

El mérito mensurado

El Reino Unido da a conocer una mega evaluación basada en *peer review* que orientará la distribución de recursos para las universidades

FABRÍCIO MARQUES

Publicado en febrero de 2009

La comunidad académica del Reino Unido vive un momento de definiciones. A finales de 2008 se dieron a conocer los resultados del sexto Research Assessment Exercise (RAE 2008), un gran esfuerzo de evaluación de la calidad de la investigación científica que determinará de qué modo se distribuirán 2.300 millones de dólares anuales en fondos públicos entre las universidades británicas durante el período que va de 2009 a 2014. Fueron evaluados 52.400 académicos de 159 instituciones de educación superior. Se constató que el 17% de las investigaciones que éstos realizaron tiene nivel de liderazgo mundial; un 37% se ubica en la categoría de excelencia internacional; un 33% posee reconocimiento internacional; el 11%, reconocimiento nacional, y el 2% se ubica por debajo de los estándares que exige el Reino Unido. “Esto representa un notable logro y confirma que somos una de las grandes potencias mundiales en investigación científica”, declaró al anunciar los resultados David Eastwood, ejecutivo jefe del Higher Education Funding Council for England (Hefce), uno de los organismos responsables por la evaluación. “De las 159 instituciones, 150 tienen algún trabajo que se ubica en la frontera mundial.”

Si bien el RAE 2008 no difunde un *ranking* consolidado de instituciones, un análisis de los datos realizado por el Times Higher Education muestra que las mejores universidades fueron las de Cambridge y Oxford, seguidas por la London School of Economics y por el Imperial College. Algunas instituciones mejoraron su desempeño con relación al RAE anterior, realizado en 2001; tal es el caso de la Universidad de Londres Queen Mary, que subió de la 48ª a la 13ª posición. Otras experimentaron caídas, como la Universidad Warwick, que bajó del 6º al 9º puesto. Pero las universidades deberán esperar hasta el día 4 de marzo para saber exactamente cuáles de ellas van a ganar y cuáles van a perder recursos, pues la distribución tendrá en cuenta, además de la calidad de la investigación, el volumen de investigadores de cada institución sometidos a la evaluación. La Universidad de Cardiff, por ejemplo, cayó de la 8ª a la 22ª posición, pero es probable que no pierda recursos porque sometió más miembros de su *staff* al RAE 2008 que en la evaluación anterior. De cualquier modo, se esperan ondas de choque en las

FOTOS: WARWICK UNIVERSITY

UNIVERSITY OF LEEDS



instituciones que registren desempeño en caída, bajo la forma de despidos, por ejemplo, tal como sucedió en ocasión de anteriores evaluaciones.

El RAE llama la atención por la sofisticación de su metodología, basada en un sistema de evaluación por pares que involucra a consultores nacionales y extranjeros, y por su magnitud –la edición de 2008 costó 17 millones de dólares, mientras que la evaluación de 2001 había costado 8 millones de dólares. Se basa en 15 paneles que supervisan el trabajo de 67 subpaneles disciplinarios. En total, 950 revisores toman parte en el proceso. Al menos un investigador extranjero está presente en los comités principales. “La idea no es comparar la evaluación de los miembros internacionales con los nacionales, sino asegurar que los niveles de calidad que se exigen sean los apropiados”, declaró a *Pesquisa FAPESP* Ed Hughes, del Hefce, el gerente del RAE 2008. “En muchos casos, los miembros internacionales ayudan a establecer parámetros. Ellos cumplen el importante rol de asegurar que el análisis de los paneles cuente con credibilidad internacional.”

Consecuencias - A los efectos de establecer una comparación, el modelo británico tiene significativas diferencias con relación al sistema que desarrolló en Brasil la Coordinación Nacional de Perfeccionamiento del Personal de Nivel Superior (Capes), que desde la década de 1970 evalúa el nivel de los programas de posgrado *stricto sensu*, empezando por los propósitos y las consecuencias de los procesos de evaluación. En el caso brasileño, la evaluación trienal de las carreras de maestría, maestría profesional y doctorado apunta no solamente a medir la calidad de los programas, sino también a estimular su desarrollo, toda vez que orienta la financiación de becas y reconoce la excelencia de los respectivos grupos de investigación. El cierre de carreras mal evaluadas se hace efectivo únicamente en situaciones extremas, y programas con nota regular preservan el derecho de formar magísteres y doctores, aunque con la consiguiente merma de prestigio. El RAE en tanto, tiene un impacto inmediato y a veces devastador, que va más allá de la investigación y del posgrado, pues hace las veces de referencia a la hora de

destinar una amplia dotación de recursos a las universidades del Reino Unido. Una mala evaluación redundará en menos dinero durante un largo período.

