enunciación

Enunciación

https://revistas.udistrital.edu.co/ojs/index.php/enunc
DOI: http://doi.org/10.14483/22486798.14311

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS
Acreditación Institucional de Alta Calidad

Artículo corto

Reflexiones y percepciones sobre la evaluación automatizada del discurso escrito

Perceptions on the automated assessment of written discourse

Ricardo Alberto Benítez*

Resumen

Este artículo presenta reflexiones y percepciones acerca de un potencial problema en la enseñanza y aprendizaje de la escritura en el ámbito escolar y universitario: su evaluación con instrumentos computarizados. El conflicto entre esta evaluación y la realizada por humanos puede constituir motivo de debate y análisis, como ya lo es en Estados Unidos. Allí, varios experimentos han intentado automatizar las evaluaciones de composiciones escritas. Sin embargo, los resultados han sido insatisfactorios desde el punto de vista tanto de los investigadores como de los estudiantes, a cuyos productos escritos se les ha aplicado dicha evaluación. Este artículo, por tanto, surge de la inquietud por indagar las percepciones de estudiantes universitarios acerca de su disposición a recibir retroalimentación y evaluación de sus composiciones escritas a través de un instrumento automatizado. Para recabar esta información, se les solicitó que contestaran en línea si estarían dispuestos o no a someter la evaluación de sus trabajos escritos a una computadora con la posibilidad de agregar comentarios. Las respuestas y los comentarios fueron categorizados como positivos, negativos e indecisos. Las negativas obtuvieron la mayoría, seguidas de las positivas y las indecisas. Se concluye que faltan estudios de mayor envergadura que exploren las percepciones tanto de estudiantes como de profesores de producción escrita para obtener resultados más consistentes y que las instituciones que decidan utilizar instrumentos automatizados para la evaluación de trabajos escritos deberían justificar sus ventajas y desventajas.

Palabras clave: método de evaluación, lengua escrita, retroalimentación.

Abstract

This article describes the thoughts and perceptions regarding a potential problem when teaching and learning writing skills in school and university environments: assessments of writing using computerised instruments. The conflict between this assessment and the one carried out by humans may be subject to discussion and analysis, as it has been in the United States, where several experiments have attempted to automate written compositions. However, the results have been unsatisfactory according to both the researchers and the students whose written output had been subjected to such assessment. Therefore, the present article aims to enquire into the perceptions of university students with respect to their willingness to receive feedback and evaluations on their written compositions through an automated instrument. To gather this information, students were asked to respond online as to whether they would be willing to subject their work to evaluation by a computer, including the option to add comments. Responses and comments were classified as positive, negative and undecided. Most of them responded negatively, followed by positive and undecided votes. It can be concluded that there is a need for further in-depth studies exploring the perception of both students and professors to obtain more consistent results. In addition, institutions that opt to use automated instruments to assess written assignments should explain its advantages and disadvantages.

Keywords: assessment method, written language, feedback.

* Master of Arts in the Teaching of English as a Second Language (MA in TESL), Arizona State University, Estados Unidos. Profesor adjunto de la Universidad Católica de Valparaíso, Chile. Correo electrónico: ricardo.benitez@pucv.cl

Cómo citar este artículo: Benítez, R.A. (2019). Reflexiones y percepciones sobre la evaluación automatizada del discurso escrito. *Enunciación, 24*(2), 227-240. DOI: http://doi.org/10.14483/22486798.14311
Artículo recibido: 02 de febrero de 2019; aprobado: 17 de julio de 2019.

Introducción

El presente artículo busca captar la atención tanto de profesores de composición escrita como de diseñadores de programas computacionales acerca de un aspecto de la enseñanza y el aprendizaje de esta habilidad que se mantiene latente en el campo educacional: su evaluación por computadora (EPC). Esta consiste en someter la calidad de un texto escrito a una herramienta automatizada que entrega un valor numérico al final del proceso. Existe escasa literatura en español acerca de esta situación en el contexto escolar y universitario. Sin embargo, como la tecnología avanza a pasos agigantados, la posibilidad de que el discurso escrito sea evaluado a través de instrumentos computarizados es motivo de preocupación, no solo por la práctica cultural de reproducir irreflexivamente lo que está en boga, sino también por las razones que se exponen más adelante.

La Organización para la Cooperación y Desarrollo Económicos (OCDE) sostiene que el 55 % de las actividades humanas serán reemplazadas por robots o máquinas en un futuro cercano y es un hecho que la condición humana, por medio de la robótica y del procesamiento del lenguaje natural como hito temprano de la inteligencia artificial (Carbonell, 1992), ha mejorado en ciertas áreas –hecho que es apoyado por aquellos miembros del movimiento ideológico denominado *transhumanismo* (Rodríguez, 2017a)–. Zoltan Istvan (citado por Rodríguez, 2017b), fundador del Partido Transhumanista de Estados Unidos, piensa que el cuerpo humano debería unirse a las máquinas por medio de un transmisor o de un implante en el cerebro que lo conectaría a internet o a una inteligencia artificial, mejorando de esta manera tanto los cuerpos como los cerebros de las personas. Corbo (2017) cree que estos adelantos tecnológicos pueden originar una cuarta Revolución Industrial y cita el informe del McKinsey Global Institute de 2017 donde se demuestra que la mitad de las actividades humanas pueden automatizarse. Sin embargo, detractores, como la directora de la Academia Chilena de la Lengua Adriana Valdés (citada por Rodríguez, 2017a) y el filósofo alemán Markus Gabriel (citado por Rodríguez, 2017b), creen que dichos adelantos deben ir acompañados de ética y de moral, lo que significaría tal vez un cambio en las leyes y las políticas. El filósofo Raúl Villarroel (citado por Rodríguez, 2017a, p. E3) sostiene:

> Hay fundadas razones para suponer que lo que habría que cautelar es, principalmente, que a partir de la confianza en la expansión futura de la ciencia y la tecnología no se vaya a generar una escena incluso peor de desigualdad y sometimiento humano como la que hasta ahora hemos conocido.

Marco teórico

La oposición entre una evaluación realizada por humanos y una EPC se ha mantenido por bastante tiempo en otras latitudes y no ha llegado todavía a nuestro medio con la fuerza que la caracteriza en esos lugares. Evidentemente, la EPC alivia la carga de leer, interpretar y aplicar una rúbrica, la cual siempre está sujeta a la subjetividad del lector/profesor, aun cuando esta se calibre con otros agentes evaluativos. Una EPC sirve cuando se utiliza a gran escala; por ejemplo, cuando se trata de ubicar a estudiantes de idioma en diferentes niveles (Shermis, Mzumura, Olson y Harrington, 2001), aunque esto significaría un detrimento de la escritura real de un estudiante por las razones que se explicitan más adelante.

