

# ARTÍCULO

Cuando RateMyProfessors se encuentra  
con Google Académico: evaluaciones  
de los alumnos, apariencia  
y la investigación

Rómulo A. Chumacero, Ricardo D. Paredes y Tomás Reyes

REVISTA  
CEPAL

COMISIÓN  
ECONÓMICA PARA  
AMÉRICA LATINA  
Y EL CARIBE



NACIONES UNIDAS

CEPAL

# Cuando RateMyProfessors se encuentra con Google Académico: evaluaciones de los alumnos, apariencia y la investigación<sup>1</sup>

Rómulo A. Chumacero, Ricardo D. Paredes y Tomás Reyes

Recibido: 14/11/2022  
Aceptado: 01/02/2023

## Resumen

Vinculamos más de un millón de evaluaciones de casi 200.000 profesores hechas por alumnos en el sitio web RateMyProfessors con información sobre la productividad de la investigación de casi 3.000 profesores en Google Académico (Google Scholar) a fin de ofrecer una caracterización sistemática de la relación entre las evaluaciones de los alumnos y las características de las clases, las universidades y los profesores, para comprobar si las evaluaciones de los alumnos tienen una relación condicionada con la productividad de la investigación. El estudio concluye que, si bien lo “fácil” e “interesante” que los alumnos consideran un curso son los factores más determinantes de sus evaluaciones, existe una prima de “apariencia” o “belleza”, sin un componente sistemático de raza, edad o género. Sorprendentemente, se constata que la productividad de la investigación no es significativa o está negativamente relacionada con la evaluación de las capacidades docentes de un profesor.

## Palabras clave

Enseñanza superior, calidad de la educación, maestros, evaluación, alumnos, análisis de datos, modelos matemáticos, sitios web

## Clasificación JEL

A22, C10, C20

## Autores

Rómulo A. Chumacero es Profesor Asociado del Departamento de Economía de la Universidad de Chile. Correo electrónico: rchumace@fen.uchile.cl.

Ricardo D. Paredes es Profesor Titular de la Pontificia Universidad Católica de Chile y Ministro del Tribunal de Defensa de la Libre Competencia (TDLC) de Chile. Correo electrónico: rparedes@tdlc.cl.

Tomás Reyes es Profesor Asociado del Departamento de Ingeniería Industrial y de Sistemas de la Pontificia Universidad Católica de Chile. Correo electrónico: threyes@uc.cl.

<sup>1</sup> Los autores agradecen a un árbitro anónimo y a Julio Pertuzé por los útiles comentarios y sugerencias, a Valeria Acevedo, Antonia Bezanilla, Bárbara Fredes, Joaquín Lezaeta, Esteban Miño, Lorena Moraga, Patricio Niculqueo, Leticia Ortiz, Antonio Ossa, Sara Paolini, Isabella Sassi y Hernán Scheel por su valiosa ayuda en la investigación, y a los proyectos Fondecyt 1171894, 1140980 y 1211367 por el apoyo financiero.

## I. Introducción

La mayoría de las instituciones de enseñanza superior utilizan instrumentos de evaluación de la enseñanza por parte de los alumnos para valorar el desempeño en los cursos que ofrecen (Dommeyer, Baum y Hanna, 2002; Dommeyer y otros, 2002). Estos instrumentos son importantes porque proporcionan a los profesores información y comentarios inmediatos sobre cómo perciben los alumnos sus clases, que pueden utilizar para mejorar su desempeño. También informan a los gestores académicos responsables de la asignación de cursos y a los futuros alumnos sobre las características de un determinado profesor. Estas son algunas de las razones por las que los administradores de personal han utilizado estos instrumentos para determinar los ascensos, al tiempo que las agencias de acreditación los han favorecido como indicador indirecto de una buena gestión académica.

Sin embargo, aunque los instrumentos de evaluación de la enseñanza por parte de los alumnos proporcionan una medida de sus opiniones, pueden usarse de manera incorrecta (Jones, Gaffney-Rhys y Jones, 2014). Las evaluaciones individuales pueden estar determinadas por factores distintos de los que los expertos asocian con la calidad de la enseñanza y, lo que es más importante, pueden estar sujetas a sesgos sistemáticos y falta de validez (Onwuegbuzie, Daniel y Collins, 2009). Así, los alumnos pueden valorar mejor un curso relativamente fácil que uno exigente, o favorecer una característica personal del profesor, como la apariencia, que no está necesariamente relacionada con la calidad.

En cuanto a la gestión, dado que algunos profesores también realizan investigación, una evaluación adecuada por parte de los administradores del personal docente debería considerar la relación entre ambas actividades. Los profesores se enfrentan a una disyuntiva entre el tiempo dedicado a la investigación y el dedicado a la docencia (aunque puede darse el caso de que el tiempo dedicado a la investigación sea un valioso aporte a la docencia). Es necesario comprobar empíricamente si la investigación contribuye a la enseñanza universitaria a nivel de grado, como podrían sugerir las políticas públicas que subvencionan la investigación con las matrículas de los alumnos. Carlozzi (2018), por su parte, advierte sobre el sesgo de confirmación en la investigación, algo que también puede incidir en las decisiones administrativas. Al clasificar a los investigadores por su actitud hacia los instrumentos de evaluación de la enseñanza por parte de los alumnos, descubrió que los autores con opiniones negativas sobre dichos instrumentos tenían 14 veces más probabilidades de obtener una puntuación inferior a la prevista, según las evaluaciones de RateMyProfessors (RMP).

Además, los instrumentos de evaluación de la enseñanza por parte de los alumnos suelen resumirse en una estadística descriptiva (promedio), lo que puede ocultar información relevante y llevar a conclusiones erróneas. El uso incorrecto de estos instrumentos puede afectar negativamente el entorno de trabajo y crear una inflación de las calificaciones (Abrami y Mizener, 1985; Dowd, 1988; Goldman, 1990; Blunt, 1991; Marsh y Bailey, 1993; Benson y Lewis, 1994; Johnson, 2003; Onwuegbuzie, Daniel y Collins, 2009).

Como los alumnos no suelen tener acceso a los mencionados instrumentos de evaluación, han surgido evaluaciones informales a través de sitios web abiertos. RMP quizás sea el más conocido, con más de 19 millones de evaluaciones de 1,7 millones de profesores de más de 7.500 instituciones (RateMyProfessors, 2020). Su valor como fuente de calificaciones de profesores se basa en una correlación razonablemente alta con las evaluaciones institucionales (Sonntag, Bassett y Snyder, 2009).

En este trabajo se utiliza esa base de datos, que es considerablemente más grande que las empleadas en la mayoría de los estudios, para analizar la relevancia de distintas variables en la forma en que los alumnos evalúan a sus profesores y relacionar esas evaluaciones con sus impresiones subjetivas y la productividad académica de los profesores en cuestión. El principal aporte metodológico del trabajo consiste en vincular las evaluaciones realizadas en RMP con el historial académico de los profesores a partir de Google Académico (Google Scholar). Esta asociación permite evaluar la relación

entre el papel del profesor como docente y como investigador. También se analizan los comentarios que los alumnos dejan por escrito en el sitio web de RMP para identificar respuestas positivas (afectuosas, edificantes, entre otras) o negativas (con resentimiento, enojo, entre otras) a partir del tono del lenguaje utilizado. Con esta información, podemos analizar los principales determinantes de la evaluación media obtenida por un profesor en RMP.

El resto del artículo se estructura como se establece a continuación. En la sección II se repasa brevemente la literatura sobre las evaluaciones de RMP y en la sección III se presenta un modelo teórico simple que explica los resultados obtenidos en la investigación. En la sección IV se describe el conjunto de datos utilizado y en la sección V se presentan las principales propiedades estadísticas de las evaluaciones y se llevan a cabo ejercicios estadísticos para caracterizar mejor los determinantes de las evaluaciones. Por último, en la sección VI se plantean las conclusiones.

## II. Repaso de la literatura

El debate sobre los instrumentos de evaluación de la enseñanza por parte de los alumnos cobró fuerza en los Estados Unidos y el Canadá con la creciente popularidad de las evaluaciones a través de sitios web abiertos, el más conocido de los cuales probablemente sea RMP. RMP incluye cinco variables en cada evaluación: i) “facilidad”, es decir, la capacidad de obtener buenas calificaciones sin mucho esfuerzo, ii) “disposición”, es decir, la buena voluntad de los profesores para responder preguntas, sobre todo fuera del aula, iii) “claridad” en la exposición en clase, iv) “calidad global” y v) “atractivo”, asociado a la apariencia o la belleza<sup>2</sup>. Las tres primeras variables se miden en una escala de Likert de 1 a 5. La calidad global es un promedio entre la disposición y la claridad, mientras que el atractivo es una variable ficticia asociada a la pregunta específica de si el alumno considera que el profesor es físicamente atractivo (RateMyProfessors, 2020).

RMP no ha estado exenta de críticas. Sonntag, Bassett y Snyder (2009) analizaron su validez comparando el sitio web con las evaluaciones realizadas a través de la encuesta Individual Development and Educational Assessment (IDEA). Utilizando más de 600 evaluaciones de RMP de 126 profesores de la Universidad de Lander, hallaron una correlación de Pearson de 0,69 entre las puntuaciones de calidad de la encuesta IDEA y RMP<sup>3</sup>. La dimensión de facilidad y la calificación media también presentaban una elevada correlación. Los autores concluyeron que RMP arrojaba resultados similares a los de las evaluaciones institucionales. Utilizando una muestra de 399 profesores en RMP, Otto, Sanford y Ross (2008) encontraron una correlación positiva entre claridad y disposición, y una correlación negativa entre estas variables y la varianza en la facilidad<sup>4</sup>.

