

LOS RANKINGS DE UNIVERSIDADES: UNA VISIÓN CRÍTICA

Felipe Martínez Rizo*

REVISTA DE LA EDUCACIÓN SUPERIOR
ISSN: 0185-2760.
Vol. XL (1), No. 157,
Enero - Marzo de 2011, pp. 77-97

* Universidad Autónoma de Aguascalientes.
Departamento de Educación.
Correo e: felipemartinez.rizo@gmail.com

Ingreso: 05/11/10 • Aprobado: 14/12/10

Resumen

Después de repasar la historia de los *rankings* de universidades y describir la forma en que han proliferado, el trabajo presenta las características de tres de los más importantes en el plano internacional: el de Shanghai, el del *Times Higher Education Supplement* y el de las universidades en la *web*, que se contrastan con el sistema del *Center for Higher Education Development* alemán. Se presentan los criterios que tales ordenamientos deberían satisfacer que para ser considerados consistentes, se aplican a los ordenamientos en general y a dos de los mencionados en particular, y se concluye que sus limitaciones metodológicas son tan graves que no justifican su pretensión de ser confiables para evaluar a las universidades.

Palabras clave:

- Educación superior
- Universidades
- Evaluación

Abstract

After reviewing the history of university rankings and describing their growth, the characterization of three of the most important rankings internationally is presented, namely: Shanghai, *Times Higher Education Supplement* and the *Web of World Universities*, these are contrasted with the German Center for Higher Education Development system. Criteria for such rankings to be considered consistent is then presented, these criteria are applied to rankings in general and to the two above mentioned in particular. Thereupon it is concluded that both rankings show severe methodological limitations and thus are not able to justify their pretense of being reliable methods for evaluating universities.

Key words:

- Higher Education
- Universities
- Evaluation

Introducción

La evaluación puede ser un estímulo poderoso para el mejoramiento institucional, pero si no se hace bien puede llevar también a decisiones equivocadas e injustas y a consecuencias negativas. La evaluación no es un fin en sí misma. Tiene sentido en la medida en que ofrece información que, por su calidad y pertinencia, constituya un elemento que ayude a mejorar, sea por la retroalimentación que ofrece para reorientar esfuerzos (evaluación formativa) sea por su uso para asignar estímulos o sanciones, con la prudencia debida (evaluación sumativa). Estas consideraciones son especialmente importantes cuando a los resultados de la evaluación del desempeño o la calidad institucional se asocian decisiones fuertes, por ejemplo sobre financiamiento, como está ocurriendo con las evaluaciones que consisten en los llamados *rankings* de universidades.

Los ordenamientos de instituciones o programas universitarios en general (*rankings*)

Las publicaciones que presentan listas de instituciones de educación superior (IES), o programas supuestamente ordenadas por su calidad, forman parte del panorama actual de los sistemas de educación superior en muchos países. Como en muchos otros casos, el fenómeno se presentó con mucha anticipación en los Estados Unidos pero, al menos desde la década de 1990, se ha extendido por todo el mundo.

Al hablar de *rankings* u ordenamientos de instituciones utilizaré dos definiciones. La primera es la propuesta por el principal estudioso de la historia de este tipo de evaluaciones, David S. Webster, que se refería especialmente a las clasificaciones basadas en el prestigio institucional, y subrayaba el que cada institución tuviera un lugar distinto. Según Webster, para que constituya un *ranking* de calidad académica, una lista “debe estar ordenada según algún criterio o conjunto de criterios que los autores de la lista consideren que mide o refleja la calidad académica, y debe consistir en un listado de las mejores universidades, *colleges* o departamentos de cierta área de estudio, en orden numérico según su supuesta calidad, de manera que cada escuela o departamento tenga su propio lugar (rank) por sí solo, y no forme parte simplemente de un grupo con otras escuelas en unas cuantas categorías, grupos o niveles” (1986: 5).

La segunda definición, de Morrison, Magennis y Carey (1995), citados por Bowden, se refiere en particular a las tablas de posiciones de instituciones que se comenzaron a publicar a fines de la década de 1980 en el Reino Unido. Según esos autores, dichas tablas se basan en “combinaciones ponderadas de puntajes de indicadores de desempeño, en las que el puntaje global se utiliza para ordenar (rank) instituciones como escuelas, universidades y hospitales” (Bowden, 2000: 42).

En la actualidad, los *rankings* suelen utilizar una combinación de indicadores de carácter “objetivo”, muchas veces sobre insumos (profesorado, biblioteca, presupuesto) con opiniones “subjetivas” sobre el prestigio institucional, por lo que la segunda definición se aplica mejor en este sentido, pero de la primera conviene no perder de vista la idea de que un elemento fundamental de cualquier *ranking* es que cada institución o programa ocupe un lugar diferente, y no sea simplemente parte de un grupo.

Los *rankings* no son un fenómeno nuevo, aunque su proliferación sí lo es. La idea surgió hace más de un siglo en los Estados Unidos. Con antecedentes desde 1888, la idea de clasificar instituciones basada en alguna apreciación de su calidad, fue propuesta por el psicólogo James McKeen Cattell, en 1910, y se apoyaba en la afiliación institucional de científicos destacados (Webster, 1986). Según este autor, en 1925 Hughes publicó un *ranking* basado en opiniones de un grupo de expertos, metodología que volvió a utilizar, con ajustes, en otro ordenamiento difundido en 1934 (Webster, 1983). La tendencia se consolidó a partir de 1959 y, sobre todo, de la década de 1980, extendiéndose al nivel de pregrado. La primera edición de la guía *America's Best Colleges*, que publica la revista *US News and World Report*, aparece anualmente desde 1983. *Time* y *Newsweek* comenzaron a publicar listas semejantes en 1996. Según Rachel Bowden, “una agencia que asesora a las instituciones en relación con los rankings ha enlistado 52 publicaciones americanas diferentes que publican listas con ordenamientos de universidades”. (2000).

Las listas de *US News* distinguen a las instituciones según su ubicación geográfica y tipo, con base en la clasificación de la *Fundación Carnegie* que, además de Universidades de Investigación tipo I y tipo II, con importante proporción de alumnos de posgrado, en el nivel de pregrado distingue instituciones nacionales, regionales, y de artes liberales.

El lugar que ocupa cada institución depende de su situación en cuanto a indicadores de recursos financieros y humanos (profesorado), de la selectividad de su alumnado, sus tasas de retención y graduación, los donativos que recibe de sus ex alumnos, así como su prestigio académico. El peso de cada indicador varía según el tipo de institución al que se refiera, pero el prestigio, que se basa en opiniones personales, suele representar un 25% del peso total.