Aquellos que defienden la EPC argumentan que es válida y confiable, y que arroja los mismos resultados con independencia de cuándo y dónde se aplique. Lo anterior es plausible, pues una computadora –programada para operar objetivamente– es insensible a factores externos que podrían influir en los resultados que arroja. Con la intención de eliminar la subjetividad de toda evaluación, un profesor generalmente usará una rúbrica con criterios como contenido, organización, vocabulario, variedad sintáctica y convenciones lingüísticas.

ISSN 0122-6339 • ISSNe 2248-6798 • Vol. 24, No. 2 (julio/diciembre de 2019), pp. 227-240.

Por otro lado, la defensa de la EPC se basa en la literatura que sostiene que sería mejor que la evaluación humana, pues esta última tendería a sesgar los resultados, si bien involuntariamente. Sin embargo, la EPC no considera –hasta la fecha– aspectos relevantes de un texto escrito como la calidad de las ideas, la voz propia del estudiante escritor, la creatividad en el uso del lenguaje, la situación retórica (Benítez, 2001) o la modificación creativa de la estructura textual que un estudiante haya imprimido a su texto. De acuerdo con Corbo (2018, p. 7), "las máquinas son mejores para ejecutar tareas repetitivas, bien definidas y predecibles, pero no son tan buenas para trabajos que requieren esfuerzos abstractos o no rutinarios. Tampoco para tareas que requieren de juicio o habilidades de interacción con personas". Precisamente, el trabajo de un profesor no es –o no debe ser– rutinario y necesita de la interacción con sus estudiantes para formar una comunidad discursiva.

Problemas que emanan de la evaluación asistida por sistemas computarizados

A la escritura y a la evaluación siempre se les imprime una cuota de subjetividad, a pesar de contar con estándares establecidos. La pregunta es, entonces, ¿la EpC sería un medio para eliminar la subjetividad, característica de los seres humanos? Entre los instrumentos más usados en Estados Unidos, se encuentran los siguientes (los comentarios que merece cada uno se encuentran en cursiva):

- *Project Essay Grade*. Calcula puntajes para variables representadas en composiciones ya evaluadas por lectores profesionales, pero no entrega información acerca de lo débil de un área de la escritura (contenido, organización, estilo, convenciones, creatividad). *Es difícil encontrar una computadora que dé explicaciones*.
- *e-rater v.2.0*. Usa técnicas de procesamiento de lenguaje natural basado en muestras de escritura que ya han sido evaluadas por humanos. *De todas maneras, se necesita un ser humano que las evalúe antes de ser procesadas por una computadora*.
- *e-rater v.2.0 features*. Evalúa rasgos relacionados con un sistema de retroalimentación, que analiza la escritura. *No clarifica dudas después de la retroalimentación*.
- *IntelliMetric*. Evalúa el contenido con un segmentador de procesamiento de lenguaje natural y un extractor de rasgos que identifica aproximadamente 400 rasgos semánticos, sintácticos y discursivos, con la ayuda de análisis morfológico, reconocimiento de ortografía, colocaciones gramaticales y detección lexical para construir sentido. *Se preocupa de estructuras y de las convenciones, no de lo original que puede ser un texto*.
- *Intelligent Essay Assessor*. Evalúa la calidad de composiciones con un diagnóstico de retroalimentación sobre lo adecuado del conocimiento y la expresión de ese conocimiento; incorpora medidas estadísticas de corpus y utiliza un modelo de aprendizaje automatizado de la comprensión humana: análisis semántico latente, que representa matemáticamente los significados de palabras y pasajes en el dominio del lenguaje, y el conocimiento del texto. *Considera lo que es adecuado dentro de límites estrictamente establecidos*.
- *Panlingua*. Supone que existe una lengua universal que refleja entendimiento y conocimiento, utilizando un *software* que imita el modo en que el cerebro entiende la lengua y las ideas (Whittington y Hunt, 1999). *Considera productos que se ciñen a la norma, no aquellos que se desvían de los cánones establecidos*.
- *Lexical Conceptual Structure* se basa en la idea de que una máquina puede capturar información de la estructura superficial –tales como las relaciones entre sujetos y objetos en las oraciones–, procesando especificidades del lenguaje, como la sintaxis y desviaciones del canon (Whittington y Hunt, 1999). *Solo considera elementos cohesivos*.

ISSN 0122-6339 • ISSNe 2248-6798 • Vol. 24, No. 2 (julio/diciembre de 2019), pp. 227-240.

Estos instrumentos se han popularizado porque las evaluaciones son más objetivas, son consistentes en el tiempo y en el espacio, y se aplican de forma rápida. Pero también existe la posibilidad de que el evaluador sienta que está siendo reemplazado por una máquina a la cual no podrá objetarle nada. Para que esto no ocurra, tendrían que existir instancias en las que el profesor *conversara* con la máquina.

Un punto crucial en una evaluación más íntegra es la situación retórica, que es fundamental para considerar un texto como texto y con la cual el escritor cuida el tono y el registro lingüístico apropiados a la tarea de escritura; evita que el tópico se desvíe innecesariamente o vuelva a este de manera fluida; adhiera a los cánones de un género textual determinado; y facilita el descubrimiento del propósito de su texto. Una EPC debería evaluar si existen digresiones innecesarias o son pertinentes; debería evaluar si el tono en una carta formal no es adecuado (lo cual conduciría a situaciones embarazosas si la carta tiene un tono sarcástico, arrogante, humorístico o insultante); tendría que *saber* si la longitud de una descripción es apropiada en una narración o si esa descripción es innecesariamente repetitiva. Además, ¿cómo podría una máquina diferenciar entre el propósito de un texto y la intención del autor de ese mismo texto? (Williams, 1998). ¿Se programaría con una evaluación holística, que entrega una impresión del texto como un todo? Pero ¿puede una computadora atribuir *impresiones* a un texto, si se asume que estas son capacidades humanas? ¿Podría arrojar como resultado: "Me impresionó la manera creativa en que el autor organizó su texto" y compartir esa apreciación con otra computadora? Si ese fuera el caso en muchos años más, ¿estaríamos dispuestos a relegar el trabajo de leer y evaluar una composición escrita a una máquina?

Con la EPC se pierde la interacción humana, lo cual representa una objeción a su uso si se concibe la escritura como una actividad esencialmente social (Harris, 1990; Wilson, 1996; Spivey, 1997; Blakeslee, 2001; Jones, 2003; Noël y Robert, 2004; entre otros). Con el uso de una EPC se pierde toda retórica –el arte de convencer y persuadir–; pero ¿debe un estudiante persuadir a una computadora de que lo que ha escrito es producto de su trabajo intelectual? Y, si la persuasión se concibe como *la* manera de inducir a la acción, ¿qué acción podría tomar una computadora? ¿Cómo podría esta cambiar de opinión una vez leído el texto? O ¿para qué convencer a una máquina de que mi opinión es plausible?

