

Habilidades blandas y técnicas en evaluaciones de estudiantes a docentes

Nigel Manchini, Óliver Jiménez y Natalia Ramos-Díaz

RESUMEN

¿Qué importancia asigna el estudiantado a las “habilidades blandas” o *soft skills* y a las “habilidades técnicas” o *hard skills* cuándo evalúan a sus docentes? Utilizando técnicas de análisis textual en una muestra de calificaciones y comentarios a docentes, se exploró cualitativamente y se contrastaron hipótesis cuantitativas sobre el vocabulario que acompaña a calificaciones altas y bajas. Los resultados muestran que el lenguaje de los comentarios y la calificación se encuentran significativamente relacionados, y que las referencias a *soft skills* predicen mejor la calificación que las referencias a *hard skills*; sin embargo, los análisis cualitativos sugieren que, en docentes calificados/as como excelentes, las *hard skills* no son ignoradas sino presupuestas. Además de su contribución metodológica, este estudio aporta implementos tanto a la discusión acerca de los sesgos en la evaluación docente como a la discusión acerca de qué cualidades del profesorado es deseable cultivar.

Palabras clave: actitudes del profesor, evaluación del profesor, análisis de textos, habilidades sociales, conocimiento profesional, Uruguay.

Nigel Manchini

nigelmanchini@uma.es

Uruguayo-italiano. Magister en Neuropsicología y Educación por la Universidad Internacional de La Rioja, España. Docente de Didáctica de la Filosofía, Consejo de Formación en Educación, Uruguay; investigador pre-doctoral, Universidad de Málaga, España. Temas de investigación: emociones y otros factores personales en educación, análisis cuantitativo del lenguaje y enseñanza de la Filosofía. ORCID: <https://orcid.org/0000-0002-1221-9259>.

Óliver Jiménez

oliverjjimenez@uma.es

Español. Doctor en Psicología, Universidad de Málaga (UMA), España; Máster en Terapias contextuales, Universidad de Almería, España. Profesor Interino, Departamento de Personalidad, Evaluación y Tratamiento Psicológico, Facultad de Psicología y Logopedia, UMA, España. Temas de investigación: *mindfulness* y gestión emocional en contextos terapéuticos, laborales y educativos. ORCID: <https://orcid.org/0000-0003-0314-8396>.

Natalia Ramos-Díaz

nsramos@uma.es

Española. Doctora en Psicología, Universidad de Málaga (UMA), España. Profesora titular, Departamento de Personalidad, Evaluación y Tratamiento Psicológico, Facultad de Psicología y Logopedia; Directora de Pinep Experience, S.L. Empresa de Base en el Conocimiento, UMA, España. Temas de investigación: gestión emocional consciente, Programa Inteligencia Emocional Plena (PINEP), emociones, meditación, *mindfulness*, acompañamiento emocional y Técnica de Acompañamiento Emocional Compasivo (TAEC). ORCID: <https://orcid.org/0000-0003-4269-8948>.

Habilidades Interpessoais e Técnicas nas Avaliações Aluno-Professor

RESUMO

Que importância os alunos atribuem às “habilidades interpessoais” ou *soft skills* e às “habilidades técnicas” ou *hard skills* quando avaliam seus professores? Usando técnicas de análise textual em uma amostra de notas e comentários aos professores, hipóteses quantitativas sobre o vocabulário que acompanha as notas altas e baixas foram qualitativamente exploradas e contrastadas. Os resultados mostram que a linguagem dos comentários e a classificação estão significativamente relacionadas, e que as referências a *soft skills* predizem melhor a nota do que as referências a *hard skills*; no entanto, análises qualitativas sugerem que, em professores classificados como excelentes, as *hard skills* não são ignoradas, mas pressupostas. Além de sua contribuição metodológica, este estudo contribui tanto para a discussão sobre vieses na avaliação de professores quanto para a discussão sobre quais qualidades dos professores é desejável cultivar.

Palavras chave: atitudes docentes, avaliação docente, análise de textos, habilidades sociais, conhecimento profissional, Uruguai.

Soft Skills and Technical Skills in Teacher Evaluations Carried out by Students

ABSTRACT

What is the significance of both soft skills and hard skills in teacher evaluations conducted by students? Through the application of textual analysis methods to a set of teacher ratings and comments, the researchers conducted a qualitative examination and tested quantitative hypotheses regarding the language used in high and low ratings. The findings indicate a significant relationship between the language used in the comments and the ratings. It was observed that references to soft skills serve as better indicators of the ratings compared to references to hard skills. However, qualitative analyses suggest that in the case of teachers rated as excellent, hard skills are not overlooked but rather presumed. In addition to its methodological contribution, this study adds to the discourse on biases in teacher evaluation and the essential qualities to be nurtured in teachers.

Key words: teacher attitudes, teacher assessment, textual analysis, social skills, professional knowledge, Uruguay.

Recepción: 03/08/23. **Aprobación:** 22/04/24.



Introducción

Las características del cuerpo docente son condicionantes fundamentales tanto del aprendizaje como del bienestar y el clima escolar (Bardach *et al.*, 2022; Burroughs *et al.*, 2019; Hattie, 2009; Thapa *et al.*, 2013). La *excelencia docente* es un asunto relevante a nivel práctico, sin embargo, es un fenómeno complejo, multidimensional y de definición problemática. Este trabajo aborda una de sus dimensiones: la percepción del estudiantado de las cualidades docentes. Para hacerlo, tuvo por objetivos preparar una base de datos de evaluaciones a docentes, analizarla cuantitativa y cualitativamente, y determinar la relación entre la calificación y la proporción de vocabulario referido a *soft skills* (SS) y *hard skills* (HS), comparando el poder predictivo de ambas categorías.

Cualidades docentes y excelencia

Los trabajos de revisión sobre la excelencia docente —operacionalizada mediante medidas institucionales, de autorreporte y de reporte de estudiantes, superiores, colegas, etcétera— muestran la importancia de considerar una amplia gama de cualidades (Bardach *et al.*, 2022; Bardach y Klassen, 2020; Cornelius-White, 2007; Hattie, 2009; Kaya y Selvitopu, 2019; Kim *et al.*, 2019; Klassen y Tze, 2014; Roorda *et al.*, 2011; Lei *et al.*, 2017).

Factores cognitivos —como el cociente intelectual, calificaciones académicas o resultados de pruebas estandarizadas— han mostrado una asociación pequeña o nula con la efectividad docente (Bardach y Klassen, 2020; cf. Harris y Rutledge, 2010). Otras medidas, como el conocimiento sobre la materia impartida o la formación docente, han mostrado efectos significativos pero bajos, aunque variables asociadas —como el desarrollo profesional, las posibilidades de actualización en el ejercicio de la profesión— han mostrado una influencia mayor (Hattie, 2009; Sancar *et al.*, 2021).

Metaanálisis recientes sugieren que características como la personalidad y la autoeficacia se relacionan

con variables importantes como la efectividad docente (Kim *et al.*, 2019; Klassen y Tze, 2014), el manejo de clase (Kaya y Selvitopu, 2019) y el *burn-out* (Ziyan *et al.*, 2022). La revisión integrativa de Bardach *et al.* (2022) sugiere que varias características psicológicas (como la autoeficacia, la personalidad, la atención plena, el entusiasmo y la inteligencia emocional, entre otras) se relacionan con la efectividad, la calidad de las relaciones interpersonales y el bienestar docente. La relación docente-estudiante ha demostrado ser importante para resultados cognitivos, afectivos y comportamentales como el éxito académico, el compromiso con la escuela y las emociones académicas (Cornelius-White, 2007; Lei *et al.*, 2017; Roorda *et al.*, 2011).

Habilidades blandas y técnicas

En la discusión educativa reciente, el contraste entre cualidades intelectuales/técnicas y personales/sociales ocupa un lugar central, tanto por su relevancia práctica como por las controversias que genera (Scheerens *et al.*, 2020). Aunque el vocabulario de la discusión sea reciente (Manchini *et al.*, 2022), la tensión no es nueva: puede retrotraerse, al menos, a la distinción de Aristóteles (ca. 350 ANE/2005) entre excelencias o virtudes intelectuales (*dianoéticas*, como el conocimiento y la técnica) y del carácter (*éticas*, como la moderación o la amabilidad). Para abordar este asunto específicamente en docentes, este trabajo se centra en las revisiones sistemáticas de Jerez *et al.* (2016) y Manchini *et al.* (2022).