En el Reino Unido el *ranking* más antiguo fue difundido en 1992 por *The Times*, que lo hace anualmente desde entonces. A partir de 1998 comenzaron a aparecer varios productos del mismo tipo en otros periódicos (*Financial Times*, *Sunday Times*) así como en libros especiales, como *The PUSH Guide to Which University*, *NatWest Guide* y *Virgin Alternative Guide* (Bowden, 2000).

La autora de la que se toma la información anterior presenta el caso de un *ranking* alternativo que se difundió únicamente vía internet, en 1998: *Red Mole*. Este *ranking* no se basa en indicadores de recursos o prestigio entre académicos o empleadores, sino en la opinión de los alumnos mismos, recogida mediante una encuesta voluntaria, igualmente por internet, en la que se les preguntó por la calidad de la vida nocturna, los dormitorios, los maestros y las instalaciones deportivas.

Aunque el número de los que respondieron fue reducido (alrededor de 1% del total), y lo hicieron en forma voluntaria, por lo que existe autoselección, el resultado no puede menos que llamar la atención, ya que las instituciones que resultan clasificadas en los primeros lugares no son las que aparecen en esos sitios en todas las otras clasificaciones, que en este caso ocuparon lugares bastante inferiores: *Cambridge* el 42°, *Oxford* el 35° y el *Imperial College* el 31°. (Bowden, 2000).

Los *rankings* que publica en Canadá la revista *Maclean's* comenzaron a aparecer en 1990. En este caso se distinguen las instituciones que ofrecen principalmente estudios de doctorado, incluyendo las escuelas de medicina, las que ofrecen tanto estudios de pregrado como de posgrado (*comprehensive*), y las que se concentran básicamente en el nivel de pregrado (*undergraduate*).

Se presentan ordenamientos separados para cada una de estas tres categorías, con base en indicadores que tienen que ver con los resultados académicos de los estudiantes; la *ratio* de alumnos por profesor, en general o sólo con definitividad (*tenured*); la calidad del profesorado; el presupuesto total, y la parte que se dedica a servicios estudiantiles y becas; la calidad de la biblioteca; y el prestigio entre egresados, académicos y empresarios.

Según un recuento reciente, es muy grande el número de países en los que se hacen *rankings* de instituciones de educación superior, sea por parte de algunos medios de comunicación, promovidos por los ministerios de educación o desarrollados por algunas universidades. Además de los países anglosajones ya considerados, a los que debe añadirse Australia, la lista incluye en Europa a Alemania, Eslovaquia, España, Francia, Italia, los Países Bajos, Polonia, Portugal, Rumania, Rusia, Suecia, Suiza y Ucrania; en Asia a China y Hong Kong, Corea del Sur, Japón, Kazajstán, la India, Malasia, Pakistán y Tailandia, y en África a Nigeria y Túnez (Marginson, 2010: 546).

El fenómeno se presenta también en América Latina, donde periódicos importantes de Argentina, Brasil, Chile, Colombia y otros países difunden *rankings*. En México lo han hecho la revista *Selecciones* y los diarios *Reforma* y *El Universal*. La Dirección General de Evaluación Institucional de la Universidad Nacional Autónoma de México, además de analizar algunos *rankings* internacionales (Ordorika y Rodríguez, 2008), desarrolla un sistema propio para clasificar las instituciones mexicanas según su producción de investigación, el *Estudio Comparativo de las Universidades Mexicanas* (ECUM) (Ordorika et al., 2009; Márquez, 2010).

Los *rankings* internacionales

Los *rankings* internacionales de universidades también tienen antecedentes relativamente antiguos, aunque los que hoy atraen mucho la atención tienen menos de diez años. Entre 1967 y 1983 Jack Gourman publicó listados en los que pretendía clasificar según su calidad centenares de programas de unas 1,500 universidades estadounidenses y 700 universidades de todo el mundo. Estos *rankings*, sin embargo, no explicaban la metodología

utilizada y presentaban rasgos extraños, como distinguir diferencias mínimas, de centésimas de punto, al calificar instituciones muy diversas.

En un caso se asignaba una puntuación de 4.73 a La Sorbona, 4.72 a Oxford y 4.71 a la Universidad Lomonosov de Moscú. Es inverosímil la circunstancia de que se repitiera generalmente un patrón similar, según el cual decenas de instituciones se ordenaban rigurosamente a la distancia de una centésima de punto, sin ningún empate o hueco (Webster, 1985; citado en Martínez Rizo, 1992: 43).

En la primera década del siglo XXI los rankings internacionales de universidades adquirieron una visibilidad sin precedentes, seguramente propiciada por la facilidad de su difusión vía internet, en particular en el caso de los tres que se presentan a continuación, el de Shanghai, el del *Times Higher Education Supplement*, y el de las universidades en la web.

El ranking de Shanghai

A partir de 2003, el Instituto de Educación Superior de la Universidad Jiao Tong, de Shanghai, publica anualmente el llamado *Academic Ranking of World Universities*. Este trabajo recibe gran atención por parte de los medios de comunicación del mundo, con crecientes repercusiones en lo relativo a toma de decisiones por parte de dirigentes universitarios y ministerios de educación, así como de la clase política y el público. Pese a las severas críticas que se le han hecho, su importancia parece aumentar, influyendo en las políticas educativas de algunos países al grado de que una meta importante llega a ser tener cierto número de las IES de un país en los lugares más altos del ranking.

Cuadro 1
Dominios y criterios del ranking de Shanghai

Dominio	Criterio y peso	Observaciones
Calidad de la educación	Egresados con Premio Nobel o Medalla Fields (10%)	De pregrado o posgrado; si uno estudió en varias IES todas reciben puntos. Más peso recientes
Calidad del profesorado	Académicos con Premio Nobel o Medalla Fields (20%)	Si uno trabaja en varias IES todas reciben puntos. Más peso recientes. No claro qué es "trabajar en"
	Académicos muy citados (20%)	De lista de 250 más citados en c/u de 21 áreas de Thomson Scientific
Productividad de investigación	Papers Nature y Science (20%)	5 últimos años; no cuenta para IES de ciencias sociales; peso según rango de autor
	Total papers publicados (20%)	En la base de datos de Thomson Scientific; se da doble peso a los de ciencias sociales
Productividad	Total de los 5 criterios previos entre profesores FTE (10%)	FTE = Full Time Equivalent. Si no se tiene el dato este criterio se ignora

Fuente: Billaut *et al.*: 2009

Para una descripción más detallada del *Academic Ranking of World Universities* pueden verse los trabajos de Billaut *et al.*, 2009 y de Ordorika, Rodríguez *et al.*, 2008.

El World University Ranking del Times Higher Education Supplement (THE)

En un artículo en el que presenta los cambios más recientes que se han hecho a la metodología de este ranking, Phil Baty plantea, en forma retórica, una pregunta que luego responde de dos formas:

¿Cómo medir algo tan intangible como la calidad de una Universidad? La respuesta corta, por supuesto, es que no se puede. Lo que se puede hacer, sin embargo, lo que tratamos de hacer con estos *rankings*, es tratar de capturar los elementos más tangibles y medibles que hacen una universidad moderna de clase mundial (Baty, 2009).