Boswell (2016) comparó los efectos de las evaluaciones de los alumnos en RMP con las evaluaciones administradas por las universidades sobre la autoeficacia de los profesores, definida como el conjunto de creencias que determinan lo bien que una persona puede ejecutar un plan de acción en determinadas situaciones, y concluyó que, aunque los profesores pensaban que RMP tenía un menor grado de precisión y seriedad, los comentarios recibidos a partir de las dos evaluaciones tenían en ellos la misma incidencia.

<sup>2</sup> RMP fue fundada en mayo de 1999 por John Swapceinski. En 2005 fue adquirida por Patrick Nagle y William DeSantis, que la vendieron a Viacom en 2007. Cheddar adquirió RMP en 2018. Durante este tiempo, RMP ha introducido cambios en el formato de su cuestionario. Desarrollamos nuestra base de datos a partir de información disponible hasta 2018.

<sup>3</sup> Un valor de 1 en el índice de correlación de Pearson representa una correlación positiva perfecta y un valor de -1 representa una correlación negativa perfecta.

<sup>4</sup> En una literatura relacionada se evaluaron similitudes y diferencias en las evaluaciones de los cursos presenciales y en línea (Rovai y otros, 2006; Kelly, Ponton y Rovai, 2007; Dziuban y Moskal, 2011).

Algunos investigadores han informado de un “efecto halo” en las evaluaciones de RMP, lo que significa que solo participan alumnos con opiniones extremas (Felton y otros, 2008; Clayson, 2013). El momento de las evaluaciones también parece ser una variable importante. Mediante el uso de una muestra de 25 profesores de una universidad de tamaño medio de los Estados Unidos, Legg y Wilson (2012) descubrieron que las evaluaciones realizadas al principio de un curso arrojaban puntuaciones más bajas que las efectuadas al final, y que los resultados de “facilidad” cambiaban en función del momento de la evaluación.

Utilizando las evaluaciones de 94 profesores de la Universidad de Texas, Hamermesh y Parker (2005) descubrieron que la belleza tiene un efecto muy significativo en las puntuaciones. De manera similar, Bonds-Raacke y Raacke (2007), Johnson y Crews (2013), Freng y Webber (2009), y Riniolo y otros (2006) hallaron una correlación condicional positiva entre la calidad global de la clase y la belleza. De hecho, Freng y Webber (2009) manifestaron que el “atractivo” explicaba buena parte de la varianza en las calificaciones de calidad.

Mediante la consideración de una muestra más amplia, Felton y otros (2008) repitieron estudios anteriores y hallaron una correlación de 0,6 entre el atractivo y la calidad global de la clase, que es más del doble de la identificada en trabajos anteriores (Felton, Mitchell y Stinson, 2004). Sen, Voia y Woolley (2010) utilizaron una muestra de profesores de economía de 16 universidades y su información de RMP para evaluar los efectos del “atractivo” en los salarios, la calidad de la enseñanza y la productividad de la investigación. Descubrieron que el “atractivo” generaba importantes primas salariales y tenía una alta correlación con la productividad de la enseñanza, pero no con la productividad de la investigación.

Green, Mixon y Treviño (2005) presentaron pruebas de autoselección, según las cuales los profesores que pueden considerarse más atractivos tienden a elegir facultades o universidades más liberales y orientadas a las artes que facultades o universidades orientadas a la investigación. Mixon y Smith (2013) tomaron un subconjunto aleatorio de 200 profesores a partir de los datos de RMP y concluyeron que los profesores más atractivos aprovechaban su apariencia y ofrecían cursos más exigentes.

En cuanto a la relación entre investigación y docencia, existen diferentes hipótesis. Una de ellas hace hincapié en la escasez de recursos, que se traduce en una distribución entre el tiempo dedicado a la investigación y el utilizado para preparar las clases (véanse, por ejemplo, Fox, 1992; Cretchley y otros, 2014; Walstad y Allgood, 2005; Arnold, 2008). Una correlación negativa entre la calidad de la investigación y de la enseñanza también podría ser consecuencia de la especialización, ya que algunos profesores se especializan en la enseñanza y otros en la investigación, en función de sus ventajas comparativas (véase, por ejemplo, Hollywood y otros, 2020).

Por el contrario, otros autores sostienen que la mayoría de los profesores consideran que la investigación es un aporte valioso para una buena enseñanza. Barnett (1992) afirma que la investigación, en un contexto de descubrimiento, es vital para una buena enseñanza (véanse también Neuman, 1992; Becker y Kennedy, 2005). Uz Zaman (2004) defiende la complementariedad entre investigación y docencia, debido, entre otras cosas, al aumento de la capacidad de pensamiento crítico que genera la investigación. En la misma línea, McCaughey (1994) señala que la mayoría de los académicos reconocen que su agenda de investigación determina sus programas de enseñanza.

Ramsden y Moses (1992) constatan que el compromiso con la enseñanza se correlaciona de manera negativa con la investigación. Sin embargo, las divergencias de las fuentes de información y el pequeño tamaño de las muestras hacen que la relación empírica entre docencia e investigación no sea concluyente. Feldman (1987) analiza 43 estudios y encuentra una débil correlación positiva entre la enseñanza evaluada por los alumnos y la productividad de la investigación. Hattie y Marsh (1996), en un metanálisis basado en 58 estudios, concluyen que no existe correlación entre investigación y docencia. Cadez, Dimovski y Zaman Groff (2017) encuentran que la productividad de la investigación no se relaciona con la calidad de la enseñanza, pero que la calidad de la investigación sí se relaciona de

manera positiva con la calidad de la enseñanza. Para el caso de los Países Bajos, Palali y otros (2018) encuentran que un mejor historial de investigación no se refleja en las evaluaciones de los estudiantes de maestría y que los estudiantes de licenciatura dan puntuaciones más bajas a los profesores que realizan mejores investigaciones.

Por último, la edad y el género de los profesores también han demostrado ser factores importantes en las evaluaciones de la enseñanza. Stonebraker y Stone (2015) descubrieron que la edad tenía una incidencia negativa en las evaluaciones de la enseñanza. Sin embargo, cuando restringieron la muestra a los profesores que los alumnos consideraban atractivos, el efecto de la edad desapareció. Por su parte, Clayson (2020), utilizando cuestionarios de alumnos, descubrió que estos aprendían más de los profesores de más edad, pero que los más jóvenes tenían mayor disposición. También encontró que los alumnos preferían los profesores a las profesoras, sobre todo cuando se centraban en el aprendizaje.

### III. Un modelo simple

En esta sección se presenta un modelo simple para ilustrar las posibles interacciones entre el tiempo dedicado a las actividades de investigación y de docencia.

Consideremos un profesor interesado en maximizar su utilidad, que depende del consumo ( $c$ ) y el ocio ( $l$ ):

$$\max u(c,l) \quad (1)$$

sujeto a las restricciones:

$$\begin{aligned} c &\leq y(t, r, w, x), \\ 1 &= l + t + r, \\ t &\geq t_0, \\ r &\geq r_0 \end{aligned} \quad (2)$$

donde  $y$  es el ingreso generado por las actividades de docencia ( $t$ ) e investigación ( $r$ ), mientras que  $w$  y  $x$  son sus respectivas retribuciones. También se supone que la persona dispone de una unidad de tiempo que debe dividir en ocio, docencia e investigación. Por último,  $t_0$  y  $r_0$  son el tiempo mínimo que el administrador establece que se debe dedicar a cada actividad<sup>5</sup>.

A partir de las condiciones de primer orden del problema de maximización, obtenemos:

$$u_c y_t + k_t = u_c y_r + k_r \quad (3)$$

donde  $v_z$  representa la derivada de la función  $v$  ( $v = u, y$ ) con respecto a la variable  $z$  ( $z = c, t, r$ ) y  $k_z$  es el multiplicador asociado a la restricción del tiempo dedicado a la actividad  $z$  ( $z = t, r$ ).

Este modelo simple tiene interesantes implicaciones a la hora de decidir cómo distribuir el tiempo entre la docencia y la investigación. Por ejemplo, consideremos que el ingreso  $y$  está determinado por:

$$y(t, r, w, x) = wt + xr \quad (4)$$

<sup>5</sup> En lugar de considerar  $t$  y  $r$  como el tiempo dedicado a cada actividad, estas variables podrían reflejar su respectiva calidad.

Además, supongamos que  $w > x$ ; es decir, la remuneración de las actividades de docencia supera a la del tiempo dedicado a la investigación. Como no se asigna ningún valor especial (en términos de preferencias) al tiempo dedicado a la investigación, el profesor dedica  $r_0$  de su tiempo (el mínimo necesario) a la investigación. En ese caso, el tiempo dedicado a la docencia debe satisfacer:

$$u_c w + k_t = k_r > 0 \quad (5)$$

Si el tiempo óptimo dedicado a la docencia (sujeto a la restricción de que  $r^* = r_0$ ) es tal que supera el mínimo necesario ( $t^* > t_0$ ), puede determinarse maximizando (1) sujeto a (4) y  $r^* = r_0$ . De lo contrario, la optimización restringida implicaría que  $t^* = t_0$ . Por el contrario, si  $w < x$ , entonces  $t^* = t_0$  es el tiempo dedicado a la docencia, y lo que hay que determinar es si el tiempo óptimo dedicado a la investigación es  $r_0$  o lo supera.

En cualquier caso, si la remuneración es mayor para una actividad que para la otra, cualquier cantidad de tiempo que supere el tiempo mínimo necesario de docencia e investigación se dedicará a la actividad con mayor remuneración. En caso de que  $w = x$ , como el profesor no tiene preferencia entre las dos actividades, dedicará el mínimo necesario a ambas (si las restricciones de tiempo son vinculantes) o se conformará con cualquier combinación de ambas, ya que en este caso son sustitutos perfectos.

Los ejemplos anteriores demuestran lo obvio: el tiempo dedicado a una actividad supone menos tiempo para la otra. Si dedicar más tiempo a una actividad hace que los profesores mejoren en dicha actividad, el hecho de que exista una compensación entre ambas significa que la calidad de su enseñanza y, por tanto, sus evaluaciones pueden verse afectadas como consecuencia de ello.