Jerez *et al.* (2016) encuentra que los 64 atributos identificados en la literatura pueden categorizarse en seis grupos, reunidos a su vez en tres grandes competencias docentes: *genéricas*, *pedagógicas* y *disciplinarias*. Las competencias *genéricas* —subdivididas en características actitudinales, personales y comunicativas— son cualidades transversales con muchos ámbitos de aplicación, como la asertividad, la escucha activa, la inteligencia emocional, la integridad y la responsabilidad. Las competencias *pedagógicas*

—divididas en estrategias de enseñanza/aprendizaje y de planificación/gestión— son específicas del contexto educativo; ejemplos de estas competencias serían dar explicaciones claras, potenciar la autonomía y el pensamiento crítico, la coherencia, actividades/evaluación y la capacidad de preparar, organizar y gestionar el tiempo pedagógico. Las competencias *disciplinares* son las competencias relacionadas con el área disciplinar que se enseña; por ejemplo, el ser experto en el área de conocimiento, la actualización, la experiencia profesional y la capacidad investigadora.

El estudio de Manchini *et al.* (2022) revisa la literatura empírica sobre Habilidades Blandas o *soft skills* (SS) en el campo de la docencia. Mediante análisis de definiciones y ejemplos, el trabajo identifica que el término *soft skills* es usado principalmente para referirse a cualidades personales, actitudinales y sociales; algunos ejemplos serían la comunicación efectiva, el trabajo en equipo, el pensamiento crítico y la inteligencia emocional. Se diferencian por su transversalidad, mientras que el término *hard skills* (HS) se refiere a conocimientos y habilidades técnicas, propias de la profesión docente y la disciplina enseñada, como el uso de estrategias pedagógicas y el saber disciplinar.

Considerando estos antecedentes, la multiplicidad de cualidades docentes puede ordenarse jerárquicamente, en niveles crecientes de abstracción:

- En el nivel más bajo se ubican los ejemplos concretos de cualidades docentes: responsabilidad, conocimiento, gestión del tiempo, empatía, puntualidad, etcétera.
- A un nivel medio, esos ejemplos se pueden clasificar en los seis grupos propuestos por Jerez *et al.* (2016): características de actitud, características de personalidad, características de comunicación, estrategias de enseñanza-aprendizaje, estrategias de planificación-gestión y saber disciplinar.
- En el nivel superior de abstracción se ubican las categorías generales *soft skills* y *hard skills*:

SS engloba a las cualidades personales de la/el docente (*características de actitud, personalidad y comunicación*), mientras que HS comprende a las cualidades profesionales específicas de la docencia (*estrategias de enseñanza-aprendizaje, estrategias de planificación-gestión y saber disciplinar*).

Evaluación de estudiantes a docentes

Aunque controvertida, la evaluación realizada por el estudiantado ha demostrado ser una fuente rica para evaluar las cualidades docentes —incluso cuando se trata de juicios globales (Hattie, 2009; Loureiro *et al.*, 2016)— y comprender, a nivel general, qué atributos hacen a la docencia de calidad (Jerez *et al.*, 2016). En las últimas décadas se han popularizado un conjunto de plataformas *online* (ejemplos, RateMyProfessor.com, MisProfesores.com) que permiten calificar y realizar comentarios anónimamente a docentes de distintos niveles educativos. Aunque es posible que esta modalidad de evaluación tenga sus propias dificultades —por ejemplo, dificultad para identificar evaluaciones falsas o realizadas con poca seriedad (Davison y Price, 2009)—, su funcionamiento parece ser equivalente a las evaluaciones docentes tradicionales (Brown *et al.*, 2009; Colardaci y Kornfield, 2007) y mostrar un poder predictivo no despreciable (Rosen, 2018; Sonntag *et al.*, 2009; Timmerman, 2008; Kindred y Mohammed, 2005). Por otro lado, también ha mostrado compartir con la evaluación docente tradicional la presencia constatable de sesgos de género, raciales y de atraktividad física (Arceo y Campos, 2019; Kreitzer y Sweet-Cushman, 2022; Murray *et al.*, 2020; Reid, 2010; Stuber *et al.*, 2009).

Estos antecedentes muestran que debe cuestionarse la validez de estas evaluaciones como indicador “objetivo” de la excelencia docente. Sin embargo, también sugieren que es posible utilizar las evaluaciones para investigar los sesgos y valoraciones del estudiantado. Como plantea Gregory (2012), el análisis del lenguaje puede utilizarse para comprender la “pedagogía intuitiva” o *folk pedagogy* del estudiantado



—sus creencias, valoraciones y percepciones implícitas acerca de una “buena” y “mala” docencia—. Concretamente, desde el enfoque de las palabras como atención (Boyd y Schwartz, 2021), el vocabulario puede utilizarse como un indicador de aquello a lo que las personas atienden. Independientemente de su conexión con indicadores externos de excelencia docente, entender la perspectiva del estudiantado se considera un bien en sí mismo —por ejemplo, para comprender mejor las dinámicas de la relación docente-estudiante (Quin, 2017; Split *et al.*, 2011).

En síntesis, parece posible analizar la relación entre la calificación de un docente y el lenguaje utilizado por el estudiantado en sus comentarios, obteniendo información cuantitativa y cualitativa sobre qué cualidades del docente son atendidas por el estudiantado.

Objetivos e hipótesis

Considerando todo lo comentado, este estudio se propuso abordar el vocabulario utilizado en MisProfesores.com Uruguay (<https://uruguay.misprofesores.com>), siendo —según nuestro conocimiento— el primer estudio que realiza un análisis sistemático de evaluaciones públicas *online* en español variante rioplatense. Se puso el foco en el vocabulario referente a *SS* (características personales, actitudinales y comunicacionales) y *HS* (estrategias de enseñanza-aprendizaje, de gestión-planificación y saber disciplinar). Los objetivos e hipótesis se pre-registraron en enero de 2022 (www.doi.org/10.17605/OSF.IO/SJH7N).

Los objetivos fueron:

- Preparar una base de datos de evaluaciones públicas *online* a docentes.
- Analizar cuantitativa y cualitativamente el vocabulario utilizado en comentarios, creando y validando un diccionario para categorizarlo considerando la distinción *SS/HS*.
- Analizar la relación entre la calificación y la proporción de vocabulario referido a *SS/*

HS, considerando el poder predictivo de cada categoría.

Las hipótesis pre-registradas fueron:

- H1. El diccionario mostrará propiedades aceptables (acuerdo interjueces, capacidad de captar el vocabulario frecuente, correlación alta con otros diccionarios, poder predictivo sobre la calificación).
- H2. Se encontrarán correlaciones significativas y moderadas entre la calificación y la proporción de términos referentes a *SS*.
- H3. El modelo de regresión lineal basado en las categorías del diccionario mostrará índices aceptables de bondad de ajuste.

Además del contraste de estas hipótesis, se reporta una serie de análisis cualitativos exploratorios que permiten interpretar con mayor profundidad los resultados cuantitativos.

Método Diseño

Se utilizó un diseño mixto, de análisis automatizado de texto con un enfoque *top-down* basado en diccionarios (Boyd y Schwartz, 2021; Boyd, 2017; Kennedy, 2021). El estudio fue pre-registrado en enero de 2022. El registro, los datos anonimizados, los materiales y los códigos se encuentran disponibles en www.osf.io/ymzp2.

Muestra

La muestra se compone de 3 946 docentes (con un total de 12 935 comentarios), de los cuales la mayoría trabaja en educación universitaria o equivalente ($n = 3262$), con una proporción menor perteneciente a educación secundaria ($n = 549$) y educación técnica ($n = 135$). Aunque corresponden a 193 instituciones distintas, el 53.6 % de los/as docentes pertenecen a las cinco instituciones más frecuentes: Facultad de

Derecho (Universidad de la República (UdelaR), $n = 730$, 18.5 %), Instituto de Profesores Artigas (Consejo de Formación en Educación, $n = 544$, 13.8 %), Facultad de Ingeniería (UdelaR, $n = 438$, 11.1 %), Facultad de Psicología (UdelaR, $n = 285$, 7.2 %) e Instituto Tecnológico Superior (Dirección General de Educación Técnico-Profesional, $n = 117.3$ %). Para cada docente se dispone de los siguientes datos: calificación, comentarios, institución, nivel educativo (universitario, secundario, entre otros) y área de conocimiento (Ciencias Sociales, Educación, Salud, etcétera). De la totalidad de la muestra, 281 casos se descartaron por no tener ningún término del diccionario, restando una muestra de $n = 3665$. Considerando que la puntuación se computa con base en la proporción de términos en cada categoría en los comentarios, para calcular el coeficiente de correlación y crear los modelos de regresión se seleccionan aquellos sujetos que en sus comentarios tienen al menos cuatro términos presentes en el diccionario ($n = 2367$).

