

MIDE no mide lo que dice medir

Jorge Mahecha G.
Profesor, Instituto de Física, Universidad de Antioquia
jorge.mahecha@udea.edu.co

Medellín, Julio 22 de 2015

Se argumenta que para que MIDE de cuenta de la calidad de las instituciones, los resultados que arroje deben reflejar propiedades objetivas de las mismas. Las características y el ordenamiento de las instituciones por su supuesto "grado de calidad" deben ser independientes del modelo. Un análisis de MIDE muestra fácilmente que el ranking de universidades es un artificio que depende del modelo, por lo tanto es subjetivo e inválido para el fin alegado. Un análisis detallado de las variables usadas por el MIDE, también permite concluir que, independientemente del modelo con el cual se elabore el ranking, tales variables son cuestionables. Se muestra que la implementación del MIDE y la publicación de los resultados se salta el debido proceso jurídico y las buenas prácticas investigativas. Finalmente, se enfatiza el carácter subjetivo de este tipo de mediciones de la "calidad", y de los ranking en general, con lo cual resultan en contra vía de los alegados propósitos.

Recientemente, los medios de comunicación informaron de la existencia del Modelo de Indicadores del Desempeño de la Educación Superior MIDE, elaborado con el apoyo del Ministerio de Educación de Colombia, y de la publicación de un ranking de las universidades colombianas según su grado de calidad elaborado con base en el MIDE. El modelo evalúa una función que tiene como entrada un conjunto de datos que corresponden a cada universidad estudiada y como salida un único índice que determina su calidad, llamado "puntaje nacional".

La modelación es de gran utilidad y se usa ampliamente en muchos campos. Un modelo se apoya en teorías y experimentos, constituye una "simulación", o representación matemática de un sistema. Un modelo válido debe ser "estable" o "robusto", eso significa que sus parámetros se deben ajustar "en un punto de equilibrio del espacio de parámetros". El índice que evalúa MIDE es la suma de tres índices, uno que toma valores entre 1 y 40 corresponde a los estudiantes, otro de 1 a 40 a los docentes y otro de 1 a 20 al entorno.

En MIDE, los valores 40-40-20 no fueron obtenidos con el criterio de estabilidad, porque es imposible que este modelo tenga equilibrios. Por lo tanto, si se cambian esos valores, por ejemplo a 37-42-21, o 45-36-19, el ranking cambiará (cualquier tripleta de números tomados al azar, cuya suma sea 100, definiría un modelo diferente que conduciría a un ranking final diferente); esto es suficiente para invalidar el "modelo" (porque los ordenamientos que arroja no son propiedades de la realidad que pretende describir sino "*model-dependent artifacts*"). Así, la evaluación resulta dependiente del modelo y subjetiva, por lo tanto inaceptable.

El modelo desglosa cada campo en dos subcampos. Así, aparecen seis nuevos

valores: 25% desempeño, 15% graduados, 20% docencia, 20% investigación de los docentes, 10% presencia y atracción, 10% internacionalización. Los pesos asignados son igualmente arbitrarios. Finalmente, estos seis subcampos se dividen en 18 sub-sub-campos. Es decir, el modelo tiene 18 variables (los datos de los sub-sub-campos) y 18 parámetros (los factores de peso de cada variable).

Se pretende usar fuentes de datos para asignarles valores a las 18 variables de sub-sub-campos que, supuestamente, son "objetivas": de las citas, de la empleabilidad, de las obras artísticas, del paso a posgrado, etc.

El modelo MIDE habla de la calidad como si fuera un observable, pero no la define ni establece la forma de medirla. Por lo tanto no considera las barras de error de la determinación experimental de "la calidad".

Varios indicadores se expresan con tres cifras decimales, implicando con ello que los autores creen que los errores comienzan a actuar a partir de la cuarta cifra; así, en "puntaje en docencia" varias instituciones aparecen "empataadas" en el puesto 168 con 33,725 puntos (son 21 instituciones "cuya docencia tiene una calidad tan pareja" que las "diferencias" sólo se empiezan a "notar" a partir de la cuarta cifra decimal). La diferencia de puntaje en "docencia" entre las instituciones del puesto 168 y la primera en este subcampo (Seminario Bíblico Colombiano) es de 32 puntos, con lo cual la distancia promedio entre dos puntajes consecutivos de este subcampo es de 0.2, por eso no deja de llamar la atención el múltiple "empate", un extraño fenómeno natural. Las diferencias por nivel de "internacionalización" (Inglés Saber Pro y coautoría con extranjeros) entre Reformada y Barraquer, ubicadas consecutivamente en este subcampo, se expresan diciendo que la primera tiene 15,02 y la segunda 15,23, se "notan" en las décimas, aquí "ya no se puede decir que están empataadas", es curioso que el ministerio de educación de un país elabore un reporte para contar que una de esas instituciones "le gana en internacionalización" a la otra.