Baty añade:

En 2004, cuando el THE concibió su primer *World University Ranking* junto con la firma Quacquarelli Symonds (QS), identificamos cuatro “pilares” que constituyen el fundamento de cualquier institución líder internacionalmente, y que son poco discutibles: investigación de gran calidad, docencia de gran calidad, destacado acceso al empleo de los egresados y enfoque internacional. Mucho más discutibles son las mediciones usadas en nuestros *rankings* y el balance entre mediciones cuantitativas y cualitativas (Baty, 2009).

Las dimensiones del concepto de calidad del ranking del THE, la forma de medirlas y el peso que recibe cada una en el puntaje total utilizado para hacer los ordenamientos se sintetizan como sigue:

Cuadro 2
Factores del ranking del Times Higher Supplement 2004-2009

Factor	Descripción	Peso	Observaciones
Investigación de excelencia	Citas de artículos publicados por personal de la institución entre total de personas	20%	Según bases de datos de Scopus-Elsevier
Docencia de excelencia	Ratio de estudiantes por profesor	20%	Pobre pero disponible
Enfoque internacional	% de profesorado de otros países	5%	Según datos de cada institución
	% de alumnado de otros países	5%	
Acceso a empleo egresados	Opiniones de académicos según encuesta	40%	Cualitativos; valen 50% del total
	Idem de empleadores de egresados	10%	

Fuente: Baty, 2009

Las cuatro dimensiones y sus seis indicadores se usaron en el *ranking* del THE desde 2004 hasta 2009. Las listas incluyen instituciones que tengan estudiantes de pregrado y posgrado y cubran al menos dos de cinco áreas: ciencias naturales; de la vida y la salud; ingenierías y tecnologías de la información; ciencias sociales; artes y humanidades. Por las limitaciones de los indicadores citados, en noviembre de 2009 el THE anunció una importante modificación de la metodología para la construcción de su *ranking*, además de que ya no trabajará con la empresa QS (que seguirá publicando sus listados como *QS World University Rankings*), sino con Thomson Reuters, por lo que usará las bases de ISI Thomson para información sobre los indicadores de productividad de investigación. El THE anunció que consultó a un grupo de expertos en educación superior, así como a unos 40 rectores, sobre su nueva metodología. Los factores considerados en la etapa preparatoria aparecen en el cuadro siguiente, y se anuncia que los definitivos se difundirán antes de que se publique el primer ejercicio basado en la nueva metodología, en el otoño de 2010 (Baty, 2010).

Cuadro 3
Factores del *ranking* del Times Higher Education Supplement a partir de 2010

Factor	Descripción	Peso
Actividad económica/innovación	Ingresos p/investigación provenientes de industria/académicos	10%
Diversidad internacional	<i>Ratio</i> alumnos internacionales/locales	10%
	<i>Ratio</i> académicos internacionales/locales	
Indicadores institucionales	Nuevo ingreso pregrado/académicos	25%
	Graduados de pregrado/doctorados	
	Graduados de doctorado	
	Prestigio de docencia según encuesta	
Indicadores investigación	Presupuesto institucional/alumnado	55%
	Artículos publicados/académicos	
	Impacto según citas normalizado por área	
	Presupuesto de investigación/investigadores	
	Presupuesto investigación de fuente externa gobierno-industria	
	Prestigio de investigación según encuesta	

Fuente: Baty, 2010

El ranking mundial de universidades en la web

Estas listas son preparadas y difundidas por el Laboratorio de Cibermetría del Consejo Superior de Investigaciones Científicas de España, y se basan en la idea de que la importancia que ha adquirido la internet es tal que se puede apreciar qué tan buena es una universidad analizando la presencia que tiene en la web la producción de sus académicos y la frecuencia con la que se consultan los productos respectivos y, en general, la página institucional. En consecuencia, la unidad de análisis en este caso es *el dominio web institucional*, por lo que sólo aquellas universidades y centros de investigación con un *dominio web independiente* son considerados. Se diseñaron cuatro indicadores a partir de los resultados cuantitativos obtenidos de los principales motores de búsqueda (Google, Yahoo, Live Search y Exalead).

Cuadro 4
Factores del ranking mundial de universidades en la web

Factor	Descripción	Peso
Visibilidad (<i>Visibility</i>)	Total de enlaces externos únicos recibidos por un sitio (<i>inlinks</i>).	50%
Tamaño (<i>Size</i>)	Número de páginas recuperadas.	20%
Archivos Ricos (<i>Rich Files</i>)	<i>Acrobat</i> y <i>PostScript</i> (.pdf y .ps), <i>Word</i> y <i>Powerpoint</i> (.doc y .ppt).	15%
Académico (<i>Scholar</i>)	Artículos y documentos académicos y citas en <i>Google Scholar</i> .	15%

Fuente: Página Web del Laboratorio de Cibermetría del csic.

El Factor de Impacto Web (WIF por sus siglas en inglés) combina el número de enlaces externos entrantes con el número de páginas web de un dominio, siguiendo una relación 1:1 entre visibilidad y tamaño. Esta relación se usa para hacer el *ranking*, añadiendo los otros dos indicadores: el número de *archivos ricos* que contiene un dominio web, y el número de publicaciones incluidas en la base de datos de *Google Scholar*. Por su relevancia en las actividades académicas y de publicación y su volumen de uso, se consideran *archivos ricos* los mencionados en el cuadro 4, con terminaciones .pdf, .ps, .doc y .ppt. La información fue tomada de la página web de este sistema (Aguillo *et al.*, 2006 y 2008).

Los rankings del CHE

El Center for Higher Education Development se define como “un grupo de trabajo para la reforma de la educación superior” (*a reform think tank for higher education*). Fue establecido en mayo de 1994, con recursos de la Fundación Bertelsmann y apoyo de la Conferencia de Rectores Alemanes, y tiene su sede en Gütersloh, en el estado federal alemán del Norte del Rin y Westfalia.

El principal producto del CHE es un *ranking* de concepción muy diferente de los anteriores, el *CHE Hochschul Ranking* o *CHE University Ranking*. Las principales características que lo hacen distinto se sintetizan a continuación, con información tomada de la página del CHE (<http://ranking.zeit.de>).