Procedimiento

Los datos se descargaron en enero de 2022 y fueron procesados en R (R Core Team; RStudio Team, 2022), utilizando los principios de Wickham *et al.* (2019) y Silge y Robinson (2016), y un enfoque de conteo de palabras o *word-count* (Boyd, 2017; Boyd y Schwartz, 2021; Pennebaker, 2018; Ramírez-Esparza *et al.*, 2007; Tausczik y Pennebaker, 2010; Zhou y Ye, 2020). Se empleó un “diccionario de comentarios”, elaborado *ad hoc*, donde cada término (una palabra o un conjunto de ellas) se encuentra categorizado considerando a qué tipo de cualidad se refiere y su valencia (positiva/negativa). Por ejemplo, el

término “no sabe nada” se refiere al *saber disciplinar* y tiene valencia negativa. Para cada sujeto se calcula la proporción de términos pertenecientes a cada categoría en sus comentarios, es decir, si un sujeto tiene una puntuación de 0.17 en términos positivos referentes a la actitud y personalidad, significa que el 17 % de los términos captados por el diccionario en sus comentarios pertenecen a esa categoría y tienen una valencia positiva.

Instrumento

El diccionario de comentarios cuenta con 2 287 términos, incluyendo unigramas (por ejemplo, “explica”, “sabe”) y n-gramas de hasta seis palabras (por ejemplo, “sabe mucho”, “buena persona”, “no se entiende nada”), anotados según su valencia (“positiva”, “ambivalente”, “negativa”) y el tipo de cualidad a la que refieren. Como se ve en la tabla 1, se utilizó un modelo que considera dos categorías de alto nivel –“*soft skills*”, “*hard skills*”– cuatro categorías de nivel básico –“actitud-personalidad”, “comunicación”, “pedagógico-organizacional”, “saber disciplinar”– y una categoría de “calificadores”, términos que expresan una valoración general pero no son asimilables a ninguna de las categorías anteriormente mencionadas. Los términos fueron seleccionados manualmente (en cinco fases iterativas) de la lista de palabras y n-gramas con frecuencia > 2 en los comentarios. Con base en Jerez *et al.* (2016) y Manchini *et al.* (2022) se elaboró una guía de codificación que fue aplicada por dos codificadores independientes, con un tercer codificador resolviendo los desacuerdos mediante consulta a la guía y eliminando las palabras en caso de ambigüedad. El diccionario y una descripción detallada de su elaboración puede verse en www.osf.io/ymzp2.



Tabla 1. Categorías del diccionario

Categoría básica	Categoría alto nivel	Descripción	Ejemplos
Actitud-personalidad	Soft skills	Características actitudinales y de personalidad	falta, buena_persona, buena_onda, muy_exigente, responsable, soberbia, amable, comprometida.
Comunicación	Soft skills	Competencia comunicativa	transmite, no_se_le_entiende, escucha, contesta, sabe_transmitir, explica_bien, entendes.
Pedagógico-organizacional	Hard skills	Competencias de enseñanza-aprendizaje y de planificación-gestión	didáctica, entretenidas, práctico, dinámicas, lee, aburrida, no_enseña, calificación, sabe_enseñar, dicta.
Saber disciplinar	Hard skills	Conocimiento específico de la disciplina enseñada	sabe, sabe_mucho, conocimiento, conocimientos, sabe_muchísimo, sabe_nada, profesión, conoce, sabe_pero, sabe_pila
Calificador		Términos con valencia pero que no refieren a una cualidad concreta	excelente, recomendable, desastre, mala, genio, crack, horrible, gusto, lamentable, capo

Fuente: elaboración propia.

El coeficiente Kappa de Fleiss (k) indica una excelente fiabilidad interjueces ($k = 0.92$, $p < 0.001$), con valores de Kappa oscilando entre 0.90 y 0.96 para todas las categorías. La valencia del diccionario muestra una correlación muy alta ($\rho = 0.82$, $p < 0.001$) con la valencia del diccionario Stadthagen-Gonzalez *et al.* (2017). El diccionario tiene una correlación alta ($\rho = 0.63$, $p < .001$) con la puntuación media de cada palabra (la media de la calificación de todos los sujetos donde esa palabra aparece). Se calculó la correlación entre la calificación y la proporción de calificadores positivos ($r = 0.37$, $p < 0.001$) y negativos ($r = -0.29$, $p < 0.001$) para una pequeña muestra ($n = 214$) de docentes argentinos/as (donde se utiliza la misma variante del español), extraída de MisProfesores.com Argentina (<https://argentina.misprofesores.com/>). Tomados en conjunto, estos datos sugieren la validez del diccionario.

Análisis

Se calculó la correlación (r) entre la calificación y la proporción de cada término en cada categoría; la interpretación del tamaño del efecto se hizo siguiendo a Funder y Ozer (2019) y Hemphill (2003). Se realizaron regresiones lineales múltiples considerando el conjunto de las variables y a cada una

independientemente; siguiendo a Ozili (2023), valores de R^2 entre 0.1 y 0.5 se consideran aceptables (en el contexto de las ciencias sociales) si todas las variables explicativas del modelo son estadísticamente significativas.

Para abordar cualitativamente el vocabulario se utilizaron técnicas de minería de textos (Silge y Robinson, 2016). Se calculó, para cada término, la frecuencia, la media (M) y la desviación estándar (DE) de la calificación de los sujetos donde el término aparece, utilizándose M como indicador de la valencia de ese término, y la DE como indicador de la ambigüedad de esa valencia. Para explorar el vocabulario distintivo de docentes con calificación alta (centil > 0.66) y baja (centil < 0.33), se calculó el estadístico TF-IDF (*Term Frequency - Inverse Document Frequency*) (Qaiser y Ali, 2018; Silge y Robinson, 2016), que permite identificar palabras especialmente relevantes para un conjunto de textos (en este caso, los comentarios de docentes con alta/baja calificación). Para determinar la co-ocurrencia de los distintos términos con calificadores, se calculó el coeficiente de correlación Φ (ϕ_i), equivalente al coeficiente r de Pearson para variables binarias (Silge y Robinson, 2016). Dado el carácter exploratorio de estos análisis cualitativos, no se reporta la significatividad estadística (p -valor), siguiendo a Murray *et al.* (2020).

Resultados Descriptivos

Después de la aplicación del diccionario, se conservan los datos de 3 665 sujetos que tienen entre 1 y 51 ($M = 7.05$, $DE = 6.12$) términos captados por el diccionario; para los análisis cuantitativos se

seleccionan sujetos con al menos cuatro términos ($n = 2367$). La tabla 2 sintetiza los estadísticos descriptivos para cada una de las variables. Se observa que en todos los casos el vocabulario positivo es más frecuente que el negativo.

Tabla 2. Estadísticos descriptivos

Categoría básica	Frecuencia	Sujetos con mención	M (n = 3665)	DE (n = 3665)	M (n = 2367)	DE (n = 2367)
Calificación			6.91	2.30	6.79	2.20
Calificadores positivos	4 265	2 101	0.20	0.27	0.17	0.17
Calificadores negativos	3 001	1 462	0.11	0.19	0.12	0.15
Calificadores ambivalentes	100	94	0.01	0.04	0.00	0.03
Actitud-personalidad positivo	4 306	2 055	0.17	0.22	0.17	0.16
Actitud-personalidad negativo	3 056	1 451	0.11	0.19	0.11	0.15
Actitud-personalidad ambivalente	1 009	760	0.04	0.11	0.04	0.07
Comunicación positivo	1 121	831	0.04	0.11	0.04	0.08
Comunicación negativo	982	717	0.04	0.10	0.04	0.08
Comunicación ambivalente	747	611	0.03	0.09	0.03	0.07
Pedagógico-organizacional positivo	2 384	1 412	0.05	0.12	0.06	0.09
Pedagógico-organizacional negativo	1 364	1 070	0.08	0.15	0.09	0.11
Pedagógico-organizacional ambivalente	1 599	913	0.05	0.13	0.05	0.09
Saber disciplinar positivo	873	671	0.03	0.10	0.03	0.07
Saber disciplinar negativo	189	168	0.01	0.05	0.01	0.04
Saber disciplinar ambivalente	858	658	0.03	0.08	0.03	0.07
ss positivo	5 427	2 886	0.21	0.25	0.21	0.18
ss negativo	4 038	2 168	0.14	0.21	0.15	0.17
hs positivo	3 257	2 083	0.12	0.18	0.12	0.13
hs negativo	1 553	1 081	0.06	0.14	0.06	0.10

Nota: frecuencia: total de términos en cada categoría; sujetos con mención: total de sujetos con al menos un término en la categoría; m: Media; DE: Desviación Estándar. Fuente: elaboración propia.



