entre las puntuaciones que obtienen las universidades en el U-Ranking y el porcentaje de éxito que alcanzan los estudiantes de dichas instituciones en las pruebas de selección para los aspirantes a las plazas de formación sanitaria especializada en Medicina, Farmacia, Enfermería y Psicología.

Bajo la premisa de que la calidad de la docencia que reciben los estudiantes universitarios debe tener una repercusión directa sobre los aprendizajes que estos adquieren, la capacidad predictiva del U-Ranking quedaría probada si las puntuaciones obtenidas por las IES en el índice de docencia de este *ranking* se relacionasen positivamente con las tasas de aprobados de dichas universidades en las pruebas selectivas para el acceso a plazas de formación sanitaria especializada. Sin embargo, la correlación entre ambas variables no resulta significativa y estos resultados podrían estar evidenciado un problema de validez inherente al U-Ranking. Concretamente, se trataría de un problema asociado al sesgo en la predicción, entendiendo este sesgo como un error sistemático al predecir el rendimiento en un criterio a través de características que no resultan relevantes en relación al mismo, perjudicando o beneficiando en las predicciones a grupos o entidades particulares (AERA *et al.*, 2014).

Llegados a este punto cabría preguntarnos: ¿existe una verdadera relación entre dicho índice de docencia y los aprendizajes adquiridos por los estudiantes? A partir de este trabajo se ha pretendido arrojar luz sobre esta cuestión, observándose, que dicho índice no

se relaciona con los resultados obtenidos por los estudiantes en unas pruebas objetivas, de carácter nacional, que abordan contenidos relativos a la titulación y que presentan unas propiedades psicométricas adecuadas (Baladrón, Curbelo, Sánchez-Lasheras, Romeo-Ladrero, Villacampa y Fernández-Somoano, 2016). En cualquier caso, es importante señalar que una respuesta precisa a esta cuestión requeriría de un estudio más profundo de la relación entre ambas variables.

En este sentido, y a pesar del valor informativo de los resultados obtenidos por los egresados que se presentan a estas pruebas selectivas, conviene subrayar algunas de las principales limitaciones que conlleva el trabajo con estas puntuaciones, y que han sido puestas de manifiesto en estudios previos que han analizado en profundidad los resultados de las universidades en estas pruebas (López-Valcárcel, Ortún, Barber, Harris y García, 2013). Entre dichas limitaciones, destaca el hecho de que no todos los egresados de estas instituciones se presentan a este examen, y la dificultad para controlar el efecto de algunas variables que podrían influir sobre los resultados de los aspirantes, pero que no están directamente relacionadas con la calidad del programa de estudio. Entre estas variables se encontrarían, por ejemplo, las características individuales de los aspirantes (motivación, esfuerzo, número de veces que se han presentado a las pruebas, etc.) o de la institución (nota media de acceso a la universidad, tamaño de la promoción, etc.). A este respecto, los resultados del estudio realizado por Sentí, Pérez y Baños (2016) muestran cómo estas dos características institucionales (nota de acceso y tamaño de la promoción) parecen no relacionarse con los resultados obtenidos por los titulados en Medicina en esta prueba, sin embargo, el número de egresados que se presentan fuera de su año “natural”, es decir, el año posterior a su graduación (recirculación) sí podría constituir un factor explicativo de las diferencias entre las universidades.

Conclusiones

La repercusión social de los *rankings*, así como su influencia sobre el proceso de toma de decisiones (individuales o administrativas), ponen de manifiesto la necesidad de analizar evidencias que apoyen la validez de los instrumentos de medida utilizados en su construcción. En el caso de los índices calculados por el U-Ranking y, de acuerdo con los resultados de este estudio, los procesos de análisis y toma de decisiones basados en estas medidas parecen no ser suficientemente robustos. Una posible solución a los problemas de validez predictiva detectados pasaría por el estudio de la validez incremental de este tipo de mediciones (Hunsley y Meyer, 2003), es decir, el análisis de variables adicionales que pudiesen mejorar la capacidad predictiva de la medida.