El punto central es que no se trata de un ordenamiento global simple sino de un análisis detallado, ya que se evita expresamente agregar datos parciales para producir un puntaje global. Esta decisión se basa en la convicción de que no existe institución alguna de educación superior que se pueda considerar sencillamente *la mejor*, ya que una puede ser más fuerte en unos aspectos o campos y otra en otros. *En lugar de coronar a un presunto ganador universal, ofrecemos un ordenamiento multidimensional.*

Relacionada con lo anterior está la decisión de no asignar lugares singulares a cada institución, sino sólo situar a cada una en un grupo: alto, medio o bajo. Se evita así el problema de otros *rankings*, de que el lugar de una institución cambia mucho de un año a otro, *por fluctuaciones azarosas que se interpretan como si fueran diferencias reales*. Las diferencias entre instituciones de uno de los tres grupos y de los otros dos son suficientemente grandes para ser significativas, mientras que no lo son dentro de cada grupo, por lo que en cada uno de estos se presentan en orden alfabético.

También debido a la característica mencionada en primer término, los ordenamientos del CHE son únicamente específicos por área, y no se da un ordenamiento global, ya que la calidad de las instituciones en disciplinas particulares varía mucho, y un *ranking* agregado no ofrece información útil para orientar a los futuros estudiantes ni toma en cuenta decisiones institucionales de cultivar ciertas áreas con especial cuidado.

Otro rasgo es la diversidad metodológica, ya que se utiliza información de diferentes fuentes y perspectivas, incluyendo datos sobre departamentos y programas, pero también opiniones de alumnos, maestros y egresados, lo que permite contrastar las opiniones de diferentes actores y también los indicadores basados en datos objetivos.

El sistema considera 34 áreas de estudios en el nivel universitario, de Administración a Trabajo Social, pasando por Arquitectura, Biología, Ciencias del Deporte, Computación, Comunicación, Derecho, Economía, Educación, Enfermería, Física, Geología, Historia, Ingeniería Eléctrica, Industrial y Mecánica, Matemáticas, Mecatrónica, Medicina, Psicología, Química o Sociología.

Con los datos sobre diversos aspectos que se obtienen se construyen alrededor de 30 indicadores numéricos que pueden usarse para ordenar los programas, y mucha información adicional que se ofrece a los interesados sin que se use para los ordenamientos. Los criterios generales que se manejan son nueve; cada uno incluye diversos aspectos particulares, que varían según el área de que se trate. En seguida se dan ejemplos de algunos y se señala el número total aproximado.

- Mercado laboral y orientación de la carrera (20): relación teoría-práctica, preparación para el trabajo, prácticas profesionales...

- Equipamiento (18): instalaciones, laboratorios, computadoras, biblioteca...
- Investigación (13): publicaciones, citas...
- Opiniones sobre calidad y prestigio del programa (4): de alumnos, profesores y egresados...
- Orientación internacional (13) estudiantes extranjeros, exigencias de idiomas...
- Resultados (10): graduados al año, tiempo para graduarse, calificaciones promedio...
- Ciudad y universidad (26): diversidad de opciones de la institución, servicios que ofrece la ciudad en que se sitúa...
- Alumnos (6): requisitos de ingreso, proporción de admitidos, de mujeres...
- Apoyo académico y docencia (37): tamaño de los grupos, *ratios* alumnos/profesor, métodos de enseñanza, asesoría individualizada, evaluación de la docencia...

Inicialmente el *ranking* del CHE consideró solamente instituciones y programas de Alemania; luego se ha extendido a países de lengua germana vecinos, en particular Austria y Suiza. A partir de 1998 se han incluido más de 130 universidades y más de 150 universidades de ciencias aplicadas (Fachhochschulen). En la última versión se consideraron las opiniones de unos 200,000 estudiantes y 15,000 profesores (<http://ranking.zeit.de>).

Los resultados del trabajo del CHE son difundidos por el periódico *Die Zeit*, pero no en la forma de listados como en los otros *rankings*, sino en forma consistente con lo expresado antes, mediante un sistema interactivo en línea que permite a cada usuario (futuros estudiantes u otras personas) personalizar los criterios para ordenar las opciones con base en ellos (Marginson, 2010: 547).

A partir de esta experiencia el CHE, con el Center for Higher Education Policy Studies de la Universidad de Twente (Países Bajos) y otras, está desarrollando un *ranking* con un concepto similar y mayor alcance internacional, el *U-Multirank*, que se encuentra en fase piloto en alrededor de 150 instituciones de Estados Unidos y Canadá, Australia, Asia, África y América Latina.

Hacia una valoración de los *rankings*

En algunos casos los *rankings* son producidos por instituciones especializadas, en relación con un medio de comunicación, como el *ranking* del THE y el del CHE con *Die Zeit*. Otras veces los hacen instituciones académicas, como el *ranking* de Shanghai o el de universidades en la web del Laboratorio de Cibermetría del Consejo Superior de Investigación Científica de España.

En muchos casos, sin embargo, son producidos sin apoyo calificado por medios de comunicación, que aducen como motivación su pretensión de que contribuyan a que el mercado de la educación superior funcione con más información, dando elementos a padres de familia y futuros estudiantes para escoger una institución o un programa. También es clara la intención de aumentar la venta del periódico o revista de que se trate, dado el interés

que suscitan los *rankings*. De hecho las publicaciones que los difunden consiguen jugosas ventas cada año, al llegar la temporada en que los jóvenes que terminan la enseñanza media deben decidir a cuál o cuáles instituciones y programas dirigirse para buscar ser admitidos, en la que para muchos es la primera decisión de repercusiones importantes para la vida.

Pero por encima de posibles intereses más o menos legítimos, las preguntas fundamentales a responder para tener una opinión sobre los ordenamientos se debe referir a su propia calidad, ya que los juicios al respecto serán muy diferentes si la información que difunden es buena o no.

Como toda evaluación, un *ranking* implica identificar ciertos *evaluandos*; definir en qué consiste su calidad y precisar el referente con que se contrastará la situación de los *evaluandos*; operacionalizar el concepto de calidad precisando dimensiones e indicadores y medirlos confiablemente; contrastar el resultado de la medición con el referente y formular un juicio de valor ponderado. En concreto, para que un *ranking* se pueda considerar de buena calidad, deberá atender los siguientes puntos:

- Identificación clara de los *evaluandos*: las instituciones, unidades o programas a evaluar.
- Definición precisa de la noción de *calidad* que se espera tengan los *evaluandos* para que se les pueda considerar buenos, precisando sus dimensiones y los indicadores de cada dimensión, para que sea posible apreciar si los *rankings* tienen *validez*.
- Calidad de la información que se utilice para cada indicador, identificando las fuentes y formas en que se obtiene, para poder apreciar si las mediciones tienen *confiabilidad* y si, con base en ellas, se pueden comparar de manera consistente diferentes *evaluandos*.
- Justificación sólida de la forma en que se agregarán las mediciones de los indicadores de modo que se generen puntajes en una sola escala en que se ubicarán los *evaluandos*, para apreciar si se justifica el tratamiento de la calidad como unidimensional y si se hace de manera consistente.
- Suficiente *precisión de las mediciones*, de manera que sea posible apreciar si la diferencia que distingue a unos *evaluandos* de otros es o no significativa y consistente, de manera que tenga sentido hacer *rankings* en los que se distingue uno a uno el lugar que ocupan los *evaluandos*.
- Identificación de los *estándares de referencia*, en comparación con los que se juzgará si un *evaluando* debe o no ser considerado de buena calidad.
- Consistencia de los resultados y las tendencias resultantes.
- Correcta formulación de los juicios a que se llega por el hecho mismo de hacer un *ranking*, cuidando que se tenga en cuenta el contexto de las instituciones que se comparan para cuidar la equidad, y cómo se tienen en cuenta las consecuencias, previstas o no, de los *rankings*.