Correlación con la calificación

Como se ve en la tabla 3, todas las variables tuvieron correlaciones estadísticamente significativas con la calificación, con un tamaño del efecto oscilando entre la correlación muy débil del *saber disciplinar* ($r =$

0.06) y la correlación fuerte de los *calificadores positivos* ($r = 0.50$). Exceptuando el caso de los calificadores (donde la correlación es virtualmente equivalente), en todas las variables la correlación es más fuerte para el vocabulario negativo que para el positivo.

Tabla 3. Correlaciones con la calificación ($n = 2\ 367$)

Variable	r	p	95 % CI	
Calificador negativo	-0.49	<0.001	-0.52	-0.46
Calificador positivo	0.50	<0.001	0.47	0.53
Actitud-personalidad negativo	-0.41	<0.001	-0.44	-0.37
Actitud-personalidad positivo	0.36	<0.001	0.33	0.40
Comunicación negativo	-0.25	<0.001	-0.29	-0.21
Comunicación positivo	0.13	<0.001	0.09	0.17
Pedagógico-organizacional negativo	-0.28	<0.001	-0.31	-0.24
Pedagógico-organizacional positivo	0.20	<0.001	0.16	0.24
Saber disciplinar negativo	-0.17	<0.001	-0.20	-0.13
Saber disciplinar positivo	0.06	0.004	0.02	0.10
SS negativo	-0.47	<0.001	-0.50	-0.44
SS positivo	0.39	<0.001	0.35	0.42
HS negativo	-0.32	<0.001	-0.35	-0.28
HS positivo	0.21	<0.001	0.17	0.25

Fuente: elaboración propia.

Regresiones lineales múltiples

Se realizaron regresiones lineales múltiples para evaluar qué porcentaje de la varianza se explica considerando separadamente las categorías de alto y de bajo nivel. Como se observa en la tabla 4, el conjunto de todas las variables (Calificadores, *SS* y *HS*) puede explicar un 48 % de la varianza en la calificación; contemplar únicamente el uso de calificadores ya tiene la capacidad de explicar el 36 % de la varianza. Dentro de las cualidades docentes, el

modelo que estudia solo *SS* tiene un mejor ajuste (explicando el 27 % de la varianza) que el que considera *HS* (que explica el 12.7 %). Contempladas aisladamente, las referencias a la *actitud-personalidad* explican el 23 %, la competencia *pedagógico-organizacional* un 11 %, la *comunicación* el 8 % y el *saber disciplinar* un 3 % de la varianza. Cabe mencionar que en estas últimas dos variables no se alcanza el nivel mínimo de varianza explicada (0.1) para considerar que el modelo es aceptable (Ozili, 2023).

Tabla 4. Regresión lineal múltiple

Modelo	R ² (ajust.)	F	df	p	Variables	β	t	p
Soft skills	0.272	442.4	2364	<0.001	ss negativo	-4.98	12.72	<0.001
					ss positivo	2.99	-19.88	<0.001
Hard skills	0.127	173.6	2364	<0.001	hs negativo	-6.48	8.64	<0.001
					hs positivo	2.83	-15.16	<0.001
Actitud-personalidad	0.229	353.3	2364	<0.001	A-P negativo	-4.94	-17.38	<0.001
					A-P positivo	3.62	14.05	<0.001
Comunicación	0.078	100.8	2364	<0.001	Com. negativo	-7.17	-12.61	<0.001
					Com. positivo	3.34	6.29	<0.001
Pedagógico-organizacional	0.11	146	2364	<0.001	P-O negativo	-6.36	-13.56	<0.001
					P-O positivo	3.54	9.04	<0.001
Saber disciplinar	0.031	38.55	2364	<0.001	SD negativo	-9.74	-8.29	<0.001
					SD positivo	1.70	2.74	0.006
Calificadores	0.362	674.7	2364	<0.001	Cal. negativo	-5.34	-20.54	<0.001
					Cal. positivo	4.67	21.12	<0.001
Soft Skills, Hard Skills y Calificadores	0.482	368.7	2360	<0.001	Cal. negativo	-4.16	-13.96	<0.001
					Cal. positivo	3.02	11.13	<0.001
					ss negativo	-3.13	-11.03	<0.001
					ss positivo	1.59	5.94	<0.001
					hs negativo	-3.50	-9.00	<0.001
					hs positivo	1.05	3.21	0.001

Fuente: elaboración propia.

Análisis cualitativo

Para abordar de manera cualitativa el vocabulario utilizado, se calculó para cada palabra la calificación media (M) de los/as docentes en cuyos comentarios aparecía esa palabra y su desviación estándar (DE). De esa manera, podemos utilizar M como una medida *bottom-up* de la valencia, y DE como un indicador de la ambigüedad de esa valencia; para simplificar la interpretación, los valores de la DE se estandarizaron (calculando los valores z), y se consideraron

términos poco ambiguos aquellos con una DE estandarizada < -0.2 .

En la figura 1 puede verse el vocabulario más frecuentemente utilizado y su calificación media. Como se observa, las referencias a las explicaciones, el conocimiento, el saber y la actitud ante la enseñanza son las más frecuentes (figura 1, izquierda). Sin embargo, aunque algunas de estas palabras tienen una valencia bastante polarizada, todas tienen un nivel de ambigüedad importante. Cuando seleccionamos

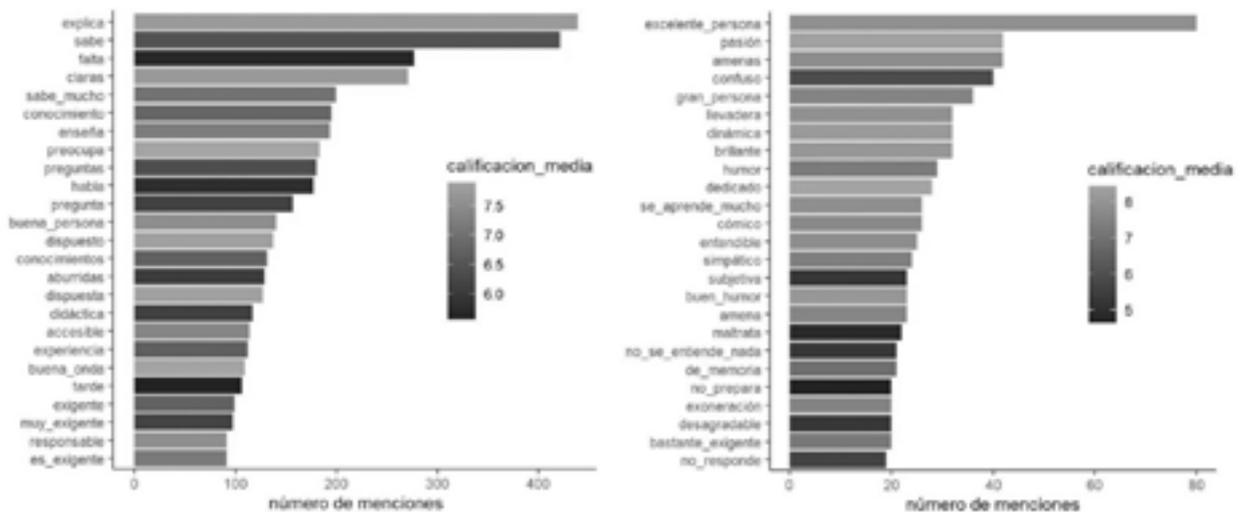


el vocabulario menos ambiguo (DE estandarizada < -0.2), encontramos que el rango de calificación aumenta y se observa un vocabulario más específico de docentes excelentes como “excelente persona”, “pasión”, “amenas”, “gran persona” y “brillante”, o docentes nada excelentes, como “confuso”, “subjetiva”, “maltrata”, “no se entiende nada” o “no prepara” (figura 1, derecha).

La figura 2 sintetiza información relevante para

entender cualitativamente la variedad del vocabulario utilizado dentro de cada una de las categorías. Se aprecia que las referencias a la *actitud-personalidad* y la competencia *pedagógico-organizacional* tienden a mostrar una mayor cantidad de matices, mientras que las referencias a las características de *comunicación* y el *saber disciplinar* tienden a presentarse como variaciones de algunas fórmulas comunes (respectivamente “explica”/“se entiende” y “sabe”/“conoce”).