Con el objetivo de proponer nuevos indicadores que permitan compensar las deficiencias de las métricas existentes, se han puesto en marcha algunas iniciativas como el proyecto AHELO (Assessment of Higher Education Learning Outcomes), dirigido por la Organización para la Cooperación y el Desarrollo Económicos (OCDE). Este proyecto, cuyo estudio de viabilidad se inició en 2008, persigue establecer una medida internacional de los resultados de aprendizaje en educación superior (Tremblay, Lalancette y Roseveare, 2012). A nivel nacional, sería aconsejable que las Administraciones Públicas adoptaran el conjunto de buenas prácticas llevadas a cabo por el Ministerio de Sanidad, Servicios Sociales e Igualdad, tanto en el proceso de recogida de información como en el análisis y publicación en abierto de los resultados de las pruebas selectivas que convocan. En este sentido, las Administraciones Públicas deberían incluir en el modelo 790 para la inscripción en pruebas selectivas, un campo para indicar la universidad y el país de expedición del título. La adición de este campo permite conocer el rendimiento de los egresados en estas pruebas selectivas e identificar las universidades con mayor impacto social en cada área de

conocimiento (Expósito-Casas, González-Benito, López-Martín y Moreno-Pulido, 2016). Asimismo, sería recomendable que los datos recogidos sean tratados y publicados en abierto, en un ejercicio de transparencia que al mismo tiempo haría posible el diseño de nuevos indicadores. Esta recomendación puede hacerse extensible al sector privado, donde las unidades de recursos humanos de las diferentes organizaciones deberían seleccionar, examinar y difundir la información relacionada con la procedencia de su plantilla, incidiendo especialmente en las nuevas incorporaciones.

Junto con la necesidad de recabar información de interés que permita incrementar la validez de los *rankings*, resulta oportuno destacar la conveniencia de replantear el modo en el que se construyen los mismos. Como se ha indicado anteriormente, en la construcción de instrumentos de medida es necesario, en primer lugar, definir con precisión el constructo a evaluar, en este caso, la calidad docente y, a continuación, asegurar la relevancia y la representatividad de los indicadores en los que dicho constructo se concreta. De esta forma, si se considera la calidad docente universitaria como la coherencia entre los resultados educativos alcanzados por los estudiantes y las condiciones o recursos educativos que proporciona la institución para alcanzar dichos resultados, su evaluación requerirá llevar a cabo una exhaustiva revisión de la teoría existente que facilite la definición y la operativización tanto de los procesos como de los productos educativos (la definición operativa de un modelo de evaluación de la calidad universitaria puede encontrarse en el trabajo de De la Orden *et al.* [1997]).

Sin embargo, en ocasiones, se observa cómo, lejos de ser la teoría la que guíe el proceso de construcción de los *rankings*, es la disponibilidad de los datos existentes la que condiciona la medida. Este hecho de medir lo que se puede, lo fácil, en lugar de lo que se debe, lo relevante, “no solo sería inútil, sino, en muchos casos también peligroso, dado que el poder de la evaluación para conformar su propio objeto podría

en tales circunstancias desviar los objetivos, procesos y productos de la Universidad” (De la Orden *et al.*, 1997: 2).

Conscientes de la dificultad que entraña la medida y evaluación de aspectos sociales, así como su fuerte repercusión en la toma de

decisiones, es preciso hacer una llamada a la responsabilidad pues, ante todo, se debe garantizar la idoneidad de las medidas y, de no ser posible, eludir su utilización, a fin de evitar los efectos adversos que su consideración pudiera tener en el proceso de toma de decisiones tanto individuales como colectivas.