Por último, consideremos el caso en que el tiempo dedicado a la investigación puede contribuir a mejorar la calidad de la enseñanza, es decir:

$$y(t, r, w, x) = w(r)t + xr \quad (6)$$

donde  $w(r)$  depende de  $r$ . Así, la productividad de la enseñanza también depende de la investigación.

La maximización de (1) sujeta a (2) y (6) se traduce en:

$$u_c w + k_t = u_c (w_r + x) + k_t \quad (7)$$

donde  $w_r$  es la derivada de  $w$  con respecto a  $r$ . Si  $w_r > 0$ , es posible tener soluciones internas para el tiempo dedicado a la docencia y la investigación, incluso si  $w > x$ . En este caso, el vínculo entre el tiempo dedicado a la investigación y la "calidad" de la enseñanza puede atenuar la compensación de tiempo.

En suma, en esta sección se presenta un modelo simple que establece la relación entre el tiempo dedicado a la docencia, su calidad y el tiempo dedicado a la investigación. Como es evidente, hay una compensación en términos de tiempo dedicado a cada actividad (y presumiblemente también hay una compensación en términos de calidad). Es posible que exista una asociación positiva entre el tiempo dedicado a la investigación y la calidad de la enseñanza, ya que la experiencia adquirida al investigar mejora la calidad de la enseñanza y compensa con creces la dedicación de tiempo.

## IV. Los datos

En esta sección se describen los datos y la elaboración de la muestra. Los datos se tomaron de dos fuentes principales: el sitio web RateMyProfessors.com y los perfiles académicos de Google Académico.

## 1. La base de datos RateMyProfessors (RMP)

RMP es una conocida plataforma en la que los alumnos evalúan a sus profesores. Las evaluaciones y comentarios están a disposición del público en el sitio web. Utilizamos un rastreador web para descargar información sobre 1.281.193 profesores calificados en RMP. Obtuvimos el nombre de todos los profesores, así como el nombre y la ubicación de la universidad en la que enseñaban, el departamento al que estaban adscritos, el número de evaluaciones de los alumnos y si estos los consideraban atractivos.

Muchos de estos profesores tenían muy pocas evaluaciones, por lo que nos quedamos solo con los que tenían 20 o más para obtener indicadores representativos<sup>6</sup>. Este criterio dejó 197.037 profesores en la muestra, y descargamos datos específicos de la base de datos RMP para cada uno de ellos, como la calidad global del profesor, lo fáciles e interesantes que los alumnos consideraban sus clases, y los comentarios escritos.

A continuación, utilizamos el programa informático Linguistic Inquiry and Word Count (LIWC) para analizar cada comentario escrito. El programa proporciona una puntuación para diferentes dimensiones lingüísticas, según el porcentaje de palabras relacionadas con cada dimensión. Las dimensiones incluyen “enojo” (identificado a partir de palabras como “odio” y “molesto”), “positividad” (palabras como “agradable” y “dulce”) y “negatividad” (palabras como “herido” y “desagradable”).

Clasificamos el departamento de cada profesor en facultades o escuelas normalizadas. Los 930 departamentos diferentes a los que pertenecían los profesores se clasificaron manualmente en categorías como arquitectura y planificación, ingeniería, humanidades, artes, ciencias sociales, y negocios y gestión<sup>7</sup>. Dado que RMP no proporciona información sobre el género, utilizamos un algoritmo para asignar un género a cada profesor basándonos en un análisis de texto de su nombre<sup>8</sup>.

Completamos los datos de RMP con otras dos fuentes. En primer lugar, encontramos que los 197.037 profesores identificados a partir de RMP pertenecían a 2.583 universidades distintas, y cotejamos cada una de estas universidades con la clasificación *World Ranking Web of Universities* (WRWU)<sup>9</sup>. En segundo lugar, buscamos a cada profesor en Google Académico para ver si tenía un perfil académico disponible, y, en caso afirmativo, procedimos como se describe en la sección IV.2<sup>10</sup>.

## 2. La base de datos de Google Académico

Complementamos los datos de RMP con información de Google Académico ([scholar.google.com](https://scholar.google.com)), donde los profesores pueden crear perfiles y ponerlos a disposición para hacer un seguimiento de sus citas y gestionar sus artículos académicos.

Utilizamos una secuencia de comandos para automatizar el proceso de emparejar a cada profesor que aparece en RMP con un perfil de Google Académico<sup>11</sup>. Esta metodología arrojó 2.401 coincidencias

<sup>6</sup> Dado que nuestro análisis se centra en la puntuación media recibida por cada profesor, 20 es un número de observaciones lo suficientemente grande como para proporcionar una estimación fiable de la tendencia central de las evaluaciones.

<sup>7</sup> No pudimos clasificar 18 departamentos (por ejemplo, “Honores” y “Estudios de posgrado”).

<sup>8</sup> El algoritmo se basa en el diccionario del programa Gender escrito por Jörg Michael, que contiene más de 40.000 nombres de pila, que abarcan la mayoría de los nombres utilizados en los Estados Unidos, Europa, China, la India y el Japón. No obtuvimos coincidencias para el 7,6% de los nombres y en esos casos se consideró que el género era desconocido (por ejemplo, “G.R. Williams”, “Profesor Agarwal”, “Yartz”).

<sup>9</sup> Encontramos 1.925 coincidencias exactas, en las que los nombres de las universidades indicados por RMP y WRWU coincidían perfectamente. Para los 658 nombres de universidades restantes, buscamos manualmente la mejor coincidencia posible. No encontramos coincidencias para 182 universidades.

<sup>10</sup> La lista de variables y sus definiciones para la muestra de 197.037 profesores de RMP figuran en el anexo A1.

<sup>11</sup> Emparejar la información de las dos bases de datos no fue fácil, ya que los nombres de los profesores podían estar escritos de forma diferente. Para cada profesor, la secuencia de comandos buscó primero en Google Académico el nombre que figuraba en RMP. A continuación, comparó los dos nombres y su universidad, según los datos de RMP y Google Académico, y asignó puntuaciones en función de su similitud.



perfectas en las que tanto el nombre del profesor como el de la universidad estaban escritos de forma idéntica. La metodología también arrojó múltiples coincidencias imperfectas, con lo que comprobamos manualmente las 10.000 más probables y encontramos 434 coincidencias adicionales<sup>12</sup>.

La lista final de profesores sobre los que se disponía de información tanto en RMP como en Google Académico incluía a 2.835 personas. Para cada uno de ellos, obtuvimos diversa información de Google Académico, como el número de artículos que el profesor había publicado, la cantidad de veces que esos artículos se habían citado y los índices *h* e *i10* de los profesores<sup>13</sup>.

Por último, completamos estos datos con información recogida manualmente sobre la apariencia de cada profesor. Para cada uno de los 2.835 profesores, tomamos dos de las fotos más recientes que aparecían en Internet y evaluamos atributos físicos como el color de la piel, el color de ojos, la edad y el peso. Aunque existen diferentes enfoques para evaluar la belleza, incluidos aspectos relacionados como la dominancia (véase Paredes, Pino y Díaz, 2019), seleccionamos diez asistentes de investigación para compartir esta tarea que lleva mucho tiempo y empleamos varios mecanismos para mantener un criterio uniforme durante todo el proceso<sup>14</sup>. Cada asistente de investigación recibió una lista de profesores y un manual con instrucciones específicas sobre cómo tabular cada atributo físico<sup>15</sup>. Proporcionamos respuestas a las preguntas más comunes que podían plantearse los asistentes de investigación y pedimos a cada uno que revisara primero diez profesores. A continuación, revisamos su trabajo inicial, les dimos nuestra opinión y les pedimos que continuaran y dejaran un comentario sobre cualquier atributo de un profesor que les generara dudas. Después tomamos una muestra aleatoria de 150 profesores para verificar su trabajo. Observamos que la mayoría de las incoherencias se referían a atributos marcados con un comentario, por lo que decidimos revisar todos los atributos marcados de esa manera en la muestra completa<sup>16</sup>.

## V. Resultados

### 1. Resultados de la base de datos RMP

La base de datos RMP proporciona información útil para estimar cómo la “calidad global de la enseñanza” medida por las evaluaciones de los alumnos se relaciona con las características personales y académicas de los profesores, y con las universidades y facultades en las que enseñan. Nos interesan especialmente dos aspectos incluidos en la base de datos RMP: la valoración que hacen los alumnos de la “facilidad” y de lo “interesante” que es un curso.

En primer lugar, consideramos la base de datos completa obtenida del sitio web de RMP (base de datos RMP). La principal ventaja de esta base de datos es su tamaño, mientras que su principal inconveniente es que no dispone de información detallada sobre las características personales y académicas de los profesores. Por ejemplo, aunque RMP pide a los alumnos que evalúen si un profesor es “atractivo”, no informa de otras características básicas, como el género del profesor.

<sup>12</sup> Estas 434 coincidencias aceptadas manualmente no fueron reconocidas correctamente por la secuencia de comandos, debido principalmente al uso de abreviaturas en los textos de entrada (por ejemplo, “Sam” en lugar de “Samuel”, “Penn State University” en lugar de “Pennsylvania State University”).

<sup>13</sup> Ambos índices están diseñados para captar la cantidad y calidad de las publicaciones. El índice *h* es una medida a nivel del autor que mide tanto la productividad como el impacto de las citas de las publicaciones. Se basa en el conjunto de los trabajos más citados y el número de citas que han recibido en otras publicaciones. El índice *i10* de un autor mide el número de publicaciones que tienen al menos diez citas.

<sup>14</sup> La evaluación exógena de la apariencia de un profesor por parte de los asistentes de investigación proporciona una medida de “belleza” que posiblemente no está contaminada por otras características del profesor que puedan afectar las evaluaciones de los alumnos que lo conocen.