Figura 1. Vocabulario frecuente (izquierda) y vocabulario frecuente con baja ambigüedad en su calificación media (derecha)



Fuente: elaboración propia.

Figura 2. Términos frecuentes por categoría



Nota: el tamaño indica la frecuencia del término en la categoría; la intensidad del color indica la calificación media del término (mayor intensidad, menor calificación). Fuente: elaboración propia.

Para explorar con más precisión qué términos distinguen a docentes de mayor y menor puntuación (pertenecientes al tercio superior e inferior en cuanto a puntuación) se calculó el estadístico TF-IDF. Los términos con mayor TF-IDF para el grupo de docentes con calificación baja fueron “no la recomiendo” (TF-IDF = 0.00033), “confusa” (TF-IDF = 0.00013), “irrespetuosa” (TF-IDF = 0.00011), “habla mal” (TF-IDF = 0.00009) y “mínimo” (TF-IDF = 0.00009). Para los/as docentes con mayor calificación los términos fueron “amamos” (TF-IDF = 0.00012), “admirable” (TF-IDF = 0.000082), “útiles” (TF-IDF = 0.000082), “apasionado” (TF-IDF = 0.000073) y “empática”

(TF-IDF = 0.000065). Debe considerarse que en todos los casos estos términos son poco frecuentes por lo cual hay que interpretarlos con cautela.

Para tener una aproximación desde los propios textos se calculó el coeficiente Φ , que expresa la correlación entre dos variables binarias (en este caso, la presencia o ausencia de un término en los comentarios), y se seleccionaron los términos que mayor correlación tienen con los diez calificadores más frecuentes, positivos (como “excelente”, “recomendable”, “genia”) y negativos (por ejemplo, “desastre”, “lamentable”, “no la recomiendo”).

Tabla 5. Términos correlacionados con calificadores positivos (izquierda) y negativos (derecha)

Término	Φ	Término	Φ
accesibles	0.16	mala_persona	0.16
claras	0.15	tarde	0.14
accesible	0.13	maltrata	0.13
se_aprende_mucho	0.11	densas	0.13
placer	0.11	vergüenza	0.12
se_aprende	0.11	pregunta	0.12
humano	0.11	pésima	0.12
dispuesto	0.11	anécdotas	0.12
pregunta	0.11	favoritos	0.12
motivación	0.10	soberbia	0.11
sabe_mucho	0.09	confusa	0.11
te_explica	0.09	interesada	0.11
dispuesta	0.09	pesadas	0.11
salvas	0.09	no_respeto	0.10
entendibles	0.09	exoneré	0.10

Fuente: elaboración propia.



Discusión

Cumpliendo los objetivos planteados, este estudio preparó una base de datos de evaluaciones a docentes y la analizó con un enfoque de conteo de palabras utilizando un diccionario *ad hoc*. Los resultados obtenidos sugieren que el análisis de datos textuales es un método viable para abordar la perspectiva de los/as estudiantes acerca de la docencia de calidad. El poder predictivo y la presencia de correlaciones de fuerza considerable (Funder y Ozer, 2019; Hemphill, 2003) indica que los efectos son lo suficientemente robustos como para ser captados, incluso utilizando un enfoque simplificador como el de contar palabras ignorando su contexto (Boyd y Schwartz, 2021).

El diccionario muestra propiedades aceptables (fiabilidad interjueces, correlaciones con otros diccionarios y poder predictivo de la calificación en la muestra uruguaya y una muestra argentina), consistentemente con la hipótesis 1; su disponibilidad pública allana el terreno para utilizar y refinar este instrumento en futuros estudios en la variante rioplatense del español. En línea con la hipótesis 2, las referencias a *SS* mostraron correlaciones significativas con la calificación, con una fuerza moderada ($r = 0.39$) en el vocabulario positivo y moderada-fuerte ($r = -0.47$) en el vocabulario negativo. El modelo de regresión lineal múltiple contemplando las categorías del diccionario muestra una bondad de ajuste aceptable, explicando el 48 % de la varianza de la calificación, consistentemente con la hipótesis 3.

En coincidencia con trabajos anteriores (Jerez *et al.*, 2016; Murray *et al.*, 2020), los resultados sugieren que las *SS* son más atendidas que las *HS* por el estudiantado a la hora de calificar a sus docentes: las referencias a *SS* (R^2 ajust. = 0.27) muestran un poder predictivo considerablemente mayor que las referencias a *HS* (R^2 ajust. = 0.13).

Estos resultados generales toman matices interesantes e informativos cuando se les contempla con mayor detalle a la luz de los resultados cualitativos. En las siguientes secciones la discusión se organiza

en torno a las categorías de bajo nivel (*calificadores*, *saber disciplinar*, competencia *pedagógico-organizacional*, características de *actitud-personalidad* y de *comunicación*).

Vocabulario positivo/negativo y calificadores

Los calificadores, predicablemente, muestran una relación fuerte con la calificación. Sus resultados, fácilmente interpretables, pueden tomarse como referencia para interpretar sustantivamente los tamaños de efecto, complementariamente a guías más generales (Funder y Ozer, 2019). Por ejemplo, una correlación como la de los términos positivos asociados con la actitud y personalidad ($r = 0.36$) sería tradicionalmente considerada como moderada (Hemphill, 2003); sin embargo, esa relación luce fuerte a la vista de que términos tan claramente asociados con la calificación como “no la recomiendo”, “desastre”, “lamentable” y “horrible” (calificadores negativos) tienen una correlación de $r = -0.49$.

Al respecto de la valencia, se constata una tendencia consistente: aunque el uso de vocabulario positivo es más frecuente que el negativo (tabla 2, figura 1), los términos negativos muestran una mayor fuerza de correlación y poder predictivo (tablas 3 y 4).

Saber disciplinar

El caso del vocabulario asociado al conocimiento es especialmente interesante: el vocabulario positivo tiene una relación muy baja con la calificación ($r = 0.06$, $p = 0.004$, 95 % CI (*Confidence Interval*) [0.02, 0.1]), mientras que el negativo muestra una relación baja pero no despreciable ($r = -0.17$, $p < 0.001$, 95 % CI [-0.20, -0.13]). Que se diga que un docente “no sabe” sugiere que no es excelente; que se diga que “sabe mucho” no alcanza para afirmar la excelencia. Igualmente, se observa que aunque los términos vinculados al saber y el conocimiento se encuentran entre los más frecuentes (figura 1; fenómeno explicado en parte por la poca variabilidad en el vocabulario al respecto, figura 2), muchos de ellos son ambiguos

en cuanto a su valencia: aparecen en docentes con diversas calificaciones.

En concordancia con lo que sugieren otros enfoques (Bardach y Klassen, 2020; Hattie, 2009), estos datos parecen consistentes con la afirmación de que la posesión del conocimiento es una condición necesaria, pero no suficiente, para la excelencia. En ese sentido, la baja frecuencia y el escaso poder predictivo de términos vinculados al *saber disciplinar* (tablas 2, 3 y 4), junto a su presencia entre los términos más frecuentes puede interpretarse de la siguiente manera: el saber docente es aceptado implícitamente y solo se lo atiende cuando parece fallar. Cuando se califica excelentemente a un/a docente y se dice que es “excelente persona”, tiene “pasión” y sus clases son “amenas” (figura 1), probablemente se está dando por supuesto que, además, tiene un conocimiento disciplinar suficiente. Estos resultados concuerdan con los hallazgos de Murray *et al.* (2020), según los cuales las evaluaciones positivas tienden a enfatizar atributos personales, mientras que las negativas tienden a enfatizar atributos profesionales.

Competencia pedagógico-organizacional

Las competencias vinculadas con la enseñanza, el aprendizaje y la gestión de lo educativo, representan la segunda categoría con mayor correlación y poder predictivo sobre la calificación. Estos datos van en línea con la literatura que sugiere que variables vinculadas al saber enseñar (como la experiencia, el desarrollo profesional, la certificación como profesor/a y el conocimiento pedagógico específico) y las decisiones pedagógicas concretas, como el tiempo dedicado a tareas de aprendizaje en relación con el tiempo utilizado para otras tareas (pasaje de lista, tareas administrativas, preparación de clima de aula, etcétera), tienen efectos consistentes en el aprendizaje y la efectividad percibida (Burroughs, 2019; Bruns y Luque, 2014; Hattie, 2009). También la literatura centrada en el reporte del estudiantado sugiere que las competencias pedagógicas son centrales en

la percepción del personal docente como excelente (Jerez *et al.*, 2016).