Notas

¹ <http://www.umultirank.org#!/home?trackType=home>

² <http://dicits.ugr.es/rankinguniversidades/>

³ <http://www.rankingcyd.org/>

⁴ <http://www.iune.es/>

⁵ <http://www.u-ranking.es/index2.php#>

⁶ <http://www.u-ranking.es/construye.php>

⁷ Datos disponibles en: <http://sis.msssi.es/fse/>

Referencias bibliográficas

- Abad, F. J., Olea, J., Ponsoda, V. y García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- Alva-Chávez, K., Díaz-Varela Arrese, M., Franco-Romo, D. y Ballesteros-Carrasco, B. (2014). *Estudio de la posición de las universidades españolas en algunos rankings internacionales*. Recuperado de <http://universidad.es/sites/default/files/informe-ficue-rankings.pdf>
- American Educational Research Association (AERA), American Psychological Association (APA) y National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing*. Washington: APA.
- Anastasi, A. y Urbina, S. (1998). *Tests psicológicos*. México: Pearson Educación.
- Baladrón, J., Curbelo, J., Sánchez-Lasheras, F., Romeo-Ladrero, J. M., Villacampa, T. y Fernández-Somoano, A. (2016). El examen al examen MIR 2015: aproximación a la validez estructural a través de la teoría clásica de los tests. *FEM: Revista de la Fundación Educación Médica*, 19(4), 217-226.
- Buela-Casal, G., Quevedo-Blasco, R. y Guillén-Riquelme, A. (2015). Ranking 2013 de investigación de las universidades públicas españolas. *Psicothema*, 27(4), 317-326. doi: <http://dx.doi.org/10.7334/psicothema2015.140>
- Cizek, G. J. (2012). Defining and distinguishing validity: interpretations of score meaning and justifications of test use, *Psychological Methods*, 17(1), 31-43. doi: <http://dx.doi.org/10.1037/a0026975>
- Daraio, C. y Bonaccorsi, A. (2016). Beyond university rankings? Generating new indicators on universities by linking data in open platforms. *Journal of the Association for Information Science and Technology*. doi: <http://dx.doi.org/10.1002/asi.23679>
- Daraio, C., Bonaccorsi, A. y Simar, L. (2015). Rankings and university performance: A conditional multidimensional approach. *European Journal of Operational Research*, 244(3), 918-930. doi: <http://dx.doi.org/10.1016/j.ejor.2015.02.005>

- De la Orden, A., Asensio, I., Carballo, R., Fernández Díaz, M. J., Fuentes, A., García Ramos, J. M. y Guardia, S. (1997). Desarrollo y validación de un modelo de calidad universitaria como base para su evaluación. *RELIEVE*, 3(1), 1-27. Recuperado de: http://www.uv.es/RELIEVE/v3n1/RELIEVEv3n1_2.htm
- Delgado-López-Cózar, E. (2012). Cómo se cocinan los rankings universitarios. *Dendra Médica*, 11(1), 43-58.
- Docampo, D. (2015). Universidades españolas según el ranking de Shanghai. Recuperado de: http://dicits.ugr.es/rankinguniversidades/?page_id=671
- Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, 15(2), 315-321. Recuperado de: <http://www.psicothema.com/pdf/1063.pdf>
- Expósito-Casas, E., González-Benito, A., López-Martín, E. y Moreno-Pulido, A. (2016). Rendimiento de los titulados en Psicología en las pruebas de acceso al PIR [Performance of graduates in the Internship in Clinical Psychology entrance exams]. *Acción Psicológica*, 13(1), 145-158. doi: <http://dx.doi.org/10.5944/ap.13.1.17429>
- González-Riaño, M. G., Repiso, R. y Delgado-López-Cózar, E. (2014). Repercusión de los rankings universitarios en la prensa española. *Revista Española de Documentación Científica*, 37(3), e055. doi: <http://dx.doi.org/10.3989/redc.2014.3.1128>
- Grau F. X. (2013). ¿Cómo están las universidades españolas en los rankings? La visión académica: La preponderancia del impacto científico en los ránquings de universidades. En: V. Climent, F. Michavila y M. Ripollés (eds.). *Los rankings universitarios, mitos y realidades*. Madrid: Tecnos. Recuperado de: https://www.academia.edu/3391556/Los_R%C3%A1nkings_Universitarios_mitos_y_realidades
- Grau, F. X. (2015). *Rankings, impacto científico y sistemas universitarios*. Madrid: CRUE. Recuperado de: http://www.crue.org/Publicaciones/Documents/MONO_RANKING_WEB.pdf
- Hunsley, J. y Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. *Psychological Assessment*, 15(4), 446-455. doi: <http://dx.doi.org/10.1037/1040-3590.15.4.446>
- Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema*, 26(1), 127-137. doi: <http://dx.doi.org/10.7334/psicothema2013.258>
- López-Valcárcel, B. G., Ortún, V., Barber, P., Harris, J. E. y García, B. (2013). Ranking Spain's Medical Schools by their performance in the national residency examination. *Revista Clínica Española (English Edition)*, 213(9), 428-434.
- Mauri, T., Coll, C. y Onrubia, J. (2008). La evaluación de la calidad de los procesos de innovación docente universitaria. Una perspectiva constructivista. *Revista de Docencia Universitaria*, 1(1), 1-11.
- Orden SSI/1892/2015, de 10 de septiembre, por la que se aprueba la oferta de plazas y la convocatoria de pruebas selectivas 2015 para el acceso en el año 2016, a plazas de formación sanitaria especializada para Médicos, Farmacéuticos, Enfermeros y otros graduados/licenciados universitarios del ámbito de la Psicología, la Química, la Biología y la Física. *Boletín Oficial del Estado*, 224, 82031-82318. Recuperado de: http://boe.es/diario_boe/txt.php?id=BOE-A-2015-10053
- Orduña-Malea, E. (2012). *Propuesta de un modelo de análisis redinformétrico multinivel para el estudio sistémico de las universidades españolas (2010)* (tesis doctoral). Valencia: Universidad Politécnica de Valencia. Recuperado de: <https://riunet.upv.es/handle/10251/14420>
- Pérez, F., Aldás, J., Aragón, R. y Zaera, I. (2016). *U-Ranking 2016: indicadores sintéticos de las universidades españolas*. doi: http://dx.medra.org/10.12842/RANKINGS_SP_ISSUE_2016
- Ramírez Garzón, M. I. y Montoya Vargas, J. (2014). La evaluación de la calidad de la docencia en la universidad: Una revisión de la literatura. *REDU. Revista de Docencia Universitaria*, 12(2), 77-95.