<sup>15</sup> Véase este material en el anexo A2.

<sup>16</sup> Las variables adicionales disponibles para la muestra de 2.835 profesores con información procedente de Google Académico y sus definiciones figuran en el anexo A3.

Para evaluar las características académicas de los profesores, los buscamos en Google Académico y los cotejamos con la información allí encontrada. Google Académico cuenta con información personal más precisa sobre los profesores que allí figuran, además de sus logros académicos, como el factor de impacto de la investigación y los artículos publicados. Por ejemplo, si hay una fotografía del profesor, la utilizamos para obtener más características físicas de las que se pueden deducir de la pregunta dicotómica de la base de datos RMP sobre si es “atractivo o no”.

En el cuadro 1 se presentan estadísticas resumidas de las variables consideradas determinantes de las evaluaciones de los profesores en la base de datos RMP. Además de las estadísticas resumidas habituales, en la última columna se incluye la correlación de Pearson de cada variable con la calidad media de los profesores.

**Cuadro 1**  
Estadísticas descriptivas de las calificaciones de los profesores  
en la base de datos RateMyProfessors

	Media	Mediana	Desviación estándar	Correlación
<b>Características de la universidad</b>				
Estados Unidos	0,903	1,000	0,295	0,051
Canadá	0,096	0,000	0,295	-0,051
Clasificación	2 767,512	1 809,000	2 686,703	0,097
50 principales	0,070	0,000	0,255	-0,013
<b>Características personales</b>				
Hombre	0,658	1,000	0,474	0,005
Atractivo	0,077	0,000	0,175	0,402
<b>Características académicas</b>				
Google Académico	0,014	0,000	0,119	-0,007
<b>Calificaciones</b>				
Calidad (media)	3,676	3,791	0,827	1,000
Facilidad (media)	3,104	3,103	0,773	0,601
Interesante (media)	3,357	3,368	0,540	0,526
Enojo (media)	0,497	0,353	0,545	-0,398
Ansiedad (media)	0,368	0,227	0,519	-0,421
Certeza (media)	2,159	2,084	0,717	0,119
Negativo (media)	2,250	1,894	1,474	-0,669
Positivo (media)	8,023	7,680	3,256	0,717
Tristeza (media)	0,268	0,217	0,269	-0,240
Incertidumbre (media)	3,047	3,026	0,827	-0,328
Cantidad de palabras (media)	37,059	37,911	8,319	-0,123
Palabras por oración (media)	12,922	12,818	2,321	-0,271

**Fuente:** Elaboración propia.

**Nota:** La correlación indica el coeficiente de correlación de Pearson entre una variable y la calidad global media de la enseñanza del curso.

Como puede verse en el cuadro 1, el 90,3% de los profesores evaluados imparten clases en los Estados Unidos, el 9,6% enseña en el Canadá y la mayor parte de los demás enseña en el Reino Unido. Las evaluaciones de los profesores tienen una correlación negativa con la clasificación de su universidad: los profesores de las universidades mejor clasificadas tienen (en promedio) calificaciones marginalmente peores<sup>17</sup>. Este resultado también se mantiene cuando consideramos a los profesores de las 50 universidades mejor clasificadas. Los profesores hombres están (marginalmente) mejor evaluados y los profesores que aparecen en Google Académico están (marginalmente) peor evaluados, aunque solo representan el 1,4% de los profesores de la base de datos.

<sup>17</sup> Una institución mejor clasificada tiene asociado un número menor para la variable Clasificación.

Resulta interesante señalar que el “atractivo” del profesor es una característica personal que guarda una estrecha correlación con la evaluación de la calidad de un curso. Del mismo modo, los cursos “más fáciles” tienden a ser mejor evaluados. Sin embargo, el grado de interés que los alumnos perciben en un curso también parece contribuir mucho a las evaluaciones positivas.

También dispusimos de los comentarios dejados por los alumnos en el sitio web de RMP, y comprobamos que cuanto más negativas eran las respuestas (por ejemplo, cuanto más enojo o ansiedad transmitían, o cuanto más dubitativas eran), peor era la valoración media de los alumnos. Los comentarios positivos (que incluyen expresiones positivas o palabras que transmiten certeza), en tanto, tenían una correlación positiva con las evaluaciones de los alumnos.

Sin embargo, las correlaciones no bastan para describir los factores determinantes de las evaluaciones de los alumnos sobre la calidad de los cursos. Además, dado que algunas de estas variables están correlacionadas entre sí, las correlaciones no sirven para determinar su importancia relativa.

En el cuadro 2 se presentan los resultados de una regresión lineal, con la calificación media global de la calidad en las evaluaciones de los alumnos como variable dependiente. Esta regresión muestra que la ubicación geográfica de la universidad no es una característica relevante en las evaluaciones de los alumnos, pero su clasificación sí. La relación no es lineal, ya que, si bien las universidades mejor clasificadas en un principio tienden a asociarse con peores evaluaciones, a partir de cierta clasificación (aproximadamente 6.000, de casi 12.000) la relación se invierte.

### Cuadro 2

Determinantes de las calificaciones globales de la calidad de la enseñanza en la base de datos RateMyProfessors

Variable	Coficiente	Error estándar	Estadístico t	Probabilidad
Constante	0,8785	0,1869	4,699	0,000
Estados Unidos	0,1129	0,1350	0,836	0,403
Canadá	-0,0269	0,1350	-0,199	0,842
Clasificación	9.69E-06	1.18E-06	8,239	0,000
Clasificación al cuadrado	-8.43E-10	1.29E-10	-6,536	0,000
50 principales	-0,0191	0,0042	-4,546	0,000
Hombre	0,0117	0,0021	5,686	0,000
Atractivo	0,4864	0,0080	60,614	0,000
Google Académico	0,0134	0,0083	1,617	0,106
Facilidad	0,8672	0,0455	19,060	0,000
Facilidad al cuadrado	-0,1735	0,0148	-11,705	0,000
Facilidad al cubo	0,0141	0,0016	9,035	0,000
Interesante	-0,3356	0,1123	-2,989	0,003
Interesante al cuadrado	0,2971	0,0341	8,705	0,000
Interesante al cubo	-0,0400	0,0034	-11,750	0,000
Enojo	-0,0120	0,0028	-4,245	0,000
Ansiedad	-0,0630	0,0033	-18,860	0,000
Certeza	0,0438	0,0015	29,106	0,000
Negatividad	-0,1208	0,0016	-77,031	0,000
Positividad	0,0949	0,0006	154,459	0,000
Tristeza	-0,0382	0,0049	-7,861	0,000
Incertidumbre	-0,0723	0,0014	-50,469	0,000
Cantidad de palabras	0,0141	0,0002	78,864	0,000
Palabras por oración	-0,0285	0,0006	-51,046	0,000
Arquitectura	-0,1058	0,0208	-5,086	0,000
Ingeniería	-0,0370	0,0078	-4,722	0,000
Humanidades	0,0490	0,0060	8,192	0,000
Negocios	0,0777	0,0066	11,758	0,000
Ciencias	0,1326	0,0062	21,236	0,000

Variable	Coefficiente	Error estándar	Estadístico t	Probabilidad
Salud	-0,0160	0,0095	-1,682	0,093
Otros	-0,0465	0,0127	-3,647	0,000
No clasificado	0,0681	0,0361	1,888	0,059
R cuadrado	0,733	Variable dependiente promedio		3,676
R cuadrado ajustado	0,733	Variable dependiente de desviación estándar		0,827
Error estándar medio de la regresión	0,427	Criterio de información de Akaike		1,138
Suma residual de cuadrados	35,990	Criterio de Schwarz		1,140
Probabilidad logarítmica	-112 085	Criterio de Hannan-Quinn		1,139
Estadístico F	17 448	Probabilidad (estadístico F)		0,000

**Fuente:** Elaboración propia.

**Nota:** Errores estándar medios compatibles con heterocedasticidad de White. Número de observaciones incluidas: 197.029.

Los profesores hombres tienden a estar mejor evaluados, pero, aunque el coeficiente es significativo, su magnitud no es tan grande (0,0117 puntos más en una escala de 1 a 5). Aparecer en Google Académico no es estadísticamente significativo. En contraste con estos efectos menores, el “atractivo” es un importante factor de predicción de las calificaciones de los alumnos. La diferencia en la valoración media de la calidad entre una situación en la que ningún alumno considera “atractivo” a un profesor y otra en la que todos lo consideran “atractivo” es de 0,4864 puntos. Se ha discutido mucho en la literatura si esto significa que los profesores “guapos” son “mejores” enseñando o que hay una prima de “belleza” (Hamermesh, 2011). En cualquier caso, la prima de “belleza” parece ser más importante que las otras características personales consideradas.

En lo que respecta a la “facilidad” y lo “interesante” de los cursos, también encontramos una relación no lineal. Aunque los cursos “más fáciles” están mejor calificados, la prima de “facilidad” disminuye con esta. Por el contrario, la prima para lo “interesante” que es un curso va en aumento, hasta cierto punto, y después disminuye<sup>18</sup>. Lo que es más importante es que la “facilidad” contribuye marginalmente más a una buena evaluación que lo “interesante” que sea un curso.

En cuanto a los sentimientos y emociones expresados en los comentarios escritos, como era de esperar, los comentarios negativos tienden a asociarse con una menor calidad media y los positivos con lo contrario. Es de suponer que los alumnos que escriben comentarios son los que tienen opiniones más extremas. Por ejemplo, en promedio, una palabra más que señala enojo se asocia con 0,012 puntos menos en la evaluación media. Curiosamente, los comentarios negativos son “numéricamente” más importantes en valores absolutos. Así, un comentario negativo perjudica más una evaluación de lo que un comentario positivo la mejora<sup>19</sup>.