La persona evaluadora es la única que puede posicionarse como audiencia invocada o abordada (Ede y Lunsford, 1988). Si bien existen sistemas que retroalimentan, ¿cuán bien retroalimentan? ¿Qué acción tomaría un estudiante si una computadora no pudiera responder preguntas específicas acerca del texto que ha escrito? O, si puede responder, ¿sería capaz de interactuar con cada pregunta que el estudiante escritor pueda plantearle? ¿Cómo respondería a preguntas específicas acerca del proceso de escribir mismo? Puede que el trabajo de Iván Schuller (citado por Marcano, 2018) sea la respuesta a estos interrogantes, ya que espera que la computación neuromórfica opere como lo haría un cerebro, con sus neuronas y sinapsis.

Condon (2013) sostiene que los problemas más importantes se relacionan con aspectos del constructo de la escritura tal como lo entiende la comunidad discursiva, y añade que la EPC es ineficaz al predecir el éxito de los estudiantes cuando estos deben rendir pruebas para ser admitidos en una institución o para ser ubicados en un nivel de desempeño (básico, intermedio, avanzado).

No obstante, la EPC podría emplearse a gran escala, por ejemplo, en una prueba de selección universitaria, ya que la habilidad de escritura generalmente se percibe disminuida en los estudiantes universitarios de primer año. Serviría para diagnosticar el desarrollo de esta habilidad y, así, tomar medidas remediales en los primeros años de estudio. Lo anterior conduciría a un análisis previo de las condiciones escolares contextuales en las que se ha desarrollado la producción escrita.

Collier (2012) publicó un artículo donde muchos expertos admiten que la EPC sirve para ahorrar

tiempo y dinero. Dichos expertos citan un ejemplo concluyente: se aplicó un puntaje máximo de 6 al discurso de Gettysburg de Abraham Lincoln[1]; el resultado fue de 2 y manifiestan cuán errónea puede llegar a ser la aplicación de una EPC. El artículo advierte sobre los peligros que pueden esconderse detrás del uso de la EPC; entre otros, puede disminuir la confianza en los estudiantes pues se identifican errores que no son tales; se pierde la instancia en que el profesor asiste individualmente al estudiante que tiene dudas acerca de su producción escrita; se pierde la oportunidad de que los profesores se desarrollen profesionalmente al no tener que analizar lo que significa una escritura comunicativa; se desvaloriza la lectura y la escritura, pues el texto se dirige a una máquina; se desincentiva la interacción comunicativa cuando el estudiante sabe que está escribiendo para una máquina y no para una persona; y se puede producir una especie de juego entre el estudiante que escribe y la máquina al querer vencerla.

Plagio, retroalimentación e interrogantes derivadas de una EPC

Esta sección presenta otros problemas que merecen reflexión sobre el uso de una EPC y también algunos problemas adicionales que no emanan del estudio de Collier.

La EPC carece de aplicaciones eficaces para detectar discurso plagiado. Si llega a detectarlo, ¿"Sabrán" las computadoras cuánto de lo plagiado es susceptible de ser penalizado y cuánto no? ¿"Sabrán" si la cita de un autor es apropiada para el contexto? ¿"Sabrán" si la longitud de esa misma cita es la correcta u "ofrecerán" sugerencias para acortarla? Si la EPC considera la posibilidad de plagio, ¿"sabría" una computadora que el estudiante escritor plagió su texto, ya sea el texto total o partes de este? La EPC tendría que someter todos los textos al escrutinio de un motor de búsqueda en internet. Probablemente lo harían con mucha rapidez, pero ¿y si el texto solo tiene partes plagiadas? Y de esas partes, ¿cuántas serían suficientes para penalizar al estudiante? ¿Cómo podría defenderse el estudiante ante una acusación de plagio formulada por una máquina? ¿Estaría el profesor dispuesto a delegar la función de penalizar a una máquina?

Asimismo, surgen los siguientes interrogantes sobre las capacidades de una computadora:

- ¿Retroalimentar con la experticia de un profesor? Estudiante: "¿Por qué este adverbio debe colocarse aquí en la oración y no donde yo lo puse?". O, "¿No entiendo por qué partes de mi introducción son apropiadas para la conclusión?".
- ¿Detectar disonancia cognitiva (falta de concordancia entre lo que se escribe y lo que se quiere decir) y dar una respuesta satisfactoria para transformarla en consonancia? Estudiante: "Tengo una idea de lo que quiero decir, pero no sé cómo".
- ¿Interpretar unívocamente el significado de ciertas expresiones con sus connotaciones y frases idiomáticas? Estudiante escribió: "… pero no nos vayamos por las ramas".
- ¿Verificar la veracidad o falsedad de hechos? Estudiante escribió: "El gobierno boliviano nunca reclamó soberanía marítima en la Corte Internacional de Justicia".
- ¿"Reaccionar" ante frases muy cortas terminadas en un punto seguido con el fin de dar suspenso a una trama narrativa? Estudiante escribió*:* "Llegó. La puerta. Crujió. Subió. El baño. La sangre. El cadáver en la tina…".
- ¿Detectar y juzgar si la verbosidad es esencial en una narración donde uno de los personajes se caracteriza por ser locuaz?
- ¿"Conversar" con otra computadora tal como lo harían dos expertos en un intercambio profesional con el propósito de interpretar un texto?
- ¿Explicar el bajo puntaje que ha obtenido el texto de un estudiante?

1 En 1863, Abraham Lincoln, lamentaba la muerte de los soldados durante la Guerra Civil de Estados Unidos con palabras que aludían a que el gobierno del pueblo, por el pueblo y para el pueblo no desaparecerá de la Tierra.

ISSN 0122-6339 • ISSNe 2248-6798 • Vol. 24, No. 2 (julio/diciembre de 2019), pp. 227-240.

- ¿Otorgar audiencia a un estudiante para que este reciba una respuesta por ese bajo puntaje?
- ¿Contribuir al diseño de una rúbrica, asignando puntajes a los criterios?
- ¿Manejar la multimodalidad (ilustraciones, fotografías, videoclips, gráficos) y la interpretación de esta?
- ¿Evaluar lo (in)apropiado del uso metafórico del lenguaje?
- ¿Verificar si el estudiante progresa en su habilidad?

Informes de los expertos

La plataforma Moodle.org ayuda a diseñar diversos tipos de preguntas –por ejemplo, relacionar columnas, arrastrar y soltar el texto, seleccionar entre alternativas múltiples, arrastrar imágenes, identificar si un enunciado es verdadero o falso–, y todas pueden ser susceptibles de EPC. Esta plataforma también asiste en la elaboración de preguntas tipo ensayo en su enlace https://docs.moodle.org/all/es/Tipo_de_pregunta_de_ensayo. Sin embargo, al abrirlo, el lector se encuentra con lo siguiente:

> Las preguntas de Ensayo son creadas de la misma forma en que se crean los otros tipos de preguntas. La diferencia está en que las preguntas de Ensayo deben de ser calificadas manualmente, y el estudiante no tendrá una calificación final hasta que el profesor no haya calificado los ensayos. A la pregunta de ensayo no se le asignará calificación hasta que haya sido revisada por un profesor que la haya calificado manualmente. Hasta que esto suceda, la calificación del estudiante será de 0.