“Con base en el RAE, las universidades pueden decidir cerrar determinados departamentos que anduvieron mal en la evaluación, lo que por cierto sucedió con el primer RAE”, dice Lea Velho, docente del Departamento de Política Científica y Tecnológica del Instituto de Geociencias de la Universidad Estadual de Campinas (Unicamp). “Las consecuencias para los departamentos que salieron mal en la evaluación son reales”, afirma.

Aunque ambos modelos tienen en cuenta datos cuantitativos y la evaluación por pares, las metodologías que adoptan la Capes y el

Hefce británico guardan escasas semejanzas. El RAE evalúa la calidad tan sólo de una parte de la producción académica de las universidades, aquella que, por decisión de cada departamento, se juzga que sea más relevante. Cada investigador puede declarar lo sumo cuatro líneas de investigación en las cuales haya tomado parte durante el período. En tanto, en el modelo de la Capes los programas de maestría y doctorado *stricto sensu* deben declarar anualmente un amplio espectro de informaciones relacionadas con la producción científica de docentes y estudiantes, con la formación del cuerpo docente y con la calidad de la formación de los alumnos –y ese conjunto de datos apuntala la evaluación trienal.

En el ejemplo británico, el *peer review* marca el tono. Los revisores tienen la obligación de leer los trabajos científicos que cada departamento destaca para formar su opinión. En casos excepcionales, se les permite que algunos comités se abstengan de analizar en detalle un determinado trabajo, siempre y cuando puedan basar su análisis en revisiones realizadas anteriormente por otros expertos, aunque no en datos bibliométricos. El capítulo de los criterios de evaluación dice textualmente que “ningún panel empleará factores de impacto de publicaciones como medida sustituta de la evaluación de calidad”.

El análisis se efectúa con base en tres elementos. El primero está constituido por los resultados de la investigación académica en forma de artículos, libros, informes técnicos y patentes, entre otros. El segundo es el ambiente de investigación y se basa en datos tales como la cantidad de becas, el volumen de recursos obtenidos o el apoyo institucional a la investigación. El tercero abarca los indicadores de prestigio –a lo sumo cuatro por cada investigador– tales como premios y distinciones obtenidas, organización de congresos y participación en el comité editorial de publicaciones científicas, entre otros. Cada comité juzga la calidad de ese conjunto de datos y la composición de los resultados de los tres elementos brinda el perfil general de calidad, que puede tener cinco clasificaciones: 0 (debajo de los estándares nacionales); 1 (reconocido nacionalmente); 2 (reconocido internacionalmente); 3 (de excelencia internacional), y 4 (con investigación de punta a nivel mundial). Esta metodología reemplazó a la de RAE anteriores, que sumaban puntos



obtenidos en diversos apartados. “El objetivo es evitar que se repitan distorsiones en la distribución de fondos, con un departamento evaluado con mención 5* que recibe muchos más fondos que otro evaluado con 5, aunque las diferencias entre ambos no sean grandes”, dijo Ed Hughes.

En el caso de la Capes, los criterios bibliométricos tienen un gran peso, aunque la evaluación esté a cargo de comités de expertos. Las principales revistas científicas fueron registradas por la agencia de acuerdo con su calidad (léase factor de impacto) y el alcance de la circulación (local, nacional e internacional). Ese sistema, denominado Qualis, sirve para evaluar los artículos científicos de los investigadores y fundamenta una parte significativa del proceso de evaluación, sobre todo en áreas cuya producción académica se expresa en artículos publicados en revistas. De esta forma, una producción modesta, pero divulgada en publicaciones de alto impacto, adquiere mayor peso en las fórmulas que adoptan los comités de evaluación que una producción más vasta divulgada en periódicos con factor de impacto restringido. Los datos recabados se elevan a los comités de área y cada uno de éstos utiliza criterios específicos para analizar las informaciones. Los programas se evalúan con notas que van de 1 a 5. Este trabajo produce planillas comunes a todos los programas, que tienen como objetivo la transparencia y exigen que los comités tengan en cuenta una serie estandarizada de informaciones tales como contingente de docentes, cantidad de tesis y tesinas defendidas, artículos publicados en periódicos científicos nacionales e internacionales, trabajos publicados en anales de eventos nacionales e internacionales, libros y capítulos de libros. Pero un análisis cualitativo puede requerirse en apartados tales como la evaluación de libros o capítulos de libros, por ejemplo, más comunes en la producción de las ciencias humanas, dada la ausencia de indicadores capaces de evaluar su calidad.