Por último, el cuadro 2 también muestra las facultades en las que enseñan los profesores. Dado que las categorías utilizadas son cualitativas, los resultados presentados toman como base de comparación a los profesores de las facultades de educación. Los datos muestran que los profesores de las facultades de negocios, humanidades y ciencias tienden a estar mejor evaluados, y que los de las facultades de ingeniería, arquitectura y salud están marginalmente peor evaluados.

En suma, cuando se consideran en conjunto todos los observables, algunos factores son más importantes que otros a la hora de describir lo que hay detrás de la percepción que tienen los alumnos respecto de la calidad de los cursos que realizan. Por lejos, los factores determinantes más importantes son, en primer lugar, lo “fácil” y, en segundo lugar, lo “interesante” que es un curso. Dejando de lado el “atractivo”, los otros factores no son estadísticamente significativos o solo tienen una débil correlación con la valoración media. En particular, ni la investigación, medida por una variable ficticia que indica

<sup>18</sup> La valoración de la calidad media se maximiza cuando la evaluación media es de aproximadamente 4,5 (sobre 5).

<sup>19</sup> Podría ocurrir que la dirección de la causalidad fuera la contraria, es decir, que las emociones expresadas al describir un curso fueran consecuencia de su calidad.

si el profesor aparece en Google Académico, ni una mejor clasificación de la universidad son factores importantes para predecir cómo evalúan los alumnos la calidad de la enseñanza<sup>20</sup>.

## 2. Resultados de la fusión de la base de datos RMP con la base de datos de Google Académico

Además de la evaluación de todos los profesores de la base de datos RMP que ya se ha considerado, dos de los aspectos mencionados anteriormente ameritan un análisis más profundo: qué importancia tiene la apariencia (“atractivo”) y cuál es la relación entre docencia e investigación. La base de datos RMP muestra que el “atractivo” es relevante y más importante que, por ejemplo, la clasificación de la universidad o si el profesor lleva a cabo actividades de investigación.

Para analizar en profundidad estos aspectos, a continuación se considerará únicamente la muestra de profesores que aparecen en Google Académico. Esto reduce el número de observaciones a 2.835 profesores de los 197.037 identificados al comienzo. Una de las ventajas de esta reducción es que podemos medir con mayor precisión la productividad académica de los profesores (medida por los factores de impacto de sus investigaciones). También tenemos acceso a más información sobre su apariencia a través de sus perfiles y fotos. Los factores de impacto pueden ayudar a comprender mejor la relación entre investigación y docencia para los profesores que participan activamente en la primera. Las fotografías pueden proporcionar una impresión aproximada de la “apariencia” que no esté contaminada por otras características de la persona (como la personalidad y la confianza en sí misma) que influyen en la percepción que tienen los alumnos. También permiten conocer mejor las características físicas de los profesores.

En el cuadro 3 se muestran las estadísticas descriptivas de las variables consideradas, al tiempo que en el cuadro 4 se presentan los resultados de una regresión que incluye variables que capturan los atributos físicos y las características académicas (publicaciones y factores de impacto) como determinantes de la calidad global media.

**Cuadro 3**

Estadísticas descriptivas extraídas de la base de datos de Google Académico (Google Scholar)

	Media	Mediana	Desviación estándar	Correlación
<b>Características de la universidad</b>				
Estados Unidos	0,799	1,000	0,401	0,061
Canadá	0,201	0,000	0,401	-0,063
Clasificación	854,160	370,000	1 378,762	0,081
50 principales	0,155	0,000	0,362	-0,046
<b>Características personales</b>				
Hombre	0,799	1,000	0,401	0,051
Menor de 40 años	0,193	0,000	0,394	0,080
Entre 40 y 55 años	0,534	1,000	0,499	0,069
Calvo	0,144	0,000	0,351	-0,019
Amigable	0,782	1,000	0,413	0,078
Ojos claros	0,404	0,000	0,491	0,074
Caucásico	0,854	1,000	0,353	0,124
Afroamericano	0,018	0,000	0,132	-0,053
Indio	0,070	0,000	0,255	-0,036
Piel clara	0,621	1,000	0,485	0,050
Piel morena	0,303	0,000	0,460	-0,020
Pelo negro	0,251	0,000	0,433	-0,050

<sup>20</sup> Por supuesto que los alumnos de universidades con clasificaciones diferentes no son directamente comparables.

	Media	Mediana	Desviación estándar	Correlación
Pelo rubio	0,105	0,000	0,306	0,013
Pelo castaño	0,359	0,000	0,480	0,078
Pelo gris	0,259	0,000	0,438	-0,064
Delgado	0,061	0,000	0,239	-0,012
Complexión media	0,759	1,000	0,428	0,027
Alto	0,131	0,000	0,338	0,081
Altura media	0,194	0,000	0,396	0,045
Buena apariencia	0,168	0,000	0,374	0,120
Apariencia media	0,668	1,000	0,471	-0,012
Atractivo	0,082	0,000	0,177	0,404
<b>Características académicas</b>				
Número de publicaciones	82,481	43,000	187,503	-0,123
Citas	2 166,393	696,000	5 119,203	-0,067
Índice H	16,150	13,000	13,839	-0,125
Índice I10	28,378	15,000	49,647	-0,108
<b>Evaluaciones</b>				
Calidad (media)	3,631	3,722	0,765	1,000
Facilidad (media)	2,928	2,933	0,661	0,521
Interesante (media)	3,399	3,419	0,496	0,489
Enojo (media)	0,494	0,363	0,532	-0,345
Ansiedad (media)	0,385	0,233	0,559	-0,406
Certeza (media)	2,124	2,027	0,726	0,123
Negativo (media)	2,428	2,096	1,514	-0,632
Positivo (media)	8,118	7,671	3,294	0,668
Tristeza (media)	0,267	0,213	0,274	-0,271
Incertidumbre (media)	3,148	3,137	0,854	-0,298
Cantidad de palabras (media)	36,708	37,700	8,921	-0,113
Palabras por oración (media)	12,635	12,602	2,249	-0,277

**Fuente:** Elaboración propia.

**Nota:** La correlación indica el coeficiente de correlación de Pearson entre una variable y la calidad global media de la enseñanza del curso.

**Cuadro 4**  
Determinantes de las calificaciones globales de la calidad de la enseñanza  
en la base de datos Google Académico

Variable	Coficiente	Error estándar	Estadístico t	Probabilidad
Constante	0,3558	0,3081	1,155	0,248
Estados Unidos	0,0754	0,0854	0,883	0,377
Canadá	-0,0845	0,0879	-0,961	0,337
Clasificación	3,43E-05	1,63E-05	2,110	0,035
Clasificación al cuadrado	-2,88E-09	1,84E-09	-1,561	0,119
50 principales	-0,0257	0,0263	-0,978	0,328
Hombre	0,0521	0,0235	2,217	0,027
40 años o menos	0,1157	0,0307	3,768	0,000
Entre 41 y 55 años	0,0874	0,0236	3,709	0,000
Calvo	0,0301	0,0253	1,190	0,234
Piel morena	-0,0133	0,0422	-0,314	0,753
Pelo negro	-0,0665	0,0510	-1,302	0,193
Pelo rubio	-0,1462	0,0523	-2,794	0,005
Pelo castaño	-0,0392	0,0478	-0,821	0,412
Pelo gris	-0,0731	0,0500	-1,461	0,144
Ojos claros	0,0372	0,0192	1,940	0,053
Caucásico	0,1689	0,0418	4,040	0,000
Afroamericano	-0,0380	0,0833	-0,455	0,649

Variable	Coficiente	Error estándar	Estadístico t	Probabilidad
Indio	0,0941	0,0526	1,789	0,074
Piel clara	0,0043	0,0434	0,099	0,921
Delgado	0,0220	0,0396	0,557	0,578
Peso normal	0,0300	0,0229	1,310	0,190
Alto	0,0381	0,0265	1,435	0,151
Altura normal	0,0198	0,0222	0,893	0,372
Amigable	0,0549	0,0222	2,473	0,014
Atractivo	0,4449	0,0502	8,859	0,000
Buena apariencia	-0,0102	0,0324	-0,315	0,753
Apariencia normal	0,0245	0,0251	0,972	0,331
Publicaciones	-0,0003	0,0002	-2,306	0,021
Publicaciones al cuadrado	8,81E-08	5,54E-08	1,590	0,112
Citas	-1,70E-06	2,43E-06	-0,700	0,484
Índice H	0,0011	0,0015	0,714	0,475
Índice I10	-0,0001	0,0004	-0,137	0,891
Facilidad	0,3124	0,0984	3,175	0,002
Facilidad al cuadrado	-0,0200	0,0158	-1,262	0,207
Interesante	0,7553	0,1652	4,571	0,000
Interesante al cuadrado	-0,0689	0,0244	-2,823	0,005
Enojo	0,0040	0,0218	0,183	0,855
Ansiedad	-0,0711	0,0262	-2,712	0,007
Certeza	0,0535	0,0128	4,168	0,000
Negatividad	-0,1143	0,0123	-9,262	0,000
Positividad	0,0852	0,0047	18,005	0,000
Tristeza	-0,0255	0,0415	-0,615	0,538
Incertidumbre	-0,0673	0,0121	-5,572	0,000
Cantidad de palabras	0,0118	0,0016	7,276	0,000
Palabras por frase	-0,0201	0,0054	-3,727	0,000
Arquitectura	-0,2388	0,1077	-2,218	0,027
Ingeniería	-0,1230	0,0598	-2,057	0,040
Humanidades	-0,0826	0,0552	-1,497	0,134
Negocios	0,0152	0,0581	0,261	0,794
Ciencias	-0,0867	0,0566	-1,531	0,126
Salud	-0,1119	0,0991	-1,130	0,259
Otros	-0,2930	0,1758	-1,666	0,096
No clasificado	0,0186	0,0949	0,196	0,845
R cuadrado	0,706202	Variable dependiente promedio		3,630257
R cuadrado ajustado	0,700035	Variable dependiente de desviación estándar		0,771407
Error estándar de la regresión	0,422492	Criterio de información de Akaike		1,135425
Suma residual de cuadrados	450,71180	Criterio de Schwarz		1,258022
Probabilidad logarítmica	-1410,13	Criterio de Hannan-Quinn		1,179863
Estadístico F	114,5157	Probabilidad (estadístico F)		0,000

