



EVALUACION AUTOMATIZADA Y SEMIAUTOMATIZADA DE LA CALIDAD DE TEXTOS ESCRITOS: UNA REVISION SISTEMATICA

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

Fernando Lillo-Fuentes (*)

Universitat Pompeu Fabra

René Venegas

Ignacio Lobos

Pontificia Universidad Católica de Valparaíso

Resumen

La tecnología para la evaluación de la escritura se ha desarrollado desde los años 60. Actualmente, el procesamiento del lenguaje natural (Shermis, 2020) ha permitido una evolución considerable. No obstante, lo fecundo de este campo, no se han encontrado revisiones sistemáticas que abordan las cuestiones siguientes: ¿en qué país, género y niveles se han desarrollado propuestas para evaluar la calidad de la escritura?, ¿cuáles son las consideraciones didácticas, tecnológicas y teóricas de estas herramientas?, ¿cuál es el rol que desempeñan los docentes en el diseño y uso de ellas? y ¿qué resultados se han obtenido? En este artículo, se revisaron 164 investigaciones entre los años 1966 y 2022. De ello se destaca: a) que la evaluación automática ha pasado de focalizarse en puntuaciones fiables, imparciales y rápidas a una evaluación centrada en la retroalimentación, b) que el docente cumple un rol primordial en el diseño y uso de las herramientas y c) que las herramientas son un apoyo útil. Además, se identifica un escaso desarrollo de herramientas para la lengua española.

Palabras clave: Evaluación, evaluación formativa, retroalimentación, aplicación informática, escritura.

Abstract

Technology for grading and evaluating written texts has been developed since the early 1960s in English-speaking countries and, preferably, in the essay genre (Chen and Cheng, 2008; Page, 2003). The first systems focused on delivering a rating of the writing evolved, thanks to natural language processing and artificial intelligence, reaching the point of providing feedback for different discursive genres (Shermis, 2020). Despite these and other advances, these techniques still have detractors who claim that this type of evaluation would seek to replace teaching tasks, would not resemble that performed by humans or that it would only be concerned by formal aspects (Vajjala, 2018). However, many of these concerns are based on a lack of knowledge about the evolution of this field, the fear of human impersonation (Palermo and Wilson, 2020) and ignorance of the new paradigms of this type of evaluation (Shermis, 2020). Despite the fecundity of this field, to date, no systematic reviews have been found that focus on the aforementioned points, so in this study we will focus on answering the following questions: What are the didactic, technological, and theoretical considerations of the tools for assessing the quality of writing? What is the role of teachers or evaluators in the design and use of these tools? For what purposes are the tools designed and used? What are the results obtained? In what country, language, genre, and levels have tools for assessing writing quality been developed? To answer these questions, a systematic review was developed, following the guidelines for Systematic Reviews and Meta-Analyses (PRISMA, 2020). Thus, 164 scientific research articles developed between 1966 and 2022 were selected. Among the results, it is observed that the initial objectives of automatic evaluation have shifted from focusing on reliable, unbiased, and rapid scores to focus on formative evaluation centered on feedback. At the same time, it was found that the role of the teacher is paramount in 94% of the works reviewed, since the tools are presented as a support and not as a replacement of the teaching work. Based on these results, we argue that this type of review can be of great help to learn about the current state of automated assessment and writing feedback. In addition, it can be useful for those who are interested in building contributions in this field.

(*) Autor para correspondencia:

Fernando Lillo-Fuentes

(Becario ANID Folio: 72200193)

University Pompeu Fabra

Carrer Roc Boronat 138, Despacho 53.212,

Barcelona

Correo de contacto:

fernandogabriel.lillo@upf.edu

©2010, Perspectiva Educacional

[Http://www.perspectivaeducacional.cl](http://www.perspectivaeducacional.cl)

RECIBIDO: 15.09.2022

ACEPTADO: 11.01.2023

DOI: 10.4151/07189729-Vol.62-Iss.2-Art.1420

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT
OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

Keywords: Automated assessment, quality of written texts, computerized formative assessment, feedback.