Los siete primeros puntos de la enumeración anterior coinciden en tener carácter técnico, porque si los ordenamientos están mal hechos, de manera que miden mal lo que pretenden evaluar, o miden algo diferente, por muy buena voluntad que se tenga los resultados no podrán sustentar sólidamente usos adecuados.

Pero aunque las mediciones estén bien hechas, los resultados se pueden usar bien o mal, por lo que es importante considerar el punto incluido en octavo y último lugar, que tiene que ver con la contextualización de los juicios y los usos de los resultados.

La elección de estos ocho criterios para juzgar la calidad de un *ranking* se justifica por una parte con base en la literatura especializada que, de diferentes maneras, coincide en identificar como dimensiones de toda buena evaluación la validez y confiabilidad de las mediciones en que se base, su unidimensionalidad para justificar agregar datos de diferentes indicadores, su consistencia en el tiempo y su precisión (por ejemplo, Bird, 2005; Goldstein y Spiegelhalter 1996; Shavelson, McDonell y Oakes, 1989); por otra parte, el autor, en el marco del desarrollo de un sistema de indicadores por parte del Instituto Nacional para la Evaluación de la Educación, ha sistematizado también estos criterios (Martínez Rizo, 2005; 2007a y 2007b).

En los párrafos siguientes los ocho criterios anteriores se aplican a los *rankings* en general.

Definición del objeto a evaluar

Los problemas de los *rankings* de instituciones completas comienzan con la definición de su objeto. Las universidades son grandes y complejas, y su calidad es eminentemente multidimensional. Tienen fortalezas y debilidades; pueden sobresalir en investigación y ser débiles en docencia; junto a grupos de investigación muy productivos puede haber otros mediocres; junto a programas con graduados de excelente nivel, otros anquilosados que dan una formación pobre y obsoleta.

Concepto de calidad subyacente, dimensiones e indicadores

Suponiendo bien definido un *evaluando*, los puntos siguientes se refieren al concepto de calidad, los indicadores que lo operacionalizan y la información que los soporta. Es usual que haya información sobre unos aspectos de la calidad y no sobre otros y, muchas veces, la más abundante se refiere a aspectos menos importantes, en tanto que la relacionada con aspectos fundamentales es escasa o de calidad dudosa. Eso ocurre en no pocos *rankings*.

Calidad de la información utilizada y de las mediciones

Al valorar la calidad de un *ranking* se pueden distinguir dos preguntas: una, si se basa en datos objetivos o en opiniones; otra, en uno y otro caso, si la información es sólida. Los *rankings* más antiguos se basaban en la calidad del personal académico que es, sin duda, un elemento central para la calidad de una institución o programa, de docencia o de investigación; pero como no se contaba con información objetiva al respecto, los logros del profesorado o los investigadores se valoraban con base en la opinión que tenían unos de otros

los académicos de las instituciones evaluadas, por lo que esos ordenamientos se denominaban “clasificaciones de prestigio” (*reputational rankings*).

Las limitaciones de tal acercamiento son evidentes. Incluso en un sistema de educación superior en el que hay tanta movilidad, información y comunicación como el de Estados Unidos, son obvios los riesgos de alta subjetividad, falta de referentes y “efectos de halo” en la opinión de profesores, empresarios u otras personas sobre la calidad de instituciones o programas. Webster menciona el caso de un *ranking* de escuelas de pregrado en el área de negocios, en el que algunos expertos mencionaron entre las mejores universidades a Harvard, Yale y otras altamente reconocidas, pero que NO tienen pregrado en dicha área. En México la Universidad Nacional Autónoma de México (UNAM) y el Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) suelen ser mencionados como *de excelencia* en estudios basados en opiniones, pero es claro que, más allá de la calidad real de esas instituciones y de cada una de sus áreas, esa reiterada mención tiene mucho que ver con que esas instituciones tienen una visibilidad mucho mayor que otras. Es sintomático que instituciones como el Centro de Investigaciones y Estudios Avanzados (CINVESTAV) o El Colegio de México son poco mencionadas en esos estudios, aunque algunas de sus áreas son muy buenas no sólo en el plano nacional, sino internacionalmente.

Actualmente, como se ha visto, los *rankings* tienden a combinar las opiniones subjetivas sobre el prestigio con algunos datos objetivos, en general sobre recursos o insumos institucionales que se pueden obtener con facilidad, como la *ratio* de alumnos por profesor; la proporción de profesores de tiempo o de asignatura; de personal con doctorado o sin él; el número de volúmenes en biblioteca; o el puntaje mínimo que un aspirante debe obtener en la prueba de ingreso para ser admitido.

Estos datos de carácter más objetivo parecen más sólidos que los que se basan en opiniones subjetivas, pero tienen el problema de que no miden la calidad de la institución o programa al que se refieran, sino sólo su riqueza, su dotación de recursos. Sin duda hay cierta asociación entre la abundancia o escasez de recursos y la calidad de una institución o programa, pero esa relación no es unívoca ni lineal. Hay instituciones bien dotadas pero mediocres o francamente malas, y las hay modestas y excelentes, y para distinguir unas de otras no basta medir insumos. En el caso de la docencia, por ejemplo, para saber si los recursos institucionales, pocos o muchos, son usados eficientemente, serían necesarios datos confiables sobre el nivel de competencia de los egresados, después de descontar el nivel que tenían al entrar a la institución, y teniendo en cuenta la deserción. Todos estos datos no están disponibles actualmente, desde luego, ni es probable que lo estén en una escala que rebase la de unos cuantos proyectos de investigación bastante complejos.

Justificación de la agregación de factores y forma de hacerlo

Un principio metodológico básico es que se debe verificar la dimensionalidad de las mediciones que se refieran a cualquier constructo latente, como lo es la calidad. La agregación de datos de varios indicadores sólo se justifica en el caso de que se refieran a una misma dimensión, y en ese caso las formas de hacerlo son varias, todas con pros y contras, pero ninguna tan simple e ingenua como la que se maneja en muchos *rankings*, con el agravante de que frecuentemente se agregan indicadores de dimensiones claramente distintas, como la calidad de la investigación y la de la docencia, o la de diferentes unidades dentro de una misma institución.