**Fuente:** Elaboración propia.

**Nota:** Errores estándar compatibles con heterocedasticidad de White. Número de observaciones incluidas: 2.579.

Los resultados coinciden aproximadamente con los de toda la muestra. No hay diferencias estadísticas en la calidad evaluada entre países, las calificaciones medias tienden a ser más bajas en las universidades mejor clasificadas y las emociones negativas tienden a estar más correlacionadas con peores evaluaciones de lo que las emociones positivas se correlacionan con las mejores evaluaciones. Además, la “facilidad” tiene un impacto positivo pero decreciente en la evaluación de la calidad de un curso. Sin embargo, una diferencia importante es que, en esta muestra de profesores, cuanto más “interesante” es un curso, más alta es su valoración (a una tasa decreciente), y esto es lo que más contribuye a la valoración global de la calidad del curso. Existe entonces una diferencia de prioridades

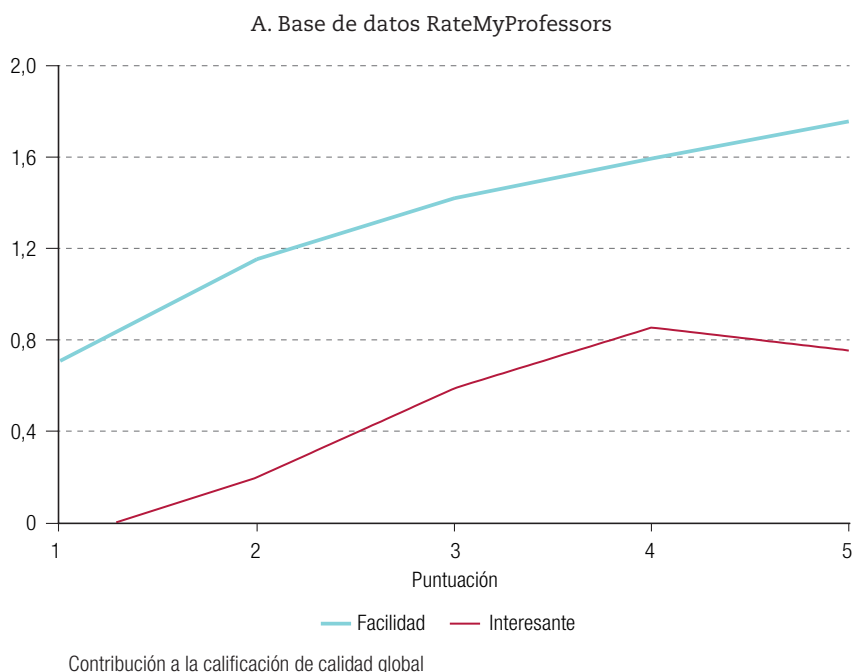
entre “facilidad” e “interesante” en el caso de los profesores que aparecen en Google Académico y presumiblemente son más activos en las tareas de investigación.

En cuanto a las características personales, los profesores tienen puntuaciones ligeramente mejores que las profesoras y el “atractivo” sigue siendo un factor importante. Sin embargo, muy pocas de las características físicas específicas mencionadas en el cuadro 4 ayudan sistemáticamente a predecir la calidad de un curso y, en general, las magnitudes son muy pequeñas.

La productividad académica, ya sea que se mida por el número de publicaciones, citas o factores de impacto, es irrelevante o perjudicial para las calificaciones. Una posible explicación es la competencia entre el tiempo dedicado a la investigación y a la docencia<sup>21</sup>. Otra posible explicación podría ser que la investigación y la docencia son complementarias en los cursos de posgrado, pero las evaluaciones de RMP abarcan principalmente cursos de licenciatura, algo que los datos no nos permiten diferenciar. En cualquier caso, los datos indican de manera rotunda que la investigación activa no necesariamente se asocia a mejores evaluaciones de la docencia.

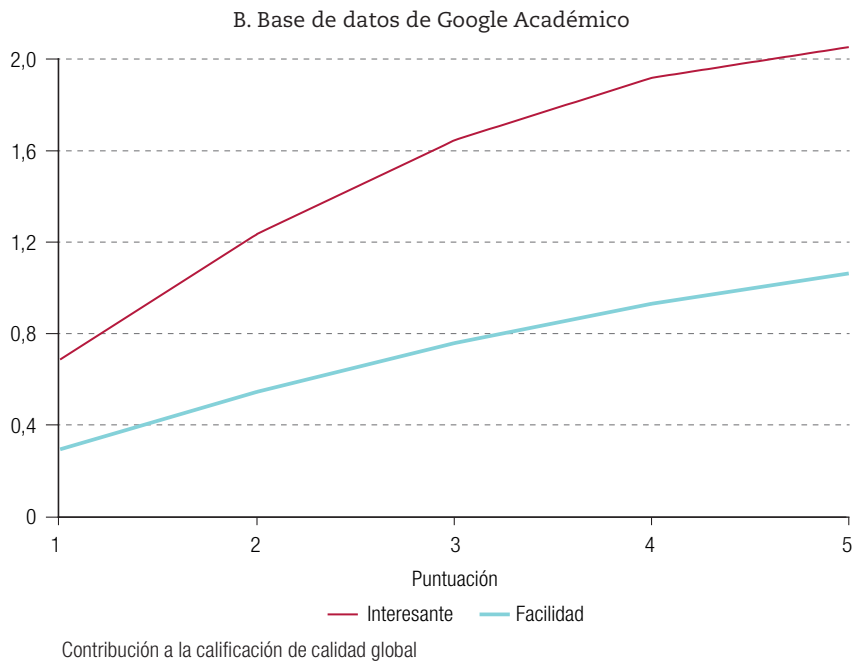
En el gráfico 1 se muestra cómo se revierte la importancia de los factores cuando se compara la base de datos RMP con la base de datos de Google Académico. Esto demuestra que la pregunta sobre cuán interesante es el curso se convierte en un factor más importante en el caso de los profesores que tienen un historial de investigación, como lo demuestra su presencia en Google Académico.

**Gráfico 1**  
Contribución de la “facilidad” y lo “interesante” a las evaluaciones de los profesores por parte de los alumnos



<sup>21</sup> De hecho, como señaló un árbitro, los académicos pueden estar más interesados en labrarse una carrera basada principalmente en la investigación y llevar a cabo la cantidad mínima obligatoria de docencia.





**Fuente:** Elaboración propia.

**Nota:** Los gráficos muestran la contribución de una puntuación determinada de cada una de las características a la calificación global. Por ejemplo, cuando se utilizan las estimaciones del gráfico 1.A (base de datos RateMyProfessors), una puntuación de 3 para “facilidad” contribuye con 1,42 puntos a la calificación de calidad global, mientras que cuando se utilizan las estimaciones del gráfico 1.B (base de datos de Google Académico), una puntuación de 3 para “interesante” contribuye con 1,65 puntos a la calificación de calidad global.

En suma, si tenemos en cuenta únicamente a los profesores cuyo historial de investigación puede rastrearse y determinamos otras características físicas, comprobamos que (cualitativamente) los principales resultados se mantienen. La “facilidad” y lo “interesante” de los cursos resultan ser los principales determinantes de la valoración de la calidad por parte de los alumnos. Existe una importante prima de “belleza” que no depende de características sistemáticas como la raza o el color de la piel, los ojos o el pelo. Para las evaluaciones de los profesores, es más perjudicial que provoquen emociones negativas de lo que resulta beneficioso que provoquen emociones positivas. Por último, existe, en todo caso, una asociación negativa entre la productividad académica y la calidad de la enseñanza según las evaluaciones de los alumnos.

## VI. Comentarios finales

Si bien se ha criticado el uso de sitios web voluntarios abiertos, y en particular RMP, hay al menos tres elementos que hacen que merezca la pena describirlos y analizarlos. En primer lugar, se ha informado de que las calificaciones globales de RMP se correlacionan razonablemente bien con las evaluaciones formales. En segundo término, la popularidad de RMP es cada vez mayor y afecta las decisiones de los alumnos, además de ser una fuente de información inmediata para los profesores. Por último, proporciona información para el análisis de un volumen y una variedad que no se encuentran en otro lugar.

En este documento, utilizamos el conjunto de datos de RMP para describir las principales características que afectan la calidad de la enseñanza, según los alumnos. A partir de las evaluaciones de casi 200.000 profesores en el sitio, casi 3.000 de las cuales podían vincularse a información en

Google Académico, descubrimos que los principales determinantes de la calidad percibida de un curso son lo “fácil” y lo “interesante” que es. La “facilidad” y lo “interesante” de un curso influyen positivamente en sus calificaciones de calidad a un ritmo decreciente y creciente, respectivamente. También encontramos una prima de “belleza” que no es sistemática con respecto a la raza, el color de la piel u otras características personales de los profesores. Por último, observamos que la productividad de la investigación es insignificante o perjudicial en las evaluaciones de la docencia.

Nuestro estudio concluye que la investigación parece competir con la calidad de la enseñanza. Si los profesores tienen incentivos principalmente para investigar, la calidad de la enseñanza puede resentirse. Además, dado que los alumnos valoran la “facilidad” en sus calificaciones, es posible que no estén recibiendo la calidad de educación necesaria para formar futuros profesionales competentes. Explorar y comprender las razones detrás de estos resultados puede tener importantes implicaciones para las políticas y decisiones académicas. Así, la existencia de una prima de belleza y la posibilidad de que el énfasis en la investigación reduzca las calificaciones de los profesores dan pie a la reflexión sobre el valor que las instituciones de educación conceden a las evaluaciones de los alumnos y los costos de centrarse exclusivamente en la investigación.