1. Introducción

La evaluación de la escritura mediada por ordenadores se ha desarrollado desde principios de los años 60 como un campo interdisciplinario que permite evaluar y puntuar los escritos de sus productores (Eid y Wanas, 2017; Vajjala, 2018). Para cumplir estos fines, en sus inicios, se empleó un enfoque estadístico en el que se otorgaba puntaje a partir de una serie de características superficiales de los textos (conteo de palabras, de oraciones, de verbos, algunos índices de legibilidad y conteo de errores gramaticales). Con el avance tecnológico, la inclusión del aprendizaje automático y la inteligencia artificial, estos sistemas se perfeccionaron y comenzaron a integrar índices que permiten retroalimentar textos escritos.

Así, hasta la fecha, se han planteado diferentes sistemas automáticos que permiten cumplir con la evaluación de los escritos. Dentro de ellos, existen algunas propuestas que entregan una puntuación o una evaluación holística, a partir de un entrenamiento previo con un conjunto de textos revisados por evaluadores humanos (Burstein, 2003; Dikli, 2006). También, existen otros sistemas que poseen objetivos similares, pero que se han centrado no solo en puntuar el escrito, sino en entregar una retroalimentación de los textos. En la actualidad, se han enriquecido los sistemas ya propuestos y se ha incursionado en diferentes géneros académicos como, por ejemplo, el ensayo, el artículo de investigación científica y el trabajo final de grado, entre otros.

Si bien las herramientas de evaluación automatizada se han pulido y han alcanzado resultados significativos con el paso de los años, aún existen algunos escépticos respecto a su alcance y efectividad. Al respecto, muchos de sus detractores mencionan que este tipo de evaluación solo busca reemplazar el trabajo de los docentes y que en ningún caso su labor se asemeja a la de un humano, pues solo se centra en aspectos formales (Vajjala, 2018). Sin embargo, muchas de estas inquietudes se basan en un desconocimiento frente a la evolución que ha tenido este campo, el miedo a la suplantación humana (Palermo y Wilson, 2020), la ignorancia de los nuevos paradigmas de este tipo de evaluación (Shermis, 2020) y el desconocimiento respecto a los objetivos que tienen las nuevas propuestas.

Por lo anterior, se hace necesario conocer el estado actual de este campo, indagando en los avances que ha tenido en el tiempo, los objetivos que persiguen sus creadores, los resultados alcanzados, el impacto de su uso y los constructos teóricos que subyacen a cada herramienta. Sin embargo, hasta la fecha, no se han hallado revisiones sistemáticas que aborden estos puntos. Si bien se han realizado revisiones en torno a las herramientas elaboradas para evaluar la escritura (Granic y Marangunic, 2019; Strobl et al., 2019), estas no han respondido a las preguntas que nos formulamos en este trabajo. Por ello, en este estudio nos enfocaremos en responder las siguientes preguntas: ¿En qué país, género y niveles se han desarrollado propuestas para evaluar la calidad de la escritura?, ¿cuáles

son las consideraciones didácticas, tecnológicas y teóricas de estas herramientas?, ¿con qué propósitos se diseñan y emplean las herramientas utilizadas para evaluar automáticamente textos escritos?, ¿cuál es el rol que desempeñan los docentes en el diseño y uso de ellas? y ¿qué resultados se han obtenido? Para resolver estas interrogantes, desarrollaremos una revisión sistemática que sigue los lineamientos para revisiones sistemáticas y metaanálisis (Urrútia y Bonfill, 2010).

En lo que sigue, se expondrán los antecedentes teóricos del trabajo y se detallará la metodología empleada. Luego, se darán a conocer los resultados obtenidos, divididos en torno a las preguntas motrices que originan la revisión. Finalmente, se expondrá una breve discusión y algunos comentarios de cierre.