No se menciona la incertidumbre de la ubicación en el ranking. No se considera el poder de resolución de su índice, por ejemplo mediante un criterio análogo al de Rayleigh. Este hecho es indispensable para reconocerle validez a la evaluación de la función que entrega el ranking: la diferencia entre los puntajes totales de dos instituciones ubicadas consecutivamente en el ranking debe ser superior a las incertidumbres de los mismos, de lo contrario tienen "empate técnico". La metrología ofrece medios para evaluar la incertidumbre de una función dependiente de datos que a su vez tienen incertidumbres. Aquí se tienen 36 incertidumbres: las 18 de cada uno de los parámetros (de los factores de peso del modelo que, recordemos, son números aleatorios aunque los autores se hayan permitido escoger ciertos valores) y las 18 de cada una de las variables de entrada (de los datos de cada universidad), con las cuales se calcularía la incertidumbre de la salida, o sea del puntaje que determina la posición en el ranking (luego de aplicar un criterio de resolución).

Pero, mucho antes de hablar de la estabilidad de un modelo respecto a variaciones de los parámetros y de incertidumbres y barras de error, incluso de "modelos", se requiere asegurarse de que las variables que caracterizan los estados del sistema que se quiere modelar si sean las apropiadas. El análisis de tales variables brilla por su ausencia en el modelo MIDE. La propiedad más elemental, el signo de la variable, no se considera. Ciertos indicadores, por ejemplo tener muchas coautorías con extranjeros, pueden ser "buenos" (signo positivo), pero también, con argumentos sólidos pueden declararse "malos" (signo negativo). Así, alrededor del 60% de los artículos de Colombia fueron hechos en colaboración internacional y alrededor del 20% de los artículos de China, Brasil y Estados Unidos fueron hechos en colaboración internacional (datos de SCImago). Por su parte, el MIDE dice que las universidades que publican más artículos con colaboradores extranjeros merecen por ello mejores puntajes que las que no lo hacen. En este aspecto, una universidad típica de Colombia sería mejor que una de los países mencionados, afirmación discutible. Este detalle simple permitiría invalidar el modelo, porque evidencia el uso de indicadores cuyo significado no se tiene claro. Pero hay otras variables que merecen descalificaciones similares.

Se tiene posible redundancia de ciertos indicadores. Por ejemplo, los de "investigación" no son indicadores mutuamente independientes. Así, "investigadores de Colciencias" y "artículos" se consideran como indicadores distintos, ¿pero son independientes? La calificación de Colciencias a las hojas de vida no se refiere a la calidad sino a aspectos formales (depende del nivel de aprendizaje y entrenamiento en el uso del software CvLAC, es subjetivo), por lo tanto su uso por parte del ministerio para calificar la calidad de los profesores es inapropiado. CvLAC utiliza el índice de impacto de las revistas y evalúa los libros con criterios formales, igualmente inapropiados. Si se pretende calificar la producción de los profesores, debería aceptarse que ésta es muy diversa; por ejemplo, los libros y las obras literarias en general se quedan por fuera, pues no se trata de artículos, obras de arte, patentes, citas o el ranking de Colciencias. Vale la pena mencionar que en la investigación (y en las otras actividades creativas del campo "docentes", como obras de arte, por ejemplo) participan no solo profesores sino estudiantes, contratistas y personas de fuera de la institución. De nuevo, se usan indicadores cuyo significado no se ajusta a su uso.

Se utilizan los números resultantes de tres "pruebas" Saber Pro para caracterizar la calidad de los estudiantes y de Inglés Saber Pro para caracterizar el grado de internacionalización. Similarmente que el índice asignado por Colciencias a los investigadores, lo que mide Saber Pro es el nivel de entrenamiento y habilidad en un juego, de ninguna manera el nivel académico de los participantes. ¿Cómo es posible que unas encuestas de marcación múltiple se utilicen para introducirles matices o sobrepasar los títulos universitarios que fueron resultado de un conjunto muy grande y diverso de materias y de evaluaciones a cargo de expertos (no propiamente con escogencia múltiple o falso-verdadero)? Y hay un

atrevimiento mayor aún: decir que las habilidades de los estudiantes para jugar Saber Pro son indicadores de calidad de las universidades en las cuales dichos alumnos se matricularon en las más diversas materias, y que la habilidad de los estudiantes para jugar Inglés Saber Pro es indicador del nivel de internacionalización de la universidad. ¿Qué tienen que ver esos números con la "alta" o "baja" calidad de las universidades? Idénticamente podrían usarse los valores de los carros de los estudiantes, o los indicadores de la salud oral, para determinar la calidad de las universidades.