Evaluar –o *calificar*, como dice la cita anterior– la escritura es arduo, a lo cual se añade la carga de trabajo que representa para el profesor. Una computadora aliviaría el proceso evaluativo, pero no sería justo para los estudiantes. La alternativa para ambos es no evaluar las composiciones escritas, como lo sugiere Thomas (2016), quien prefiere proporcionar retroalimentación y requerir que sus estudiantes revisen la mayoría de sus tareas hasta que ellos estén conformes con su trabajo. Sugiere que los roles del profesor y del estudiante deben ser revisados: el profesor como autoridad, no como persona autoritaria; el profesor como profesor/estudiante; y el estudiante como estudiante/profesor. Thomas sostiene, entre otras cosas, que la retroalimentación más poderosa es aquella en que el profesor incluye la identificación de las fortalezas clave en el trabajo de sus estudiantes, los invita a concertar citas con él durante el semestre para analizar qué evaluación merecerían sus trabajos y estimula a todos los profesores de composición escrita a dejar de lado la urgencia de evaluar con estimaciones numéricas. Sin embargo, es casi imposible pensar en una alternativa como la que plantea Thomas cuando las calificaciones cumplen un papel dominante en la promoción del estudiante al siguiente curso. Thomas sostiene que los profesores deben saber y comprender cómo la tecnología puede mejorar el proceso educativo, el cual se basa en la importante relación entre ellos y sus estudiantes. Y añade que, para educar a nuestras comunidades, es esencial comenzar con la intención de mejorar el tiempo de contacto entre profesor y estudiante, no reemplazarlo. El mismo autor advertía en 2015 que las computadoras son beneficiosas para la enseñanza de la escritura, pero que no pueden sustituir al humano en la evaluación de una composición.

A pesar de los esfuerzos que ha realizado con su *Basic Automatic B.S. Essay Language Generator* (BABEL), Les Perelman (citado por Kolowich, 2014) se opone a la EPC sosteniendo que los instrumentos como el mencionado no miden los constructos reales que tienen que ver con la escritura; no pueden leer significados ni tampoco verificar hechos, y son incapaces de diferenciar entre la verbosidad sin sentido y la escritura lúcida. Kolowich, con la *Enhanced AI Scoring Engine,* intenta hacer más humanas la EpC, haciéndola imitar los estilos evaluativos de profesores para obtener resultados más parecidos a los que podrían arrojar los humanos. El

procedimiento se inicia cuando el profesor otorga puntaje a una serie de ensayos según sus propios criterios. Luego, el *software* escanea los ensayos ya corregidos para obtener patrones y asimilarlos. La idea es crear una versión automatizada de la versión del profesor que retroalimenta una cantidad de trabajos más amplia, mejorando significativamente la evaluación objetiva y la velocidad de la evaluación formativa. Sin embargo, ¿qué hicieron exactamente los profesores cuyos estilos de evaluar fueron imitados? ¿Y si lo único que hicieron fue fijarse en las convenciones del lenguaje?

Según Hesse (2013), la EPC debe considerar la audiencia y el propósito de un texto (dos componentes esenciales de la situación retórica). Las audiencias varían en cuanto a experticia, expectativas, necesidades, circunstancias, creencias y relaciones con el lector. Además, el profesor tiene que juzgar si los textos logran su propósito (explicar, analizar, interpretar, sintetizar, cambiar la opinión de las personas, mover a la acción, hacer que los lectores asimilen ciertos puntos de vista, entretener, demostrar conocimiento, expresar pensamientos, actitudes o creencias, o sostener una relación). Hesse dice que todas estas son razones válidas para escribir y si se tiene éxito en un propósito, en otro puede fallar. Pero ¿están diseñadas las computadoras para juzgar si un texto cumple o no con su propósito?

Haimson (2016) sostiene que la inhabilidad de la EPC puede degenerar la enseñanza, ya que los profesores entrenan a los estudiantes para que escriban con verbosidad a lo cual la computadora le otorgará puntajes altos, engañando así al sistema; así, la EPC estimularía a los estudiantes a que sean comunicadores deficientes. El trabajo de Haimson se basó en las conclusiones de un análisis realizado por el Eduactional Testing Service (ETS): la actual EPC no puede evaluar aspectos cognitivos exigentes (conciencia de la audiencia, argumentación, pensamiento crítico y creatividad), y que podría manipular pruebas al intentar obtener una ventaja injusta, usando, por ejemplo, palabras complicadas, frases fijas pero incoherentes o aumentando artificialmente la longitud del texto para mejorar los puntajes.

Collier (2012) constató lo anterior al informar acerca del experimento de Les Perelman con la aplicación de su *e-rater* para el cual creó 17 ensayos estudiantiles, identificando errores que no lo eran, otorgando mayor puntaje cuando copiaba partes inconexas y cuando usaba palabras poco comunes, como "atroz" en vez de "malo" o "plétora" en vez de "muchos". Collier insiste en que se quiebra la comunicación, se trastoca el concepto de *escritura* y se desmotiva la creatividad cuando el estudiante sabe que su audiencia es una máquina.

El Consejo Nacional de Profesores de Inglés (NCTE, por su sigla en inglés) de Estados Unidos, en su *Position Statement on Machine Scoring* (2013), expresa que la EPC es tentadora para las grandes corporaciones privadas, pero al considerar lo que se pierde a causa de esta, los ahorros se transforman en costos para el estudiante, para las instituciones educacionales y para la sociedad. Dice el NCTE que las computadoras

- No pueden juzgar elementos asociados a la buena escritura: lógica, claridad, exactitud, ideas relevantes, estilo innovador, apelaciones a la audiencia eficientes, diferentes formas de organización, tipos de persuasión, calidad de la evidencia, tonos, y repeticiones apropiadas.
- Usan métodos más rústicos que los que usan lectores humanos: algunas miden la sofisticación del vocabulario por la longitud promedio de palabras y la frecuencia de estas en un corpus de textos; o miden el desarrollo de ideas contando el número de oraciones.
- Se programan para dar puntajes basados en estímulos muy específicos, sin incentivar instancias creativas ni para la escritura ni para la evaluación.
- Reducen las instrucciones para las tareas de escritura, y esto no representa la variedad de tareas más compleja que los estudiantes encuentran en la educación superior.