2. MARCO TEORICO

2.1. Evaluation de la escritura

La evaluación forma parte fundamental del proceso de enseñanza-aprendizaje (Harvey, 2002; Fernández, 2017). En efecto, su relevancia se ha traducido en permanentes esfuerzos de académicos y docentes para comprender el fenómeno cada vez mejor, tanto en la identificación de distintos tipos de evaluación como en sus eventuales alcances. Particularmente, la evaluación formativa es definida como aquella en la que se recoge información durante un proceso de enseñanza- aprendizaje determinado para proporcionar comentarios u orientaciones que apunten a mejorar el desempeño de los estudiantes en virtud del objetivo de aprendizaje esperado (Talanquer, 2015; Cowie y Bell, 1999).

Por otro lado, la evaluación sumativa se orienta a proporcionar un juicio o valor sobre el producto del aprendizaje y no proporciona orientaciones durante el proceso de aprendizaje (Ruz, 2018). En cambio, en palabras de Ruz (2018, p. 21), la evaluación formativa “se da generalmente durante el proceso de enseñanza y aprendizaje; otorga información a estudiantes y docentes acerca de un momento particular, que debe contribuir a tomar decisiones adecuadas que mejoren las actividades de enseñanza y ayuden a los estudiantes a que potencien sus aprendizajes”.

En lo que refiere específicamente a la producción textual, existen distintas propuestas sobre los aspectos centrales que deben ser considerados como criterios al momento de evaluar la producción de un estudiante (Morales, 2004; Contreras et al., 2009; Navarro et al., 2019; Andueza, 2019; Muñoz y Pérez, 2021, entre otros). Por ejemplo, Morales (2004) sugiere considerar la ortografía, vocabulario, gramática, cohesión y estructura. Contreras et al. (2009) proponen fijarse en ideas y contenidos, organización, voz del escritor, elección de palabras, fluidez de las oraciones y convenciones ortográficas. Por su parte, Navarro et al. (2019) defienden una evaluación a través de rúbricas

holísticas que midan el cumplimiento de objetivos, superando el nivel superficial de la puntuación y ortografía. Andueza (2019) propone hacer énfasis en 4 dimensiones: coherencia textual, lenguaje académico, manejo de fuentes y dominio del género discursivo. Por último, Muñoz y Pérez (2021) hacen referencia a la medición de habilidades de escritura mediante los criterios de cohesión, coherencia y comprensión lectora para seleccionar información.

Como se puede apreciar, no existe un consenso absoluto acerca de los criterios que deben ser evaluados en la producción textual, pero sí un interés permanente de establecer con claridad los elementos que permiten que un texto sea una producción exitosa. De esta manera, se observan criterios de evaluación que apuntan a aspectos gramaticales, la organización de ideas, la coherencia, la cohesión y otro grupo que se centra en la medición de habilidades asociadas con un plano funcional del lenguaje.

Una práctica estrechamente asociada con la acción de evaluar es la retroalimentación, entendida como la devolución, comentario u orientación proporcionada a un aprendiz en el marco de un proceso evaluativo. La retroalimentación, como parte del proceso de enseñanza-aprendizaje, se posiciona como una acción importante en la tarea de alcanzar las metas propuestas para los estudiantes (Díaz López, 2018).

3. MARCO METODOLÓGICO

El presente trabajo busca determinar las investigaciones que se han producido en torno a la evaluación automática y semiautomática de la calidad de textos escritos, desde su origen hasta los tiempos actuales. Específicamente, nos interesamos en conocer las herramientas, sistemas y prototipos que se han empleado para ello, profundizando en los objetivos que subyacen a estas propuestas y sus constructos teóricos.

Las preguntas que guían este trabajo son: ¿En qué país, género y niveles se han desarrollado propuestas para evaluar la calidad de la escritura?, ¿cuáles son las consideraciones didácticas, tecnológicas y teóricas de estas herramientas?, ¿con qué propósitos se diseñan y emplean las herramientas utilizadas para evaluar automáticamente textos escritos?, ¿cuál es el rol que desempeñan los docentes en el diseño y uso de ellas? y ¿qué resultados se han obtenido? Para responder a estas preguntas y cumplir el propósito descrito, realizamos una revisión sistemática de la literatura siguiendo los lineamientos para Revisiones Sistemáticas y Metaanálisis de PRISMA (Urrútia y Bonfill, 2010).