En términos cualitativos, el vocabulario más frecuente se refiere a tópicos esperables dentro de la categoría: si las clases son “claras” o no, si son “dinámicas” o “aburridas” y si el/la docente “enseña” bien. Particularmente interesantes son las referencias a “pregunta” y “preguntas”: se encuentran entre los más frecuentes dentro de la categoría (figura 2) y dentro de los términos con mayor correlación tanto con los calificadores negativos como positivos (tabla 5). En conjunto con las frecuentes referencias a la “exigencia” (en la categoría “actitud-personalidad”, figura 2), y la referencia con valencia negativa a la “subjetividad” (figura 1), estos resultados enfatizan el lugar central que tiene la evaluación en la perspectiva de los/as estudiantes. Esta constatación es consistente con los hallazgos de Azab *et al.* (2016) en su estudio a gran escala de RateMyProfessor.com (plataforma en inglés equivalente a MisProfesores.com) donde palabras como “*exams*”, “*questions*”, “*hard*”, “*answer*” y “*tricky*” formaron parte de uno de los tópicos principales; igualmente, los análisis de Murray *et al.* (2020) muestran que las etiquetas de “*test heavy*” y “*tough grader*” se relacionan con evaluaciones negativas.

Características de comunicación

Comparativamente, las características de *comunicación* se encuentran en tercer lugar tanto al respecto de la frecuencia como en sus correlaciones con la calificación. Aunque parecería que estas cualidades —centrales en la profesión docente— deberían tener correlaciones mayores, el razonamiento aplicado al saber disciplinar probablemente puede transferirse a estas cualidades. Es probable que en muchos casos la centralidad de la comunicación se deje implícita o se sugiera en términos que fueron categorizados como de *actitud-personalidad* (como “accesible” o “dispuesta”). En este sentido, es conveniente enfatizar la provisionalidad de imponer categorías discretas a un fenómeno que implica la interacción compleja de



múltiples cualidades humanas y condiciones contextuales (Jerez *et al.*, 2016).

A un nivel mayor de detalle, si se toma el vocabulario utilizado por el estudiantado en las evaluaciones como un indicador de sus creencias implícitas sobre el fenómeno educativo —su “pedagogía implícita” (Gregory, 2012)—, se puede inferir que prepondera una visión del rol docente frontal, como transmisor de conocimiento y emisor de mensajes. La alta frecuencia de términos vinculados a “entender”, “transmitir” y “ser claro/a” enfatiza un tipo de comunicación más unidireccional frente a términos como “escucha”, “responde”, “contesta” y “entendió” que implican mayor interacción docente-estudiante.

Actitud y personalidad

Al igual que estudios anteriores (Azab *et al.*, 2016; Gregory, 2012; Murray *et al.*, 2020), los resultados sugieren que las referencias al aspecto personal, afectivo y actitudinal están claramente vinculadas a la calificación. En este caso, las referencias a la actitud personalidad son, después de los calificadores, las que mayor correlación y poder predictivo presentan. Aunque este resultado enfatiza el bien establecido rol de lo personal y actitudinal en los juicios de excelencia docente (Jerez *et al.*, 2016), es importante considerar la influencia de los sesgos de género constatados en MisProfesores.com por Arceo *et al.* (2019): siendo las referencias personales más frecuentes en las docentes, esta es una población feminizada (Bruns y Luque, 2014).

Es interesante notar que incluso en esta variable —la más personal— son frecuentes las menciones a actitudes que se manifiestan directamente en la tarea docente (“falta”, “preocupa”, “exigente”, “no ayuda”) y la relación docente-estudiante (“soberbio”, “respeto”, “arrogante”, “valora”, “comprensiva”).

Limitaciones

Existen varias limitaciones que deben enfatizarse. En

términos metodológicos, el diseño observacional no permite la inferencia de causalidad, y las características peculiares de la muestra (evaluaciones públicas *online* a docentes uruguayos) no habilita la generalización a otros contextos de evaluación. Igualmente, hay que considerar que, aunque el diccionario mostró buenas propiedades, fue elaborado *ad hoc* en el contexto de este proyecto; su utilización en otros textos, la exploración de sus propiedades y su subsecuente perfeccionamiento son altamente deseables.

A nivel teórico, se debe tener presente que este estudio se basa en el robusto, pero limitado, enfoque de las *palabras como atención*: “las palabras reflejan la atención y nada más” (Boyd, 2017; Boyd y Schwartz, 2021). Por ejemplo, la frecuencia, correlación y poder predictivo de los términos vinculados a la personalidad y la actitud no implica necesariamente que estos sean factores relevantes para ser excelente como docente; indica solamente que al reflexionar y escribir comentarios, el foco atencional del estudiantado está puesto en estas cualidades.

A un nivel más práctico, la relación entre la evaluación de estudiantes y otros indicadores de calidad docente dista de ser lineal. Tanto la noción de “excelencia docente” como las categorías utilizadas son teóricamente problemáticas: representan más un modelo simplificado para abordar empíricamente intuiciones comunes que conceptos definidos de manera clara, precisa o ideológicamente neutra (Jerez *et al.*, 2016; Manchini *et al.*, 2022; Scheerens *et al.*, 2020). Este tipo de investigación permite tener observaciones más detalladas para profundizar en la discusión, no para zanjarla.

Prospectiva

Uno de los aportes de este estudio es la creación de la base de datos anonimizada y el diccionario. En esta investigación se eligió maximizar la interpretabilidad pre-registrando hipótesis, utilizando un enfoque transparente (*word-count*) y análisis sencillos. Sería deseable la exploración de otras hipótesis y/o la

utilización de técnicas más refinadas a nivel cualitativo (por ejemplo, codificación manual de comentarios aleatorios) y cuantitativo (por ejemplo, topic-modeling, *Meaning Extraction Method* (MEM), análisis de conglomerados, *Structural Equation Modelling* (SEM)). Desde el punto de vista instrumental, la ampliación (y depurado) del diccionario —tanto mediante el agregado/eliminación de términos como mediante la creación de nuevas categorías— permitirá aumentar su utilidad. También sería conveniente explorar y abordar la ambigüedad dentro de cada categoría: por ejemplo, distinguiendo aspectos de la categoría *pedagógico-organizacional* que podrían considerarse *SS*.

Conceptualmente, la base de datos permite la contrastación de hipótesis relacionadas a sesgos de género, nivel y área de estudios. La comparación con otras medidas de excelencia —por ejemplo, autorreporte, indicadores institucionales, número de publicaciones, etcétera— permitiría evaluar la validez del reporte de estudiantes.

Conclusión

Este estudio abordó la forma en que las diversas cualidades docentes son atendidas y comentadas por los/as estudiantes en una muestra amplia de evaluaciones públicas *online*. Los resultados sugieren que el análisis de los comentarios en relación con la calificación es un método viable para hacerlo. Además de aportar información relevante para entender este fenómeno en Uruguay y la comparación con otras muestras, las propiedades aceptables del diccionario sugieren que la herramienta puede reutilizarse y perfeccionarse.

El uso del vocabulario en los comentarios tiene relaciones significativas con la calificación y, aunque el empleo de términos positivos es más frecuente, el vocabulario negativo muestra relaciones más fuertes. Aunque todas las variables se relacionan con la calificación de los/as docente, las referencias a cualidades personales, actitudinales y comunicacionales (Habilidades Blandas o *SS*) la predicen mejor que las

competencias profesionales y disciplinares (Habilidades Duras o *HS*). Sin embargo, las frecuencias, las diferencias entre el vocabulario positivo/negativo al referirse al saber disciplinar y los términos específicos utilizados al hablar de cualidades personales sugieren que las cualidades profesionales no son subvaloradas, sino que son asumidas implícitamente.

Es pertinente para concluir, explicitar la relevancia de este estudio. Desde la perspectiva de la investigación en educación, es de notar la escasez de estudios que utilicen plataformas de evaluación *online* en lengua española —con notables excepciones, como Arceo y Campos, 2019—; en ese sentido, este trabajo espera estimular la realización de estudios análogos a los que se vienen realizando desde más de una década en lengua inglesa (Brown *et al.*, 2009; Colardaci y Kornfield, 2007; Gregory, 2012; Sonntag *et al.*, 2009; Timmerman, 2008).