Los programas de doctorado que obtuvieron nota máxima (5) pueden pasar por una segunda etapa de evaluación, en este caso de carácter más cualitativo. Pueden ser reevaluados con mención 6 ó 7, dependiendo de indicadores tales como la capacidad de nuclear grupos de investigación o su inserción internacional, medida según criterios tales como la existencia de convenios internacionales, la presencia de docentes visitantes de universidades extranjeras conceptuadas como de primera línea, el intercambio de alumnos con universidades extranjeras y la participación de docentes en comités y directorios de asociaciones internacionales, entre otros.

La elección de los evaluadores tiene formatos diferentes en ambos modelos. En

el caso del RAE 2008, hubo una disputa destinada a ocupar las plazas de miembros y jefes de los paneles y subpaneles. La elección de los mismos estuvo a cargo de representantes de las agencias de financiamiento con base en 4.948 nominaciones realizadas por 1.371 instituciones y sociedades científicas (las universidades no pueden nominar miembros). En la Capes, los coordinadores de los comités, que son elegidos por la institución, tienen una cierta libertad para sugerir con quiénes van a trabajar, respetando criterios de competencia en el área. De todos modos, los nombres deben pasar por la aprobación de la dirección de evaluación de la agencia. La última evaluación trienal involucró a alrededor de 700 revisores. Existe una exigencia de que al menos el 50% de los miembros de cada comité sean sustituidos cada tres años.

Citas - El RAE 2008 será la última evaluación británica que seguirá este modelo. Para disminuir costos y dotar de mayor agilidad a la evaluación, el gobierno del Reino Unido decidió lanzar un nuevo sistema: el Research Excellence Framework (REF), que aunque no abandonará la evaluación por pares, hará un gran uso de indicadores bibliométricos, tales como la cantidad de citas de publicaciones efectuadas por los científicos. “Los elementos que se usarán y el equilibrio entre ellos variarán de acuerdo con las características de cada campo del conocimiento”, afirmó Ed Hughes. El Hefce está haciendo un estudio piloto que comprende 22 áreas del conocimiento, para comparar los resultados del RAE 2008 con lo que será la metodología del REF. El cambio es objeto de división en el seno de la comunidad científica británica, fundamentalmente porque aún no están claros los métodos que se adoptarán. “Tomadas en forma aislada, las citas se han mostrado repetidamente como una medida pobre de la calidad de la investigación”, dijo la revista *Nature* en editorial sobre los cambios en su edición del 1º de enero.

La referida publicación menciona un estudio de 1998, que comparó los resultados de dos análisis de un conjunto de artículos sobre física: una empleaba métricas como citas y la otra se basaba en *peer review*. Las divergencias llegaron al 25% de los artículos analizados. “A los formuladores de políticas no les queda otra opción a no ser reconocer que la revisión de especialistas desempeña un papel indispensable en la evaluación”, dijo *Nature*.

En un informe elevado en 2003 al Hefce, los investigadores Nick von Tunzelman, de la Universidad de Sussex, y Erika Kraemer-Mbula, de la Universidad de Brighton, informaban que, pese a las críticas al sistema británico de evaluación por pares, son raros los países que adoptan sistemas cuantitativos puros para evaluar la investigación científica, y donde se ha hecho, como ven el caso de la región de Flandres, en Bélgica, estuvo plagado de controversias. La cuestión, según Ed Hughes, es hallar un punto de equilibrio. “El nuevo sistema mantendrá elementos de la evaluación por pares, pero debemos encontrar medios de elaborar un sistema de evaluación más sencillo y eficiente, sin por ello perder el valor obtenido mediante los rigurosos procesos aplicados en los RAE”, afirmó. ■