3.1. Estrategias de búsqueda y criterios de inclusion/exclusion

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

La identificación de los trabajos que comprenden la revisión sistemática se realizó en el mes de febrero de 2022, mediante una búsqueda en las bases de datos *Scopus*, *Dialnet*, *Scielo* y *Google Scholar*. Se consideraron artículos científicos y actas de congreso escritas en inglés y español. La inclusión de actas de congreso se debió a que, en una búsqueda piloto se notó que, en el campo de la evaluación automática, muchas de las propuestas son presentadas como artículos de actas de congreso (*proceedings*).

Como palabras clave de la búsqueda se emplearon términos asociados con la evaluación automática de la escritura en inglés y español, además de operadores lógicos (AND, OR) para maximizar sus combinaciones. De esta manera, empleamos: (automática OR informatizada OR automatizada) AND (evaluación OR puntuación OR calificación) AND (escritura OR texto) o en inglés (automatic OR computerized OR automated) AND (score OR feedback OR assess* OR evaluat*) AND writ*.

Respecto a los criterios de inclusión y exclusión estos fueron los que se exponen en la Tabla 1 a continuación.

Tabla 1

Criterios de inclusión y exclusión de la revisión sistemática

Criterios de inclusión	Criterios de exclusión
Evaluación de la calidad de la escritura/textual.	Evaluación de la calidad de la oralidad, traducción, lenguaje de señas, etc.
Estudios con herramientas, prototipos o sistemas automáticos. Trabajos que los presentaran.	Revisiones sistemáticas del tema, estudios análogos.
Ámbito lingüístico, didáctico, pedagógico y educativo.	Ámbito de programación, matemático, medioambiental u otros.

Fuente: elaboración propia.

Como se observa, el primer criterio consiste en estudios que se centren en la evaluación de la calidad de la escritura, de un texto o género escrito, dejando fuera evaluaciones de la calidad del lenguaje oral, la traducción o el lenguaje de señas, entre otros. También se incluyen trabajos que emplean herramientas, sistemas o prototipos automatizados, así como también aquellos trabajos que proponen o presentan una nueva herramienta. Sin embargo, se excluyen revisiones sistemáticas del tema en cuestión o intervenciones que evalúen la calidad de la escritura, pero sin el empleo de herramientas computacionales. Finalmente, se consignan aquellos trabajos desarrollados en el ámbito

lingüístico, didáctico y pedagógico, pero se dejan fuera aquellos ligados a otros ámbitos como, por ejemplo, la evaluación automática de la calidad de los diseños de programas informáticos.

3.2. Procedimiento de la selección de estudios

La selección de los estudios considerados en esta revisión se conformó por 4 pasos, descritos en la Figura 1. De esta manera, se recolectaron 3717 trabajos en la búsqueda inicial (3002 en *Scopus*, 342 en *Dialnet*, 247 en *Scielo* y 63 en *Google Scholar*). Posteriormente se eliminaron los duplicados ($n = 843$), quedando para el cribado 2874 estudios. En esta etapa, se procedió a evaluar los trabajos según los criterios de inclusión y exclusión declarados en la Tabla 1. Así, considerando estos aspectos, primero se revisaron los títulos (excluidos $n = 1341$) y luego los resúmenes (excluidos $n = 1024$). Posteriormente, se realizó una lectura de todo el texto, eliminando ejemplares que no respondieron a los criterios de inclusión (excluidos $n = 341$). El cribado dio como resultado 168 estudios seleccionados para la revisión completa.