ISSN 0122-6339 • ISSNe 2248-6798 • Vol. 24, No. 2 (julio/diciembre de 2019), pp. 227-240.

- Favorecen lo más objetivo y superficial (gramática, ortografía, puntuación), pero los problemas en estas áreas a menudo surgen en las condiciones en que se rinde una prueba y son las que se pueden rectificar más fácilmente cuando hay tiempo para revisar y editar.
- Discriminan a los estudiantes menos familiarizados con la tecnología para rendir pruebas: la EPC pone en desventaja a las escuelas que carecen de fondos para proporcionarla y se adquiere tecnología solo para cumplir con los requisitos que imponen las pruebas.
- Eliminan el propósito de un texto: crear interacciones humanas a través de la construcción de significados complejos con consecuencias sociales.

En la misma declaración, el NCTE ofrece una bibliografía anotada de experimentos científicos que atacan la EPC; entre ellos, Klobucar *et al*. (2012) advierten que la sobredependencia de estos sistemas computarizados podría resultar en una fijación en el error y en rasgos superficiales, como la longitud. Bridgeman, Trapani y Yigal (2012) descubrieron que el texto que se desviaba ligeramente del tópico obtenía un puntaje más alto y explicaron que, para ciertos grupos, los textos tienden a obtener puntajes más altos con el *e-rater* que con los evaluadores humanos. Vojak *et al*. (2011) usaron los programas *Criterion*, *MyAccess! Essayrater*, *MyCompLab*, *Project Essay Grader* y *Calibrated Peer Review*, y encontraron evidencia de enfoques preplaneados, retroalimentación no específica, identificación de errores incorrecta, énfasis en las convenciones y tendencia a valorar longitud por sobre contenido, pero también que dichos programas "asumían" que los estudiantes exitosos reproducían estructuras convencionales. Neal (2011) sostiene que, de alguna manera, los profesionales de composición escrita han perdido la idea de cómo y por qué la gente lee y escribe en situaciones retóricas significativas y hace notar que la EPC es una solución barata a un problema que no se ha definido. Dikli (2010) dividió en dos grupos a 12 aprendices de inglés; seis recibieron retroalimentación desde una computadora y los otros seis de un profesor; los dos tipos de retroalimentación difirieron enormemente en términos de longitud, facilidad de uso, redundancia y consistencia, concluyendo que el programa no satisfacía las necesidades de los hablantes no nativos.

Para comparar los resultados del *eGrader* con jueces humanos, Byrne, Tang, Truduc y Tang (2010) usaron 33 ensayos, pero decidieron no usar este sistema porque no pudo detectar la ironía, la metáfora, los giros lingüísticos, las connotaciones y otros mecanismos retóricos, y penalizaba el pensamiento y la escritura original o diferente. Es decir, no pudieron identificar la pragmática del texto. McCurry (2010) analizó los resultados arrojados tanto por humanos como por aplicaciones automatizadas, y concluyó que estas no pueden dar puntaje a tareas de escritura abiertas y amplias con tanta confiabilidad como lo hacen los evaluadores humanos. Scharber, Dexter y Riedel (2008) usaron un sistema en línea que incluyó una opción de EPC para retroalimentación formativa en un estudio de cuatro casos, pero dicho sistema no fue lo suficientemente sofisticado para saber qué tipo de revisión específica se necesitaba. Chen y Cheng (2008) encontraron que la retroalimentación dada por el sistema *MyAccess!* era más útil durante la producción de borradores y la revisión, pero solo cuando era seguida por la retroalimentación de los pares o de los profesores. Cuando los participantes en su investigación usaron *MyAccess!* por cuenta propia, estos se frustraban, su aprendizaje era limitado, y percibieron el *software* y su retroalimentación de forma negativa.

Sandene *et al*. (2005) informaron sobre los resultados de un estudio de escritura en línea por estudiantes de octavo año, compararon los resultados con aquellos estudiantes que usaron papel y lápiz y demostraron que la EPC no coincidía con los puntajes asignados por evaluadores humanos; la EPC otorgaba puntajes medios significativamente más altos que los asignados por humanos. Pero Whithaus (2005) critica la evaluación de la escritura a gran escala, ya que estimula a los estudiantes

ISSN 0122-6339 • ISSNe 2248-6798 • Vol. 24, No. 2 (julio/diciembre de 2019), pp. 227-240.

a predeterminar cualquier material frente a ellos más que a pensar en cómo comunicarse con diferentes audiencias para diferentes propósitos y a través de diferentes modalidades; sostiene que las computadoras pueden ser apropiadas si la tarea es reproducir hechos conocidos; y añade que el *análisis semántico latente* puede servir para evaluar conocimiento reproducible o formatos textuales "muertos", como el ensayo de cinco párrafos, pero no pueden evaluar con justicia la escritura multimodal como los blogs, la mensajería instantánea o los portafolios electrónicos.

Burstein y Marcus (2003) explican cómo una máquina puede evaluar un criterio de buena escritura (organización) que se cree que no puede ser medido empíricamente y sostienen que los sistemas de análisis discursivo pueden identificar en forma confiable tesis y enunciados en las conclusiones. Sin embargo, Herrington y Moran (2001) sostienen que la EPC no trata a la escritura como una interacción retórica entre escritores y lectores; transmite el mensaje de que las lecturas realizadas por personas no son confiables, son irrelevantes y reemplazables; y que los rasgos superficiales de un idioma importan más que el contenido y las interacciones entre el lector y el texto. Por último, Anson (2006) se opone a la EPC diciendo que las máquinas no pueden leer discurso natural con la complejidad que los humanos lo hacen, aunque argumenta a favor de investigar cómo las tecnologías digitales pueden analizar prosa para proporcionar información útil a escritores en desarrollo.

Metodología

Considerando la importancia a futuro que puede tener la decisión de evaluar trabajos escritos con instrumentos automatizados, el método consistió en consultar a estudiantes universitarios la percepción que tienen de una posible EPC. La consulta contenía solo una pregunta escrita en castellano: "¿Estarías dispuesto(a) a someter la evaluación de tus trabajos escritos a una computadora?". Como se ve, la pregunta requiere una respuesta positiva o negativa, lo cual no les tomaría mucho tiempo en contestarla; sin embargo, la pregunta también ofrecía la oportunidad de escribir comentarios. Estos sirvieron para clasificar sus percepciones como positivas, negativas o indecisas. Los participantes contestaron la pregunta en forma anónima en un laboratorio de computación y se les dio 10-15 minutos para hacerlo, con el fin de brindarles la oportunidad de añadir sus comentarios. Después de contestar la pregunta, los participantes solo tenían que presionar "Enviar".

Este estudio de percepciones es eminentemente cualitativo, el cual permitió que emergieran categorías de las respuestas de los estudiantes gracias a los comentarios que generó la pregunta.