## Bibliografía

- Abrami, P. y D. Mizener (1985), “Student/instructor attitude similarity, student ratings, and course performance”, *Journal of Educational Psychology*, vol. 77.
- Arnold, I. (2008), “Course level and the relationship between research productivity and teaching effectiveness”, *Journal of Economic Education*, vol. 39, N° 4.
- Barnett, R. (1992), “Linking teaching and research: a critical inquiry”, *Journal of Higher Education*, vol. 63.
- Becker, W. and P. Kennedy (2005), “Does teaching enhance research in economics?” *American Economic Review*, vol. 95, N° 2.
- Benson, D. y J. Lewis (1994), “Students evaluations of teaching and accountability: implications from the Boyer and the ASA reports”, *Teaching Sociology*, vol. 22.
- Blunt, A. (1991), “The effects of anonymity and manipulated grades on student ratings of instructors”, *Community College Review*, vol. 18.
- Bonds-Raacke, J. y J. Raacke (2007), “The relationship between physical attractiveness of professors and students’ ratings of professor quality”, *Journal of Psychiatry, Psychology and Mental Health*, vol. 1, N° 2.
- Boswell, S. (2016), “Rate my professors is hogwash (but I care): Effects of Rate my professors and university-administered teaching evaluations on professors”, *Computers in Human Behavior*, vol. 56.
- Cadez, S., V. Dimovski, y M. Zaman Groff (2017), “Research, teaching and performance evaluation in academia: the salience of quality”, *Studies in Higher Education*, vol. 42, N° 8.
- Carlozzi, C. (2018), “Rate my attitude: research agendas and RateMyProfessor scores”, *Assessment & Evaluation in Higher Education*, vol. 43, N° 3.
- Clayson, D. (2020), “Student perception of instructors: the effect of age, gender and political leaning”, *Assessment & Evaluation in Higher Education*, vol. 45, N° 4.
- \_\_\_\_\_(2013), “What does Rate my professors.com actually rate?”, *Assessment & Evaluation in Higher Education*, vol. 39, N° 68.
- Cretchley, P. y otros (2014), “Research and/or learning and teaching: A study of Australian professors’ priorities, beliefs and behaviours”, *Higher Education Research and Development*, vol. 33, N° 4.
- Dommeyer, C., P. Baum y R. Hanna (2002), “College students’ attitudes towards methods of collecting teaching evaluations: in-class vs. online”, *Journal of Education for Business*, vol. 78, N° 1.
- Dommeyer, C. y otros (2002), “Attitudes of business faculty toward two methods of collecting teaching evaluations: paper vs. online”, *Assessment & Evaluation in Higher Education*, vol. 27, N° 5.
- Dowd, J. (1988), “Sociology is different: the misevaluation of teaching effectiveness”, *Sociological Inquiry*, vol. 58.

- Dziuban, C. y P. Moskal (2011), "A course is a course is a course: factor invariance in student evaluation of online, blended and face-to-face learning environments", *Internet and Higher Education*, vol. 14.
- Feldman, K. (1987), "Research productivity and scholarly accomplishment of College teachers as related to their instructional effectiveness: a review and exploration", *Research in Higher Education*, vol. 26.
- Felton, J., J. Mitchell y M. Stinson (2004), "Web-based student evaluations of professors: the relations between perceived quality, easiness and sexiness", *Assessment & Evaluation in Higher Education*, vol. 29, N° 1.
- Felton, K. y otros (2008), "Attractiveness, easiness and other issues: student evaluations of professors on RateMyProfessors.com", *Assessment & Evaluation in Higher Education*, vol. 33.
- Fox, M. (1992), "Research, teaching and publication productivity: mutuality versus competition in academia", *Sociology of Education*, vol. 65.
- Freng, S. y D. Webber (2009), "Turning up the heat on online teaching evaluations: does "hotness" matter?", *Teaching of Psychology*, vol. 36, N° 3.
- Green, T., F. Mixon y L. Trevino (2005), "Have you seen the new econ prof? beauty, teaching, and occupational choice", *Shaping the Learning Curve: Essays on Economic Education*, F. Mixon (ed.), iUniverse Inc.
- Goldman, L. (1990), "Student evaluations of their professors rarely provide a fair measure of teaching ability", *Chronicle of Higher Education*, vol. 8, August.
- Hamermesh, D. (2011), *Beauty Pays: Why Attractive People Are More Successful*, Princeton University Press.
- Hamermesh, D. y A. Parker (2005), "Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity", *Economics of Education Review*, vol. 24.
- Hattie, J. y H. W. Marsh (1996), "The relationship between research and teaching: a meta-analysis", *Review of Educational Research*, vol. 66, N° 4.
- Hollywood, A. y otros (2020), "Overwhelmed at first: the experience of career development in early career academics", *Journal of Further and Higher Education*, vol. 44, No.7.
- Johnson, V. (2003), *Grade Inflation: A Crisis in College Education*, Nueva York, Springer.
- Johnson, R. y A. Crews (2013), "My professor is hot! Correlates of RateMyProfessors.com ratings for criminal justice and criminology faculty members", *American Journal of Criminal Justice*, vol. 38, N° 4.
- Jones, J., R. Gaffney-Rhys y E. Jones (2014), "Handle with care! An exploration of the potential risks associated with the publication and summative usage of student evaluation of teaching (SET) results", *Journal of Further and Higher Education*, vol. 38, No.1.
- Kelly, H., M. Ponton y A. Rovai (2007), "A comparison of student evaluations of teaching between online and face-to-face courses", *Internet and Higher Education*, vol. 10.
- Legg, A. y J. Wilson (2012), "RateMyProfessors.com offers biased evaluations", *Assessment & Evaluation in Higher Education*, vol. 37, No.1.
- Marsh, H. y M. Bailey (1993), "Multidimensional students' evaluations of teaching effectiveness: a profile analysis", *Journal of Higher Education*, vol. 64, N° 1.
- Mc Caughey, R. (1994), *Scholars and Teachers: The Faculties of Select Liberal Arts Colleges and Their Place in American Higher Learning*, Nueva York, Conceptual Litho Reproductions.
- Mixon, F. y K. Smith (2013), "Instructor attractiveness and academic rigour: examination of student evaluation data", *Australasian Journal of Economics Education*, vol. 10, N° 1.
- Neumann, R. (1992), "Perceptions in the teaching-research nexus: A framework for analysis", *Higher Education*, vol. 23.
- Onwuegbuzie, A., L. Daniel y K. Collins (2009), "A meta-validation model for assessing the score-validity of student teaching evaluations", *Quality and Quantity*, vol. 43.
- Otto, J., D. Sanford y D. Ross (2008), "Does RateMyProfessors.com really rate my professor?", *Assessment & Evaluation in Higher Education*, vol. 33, N° 4.
- Palali, A. y otros (2018), "Are good researchers also Good teachers? The relationship between research quality and teaching quality", *Economics of Education Review*, vol. 64.
- Paredes, V., F. Pino y D. Díaz (2019), "Does facial structure explain differences in students' evaluations of teaching? The role of perceived dominance", *Serie Documentos de Trabajo*, N° 483 [online] <https://repositorio.uchile.cl/handle/2250/168496>.
- Ramsden, P. y I. Moses (1992), "Association between research and teaching in Australian higher education", *Higher Education*, vol. 23.
- Rate My Professors (2020), *About RateMyProfessors.com* [online] <http://www.ratemyprofessors.com/About.jsp>.

- Riniolo, T. y otros (2006), "Hot or not: do professors perceived as physically attractive receive higher student evaluations?", *The Journal of General Psychology*, vol. 133, N° 1.
- Rovai, A. y otros (2006), "Student evaluation of teaching in the virtual and traditional classrooms: A comparative analysis", *Internet and Higher Education*, vol. 9.
- Sen, A., M. Voia y F. Woolley (2010), "The effect of hotness on pay and productivity", *Carleton Economic Papers*, N° CEP 10-07.
- Sonntag, M., J. Bassett y T. Snyder (2009), "An empirical test of the validity of student evaluations of teaching made on RateMyProfessors.com", *Assessment & Evaluation in Higher Education*, vol. 34, N° 5.
- Stonebraker, R. y G. Stone (2015), "Too old to teach? The effect of age on college and university professors", *Research in Higher Education*, vol. 56, N° 8.
- Uz Zaman, M. (2004), "Review of the academic evidence on the relationship between teaching and research in higher education", *Research Report*, N° RR506, Londres, Departamento de Educación y Habilidades.
- Walstad, W. y S. Allgood (2005), "Views of teaching and research in economics and other disciplines", *American Economic Review*, vol. 95, N° 2.

## Anexo A1

### Lista completa de variables para la muestra de 197.037 profesores de RateMyProfessors

#### Características de la universidad:

- **Estados Unidos:** variable ficticia que toma el valor 1 si la universidad del profesor está en los Estados Unidos. Fuente: RMP.
- **Canadá:** variable ficticia que toma el valor 1 si la universidad del profesor está en el Canadá. Fuente: RMP.
- **Clasificación:** clasificación de la universidad del profesor; es mejor un número más bajo. Fuente: World Ranking Web of Universities (WRWU).
- **50 principales:** variable ficticia que toma el valor 1 si la variable Clasificación es menor o igual a 50. Fuente: WRWU.

#### Características personales:

- **Hombre:** variable ficticia que toma el valor 1 si el profesor es hombre. Fuente: algoritmo basado en el diccionario gender/nam\_dict escrito por Jörg Michael.
- **Atractivo:** variable ficticia que toma el valor 1 si el número de alumnos en RMP que encuentran al profesor “atractivo” menos el número de los que no lo encuentran “atractivo” es 20 o mayor. Fuente: RMP.

#### Características académicas:

- **Google Académico:** variable ficticia que toma el valor 1 si el profesor tiene un perfil académico en Google Académico. Fuente: Google Académico.