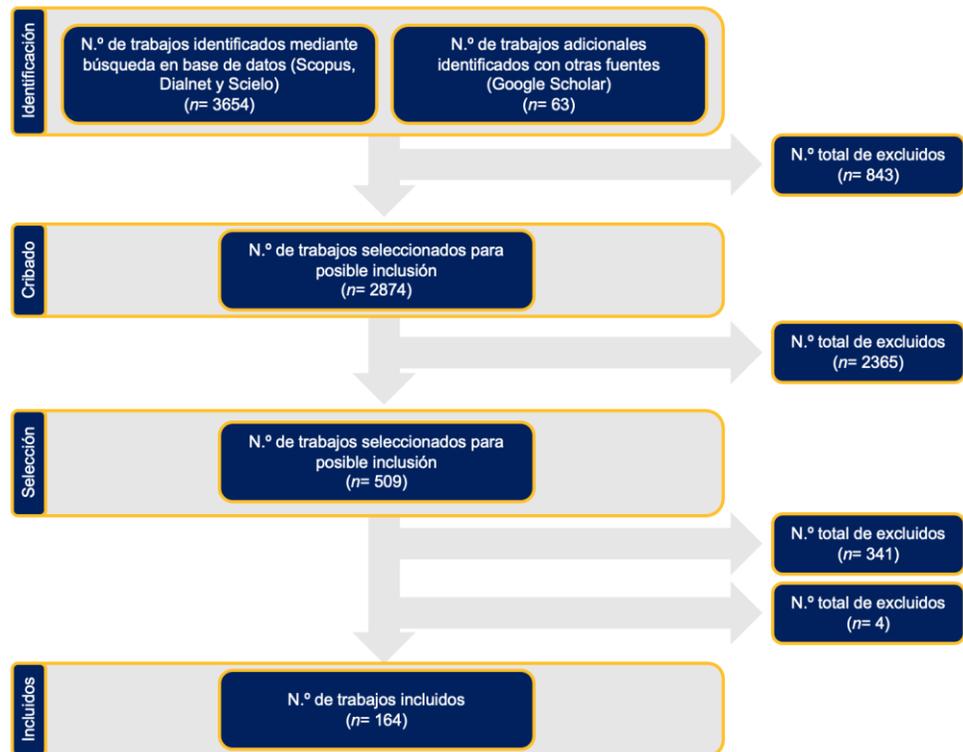


Figura 1: Procedimiento de selección de estudios basado en lineamientos PRISMA.

Fuente: Elaboración propia.

La lectura del texto completo se repitió, pero esta segunda vez atendiendo al país, idioma, los objetivos y preguntas de investigación que se proponían, las herramientas de análisis y recolección de datos y

los resultados obtenidos en el estudio. A su vez, se tuvo en consideración si los artículos permitían responder a las preguntas que dan origen a esta revisión sistemática. Finalmente, se seleccionaron 164 estudios para llevar a cabo el análisis final, eliminando 4 trabajos.

3.3. Analisis de estudios

El análisis de los 164 estudios se realizó completando una tabla en formato Excel. En ella se recogía el año de la publicación, el país en el que se realizaba el estudio, el idioma, el género, la herramienta o sistema que se utilizaba, la pregunta/objetivo de investigación, la muestra/participantes, los resultados obtenidos, los constructos teóricos de la propuesta y el rol de los docentes y participantes, entre otros. Una vez realizado el procedimiento, se tabularon los datos y se procedió con la representación gráfica de ellos.

4. Resultados

En este apartado, se expondrán los resultados de nuestro trabajo, organizados de la siguiente manera, primero, se expondrán los datos descriptivos respecto a los métodos y países en los que se han desarrollado las investigaciones y propuestas. Luego, se darán a conocer los objetivos que se persiguen en los trabajos revisados. Posteriormente, se dará cuenta de la tecnología que se emplea para construir las propuestas, así como sus constructos teóricos. Finalmente, los resultados que tienen estas herramientas desde la perspectiva de sus creadores y usuarios.

4.1. Datos descriptivos de investigaciones y propuestas de evaluación formativa automatizada y semiautomatizada

Como se observa en el Figura 2, los estudios se comienzan a producir desde 1966, cuando Page (1966) refiere a la calificación automática de ensayos. Así, en 1973 se desarrolla el primer motor de puntuación de ensayos llamado *Project Essay Grade* (Ajay et al., 1973). Posteriormente, ya en 1990, comienzan a desarrollarse una serie de trabajos que retoman los primeros aportes de Page (1966) y proponen otros motores, como, por ejemplo, *Vantage Learning* (Elliot, 1999).