- Rauhvargers, A. (2011). *Global University Rankings and their Impact*. Bruselas: European University Association. Recuperado de http://www.eua.be/Libraries/publications-homepage-list/Global_University_Rankings_and_Their_Impact.pdf?sfvrsn=4
- Rauhvargers, A. (2013). *Global university rankings and their impact: Report II*. Bruselas: European University Association. Recuperado de: http://www.eua.be/Libraries/publications-homepage-list/EUA_Global_University_Rankings_and_Their_Impact_-_Report_II
- Sentí, M., Pérez, J. y Baños, J. E. (2016). Factores predictores de resultados en la prueba MIR en las universidades públicas: análisis de la cohorte 2008-2014. *FEM: Revista de la Fundación Educación Médica*, 19(3), 155-160.
- Sireci, S. y Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100-107. doi: <http://dx.doi.org/10.7334/psicothema2013.256>
- Tomàs-Folch, M., Feixas-Condom, M., Ruiz-Ruiz, J. M. y Bernabeu-Tamayo, M. (2015). La literatura científica sobre rankings universitarios. *REDU, Revista de Docencia Universitaria*, 13(3), 33-54. Recuperado de: <http://red-u.net/redu/index.php/REDU/article/view/828>
- Tremblay, K., Lalancette, D. y Roseveare, D. (2012). *AHELO Feasibility Study Report. Volume 1- Design and Implementation*. Paris: OECD. Recuperado de: <http://www.oecd.org/education/skills-beyond-school/AHELOFSReportVolume1.pdf>
- Van Raan, A. F. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133-143.

Anexo I

TABLA 4. Indicadores y fuentes estadísticas U-Ranking 2016: Dimensión de docencia

Dimensión	Subdimensión	Indicador	Fuente	Periodo	Desagregación
DOCENCIA	Recursos (Peso: 25,4 %)	Profesor doctor (a tiempo completo) por cada cien alumnos matriculados en estudios de 1er y 2º ciclo y grado (centros propios), máster oficial y doctorado.	CRUE	2008-09, 2010-11, 2012-13 y 2013-14	Rama de enseñanza
		Presupuesto (ingresos liquidados de la Universidad) / Alumnos matriculados en estudios de 1er y 2º ciclo y grado (centros propios), máster oficial y doctorado.	CRUE	2008, 2010, 2012 y 2013	Universidad
		Profesor doctor (a tiempo completo) / Total del personal docente e investigador equivalente a tiempo completo.	CRUE	2008-09, 2010-11, 2012-13 y 2013-14	Universidad

TABLA 4. Indicadores y fuentes estadísticas U-Ranking 2016: Dimensión de docencia (cont.)