Precisión de las mediciones

La revisión de los *rankings* muestra que sus autores no suelen tener una buena formación en medición y que tampoco buscan asesoría al respecto. Por ello no sorprende que los procedimientos para cuantificar cada uno de los indicadores, y los que se utilizan para agregar los resultados parciales y construir la escala única en que se basa el ordenamiento, se hagan sin respetar reglas metodológicas básicas. Esta falla se refleja, *a fortiori*, en la dudosa precisión de las mediciones, evidenciada en que no se ofrece justificación alguna con base en la cual se pueda defender que el que una institución tenga una puntuación menor a otra, sin importar la diferencia, implique una correspondiente diferencia de calidad. No se conoce el margen de error que inevitablemente tienen los resultados y, en consecuencia, es imposible valorar el significado de las diferencias.

Definición de estándares de referencia

Lógicamente no suele contarse con este tipo de definición, indispensable para que los juicios de valor a los que se llega no sean solamente relativos, sino absolutos. Los *rankings* sólo permiten llegar a juicios relativos en el sentido de que afirman que un evaluando *es mejor o peor* que tal otro o tales otros aunque sin saber, como se ha señalado ya, si es significativamente mejor o peor, o sólo ligeramente, tanto que la diferencia no es significativa.

Pero mejor que algo/alguien no es necesariamente positivo, ni peor que algo/alguien forzosamente negativo. Ninguna universidad considerará malo ser un poco peor –o menos buena– que Harvard, ni verá con buenos ojos ser un poco mejor que una institución claramente deficiente. Todo depende del punto de comparación, que no debería ser simplemente la situación de los demás *evaluados*, sino algún parámetro más sólido, del que los *rankings* típicamente carecen (Orpee, 2010).

Consistencia de resultados y tendencias

A consecuencia de las deficiencias a las que se alude en los puntos anteriores, los *rankings* de universidades y programas suelen presentar cambios muy importantes en el lugar que ocupan algunas de los evaluados, sin que haya fundamento para pensar que esos cambios reflejan modificaciones reales de la calidad. Este tipo de instituciones no cambian fácilmente de un año a otro, a no ser que ocurran catástrofes o reformas espectaculares. Por ello la existencia de cambios bruscos debe considerarse un claro indicio de la falta de consistencia de los ordenamientos.

Contextualización y consecuencias: los usos de los resultados

A consecuencia de su fuerte difusión, es cada vez más frecuente que se asocien decisiones fuertes a los resultados de los *rankings*. Por otra parte, las mediciones en que se basan suelen carecer de las cualidades de validez, confiabilidad y comparabilidad que son indispensables para sustentar juicios de valor objetivos sobre la calidad de las instituciones evaluadas. La conclusión es que el uso de este tipo de ordenamientos llevará fácilmente a cometer injusticias graves, al no identificar verdaderamente las instituciones o programas que sobresalen realmente por su calidad y merecen estímulo; los que, pese a no tener aún resultados excelentes, hacen esfuerzos meritorios en situación desfavorable y necesitan apoyo para despegar; y los que manifiestan fallas graves reiteradas e injustificadas, que justificarían medidas correctivas.

Una apreciación de los *rankings* de Shanghai y del *Times Higher Education Supplement*

Estos *rankings* son ejemplos de metodología pobre, que se combina con fuerte impacto mediático, lo que lleva a usos que no pueden sustentar decisiones que lleven a mejoras reales de la calidad. Un repaso de los criterios metodológicos mencionados permitirá apreciar que los dos *rankings* más visibles no cumplen con ellos.

No definen bien los objetos a evaluar: Ya se ha dicho que una universidad es un todo complejo, cuya valoración no parece razonable reducir a un lugar único en un listado unidimensional. Pero además, en el caso de los *rankings* de Shanghai y el THE, no es claro qué debe considerarse una universidad ni cómo se elabora el listado de las que se consideran en el *ranking*. ¿Cómo saber si Berkeley y la UCLA, con otras ocho instituciones, que forman el sistema de la Universidad de California, deben considerarse por separado o como un todo? La misma pregunta procede en cuanto al tratamiento de las facultades de estudios profesionales de la UNAM, las unidades de la Autónoma Metropolitana y la Universidad de Guadalajara o los campus del Tecnológico de Monterrey, o las universidades en que se dividió la Universidad de París. Por lo que toca a las instituciones mexicanas es interesante señalar que en las en-

cuestas de opinión entre académicos y empleadores que constituían el 50% del peso total del *ranking* del THE únicamente se preguntaba a las personas consultadas sobre la UNAM, el ITESM y la Universidad Iberoamericana. Ninguna otra universidad podía ser considerada por los consultados.

Supone un concepto de calidad parcial y operacionalizado en forma sesgada: Al parecer se entiende por “universidad de clase mundial” la que produzca más investigación, pero dado que ésta es sólo una de tres o cuatro funciones básicas de las instituciones de educación superior, es legítimo preguntarse por qué se privilegia esa en el grado en que lo hace el *ranking* en cuestión. Por otra parte, las dimensiones y los indicadores que se utilizan tienen obviamente sesgos muy fuertes a favor de ciertos campos del conocimiento y ciertas regiones geográficas y lingüísticas.

Maneja información de baja calidad y mide mal: Aunque todos los indicadores utilizados son de tipo objetivo, las mediciones son muy pobres, por las limitaciones de las bases de datos de las que se toma la información, así como por decisiones no justificadas sobre la manera de atribuir la producción de un premio Nobel y otros científicos destacados a determinada institución, la manera de tener en cuenta la antigüedad de los trabajos, entre otras razones.

Agregación arbitraria y deficiente de factores: Un punto débil especialmente obvio. Considerar los cinco primeros factores en forma global, sin relativizar por el tamaño de la institución, hace que las más grandes por ese sólo hecho tengan una considerable ventaja, que no se corrige por el hecho de manejar ese criterio en el sexto factor, pero además el que este último parta de la agregación de los anteriores sin que se justifique el peso de cada uno resulta inadmisiblemente metodológicamente.

Muy deficiente precisión de las mediciones: No se ofrece justificación alguna con base en la cual se pueda defender que el hecho de que una institución obtenga una puntuación menor a otra, sin importar la diferencia, necesariamente implique una correspondiente diferencia de calidad. No se respetan los principios metodológicos elementales de toda medición, no se sabe el margen de error que haya en los resultados y, en consecuencia, es imposible valorar el significado de las diferencias.

Ausencia de definición de estándares de referencia: Ninguno de los dos *rankings* en cuestión explican si algunos valores de la escala que utilizan pueden interpretarse como indicativos de buena o mala calidad, lo que permitiría ir más allá de saber si una institución es mejor o peor que otra u otras, ya que esto no necesariamente es en sí mismo positivo o negativo.