Figura 2: Número de publicaciones por año.

Fuente: Elaboración propia.

Si bien existieron algunas propuestas en los años 90, el apogeo de esta tecnología se produce desde el año 2010 hasta la fecha. A su vez, como era de esperar, en el año 2021 se produce un crecimiento de los artículos publicados, lo que puede deberse a la pandemia Covid-19 y el confinamiento, pues investigaciones anteriores han establecido una relación entre la producción científica y el confinamiento, precisando que durante este periodo aumentó significativamente la producción (Perdomo, 2021). A su vez, la investigación, en dicho periodo, se centró más en los recursos digitales y tecnologías empleadas durante el proceso de enseñanza-aprendizaje, favoreciendo el aumento mencionado.

Respecto a los métodos utilizados en los trabajos revisados, se puede precisar que un 58% adopta uno cuantitativo (Saricaoglu, 2019), un 38% mixtos (Rapp y Kauf, 2018) y solo un 4% cualitativos (Davidson et al., 2014). Al respecto, se debe mencionar que todos aquellos trabajos en los que se evalúa el desempeño de una herramienta, interfaz o prototipo se emplean métodos cuantitativos o mixtos. Mientras que en los 6 estudios donde se utilizan métodos cualitativos se evalúa la usabilidad de la herramienta o la impresión que los usuarios tienen en torno a ella. Un aspecto importante que se deriva de estos datos es la prevalencia que existe de los métodos cuantitativos, pues tal como plantean López et al. (2015), la comunidad científica tiende a producir y a valorar más los trabajos que emplean estos métodos, por sobre otros.

En lo que refiere al continente en el que se han producido estas investigaciones, la mayoría de ellas fueron desarrolladas en Norte América y Asia. Así, en Estados Unidos se generaron más de la mitad de los estudios, centrados en el idioma inglés y en el género ensayo. A pesar de lo anterior, también se han realizado indagaciones en otros géneros, aunque en un número mucho menor. En cuanto al continente asiático, los países con mayor número de publicaciones son China, Taiwán y Japón, quienes alcanzan un 21,36%. Por otro lado, en Europa las publicaciones son menores en número, pero son diversos los países que las producen, por ejemplo, España, Alemania, Reino Unido y Países Bajos.

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

Respecto a los continentes australiano, sudamericano y africano, estos son los que menores trabajos poseen.

Por otro lado, el 34,14% de los trabajos revisados emplea o propone una herramienta, prototipo o modelo nuevo, diseñado por los autores, (Acosta y Nazar, 2020), el 23,11% utiliza una herramienta propia que ya había sido presentada en algún trabajo anterior (McNamara et al., 2015), el 30,3% utiliza un motor de puntuación, pero sin interfaz gráfica (Powers et al., 2002) y el 12,45% ocupa una herramienta de otro grupo de investigación o de pago (Huang y Wilson, 2021). Así, en la revisión sistemática desarrollada se han documentado 59 prototipos o herramientas desarrolladas en los últimos años (Figura 3).

Academic Vocabulary	Essay Map	SWAN
AcaWriter	Gingko	Scribo
Article Writing Tool	Grammar	Search Tool
Artext	Grammarly	Summary Street
AWSuM	Inputlog	STILUS
C-SAW	JESS	Style Writer
Calliope	Klinkende Taal	Thesis Writer
CAleSE	Larka	Turnitin (Revision Assistant)
Carnegie Mellon	LinguaKit	White Smoke
Prose style tool	Marking Mate	White&Improve
Coh Vis	My Access!	WriteCheck
Corpuscript	MY Write	WriteFull
Correct English	NC Write	WriteLab
Criterion	Open Essayist	WriteWise
De-jargonizer	Paper rater	Writer's Workbench
Deutsch-uni online	PEG Writing	WriteToLearn
DicSci	Pigiai	Writing Aid English
Editor	Rationale	Writing Pal
Escribo	Research Writing	Writing Roadmap
Estilector	Tutor	
Essay Jack	Right Writer	

Figura 3: Herramientas y prototipos de evaluación automatizada y semiautomatizada.