Desde el punto de vista de la evaluación docente, este trabajo coincide con esos antecedentes anglófonos en mostrar que las evaluaciones estudiantiles se encuentran sesgadas —en este caso, a favor de determinadas personalidades, actitudes y formas comunicativas—. La evaluación de los/as estudiantes es relevante, pero debe ser solo uno de los varios indicadores para evaluar las cualidades profesionales de los/as docentes.

Finalmente, en un contexto histórico de transformación de la formación en educación (por ejemplo: ANEP, 2023), estos hallazgos son relevantes para discutir las cualidades que hacen excelente a un/a docente y que, por tanto, es deseable cultivar. Este estudio sugiere que los/as estudiantes —coincidentalmente con la investigación en efectividad docente (Bardach *et al.*, 2022)— asignan un valor muy importante a las *SS* de sus docentes; ¿cómo pueden cultivarse estas cualidades de manera sistemática? Aunque diversos enfoques han sido evaluados (como: Body *et al.*, 2016; Vourinen *et al.*, 2021), el campo de las *SS* docentes aún se encuentra —conceptual y metodológicamente— en sus inicios (Manchini *et al.*,



2022). Crear ambientes educativos que estimulen las SS docentes es un reto contemporáneo que se muestra incluso más desafiante al considerar la necesidad de que el desarrollo de las *soft skills* se dé de manera complementaria al desarrollo de las *hard skills*.

La percepción de un/a docente como un excelente profesional depende, en gran medida, de su forma

de ser, actuar y relacionarse. Sin embargo, la percepción de esas cualidades personales se da en un contexto que presupone cualidades profesionales. Más que de una u otra cualidad aislada, la percepción de la excelencia docente parece depender de la integración de lo técnico y lo humano. ■

Referencias

- Administración Nacional de Educación Pública (ANEP) (2023), *Marco Curricular de la formación de grado de los educadores*, Montevideo, ANEP.
- Arceo, Eva y Raymundo Campos (2019), “Gender stereotypes: The case of MisProfesores.com in Mexico”, *Economics of Education Review*, núm. 72, pp. 55-65, DOI: <https://doi.org/10.1016/j.econedurev.2019.05.007>.
- Aristóteles (2005), *Ética a Nicómaco*, Madrid, Alianza.
- Azab, Mohamud, Rada Mihalcea y Jacob Abernethy (2016), “Analysing RateMyProfessors evaluations across institutions, disciplines, and cultures: The tell-tale signs of a good professor”, en E. Spiro y Y.-Y. Ahn (eds.), *Social informatics*, vol. 10046, pp. 438-453, DOI: https://doi.org/10.1007/978-3-319-47880-7_27.
- Bardach, Lisa y Robert Klassen (2020), “Smart teachers, successful students? A systematic review of the literature on teachers’ cognitive abilities and teacher effectiveness”, *Educational Research Review*, núm. 30, pp. 100312, DOI: <https://doi.org/10.1016/j.edurev.2020.100312>.
- Bardach, Lisa, Robert M. Klassen y Nancy E. Perry (2022), “Teachers’ Psychological Characteristics: Do They Matter for Teacher Effectiveness, Teachers’ Well-being, Retention, and Interpersonal Relations? An Integrative Review”, *Educational Psychology Review*, núm. 34, pp. 259-300, DOI: <https://doi.org/10.1007/s10648-021-09614-9>.
- Body, Lawrence, Natalia Ramos, Olivia Recondo y Manuel Pelegrina (2016), “Desarrollo de la Inteligencia Emocional a través del programa mindfulness para regular emociones (PINEP) en el profesorado”, *Revista interuniversitaria de formación del profesorado*, núm. 87, pp. 47-59.
- Burroughs, Nathan, Jacqueline Gardner, Youngjun Lee, Siwen Guo, Israel Tuitou, Kimberly Jansen y William Schmidt (2019), “A Review of the Literature on Teacher Effectiveness and Student Outcomes”, *IEA Research for Education, Teaching for Excellence and Equity*, pp. 7-14. Cham, Springer, DOI: https://doi.org/10.1007/978-3-030-16151-4_2.
- Boyd, Ryan (2017), “Psychological text analysis in the digital humanities”, en S. Hai-Jew (ed.), *Data Analytics in Digital Humanities*, Nueva York, Springer International Publishing, pp. 161-189, DOI: https://doi.org/10.1007/978-3-319-54499-1_7.
- Boyd, Ryan y Andrew Schwartz (2021), “Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field”, *Journal of*

- Language and Social Psychology*, núm. 40, pp. 1-21, DOI: <https://doi.org/10.1177/0261927x20967028>.
- Brown, Michael, Michelle Baillie y Schwandel Fraser (2009), "Rating Ratemyprofessors.com: A Comparison of Online and Official Student Evaluations of Teaching", *College Teaching*, núm. 57, pp. 89-92, DOI: <https://doi.org/10.3200/CTCH.57.2.89-92>.
- Bruns, Barbara y Javier Luque (2014), *Profesores excelentes: cómo mejorar el aprendizaje en América Latina y el Caribe*, Washington, Banco Mundial.
- Coladarci, Theodore y Irv Kornfield (2007), "RateMyProfessors.com versus formal in-class student evaluations of teaching", *Practical Assessment, Research, and Evaluation*, vol. 12, núm. 6, pp. 1-16, DOI: <https://doi.org/10.7275/26ke-yz55>.
- Cornelius-White, Jeffrey (2007), "Learner-Centered Teacher-Student Relationships Are Effective: A Meta-Analysis", *Review of Educational Research*, núm. 77, pp. 113-143, DOI: <https://doi.org/10.3102/003465430298563>.
- Davison, Elizabeth y Jamie Price (2009), "How do we rate? An evaluation of online student evaluations", *Assessment & Evaluation in Higher Education*, num. 34, pp. 51-65, DOI: <http://doi.org/10.1080/02602930801895695>.
- Funder, David y Daniel Ozer (2019), "Evaluating effect size in psychological research: Sense and nonsense", *Advances in Methods and Practices in Psychological Science*, num. 2, pp. 156-168, DOI: <http://doi.org/10.1177/2515245919847202>.
- Gregory, Karen (2012), "How undergraduates perceive their professors: a corpus analysis of Rate My Professor", *Journal of Educational Technology Systems*, núm. 40, pp. 169-193, DOI: <https://doi.org/10.2190/ET.40.2g>.
- Harris, Douglas y Stacey Rutledge (2010), "Models and Predictors of Teacher Effectiveness: A Comparison of Research About Teaching and Other Occupations", *Teachers College Record*, núm. 112, pp. 914-960, DOI: <https://doi.org/10.1177/016146811011200304>.
- Hattie, John (2009), *Visible learning*, New York, Routledge.
- Hemphill, James (2003), "Interpreting the magnitudes of correlation coefficients", *American Psychologist*, núm. 58, pp. 78-79, DOI: <https://doi.org/10.1037/0003-066X.58.1.78>.
- Jerez, Oscar, Cesar Orsini y Beatriz Hasbún (2016), "Atributos de una docencia de calidad en la educación superior: una revisión sistemática", *Estudios Pedagógicos*, núm. 42, pp. 483-506, DOI: <http://doi.org/10.4067/S0718-07052016000400026>.
- Kaya, Metin y Abdullah Selvitopu (2019), "A meta-analysis of the effects of some factors on teachers' classroom management skills", *International Journal of Contemporary Educational Research*, núm. 6, pp. 409-425, DOI: <https://doi.org/10.33200/ijcer.621313>.
- Kennedy, Brendan, Ashwini Ashokkumar, Ryan Boyd y Morteza Dehghani (2021), "Text analysis for psychology: Methods, principles, and practices", en M. Dehghani y R. Boyd (Eds.), *The Handbook of Language Analysis in Psychology*, Nueva York, Guilford, pp. 3-62.
- Kim, Lisa, Verena Jörg y Robert Klassen (2019), "A Meta-Analysis of the Effects of Teacher Personality on Teacher Effectiveness and Burnout", *Educational Psychology Review*, núm. 31, pp. 163-195, DOI: <https://doi.org/10.1007/s10648-018-9458-2>.
- Kindred, Jeannette y Shaheed Mohammed (2005), "'He Will Crush You like an Academic Ninja!': Exploring Teacher Ratings on Ratemyprofessors.com", *Journal of Computer-Mediated Communication*, núm. 10, pp. 314-322, DOI: <https://doi.org/10.1111/j.1083-6101.2005.tb00257.x>.
- Klassen, Robert y Virginia Tze (2014), "Teachers' Self-Efficacy, Personality, and Teaching Effectiveness: A Meta-Analysis", *Educational research review*, núm. 12, pp. 59-76, DOI: <https://doi.org/10.1016/j.edurev.2014.06.001>.
- Kreitzer, Rebecca y Jennie Sweet-Cushman (2022), "Evaluating Student Evaluations of Teaching: a Review of Measurement and Equity Bias in SETs and Recommendations for Ethical Reform", *Journal of Academic Ethics*, núm. 20, pp. 73-84, DOI: <https://doi.org/10.1007/s10805-021-09400-w>.
- Lei, Hao, Yunhuo Cui y Ming Chiu (2017), "The Relationship between Teacher Support and Students'