Dimensión	Subdimensión	Indicador	Fuente	Periodo	Desagregación	
DOCENCIA	Producción (Peso: 30,4 %)	Tasa de éxito: número de créditos aprobados (excluyendo adaptados, convalidados y reconocidos) respecto al total de créditos evaluados.	SIIU*	2009-10 a 2013-14	Rama de enseñanza	
		Tasa de evaluación: número de créditos evaluados respecto al total de créditos matriculados.	SIIU*	2009-10 a 2013-14	Rama de enseñanza	
		Tasa de abandono: alumnos matriculados en el curso t que, dos años después de haberse matriculado en el primer curso de una titulación, la abandonan sin graduarse sobre el total de alumnos matriculados en el curso t.	SIIU*	2009-10 a 2013-14	Rama de enseñanza	
			Índice de capacidad de atracción (capacidad de atracción de alumnos de otras provincias).	Indicador no considerado por motivos de disponibilidad o de calidad de la información		
	Calidad (Peso: 23,9 %)	% de estudiantes de posgrado: Alumnos matriculados en estudios de máster oficiales sobre el total de alumnos de Grado y 1º y 2º Ciclo y máster oficial.	SIIU	2008-09 a 2014-15	Rama de enseñanza	
		Notas de corte: nota de corte final de acceso a la titulación para el grupo general.	SIIU	2015-16	Grado	
		% de alumnos extranjeros: alumnos no españoles de 1er y 2º ciclo y grado y máster oficial respecto al total de alumnos de 1er y 2º ciclo y grado y máster oficial.	SIIU	2009-10 a 2013-14	Grado	
	Internacionalización (Peso: 20,3 %)	% de alumnos en programas de intercambio: alumnos españoles de 1er y 2º ciclo y grado que participan en el programa ERASMUS, sobre el total de alumnos de 1er y 2º ciclo y grado.	CRUE	2008-09, 2010-11, 2012-13 y 2013-14	Rama de enseñanza	
		% de alumnos matriculados en programas en lenguas no oficiales**	Indicador no considerado por motivos de disponibilidad o de calidad de la información			

Fuente: adaptado de Pérez et al. (2016: 69).

TABLA 2. Análisis descriptivo y de normalidad

		Porcentaje aprobados (Global)	Porcentaje aprobados 2012-13	Porcentaje aprobados 2013-14	Porcentaje aprobados 2014-15	Porcentaje aprobados 2015-16	Índice docencia	Índice (ponderación por defecto)
Enfermería	N	56	51	51	56	55	48	48
	Media	79,52%	74,88%	83,84%	77,52%	89,04%	1,008	1,021
	Desviación típica	13,47%	9,79%	8,50%	14,21%	7,25%	0,181	0,143
	Kolmogorov-Smirnov	0,164**	0,094	0,174**	0,188**	0,170**	0,139*	0,165**
Farmacia	N	19	15	16	18	19	17	17
	Media	85,38%	87,27%	78,31%	84,61%	89,11%	0,976	0,976
	Desviación típica	9,46%	12,52%	12,92%	10,32%	8,40%	0,171	0,125
	Kolmogorov-Smirnov	0,153	0,233*	0,274**	0,182	0,179	0,202	0,163
Medicina	N	34	28	29	34	34	34	34
	Media	90,69%	89,26%	85,32%	91,57%	92,72%	1,009	1,009
	Desviación típica	5,50%	5,17%	6,74%	5,56%	4,67%	0,173	0,131
	Kolmogorov-Smirnov	0,079	0,119	0,106	0,083	0,14	0,181**	0,233**
Psicología	N	40		37	36	40	38	38
	Media	81,66%		75,73%	85,31%	84,75%	1,005	1,011
	Desviación típica	12,72%		18,37%	10,26%	13,24%	0,203	0,131
	Kolmogorov-Smirnov	0,227**		0,246**	0,155*	0,193**	0,162*	0,179**

** La variable analizada no se distribuye normalmente a un nivel de confianza del 99% ($\alpha = 0,01$).

* La variable analizada no se distribuye normalmente a un nivel de confianza del 95% ($\alpha = 0,05$).

Abstract

The U-Ranking's predictive validity for postgraduate studies in the fields of Health Studies

INTRODUCTION. In recent years, university rankings have attracted significant media attention. The information –that these rankings provide about the university activities in relation to quality of teaching, research and innovation and technological development– has an important social influence and it affects the reputation and prestige of universities. In addition, these rankings increase the university attractiveness for students, conditioning the

decision-making process of potential students. This paper studies the predictive validity of the U-Ranking for graduates of the Faculties of Medicine, Pharmacy, Nursing and Psychology in Spain. **METHOD.** This paper analyzes how the scores obtained by different public and private Spanish universities may predict the level of performance achieved by their graduates in a national test for selecting health workers (using data from four calls 2012/13, 2013/14, 2014/15 and 2015/16). **RESULTS.** The results show that the position in the U-Ranking is not related to the universities' success rates in these selective tests. **DISCUSSION.** The marked trend of these results reveals a validity problem of the U-Ranking, which could be associated with a systematic error in predicting performance criteria through features that are not relevant in relation to it. Consequently, specific predictions to certain groups or entities might show a major benefit for definite units, producing significant inconsistencies with reality. Likewise, it is also important to rethink how the university rankings are built, based on a precise definition of the construct, in order to improve the quality of the measurement ensuring their suitability.