Se usan indicadores que les permite una mejor ubicación en el ranking a las universidades a las cuales usualmente asisten personas de familias influyentes. Por ejemplo, "salario de enganche" como indicador de la calidad de una universidad. Igualmente "empleabilidad". Decir que una universidad con más "ingresos propios" es de mejor calidad que una que no los tiene, favorece a las universidades privadas respecto a las públicas. Con todo esto se están comparando universidades que no pueden ser comparadas. La objeción central en este aspecto es que el ranking MIDE convierte correlación en causación, dice que si dos cosas ocurren simultáneamente, una es causa de la otra (por ejemplo, que jugar bien Saber Pro es causado por estar matriculado en una universidad de calidad).

Docentes con doctorado y docentes con posgrado son indicadores no independientes. Paradójicamente, lo que caracteriza a una universidad "de garaje" (de negocio) es depender de profesores por horas y poseer una gran nómina informal. El ranking MIDE no intentó contabilizar esos contratos. Se recalcan indicadores irrelevantes de la calidad de una universidad, se "cuantifican" cosas que son atributos de los estudiantes, sus familias y su pertenencia social y se hacen pasar como atributos de las universidades. Pero se omiten aspectos esenciales, como la nómina informal de profesores. Un aspecto esencial a considerar, pero que se omite es el siguiente: ¿en qué medida los sistemas curriculares permiten el libre desarrollo de las más diversas fortalezas individuales de los alumnos?, ¿en qué medida la dirección central de la universidad protege y estimula culturas curriculares que estandarizan y controlan el trabajo de los alumnos?, en últimas ¿cuál es el grado de libertad con el cual las personas que asisten a la institución realizan su trabajo?

La publicación de MIDE, independiente de la validez del modelo y de la calidad de la medición como tal, merece criticarse desde el punto de vista del "debido proceso". Los resultados tienen efectos de "premios" y "castigos", por lo tanto los implicados por cualquier veredicto tienen derecho a conocer las pruebas y a controvertirlas antes de la asignación oficial de la "medalla" o la "condena". La apelación, la solicitud formal de aclaración y la posibilidad real de rectificación son componentes esenciales del debido proceso.

Dos "colombianadas" típicas en el campo académico son prometer resultados de investigaciones que se quisieran realizar, y anunciar resultados sin tener

seguridad de su validez. Desde el punto de vista de las debidas prácticas investigativas, un nuevo modelo o instrumento debe someterse a pruebas antes de divulgarse y usarse oficialmente, debe haber una etapa preliminar de experimentos y validaciones antes de la publicación. También tiene su aspecto jurídico: la norma debe tener un proceso riguroso de discusión y decantación antes de oficializarse y de aplicarse. Jurgen Habermas (en *Conciencia Moral y Acción Comunicativa*) y otros autores dan como criterio de validez de una norma su aceptación por los involucrados. Nada de esto se hizo con MIDE: elaboró un procedimiento cuestionable, lo implementó y publicó los resultados saltándose el debido proceso jurídico y las buenas prácticas investigativas. Lo mismo que hace Colciencias desde tiempos inmemoriales.

MIDE recoge todo lo cuestionable de los ranking internacionales, de la "calidad" y la "medición", y lo empeora con muchos elementos de "innovación nacional". Claudine Haroche (en *O Inavaliável em uma Sociedade de Desconfiança*), presenta citas de Yves Charles Zarka (de *Qu'est-ce que Tyranniser le Savoir?* y *L'évaluation: un Pouvoir Suposto Saber*). Refiriéndose a las mediciones del saber, escribió que "Zarka menciona que "la evaluación es siempre subjetiva y relativa", pero que esta "subjetividad... debe permanecer escondida" y que, por este motivo, "procura esconderse detrás de una matemática". Insiste sobre el hecho de que "la evaluación (...) mide otra cosa". Pensemos que esa otra cosa corresponde, entre otras, a la parte imposible de observar que protege la interioridad del individuo: la evaluación realmente realiza... "una inquisición de la interioridad", con el fin de gobernar tanto a los individuos como sus comportamientos, actuaciones y reflexiones,...".