Participantes

Los participantes fueron 50 estudiantes de la carrera de Pedagogía en Inglés de la Pontificia Universidad Católica de Valparaíso (PUCV), Chile. De estos, 45 contestaron la pregunta. Al momento de hacerlo, los estudiantes cursaban la asignatura obligatoria Inglés Escrito con Propósitos Profesionales (IEPP), que se imparte en ocho horas semanales y se ubica en el primer semestre del cuarto año del currículo. Por ende, los estudiantes ya tenían experiencia con la escritura en lengua extranjera, pero sus textos nunca han sido sometidos a algún tipo de EPC.

Como parte del sílabo de IEPP, los estudiantes ya habían escrito cuatro ensayos cortos antes de contestar la pregunta. En estos, ellos manifiestan su reacción a algún tópico que les interesa y deben entregar sus ensayos cada dos semanas. El profesor proporciona retroalimentación manual en una copia impresa, siguiendo un código de corrección previamente subido a la plataforma virtual del curso y que los estudiantes utilizan para corregir sus errores. Los participantes también habían sido entrenados en escritura técnica (cartas comerciales, *curriculum vitae*, cartas de invitación, entre otros contenidos). El sílabo de dicha asignatura también incluye la práctica de escribir resúmenes, síntesis, textos expositivos y argumentativos, y un diario

ISSN 0122-6339 • ISSNe 2248-6798 • Vol. 24, No. 2 (julio/diciembre de 2019), pp. 227-240.

del escritor, un género textual donde el estudiante expresa sus actitudes hacia su propio proceso de escritura, incluyendo sus debilidades, fortalezas y experiencias de escritura en su vida.

Resultados

De acuerdo con las respuestas a la pregunta "¿Estarías dispuesto(a) a someter la evaluación de tus trabajos escritos a una computadora?", los resultados fueron clasificados en "sí", en "no" y en "indeciso(a)". Se obtuvieron 4 "sí" (8,9 %); 23 "no" (51,1 %); y 18 "indecisos(as)" (40 %). A continuación, las respuestas más representativas que se obtuvieron: las respuestas "sí" son muy concisas; en cambio, atendiendo al espacio de los otros dos tipos de respuestas se presentan aquí extractos de comentarios de los participantes.

Sí

"Sí, estaría dispuesta, pero si es que el computador está menos influenciado que el profesor".

"Sí, es más práctico y se ahorra papel".

"Creo que sería una buena instancia para analizar mi trabajo".

No

"[...] una máquina no evaluará correctamente nuestras respuestas, a menos que sea solo en pruebas estandarizadas".

"[...[el feedback personal [...] es mucho más efectivo que el de una computadora a la cual no le puedes hacer preguntas respecto a tus dudas que tienes sobre la corrección".

"[...] se terminará por desprestigiar y debilitar el rol del docente...".

"[es] una buena alternativa para simplificar el trabajo de los profesores, sin embargo [...] no es un método totalmente confiable para los estudiantes".

"[...] sería siempre más confiable para nosotros que lo revise un experto en persona".

"[...] prefiero que el feedback sea directamente desde el profesor porque de esa forma puedo entender mejor mis errores".

"Una computadora o cualquier tipo de robot no tiene capacidad de pensamiento ni de inferencia [...] no se considerarían elementos importantes como la audiencia."

"Los textos son leídos por personas, por lo cual sería más lógico que sean revisados por las mismas."

"[...] en el caso de un ensayo, el profesor tiene muchas más facultades para poder inferir o entender completamente la intención [...]".

"[...] un profesor puede ser capaz de señalar errores más específicos, como problemas de coherencia, cohesión, shift in person [cambios de elementos referenciales], entre otros".

"No me gustaría que fuera solo el programa quien evaluara".

"No sé qué tan práctico sería dejar tal responsabilidad a un computador que en cualquier momento puede fallar".

"[...] el criterio del evaluador jugará un rol importante al momento de interpretar lo que los alumnos respondan, [...]".

"[...] el profesor dará cuenta de los verdaderos errores y podrá justificarlos. Si bien la tecnología está muy avanzada, creo que no sería capaz de distinguir ciertos patrones que son más humanos".

"[...] prefiero tener profesores por sobre computadores evaluando mis trabajos escritos, ya que ellos pueden identificar errores que las computadoras no".

Indecisos

"No debería ser la única técnica de corrección sino más bien algo complementario a lo que el profesor podría corregir".

"No estaría del todo segura de otorgar todo el trabajo a una computadora, puesto que seguramente no se basará mucho en rúbricas, sino que en errores gramaticales".

"[...] debe haber una intervención humana, principalmente al momento de otorgar retroalimentación".

ISSN 0122-6339 • ISSNe 2248-6798 • Vol. 24, No. 2 (julio/diciembre de 2019), pp. 227-240.

"[...] una computadora quizás no podría analizar, como la intención, el contexto, ironía/sarcasmo, las emociones, [...] Podrá ser capaz de reconocer un argumento bien planteado?".

"Apoyo la asistencia por computadora para las evaluaciones, pero que sea bajo un supervisor".

"Depende de los criterios que sean evaluados".

"[...] prefiero que el feedback me lo entregue un profesor con el cual pueda interactuar".

"Si se está evaluando solamente gramática, tal vez. En cualquier otro caso, me parece poco apropiado".

"Sí y no, ya que, si bien un computador puede fácilmente corregir o identificar errores de la mecánica de la escritura, hasta ahora las inteligencias artificiales no son capaces de comprender la pragmática del lenguaje".

"Solo para trabajos cuyo objetivo es evaluar el uso correcto de estructuras gramaticales [...], debido a que no creo que una computadora pueda evaluar el contenido (la relevancia) de un trabajo escrito".

De los cuatro estudiantes que contestaron con un categórico "sí", solo uno de ellos escribió únicamente esa palabra; los otros tres se refirieron a lo práctico de ese tipo de evaluación, a que el profesor está más influenciado que la computadora y a que esta puede analizar sus trabajos. En todo caso, estas cuatro respuestas "sí" son muy cortas y, en una comparación con los otros dos tipos de respuestas categorizadas, se puede decir que son poco profundas, considerando que se dio la oportunidad de agregar comentarios.

Los "no" (23) fueron decisivos y la mayoría de las percepciones se refiere a la importancia que tiene el profesor en la entrega de retroalimentación, algo que una computadora difícilmente puede suplir. Tal vez al contestar la encuesta, los participantes que respondieron con un "no" se sintieron amenazados por su calidad de futuros profesores de un idioma extranjero, quienes al ejercer como tales deberán enseñar la habilidad de escritura y se proyectaron como reemplazados por una máquina, lo cual minusvaloraría su profesión.