Consistencia de resultados y tendencias: Además, según estos *rankings*, unas instituciones podrían ser mucho mejores que otras un año y mucho peores el siguiente. Algunos ejemplos bastarán para evidenciar las debilidades al respecto.

En el *ranking* de Shanghai, en 2004 la Universidad de Buenos Aires ocupó el lugar 295 y en 2005 el 279; una mejora de 16 lugares en un año es bastante sorprendente, pero en 2006 alcanzó el lugar 159, en un salto de 120 lugares que, simplemente, no es creíble. Tampoco lo es el descenso de 58 puestos

que sufrió la Universidad Autónoma de Madrid entre 2005 y 2006, al pasar del puesto 198 al 256 (Ordorika, Rodríguez *et al.*, 2009: 28). Por lo que se refiere al *ranking* del THE, la UNAM ocupó el lugar 195 en 2004; subió al 95 en 2005 y al 74 en 2006, para caer al 192 en 2007, subir al 150 en 2008 y bajar de nuevo al 190 en 2009 (Ordorika y Rodríguez, 2010: 20).

Contextualización y consecuencias: los usos de los resultados. Los *rankings* internacionales están teniendo un gran impacto mediático, al grado que decisiones importantes tienden a basarse en ellos, con la inequidad que implica su falta de confiabilidad, tanto para las instituciones penalizadas injustamente como para las beneficiadas sin fundamento real. La gravedad del impacto negativo de los *rankings* puede apreciarse si se piensa que instituciones y ministerios de educación de países importantes tienden a basar sus políticas en estos ordenamientos. Como ejemplos se pueden citar en México la Universidad de Guanajuato, que en su Plan de Desarrollo recientemente aprobado se propone llegar a estar *entre las 100 mejores del mundo*, y en Francia el anuncio del gobierno de que se propone que, en un futuro cercano, el país cuente con cuatro instituciones entre las 20 mejores.

Otros análisis llegan a conclusiones similares. Billaut señala que el primer análisis serio del *ranking* de Shanghai, hecho en 2005 (Van Raan, 2005a y 2005b y Liu, Cheng y Liu, 2005) concluyó que no debería tratarse como una verdadera evaluación, y que el problema más serio de los *rankings* consiste en que se les ve como si lo fueran “lo que es absolutamente inaceptable” (Billaut *et al.*, 2009: 33). Utilizando el enfoque de Toma de Decisiones con Múltiples Criterios (*Multiple Criteria Decision Making*), un grupo encabezado por el autor citado revisó los criterios del *ranking* de Shanghai y el método de agregación empleado. Su principal conclusión es que los criterios son irrelevantes y que el método de agregación tiene problemas graves, por lo que el *ranking* “no es una herramienta adecuada para analizar la calidad de las instituciones” (Billaut *et al.*, 2009). De estos autores se retoman dos párrafos que ilustran, de manera elocuente, las consecuencias a las que puede llevar un uso acríptico de esos listados.

“Aparentemente, el *ranking* de Shanghai pretende responder la pregunta de cuál es la mejor universidad del mundo. Para algunos lectores la pregunta misma puede parecer infantil y sin mucho interés. Nosotros estamos de acuerdo. Sin embargo, esos lectores deben tener presente que tal vez hay tomadores de decisiones perezosos que sencillamente pueden usar los resultados de un *ranking* que está a su alcance. Y más importante, tomadores de decisiones estratégicas que pueden usar los resultados para promover sus propias ideas de cómo reorganizar un sistema de educación superior. Más aún, como sucede con toda herramienta de gestión, la existencia misma de un *ranking* contribuye a modificar la conducta de los agentes involucrados, produciendo cambios a veces indeseables” (2009: 2).

“Suponga que administra una universidad y quiere mejorar su posición en el *ranking*. Es bastante sencillo. Hay áreas importantes de su universidad que no aportan para la posición en el *ranking*, como el derecho, las humanidades

y la mayoría de las ciencias sociales. Elimine esas áreas, seguramente ahorrará mucho dinero. Úselo para formar grupos de investigación que sí contribuyan a mejorar su posición. Algunos índices de Thomson Scientific son bastante útiles para ese propósito; después de todo la lista de los probables siguientes cinco premios Nobel en medicina no es tan larga y, de todas formas, si los que usted contrata no reciben el Nobel sin duda publicarán mucho en revistas arbitradas que cuentan en el *ranking*, y muy probablemente estarán en la lista de investigadores muy citados. Lo anterior tiende a promover una manera de ver la ciencia que se parece mucho a los deportes profesionales, en los que unos cuantos equipos ricos compiten por atraer a los mejores jugadores del mundo. No estamos muy convencidos de que esta sea la mejor manera de incrementar el conocimiento humano, por decir lo menos” (2009: 32).

En cuanto al *ranking* del *Times Higher Education Supplement*, hace más de diez años un análisis mencionaba los siguientes puntos, como sólo algunos de los desafíos metodológicos que se han hecho contra las tablas de posiciones:

- Estatus técnico de algunas variables;
- Inadecuada validez de constructo;
- Escalamiento de las variables;
- Cambios de un año a otro en las variables y sus pesos;
- Manipulación de los datos;
- Inconsistencia de los cambios;
- Falta de correspondencia entre el puntaje global y la calidad de cada unidad académica; y
- Distorsión del propósito institucional (Yorke, 1997: 62).

Conclusión

El análisis de los criterios utilizados por algunos de los *rankings* de universidades más importantes en el plano internacional, y la coincidencia con análisis similares, llevan a la conclusión de que las serias deficiencias metodológicas que tienen los hace sumamente inapropiados como herramientas para evaluar de manera confiable la calidad global de las instituciones a las que se refieren.

Incluso limitándose a la pretensión menos ambiciosa de evaluar solamente la calidad de la función de investigación, las deficiencias son tan serias que ni siquiera con dicha acotación los resultados son confiables y no son una buena base para sustentar decisiones por parte de las autoridades, a las que pueden inducir a errores que resulten ser muy costosos para las instituciones.

Contra lo que pretenden los medios de comunicación que patrocinan muchas de estas iniciativas, los *rankings* de programas particulares tampoco son apropiados para orientar a los futuros estudiantes en la delicada tarea de escoger la institución y programa en que tratarán de ser admitidos. Si se tiene en cuenta que no existe la mejor universidad ni el mejor programa *en absoluto*, sino que hay programas más o menos adecuados para ciertos aspirantes, se entenderá que los *rankings* más conocidos no sustituyen a un

buen sistema de orientación vocacional. Los *rankings* elaborados por el CHE resultan mucho mejores para esa función, lo que consiguen justamente porque tienen pretensiones menos ambiciosas y características apropiadas a esos propósitos.