Keywords: *University rankings, U-Ranking; Predictive validity, Impact of training, Health sciences; Residency training program.*

Résumé

Validité prédictive du classement U-Ranking dans les études universitaires en Sciences de la Santé

INTRODUCTION. Au cours des dernières années le classement académique des universités a mérité une très grande attention de la part des médias. Les informations publiées sur les universités envers la qualité de ses activités d'enseignement, de sa recherche et de l'innovation et la technologie qu'elles produisent, elles ont toutes eu un grand impact social. Ce sont les informations qui affectent la réputation et le prestige des universités, et qui influencent le processus d'attraction des étudiants. Cette recherche étudie la validité prédictive du classement U-Ranking pour les études de Médecine, Pharmacie, Soins Infirmiers et Psychologie en Espagne. **MÉTHODE.** Cet article étudie comment les scores obtenus dans ce classement par les différentes universités espagnoles, aussi bien publiques que privées, peuvent prédire le niveau de performance atteint par leurs diplômés aux épreuves nationales de formation médicale spécialisée (en utilisant les données de quatre années différentes 2012/13, 2013/14, 2014/15 et 2015/16). **RÉSULTATS.** Les résultats montrent une absence de relation entre les scores que les universités obtiennent au classement U-Ranking et le pourcentage de réussite de leurs diplômés dans les épreuves de sélection des aspirants aux placements de formation médicale spécialisée. **DISCUSSION.** La tendance observée dans les résultats met en évidence un problème de validité intrinsèque du classement U-Ranking qui est associé à une erreur systématique dans la prévision d'un critère donné. L'article souligne qu'il est nécessaire de repenser la façon dont les classements universitaires sont construits, se fondant sur la base d'une définition précise du modèle d'évaluation en vue d'améliorer la qualité de la mesure en assurant son adéquation.

Mots clés: *Classements universitaires, U-Ranking, Validité prédictive, impact de la formation, Sciences de la santé, Epreuves sélectives du personnel de santé.*

Perfil profesional de los autores

Esther López-Martín (autora de contacto)

Profesora del Departamento de Métodos de Investigación y Diagnóstico en Educación II de la Universidad Nacional de Educación a Distancia (UNED). Licenciada en Pedagogía (Premio Extraordinario) por la Universidad Complutense de Madrid (UCM), doctora en Ciencias de la Educación con mención de “Doctor Europeo” (Premio Extraordinario) por esta misma universidad. Su campo de investigación se centra en la construcción y validación de instrumentos para la evaluación de sistemas educativos y la evaluación de la eficiencia y de la productividad.

Correo electrónico de contacto: estherlopez@edu.uned.es

Dirección para la correspondencia: Departamento de Métodos de Investigación y Diagnóstico en Educación II. C/ Juan del Rosal, 14. 28040 Madrid.

Alexis Moreno-Pulido

Licenciado en Historia y en Documentación por la Universidad de Salamanca. Actualmente trabaja en la Unidad de Apoyo a la Docencia y a la Investigación de la UNED. Sus principales líneas de investigación se centran en la evaluación de la actividad investigadora mediante indicadores bibliométricos y en las innovaciones docentes en la enseñanza de las competencias informacionales.

Correo electrónico de contacto: axmoreno@pas.uned.es

Eva Expósito-Casas

Profesora del Departamento de Métodos de Investigación y Diagnóstico en Educación II de la Universidad Nacional de Educación a Distancia (UNED). Licenciada en Pedagogía por la Universidad Complutense de Madrid (UCM), con mención de premio extraordinario de licenciatura. Doctora en Educación por la UCM con mención “Doctor Internacional”. Su campo de investigación se centra en la construcción y validación de instrumentos de medida y en el estudio de problemas metodológicos asociados a la evaluación educativa.

Correo electrónico de contacto: evaexpositocasas@edu.uned.es