Otras percepciones tienen que ver con la dificultad de diseñar una computadora que detecte errores tan bien como lo haría un profesor, indicando, por ejemplo, la intención del escritor, el contexto, las ironías y las emociones, o la evaluación de un poema, coincidiendo así con partes del contenido de lo que este artículo propone.

Los 18 "indecisos(as)" son aquellos estudiantes que contestaron con un "sí" y un "no", o aquellos que en un principio contestaron con un "sí", pero después añaden "sin embargo" o "no obstante" o iniciaron sus respuestas con "depende". Al parecer, para el 40 % que ellos representan, la pregunta les resultó más sorpresiva que al resto, puesto que sus respuestas no contienen argumentos sólidos para rechazar o aceptar la evaluación automatizada de sus escritos. Generalmente, se refieren a la capacidad que tendría una computadora de detectar cierto tipo de errores, pero algunos añaden que no todos los errores serían detectados tan bien como lo haría un profesor, alineándose así con los que contestaron con un categórico "no". La mayoría de estos participantes expresa su disposición a que sus trabajos escritos sean evaluados por una computadora, pero solo como algo complementario a lo que el profesor podría entregarles como enseñanza personalizada; en particular, se refieren a la capacidad del profesor para proporcionar retroalimentación. Esta sería una forma de hacer notar que la computadora puede entregar datos objetivos, pero que se necesita además tanto la subjetividad del profesor como su criterio para evaluar un producto escrito.

Siendo estudiantes de pedagogía en inglés, ninguno de ellos hizo la diferencia entre ser evaluado en primera lengua o en segunda lengua con una computadora; puede que hayan pensado que una diferencia en el código no sería relevante.

Conclusiones

Sin duda este es un artículo que pone en alerta una potencial práctica de la EPC en contextos escolares y universitarios. Definitivamente, aún faltan

ISSN 0122-6339 • ISSNe 2248-6798 • Vol. 24, No. 2 (julio/diciembre de 2019), pp. 227-240.

estudios de mayor envergadura que exploren las preferencias y percepciones tanto de profesores y estudiantes como también de instituciones educacionales. En esa línea, se podría ampliar el espectro de participantes y los resultados serían más consistentes. Con todo, si las instituciones se deciden a usar EpC, deberían justificar sus ventajas y especificar sus desventajas. También sería necesario triangular los datos para obtener resultados generalizables, lo cual podría realizarse con entrevistas a evaluadores y a evaluados.

Es fuerte la propaganda de herramientas computarizadas para comercializar su uso en establecimientos de enseñanza en Estados Unidos y tal vez en otros países; sin embargo, las falencias basadas en investigaciones y presentadas en este artículo no deberían permitir el uso indiscriminado de ellas. En general, estos instrumentos computarizados son útiles cuando se trata de evaluar las convenciones mecánicas del lenguaje en, por ejemplo, pruebas de selección múltiple, o llenar espacios en blanco en una prueba de gramática, o identificar la ortografía correcta, o completar oraciones descontextualizadas. Evidentemente, las consideraciones inmediatamente anteriores no constituirían producción escrita como tal. Quizá la utilización de estas herramientas serviría para evaluar precisamente esos aspectos del lenguaje en forma masiva –por ejemplo, en una prueba de selección universitaria o en pruebas de certificación internacional para un segundo idioma–, pero esto aumentaría los costos en que se debería incurrir para el mantenimiento del sistema y de los equipos.

Por otro lado, en los sistemas computarizados aludidos se dificulta la posibilidad de incluir la validez del constructo de la escritura en tanto las habilidades que ponen a prueba no coinciden con las habilidades que se necesitan para componer un texto; a este respecto, un caso principalmente conflictivo es la apreciación del pensamiento crítico y del razonamiento que permite realizar tareas cuya ejecución dependerá del desarrollo intelectual de los estudiantes. A esto se debe agregar la evaluación del lenguaje figurativo (metáfora, ironía, hipérbole) y su efectividad o la pragmática de un texto.

Si aceptamos que todo acto de escritura se inserta en una situación retórica, la posibilidad de que una computadora colabore con el desarrollo de un tópico no es tan lejana, considerando que puede ampliar léxico específico, por ejemplo, o proporcionar ideas para el desarrollo de un tópico; pero en cuanto a la eventualidad que dicha máquina evalúe el propósito de un texto o la intención del autor, esta posibilidad se aleja. Ya más remota es la posibilidad de que la máquina sea una audiencia válida: estaríamos en definitiva aceptando a la vez que esta máquina es también miembro de la misma sociedad en que vivimos. Escribirle a una máquina es completamente antinatural cuando de situaciones retóricas se trata.

Además, la evaluación en todo campo del saber implica explicaciones de cómo, cuándo, dónde, quién y también por qué. Implica en último término explicación de errores y cuando no la provee un ser humano, se tiene la esperanza de que la computadora la proporcione; sin embargo, las máquinas nunca dan explicaciones, en particular cuando se trata de dar razones. También es posible diseñar una computadora que lea y evalúe con rapidez la exactitud con que un texto presenta su inserción dentro de los cánones de un género textual determinado, programándole sistemas que reconozcan ciertos rasgos de dicho género (a saber, frases fijas en un texto expositivo o las que se pueden encontrar en un texto descriptivo). No obstante, se pone en duda que el jurado de un concurso de poesía delegue sus funciones a una máquina para juzgar el mejor poema, pues no existe impacto de la escritura en los sentimientos, emociones y pensamientos de los miembros de ese jurado y, llevando lo anterior un poco más lejos, la máquina no tiene la intención de hacer cambiar de opinión a un lector ni tampoco argumentar en pro o en contra de una posición, porque las computadoras –hasta el momento del estado del arte– no opinan ni argumentan.

ISSN 0122-6339 • ISSNe 2248-6798 • Vol. 24, No. 2 (julio/diciembre de 2019), pp. 227-240.

Reconocimientos

Este artículo, que no está adscrito a ningún proyecto de investigación en curso ni cuenta con apoyo financiero, surge de la inquietud por indagar las percepciones de estudiantes universitarios acerca de su disposición a recibir retroalimentación y evaluación de sus composiciones escritas a través de un instrumento automatizado.

Referencias bibliográficas

Anson, C. (2006). Can't touch this: Reflections on the servitude of computers as readers. En P.F. Ericsson y R. Haswell (eds.), *Machine scoring of student essays* (pp. 38-56). Logan, UT: Utah State University Press**.** Recuperado de https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1138&context=usupress_pubs

Benítez, R. (2001). La situación retórica: Su importancia en el aprendizaje y en la enseñanza de la producción escrita. *Revista Signos*, *XXXIV*(48), 29-48. Universidad Católica de Valparaíso.

Blakeslee, A. (2001). *Interacting with audiences: Social influences on the production of scientific writing*. Mahwah, NJ: Erlbaum Associates, Inc.