- Academic Emotions: A Meta-Analysis”, *Frontiers in Psychology*, núm. 8, pp. 1-12, DOI: <https://doi.org/10.3389/fpsyg.2017.02288>.
- Loureiro, Silvana, Marina Míguez y Ximena Otegui (2016), “Desempeño docente en la enseñanza universitaria: análisis de las opiniones estudiantiles”, *Cuadernos de Investigación Educativa*, núm 7, pp. 55-67, DOI: <https://doi.org/10.18861/cied.2016.7.1.2576>.
- Manchini, Nigel, Óliver Jiménez y Natalia Ramos (2022, Agosto 8), “Teachers’ Soft Skills: a Scoping Review”, *PsyArxiv*, DOI: <https://doi.org/10.31234/osf.io/trjub>.
- Murray Dakota, Clara Boothby, Huimeng Zhao, Vanessa Minik, Nicolás Bérubé, Vincent Larivière y Cassidy Surimoto (2020), “Exploring the personal and professional factors associated with student evaluations of tenure-track faculty”, *PLoS ONE*, núm. 15, pp. e0233515, DOI: <https://doi.org/10.1371/journal.pone.0233515>.
- Ozili, Peterson (2023), “The acceptable R-square in empirical modelling for social science research”, *MPPRA*, <<https://mpra.ub.uni-muenchen.de/116496/>> [Consulta: marzo de 2023].
- Pennebaker, James (2018) “Telling Stories: Analyzing Text to Understand Personality, Social behavior, and Narratives”, *Microsoft Research Talks*, Washington, <<https://www.youtube.com/watch?v=e25YKwSKOo4>> [Consulta: marzo de 2023].
- Qaiser, Shahzad y Ramsha Ali (2018) “Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents”, *International Journal of Computer Applications*, núm. 181, pp. 25-29, DOI: <https://doi.org/10.5120/ijca2018917395>.
- Quin, Daniel (2017), “Longitudinal and Contextual Associations Between Teacher-Student Relationships and Student Engagement: A Systematic Review”, *Review of Educational Research*, núm. 87, pp. 345-387, DOI: <https://doi.org/10.3102/0034654316669434>.
- R Core Team (2022), *R: A language and environment for statistical computing*, Viena, R Foundation for Statistical Computing, <<https://www.R-project.org/>> [Consulta: diciembre de 2022].
- Ramírez-Esparza, Nairán, James Pennebaker, Florencia García y Raquel Suría (2007), “La psicología del uso de las palabras: un programa de computadora que analiza textos en español”, *Revista Mexicana de Psicología*, vol. 24, núm. 1, pp. 85-99, DOI: <http://hdl.handle.net/10045/25918>.
- Reid, Lándern (2010), “The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.Com”, *Journal of Diversity in Higher Education*, núm. 3, pp. 137-152, DOI: <https://doi.org/10.1037/a0019865>.
- Roorda, Deborah, Helma Koomen, Jantine Spilt y Frans Oort (2011), “The Influence of Affective Teacher-Student Relationships on Students’ School Engagement and Achievement: A Meta-Analytic Approach”, *Review of Educational Research*, núm. 81, pp. 493-529, DOI: <https://doi.org/10.3102/0034654311421793>.
- Rosen, Andrew (2018), “Correlations, trends and potential biases among publicly accessible web-based student evaluations of teaching: a large-scale study of RateMyProfessors.com data”, *Assessment & Evaluation in Higher Education*, núm. 43, pp. 31-44, DOI: <https://doi.org/10.1080/02602938.2016.1276155>.
- RStudio Team (2022), *RStudio: Integrated Development for R*, Mass, Rstudio, <<http://www.rstudio.com>>.
- Sancar, Raziye, Deniz Atal y Deniz Deryakulu (2021), “A new framework for teachers’ professional development”, *Teaching and Teacher Education*, vol. 101, pp. 103305, DOI: <https://doi.org/10.1016/j.tate.2021.103305>.
- Scheerens, Jaap, Greetje van der Werf y Hester de Boer (2020), *Soft Skills in Education: Putting the evidence in perspective*, Nueva York, Springer International Publishing, DOI: <https://doi.org/10.1007/978-3-030-54787-5>.
- Silge, Julia y David Robinson (2016), “tidytext: Text Mining and Analysis Using Tidy Data Principles in R”, *Journal of Open Source Software*, núm. 1, pp. 37-38, DOI: <http://doi.org/10.21105/joss.00037>.
- Sonntag, Michael, Jonathan Bassett y Timothy Snyder (2009), “An empirical test of the validity of student

- evaluations of teaching made on RateMyProfessors.com”, *Assessment & Evaluation in Higher Education*, núm. 34, pp. 499-504, DOI: <https://doi.org/10.1080/02602930802079463>.
- Spilt, Jantine, Helma Koomen y Jochem Thijs (2011), “Teacher Wellbeing: The Importance of Teacher-Student Relationships”, *Educational Psychology Review*, núm. 23, pp. 457-477, DOI: <https://doi.org/10.1007/s10648-011-9170-y>.
- Stuber, Jennie, Amanda Watson, Adam Carle y Kristin Staggs (2009), “Gender expectations and on-line evaluations of teaching: evidence from RateMyProfessors.com”, *Teaching in Higher Education*, núm. 14, pp. 387-399, DOI: <https://doi.org/10.1080/13562510903050137>.
- Stadthagen-Gonzalez, Hans, Constance Imbault, Miguel Pérez Sánchez y Marc Brysbaert (2017), “Norms of valence and arousal for 14,031 Spanish words”, *Behavior Research Methods*, núm. 49, pp. 111-123, DOI: <https://doi.org/10.3758/s13428-015-0700-2>.
- Tausczik, Yla y James Pennebaker (2010), “The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods”, *Journal of Language and Social Psychology*, núm. 29, pp. 24-54, DOI: <https://doi.org/10.1177/0261927X09351676>.
- Thapa, Amrit, Jonathan Cohen, Shawn Guffey y Ann Higgins-D’Alessandro (2013), “A Review of School Climate Research”, *Review of Educational Research*, núm. 83, pp. 357-385, DOI: <https://doi.org/10.3102/0034654313483907>.
- Timmerman, Thomas (2008), “On the Validity of RateMyProfessors.com”, *Journal of Education for Business*, núm. 84, pp. 55-61, DOI: <https://doi.org/10.3200/JOEB.84.1.55-61>.
- Vuorinen, Kaisa, Anna Pessi y Lotta Uusitalo (2021), “Nourishing Compassion in Finnish Kindergarten Head Teachers: How Character Strength Training Influences Teachers’ Other-Oriented Behavior”, *Early Childhood Education Journal*, núm. 49, pp. 163-176, DOI: <https://doi.org/10.1007/s10643-020-01058-0>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan y Yutani Hiroaki (2019), “Welcome to the tidyverse”, *Journal of Open Source Software*, núm. 4, pp. 1-4, DOI: <https://doi.org/10.21105/joss.01686>.
- Zhou, Jin y Jun-min Ye (2020), “Sentiment analysis in education research: a review of journal publications”, *Interactive Learning Environments*, núm. 31, pp. 1-13, DOI: <https://doi.org/10.1080/10494820.2020.1826985>.
- Ziyan Liu, Li Yingnan, Zhu Wenying, He Yuanping y Li Dongbin (2022), “A Meta-Analysis of Teachers’ Job Burnout and Big Five Personality Traits”, *Frontiers in Education*, núm. 7, pp. e822659, DOI: <https://doi.org/10.3389/feduc.2022.822659>.

Cómo citar este artículo:

Manchini, Nigel, Óliver Jiménez y Natalia Ramos-Díaz (2024), “Habilidades blandas y técnicas en evaluaciones de estudiantes a docentes”, *Revista Iberoamericana de Educación Superior (RIES)*, vol. XV, núm. 44, pp. 112-131, DOI: <https://doi.org/10.22201/iisue.20072872e.2024.44.1894>.