#### Evaluaciones:

- **Calidad:** mide la calidad del profesor percibida por los alumnos. RMP la calcula como la media entre “disposición” y “claridad”. La disposición se mide en una escala de Likert de 1 (“Nada de disposición”) a 5 (“Me salvó el semestre”). La claridad se mide en una escala Likert de 1 (“¿Qué dice?”) a 5 (“Claro como el agua”). Fuente: RMP.
- **Facilidad:** mide lo fácil que les resulta la clase a los alumnos en una escala de Likert de 1 (“Lo más difícil que he hecho”) a 5 (“Para aprobar solo hay que ir”). Fuente: RMP.
- **Interesante:** mide cuán interesante los alumnos encuentran al profesor en una escala Likert de 1 (“Regular”) a 5 (“Me encanta”). Fuente: RMP.
- **Enojo:** porcentaje de palabras en el comentario del alumno que denotan enojo (palabras como “odio” y “molesto”). Fuente: RMP y programa informático Linguistic Inquiry and Word Count (LIWC).
- **Ansiedad:** porcentaje de palabras en el comentario del alumno que denotan ansiedad (palabras como “preocupado” y “asustado”). Fuente: RMP y LIWC.

- **Certeza:** porcentaje de palabras en el comentario del alumno que denotan certeza (palabras como “siempre” y “nunca”). Fuente: RMP y LIWC.
- **Negatividad:** porcentaje de palabras en el comentario del alumno que denotan emociones negativas (palabras como “sufrimiento” y “desagradable”). Fuente: RMP y LIWC.
- **Positividad:** porcentaje de palabras en el comentario del alumno que denotan emociones positivas (palabras como “lindo” y “dulce”). Fuente: RMP y LIWC.
- **Tristeza:** porcentaje de palabras en el comentario del alumno que denotan tristeza (palabras como “llorar” y “triste”). Fuente: RMP y LIWC.
- **Incertidumbre:** porcentaje de palabras en el comentario del alumno que denotan incertidumbre (palabras como “suponer” y “quizás”). Fuente: RMP y LIWC.
- **Cantidad de palabras:** cantidad de palabras en el comentario del alumno. Fuente: RMP y LIWC.
- **Palabras por oración:** número promedio de palabras por oración en el comentario del alumno. Fuente: RMP y LIWC.

## Características del departamento:

- **Arquitectura:** variable ficticia que toma el valor 1 si el departamento del profesor forma parte de una escuela de arquitectura y planificación. Fuente: clasificación manual a partir de la denominación del departamento según se informa en RMP.
- **Ingeniería:** variable ficticia que toma el valor 1 si el departamento del profesor forma parte de una escuela de ingeniería. Fuente: clasificación manual a partir de la denominación del departamento según se informa en RMP.
- **Humanidades:** variable ficticia que toma el valor 1 si el departamento del profesor forma parte de una escuela de humanidades, arte y ciencias sociales. Fuente: clasificación manual a partir de la denominación del departamento según se informa en RMP.
- **Negocios:** variable ficticia que toma el valor 1 si el departamento del profesor forma parte de una escuela de negocios y gestión. Fuente: clasificación manual a partir de la denominación del departamento según se informa en RMP.
- **Ciencias:** variable ficticia que toma el valor 1 si el departamento del profesor forma parte de una escuela de ciencias. Fuente: clasificación manual a partir de la denominación del departamento según se informa en RMP.
- **Salud:** variable ficticia que toma el valor 1 si el departamento del profesor forma parte de una escuela de salud y bienestar. Fuente: clasificación manual a partir de la denominación del departamento según se informa en RMP.
- **Otros:** variable ficticia que toma el valor 1 si el departamento del profesor forma parte de una escuela de educación, derecho, agricultura, asesoría psicológica o deportes. Fuente: clasificación manual a partir de la denominación del departamento según se informa en RMP.
- **No clasificado:** variable ficticia que toma el valor 1 si el departamento del profesor no pudo clasificarse en ninguna de las facultades o escuelas mencionadas anteriormente. Fuente: clasificación manual a partir de la denominación del departamento según se informa en RMP.

## Anexo A2

### Instrucciones para los asistentes de investigación que participen en el proyecto

Es muy importante que lea todas estas instrucciones antes de empezar a recopilar la información necesaria.

#### Introducción

A cada asistente de investigación que colabore en este proyecto se le asignará un grupo de profesores. Usted dispondrá de información de identificación para cada profesor, como el nombre y la universidad en la que da clases, así como de un identificador numérico único asignado a ese profesor (ID). A partir de la información de identificación, tiene que encontrar dos fotos actuales del profesor y guardarlas con los nombres ID\_1 e ID\_2. A continuación, debe completar una base de datos con variables relacionadas con las características físicas del profesor. A continuación, se presenta una lista de las variables. Cuando se precise una evaluación cualitativa de la apariencia física de un profesor, sea lo más objetivo posible.

#### Variables

##### A. Género

- 0. Mujer
- 1. Hombre

##### B. Color de la piel

- 0. Clara
- 1. Morena
- 2. Oscura

##### C. Raza

- 0. Caucásico/Blanco/Latino
- 1. Africano/Caribeño
- 2. Indio/Árabe
- 3. Chino/Japonés/Asiático

##### D. Edad

- 0. Menor de 40 años
- 1. Entre 40 y 55 años
- 2. Mayor de 55 años

## E. Color de ojos

0. Negro/marrón
1. Verde/azul

## F. Color de pelo

0. Negro
1. Rubio
2. Castaño
3. Gris
4. Rojo

## G. Calvo

0. Sí
1. No

## H. Peso

0. Delgado/bajo peso
1. Complexión media
2. Sobrepeso

## I. Altura

0. Alto
1. Media
2. Bajo
3. No disponible

## J. Apariencia

0. Atractivo
1. Media
2. Poco atractivo

## K. Amigable

0. Amigable
1. No amigable



## Anexo A3

### Variables para la muestra de 2.835 profesores con un perfil en Google Académico

#### Características académicas adicionales:

- **Número de publicaciones:** número de artículos escritos por el profesor. Fuente: Google Académico.
- **Citas:** número de citas de todos los artículos escritos por el profesor. Fuente: Google Académico.
- **Índice H:** índice que mide la productividad y el impacto de un profesor. Un profesor con un índice H ha publicado H artículos, cada uno de los cuales ha sido citado al menos H veces. Fuente: Google Académico.
- **Índice I10:** número de artículos con al menos diez citas. Fuente: Google Académico.

#### Características personales adicionales:

- **Menor de 40 años:** variable ficticia que toma el valor 1 si el profesor tiene menos de 40 años. Fuente: inspección visual de dos fotografías del profesor.
- **Entre 40 y 55 años:** variable ficticia que toma el valor 1 si el profesor tiene entre 40 y 55 años. Fuente: inspección visual de dos fotografías del profesor.
- **Calvo:** variable ficticia que toma el valor 1 si el profesor es calvo. Fuente: inspección visual de dos fotografías del profesor.
- **Amigable:** variable ficticia que toma el valor 1 si el profesor tiene una apariencia amigable. Fuente: inspección visual de dos fotografías del profesor.
- **Ojos claros:** variable ficticia que toma el valor 1 si el profesor tiene ojos claros (por ejemplo, verdes o azules). Fuente: inspección visual de dos fotografías del profesor.
- **Caucásico:** variable ficticia que toma el valor 1 si el profesor tiene apariencia caucásica, blanca o latina. Fuente: inspección visual de dos fotografías del profesor.
- **Afroamericano:** variable ficticia que toma el valor 1 si el profesor tiene apariencia afroamericana o caribeña. Fuente: inspección visual de dos fotografías del profesor.
- **Indio:** variable ficticia que toma el valor 1 si el profesor tiene apariencia india o árabe. Fuente: inspección visual de dos fotografías del profesor.
- **Piel clara:** variable ficticia que toma el valor 1 si el profesor tiene piel clara o pálida. Fuente: inspección visual de dos fotografías del profesor.
- **Piel morena:** variable ficticia que toma el valor 1 si el profesor tiene piel morena u oscura. Fuente: inspección visual de dos fotografías del profesor.
- **Pelo negro:** variable ficticia que toma el valor 1 si el profesor tiene pelo negro. Fuente: inspección visual de dos fotografías del profesor.
- **Pelo rubio:** variable ficticia que toma el valor 1 si el profesor tiene pelo rubio. Fuente: inspección visual de dos fotografías del profesor.
- **Pelo castaño:** variable ficticia que toma el valor 1 si el profesor tiene pelo castaño o marrón. Fuente: inspección visual de dos fotografías del profesor.

- **Pelo gris:** variable ficticia que toma el valor 1 si el profesor tiene pelo gris. Fuente: inspección visual de dos fotografías del profesor.
- **Delgado:** variable ficticia que toma el valor 1 si el profesor tiene apariencia delgada o de bajo peso. Fuente: inspección visual de dos fotografías del profesor.
- **Complexión media:** variable ficticia que toma el valor 1 si el profesor parece tener una complexión media o un peso saludable. Fuente: inspección visual de dos fotografías del profesor.
- **Alto:** variable ficticia que toma el valor 1 si el profesor parece alto. Fuente: inspección visual de dos fotografías del profesor.
- **Altura media:** variable ficticia que toma el valor 1 si el profesor parece tener una altura media. Fuente: inspección visual de dos fotografías del profesor.
- **Buena apariencia:** variable ficticia que toma el valor 1 si el profesor parece atractivo. Fuente: inspección visual de dos fotografías del profesor.
- **Apariencia media:** variable ficticia que toma el valor 1 si el profesor tiene una apariencia media. Fuente: inspección visual de dos fotografías del profesor.



[www.cepal.org/revista](http://www.cepal.org/revista)



NACIONES UNIDAS

CEPAL

COMISIÓN ECONÓMICA PARA AMÉRICA LATINA Y EL CARIBE