Bridgeman, B., Trapani, C. y Yigal, A. (2012). Comparison of human and machine scoring of essays: Differences by gender, ethnicity, and country. *Applied Measurement in Education, 25*(1), 27-40.

Burstein, J. y Marcus, D. (2003). A machine learning approach for identification of thesis and conclusion statements in student essays. *Computers and the Humanities,* 37, 455-467.

Byrne, R., Tang, M., Truduc, J. y Tang, M. (2010). eGrader, a software application that automatically scores student essays: with a postscript on the ethical complexities. *Journal of Systemics, Cybernetics & Informatics, 8*(6), 30-35.

Carbonell, J. (1992). El procesamiento del lenguaje natural, tecnología en transición. En *Actas del Congreso de la Lengua Española* (pp. 247-250). Sevilla, 7 al 10 de octubre. Recuperado de https://cvc.cervantes.es/obref/congresos/sevilla/tecnologias/ponenc_carbonell.htm

Chen, E. y Cheng, W. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning and Technology, 12*(2), 94-112.

Collier, L. (2012). *Robo-grading of student writing is fueled by new study-But earns "F" from experts*. Recuperado de http://www.lornacollier.com/robogradingCC912.pdf

Condon, W. (2013). Large-scale assessment, locally-developed measures, and automated scoring of essays: Fishing for red herrings? *Assessing Writing* 18, 100-108.

Corbo, V. (julio de 2017). Preparándonos para el siglo XXI: Retos y oportunidades de la automatización. *El Mercurio de Santiago*, B, p. 5.

Corbo, V. (mayo de 2018). Los retos de las nuevas tecnologías para las políticas públicas. *El Mercurio de Santiago*, B, p. 7.

Dikli, S. (2010). The nature of automated essay scoring feedback. *CALICO Journal, 28*(1), 99-134.

Ede, L. y Lunsford, A. (1988). Audience addressed/audience invoked: The role of audience in composition theory and pedagogy. En G. Tate y E. Corbett (eds.), *The writing teacher sourcebook* (pp. 169-184). Oxford: Oxford University Press.

Haimson, L. (2016). Should you trust a computer to grade your child's writing on Common Core tests? *The Washington Post*. Recuperado de https://www.washingtonpost.com/news/answer-sheet/wp/2016/05/05/should-you-trust-a-computer-to-grade-your-childs-writing-on-common-core-tests/?utm_term=.b575f2f27950

Harris, J. (1990). The idea of community in the study of writing. En R. Graves (ed.), *Rhetoric and composition. A sourcebook for teachers and writers* (pp. 267-278). Portsmouth, Boynton: Cook Publishers.

Herrington, A. y Moran, C. (2001). What happens when machines read our students' writing? *College English,* 63, 480-499.

Hesse, D. (2013). Interview with Valerie Strauss. Grading writing: The art and science –and why computers can't do it. *The Washington Post*. Recuperado de https://www.washingtonpost.com/news/

answer-sheet/wp/2013/05/02/grading-writing-the-art-and-science-and-why-computers-cant-do-it/

Jones, I. (2003). Collaborative writing and children's use of literate language: A sequential analysis of social interaction. *Journal of Early Childhood Literacy*, *3*(2), 165-178.

Klobucar, A., Deane, P., Elliot, N., Raminie, C., Deess, P. y Rudniy. A. (2012). Automated essay scoring and the search for valid writing assessment. En C. Bazerman, C. Dean, J. Early, K. Lunsford, S. Null, P. Rogers y A. Stansell (eds.), *International Advances in Writing Research: Cultures, Places, Measures* (pp. 103-119). Fort Collins, CO: WAC Clearinghouse & Parlor Press.

Kolowich, S. (2014). Writing instructor, skeptical of automated grading, pits machine vs. machine. *The Chronicle of Higher Education*. Recuperado de https://www.chronicle.com/article/Writing-Instructor-Skeptical/146211

Marcano, J. (diciembre de 2018). En 25 años más existirán máquinas capaces de imitar la inteligencia humana. *El Mercurio de Santiago*, A, p. 14.

McCurry, D. (2010). Can machine scoring deal with broad and open writing tests as well as human readers? *Assessing Writing, 15*(2), 118-129.

National Council of Teachers of English (NCTE) (2013). *NCTE position statement on machine scoring*. Recuperado de http://www2.ncte.org/statement/machine_scoring/

Neal, M. (2011). *Writing assessment and the revolution in digital texts and technologies*. Nueva York: Teachers College Press.

Noël, S. y Robert, M.J. (2004). Empirical study on collaborative writing: What do co-authors do, use, and like? *Computer-Supported Cooperative Work*, 13, 63-89.

Rodríguez, J. (julio de 2017a). Tecnologías y biociencias: El futuro que ya llegó. *El Mercurio de Santiago*, E, p. 3.

Rodríguez, J. (diciembre de 2017b). La utopía del hombre-máquina: ¿El futuro será transhumanista? *El Mercurio de Santiago*, E, 2-3.

Sandene, B., Horkay, N., Bennet, R., Elliot, R., Allen, N., Braswell, J., Kaplan, B. y Oranje, A. (2005). Part II: *Online writing assessment. Online assessment in mathematics and writing: Reports from the NAEP Technology-Based Assessment Project, Research and Development Series. NCES 2005-457*. U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing.

Scharber, C., Dexter, S. y Riedel, E. (2008). Students' experiences with an automated essay scorer. *Journal of Technology, Learning and Assessment, 7*(1), 1-45. Recuperado de http://www.jtla.org

Shermis, M., Mzumara, H., Olson, J. y Harrington, S. (2001). On-line grading of student essays: PEG goes on the world wide web. *Assessment and Evaluation in Higher Education, 26*(3), 247-260.

Spivey, N. (1997). *The constructivist metaphor: Reading, writing, and the making of meaning*. San Diego, CA: Academic Press.

Thomas, P. (2016). *Not How to Enjoy Grading but Why to Stop Grading*. Recuperado de https://radicalscholarship.wordpress.com/2016/05/06/not-how-to-enjoy-grading-but-why-to-stop-grading

Vojak, C., Kline, S., Cope, B., McCarthey, S. y Kalantzis, M. (2011). New spaces and old places: An analysis of writing assessment software. *Computers and Composition,* 28, 97-111.

Whithaus, C. (2005). *Teaching and evaluating writing in the age of computers and high-stakes testing*. Mahwah, NJ: Lawrence Erlbaum.

Whittington, D. y Hunt, H. (1999). Approaches to the computerized assessment of free text responses. En *Proceedings of the Third Annual Computer Assisted Assessment Conference*. Loughborough, Inglaterra: Loughborough University.

Williams, J. (1998). *Preparing to teach writing. Research, theory, and practice*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Wilson, B. (1996). *Constructivist learning environments: Case studies in instrumental designs*. Englewood Cliffs, NJ: Educational Technology Publications.