Ninguna metodología podrá valorar bien en una sola dimensión instituciones cuya calidad, en sentido global, es eminentemente multidimensional. Más factible será evaluar unidades, funciones o aspectos de una universidad cuya calidad se pueda definir razonablemente como unidimensional. Suponiendo que se haga con una metodología adecuada no parece absurdo comparar entre sí, por ejemplo, carreras de medicina o de trabajo social, doctorados en astronomía o en antropología, o grupos de investigación sobre cáncer pulmonar o sobre las metodologías de enseñanza de la lectura.

En ausencia de evaluaciones se tiende a tratar a todas las instituciones por igual, lo que no parece correcto; pero con base en evaluaciones no adecuadas ocurre algo peor: las diferencias de trato no tienen qué ver con méritos objetivos, sino ilusorios. El resultado podrá ser que lo bueno tienda a deteriorarse, lo consolidable no madure y lo deficiente permanezca y empeore.

Evaluar bien la calidad de las instituciones de educación superior no es imposible, pero implica acercamientos que no se reduzcan al simplismo de los *rankings* comunes. Sólo con una gama de acercamientos complementarios podrá tenerse una aproximación razonablemente completa a algo tan complejo como la calidad de una universidad.

Referencias

- Aguillo, I. E., J. L. Ortega y M. Fernández (2008). "Webometric Ranking of World Universities: Introduction, Methodology, and Future Developments", en *Higher Education in Europe*, Vol. 33 N° 2-3: 234-244.
- Aguillo, I. E., B. Granadino, J. L. Ortega y J. A. Prieto (2006). "Scientific research activity and communication measured with cybermetric indicators" en *Journal of the American Society of Information Science and Technology*, Vol. 57, N° 10: 1296-1302.
- Baty, Phil (2010). *The unveils broad, rigorous new rankings methodology*. June 3. Consulta en <http://www.timeshighereducation.co.uk/storyasp?sectioncode=26&storycode=411907&c=1> en 2010-08-10.
- Baty, Phil (2009). *Rankings 09: Talking Points. The 2009 world ranking methodology*. October 8. Consulta en <http://www.timeshighereducation.co.uk/storyasp?storycode=408562> en 2010-08-10.
- Billaut, Jean-Charles, D. Bouyssou y Ph. Vincke (2009). Should you believe in the Shanghai ranking? A Multiple Criteria Decision Making view. 29 May.
- Bird, Sheila M. et al. (2005). "Performance indicators: good, bad and ugly", en *Journal of the Royal Statistical Society A*, Vol. 168, Part 1, pp. 1-27.
- Bowden, Rachel (2000). "Fantasy Higher Education: university and college league tables", en *Quality in Higher Education*, Vol. 6, N° 1, pp. 41-60.
- Center For Higher Education Development (CHE) *Hochschul Ranking o University Ranking* <http://ranking.zeit.de> Consulta en 2010-08-10
- Goldstein, Harvey y D. J. Spiegelhalter (1996). "League tables ad their limitations: statistical issues in comparisons of institutional performance", en *Journal of the Royal Statistical Society A*, Vol. 159: 385-443.

- Liu, N. C., Y. Cheng y L. Liu (2005). "Academic ranking of world universities using scientometrics: A comment to the 'fatal attraction'", en *Scientometrics*, 64: 101-109.
- Marginson, Simon (2010). "National and International Rankings of Higher Education", en Peterson, P., E. Baker y B. McGaw, *International Encyclopedia of Education*, 3rd Ed. Amsterdam, Elsevier-Academic Press. Vol. 4, pp. 546-553.
- Márquez J., Alejandro (2010). "Estudio comparativo de universidades mexicanas (ECUM): otra mirada a la realidad universitaria", en *Revista Iberoamericana de Educación Superior (RIES)*. México, IISUE-UNAM/Universia, Vol. 1 N° 1 pp. 148-156. <http://ries.universia.net>
- Martínez Rizo, Felipe et al. (2007a). *Propuestas y experiencias para desarrollar un sistema de indicadores educativos*. México. INEE.
- Martínez Rizo, Felipe (2007b). "Propuesta metodológica para desarrollar un sistema de indicadores para evaluar la calidad de la educación en México", en Varios. *Seminario Internacional de Indicadores Educativos*. Memoria. México. INEE. Julio 9.
- Martínez Rizo, Felipe (2005). "El diseño de sistemas de indicadores educativos: consideraciones teórico-metodológicas", en *Cuadernos de Investigación*, N° 14. México, INEE.
- Martínez Rizo, Felipe (1992). *La calidad de las instituciones de educación superior. Su evaluación y su promoción*. Cuadernos de Planeación Universitaria, 3^a época. Año 6, N° 2. México. UNAM.
- Morrison, H. G., S. P. Magennis y L. J. Carey (1995). "Performance indicators and league tables: a call for standards", en *Higher Education Quarterly*, 49(2), pp. 128-45.
- Observatorio Regional de Políticas de Evaluación Educativa (2010). *Los rankings de escuelas como forma de entregar información sobre calidad educativa*. Boletín monográfico N° 17, agosto, 11 págs.
- Ordorika, Imanol y R. Rodríguez Gómez (2010). "El ranking Times en el mercado del prestigio universitario", en *Perfiles Educativos*, Vol. xxxii, N° 129, pp. 8-25.
- Ordorika Imanol et al. (2009). Desempeño de universidades mexicanas en la función de investigación: estudio comparativo. Datos básicos 2007. *Cuadernos de Trabajo de la Dirección General de Evaluación Institucional*, N° 1. UNAM-DGEI, 37 págs.
- Ordorika, Imanol, R. Rodríguez Gómez et al. (2008). Comentarios al *Academic Ranking of World Universities 2008*. *Cuadernos de Trabajo de la Dirección General de Evaluación Institucional*, N° 2. UNAM-DGEI, 97 págs.
- Shavelson, Richard J., L. McDonell y J. Oakes, Eds. (1989). *Indicators for Monitoring Mathematics and Science Education. A Sourcebook*. Santa Monica, Rand.
- Van Raan, A. F. J. (2005a). "Fatal attraction: Ranking of universities by bibliometric methods", en *Scientometrics*, 62: 133-145.
- Van Raan, A. F. J. (2005b). "Reply to the comments of Liu et al.", en *Scientometrics*, 64: 111-112.
- Webster, David S. (1986). *Academic Quality Rankings of American Colleges and Universities*. Springfield, Ill., Charles C. Thomas Publ.
- Webster, David S. (1985). "How not to rank universities", en *Higher Education*, Vol. 14, N° 1, pp. 71-79.
- Webster, David S. (1983). "America's Higher-Ranked Graduate Schools, 1925-1982", en *Change*, May-June, pp. 13-24.
- Yorke, M. (1997). "A good league guide?", en *Quality Assurance in Education*, 5(2), pp. 61-72.