Fuente: Elaboración propia.

En cuanto a los destinatarios de las herramientas o intervenciones, se puede mencionar que el 59% son estudiantes de grado (McNamara et al., 2015), un 15% estudiantes de secundaria (Huang y Wilson, 2021; Ware, 2014), un 4% de postgrado (Ai, 2017) y un 22% sin indicar (Belcadhi, 2016). Respecto a estos últimos, muchos de los estudios refieren a público general, por lo que no se menciona un usuario ideal o no se refiere a ellos.

Por último, respecto a los géneros discursivos empleados, la mayoría de ellos corresponden a ensayos persuasivos y descriptivos con un 62,5% (Huang y Wilson, 2021; Wilson y Roscoe, 2020), un 6% al

género resumen (Franzke et al., 2005), un 5% al artículo de investigación científica (Caro et al., 2020), un 2,5% al Trabajo final de Grado (Rapp y Kauf, 2018; Da Cunha, 2020) y un 24% no indica el género o bien es de escritura en general (Belcadhi, 2016).

4.2. Objetivos de las investigaciones y propuestas de evaluación formativa automatizada y semiautmatizada

Los objetivos de las investigaciones y prototipos pioneros en evaluación automática se centraron en plantear herramientas que revisaran automáticamente ensayos. Así, Page (1966) se proponía como objetivo superar los altos costos humanos (tiempo y esfuerzo) que se invierten en revisar y puntuar textos a gran escala. Para ello, junto con un grupo de investigación desarrollaron el *Project Essay Grade* (PEG, por sus siglas en inglés), prototipo que evaluaba la calidad de los ensayos escritos por estudiantes en Estados Unidos.

El trabajo iniciado por Page (1966) fue continuado por Page (1968) y Ajay et al. (1973), quienes desarrollaron investigaciones en torno a PEG. En estos años, los objetivos expresados en las propuestas se mantuvieron, a pesar de que, posteriormente, se declaró que estas herramientas tenían un doble objetivo. Por un lado, el declarado por Page (1966) y, por otro, lograr puntuaciones imparciales y libres de sesgo humano (Stevenson y Phakiti, 2014; Wang et al., 2020).

En 1990, los objetivos cambiaron, pues no solo se buscaba bajar los costos humanos, sino también mejorar la fiabilidad de la calificación automática. Para ello, autores como Braun et al. (1990) centraron sus estudios en la precisión de estos sistemas, incluyendo nuevas técnicas y teniendo siempre en consideración la labor realizada por humanos. En 1993 los objetivos nuevamente se habían robustecido y ya no solo se ocupaban de igualar las calificaciones, sino también de estudiar la preferencia de los estudiantes, comparando la evaluación automática y la docente. Así, Dwyer y Sullivan (1993) indagaron en las preferencias de los estudiantes frente a los dos tipos de evaluación y sus razones.

Con el paso del tiempo y el avance de la tecnología, los ordenadores mejoraron sus funciones y capacidades, lo que permitió modificar las herramientas diseñadas previamente (Page, 2003). A su vez, los nuevos conocimientos en torno a la Lingüística y la Educación dieron prioridades a aspectos más discursivos de los textos y evaluaciones más centradas en la retroalimentación. Estos cambios de paradigma, junto a los avances tecnológicos, influyeron en los objetivos de la evaluación automática, por lo que, a partir de los 2000, se centraron en apoyar la labor humana, permitiendo que docentes tuviesen mayor tiempo para retroalimentar el contenido del texto, mientras que la máquina corregía los aspectos formales y superficiales de él (Palermo y Wilson, 2020).

Así, en los últimos 20 años, se han mejorado las herramientas existentes (Phandi et al., 2015; Reilly et al., 2014), como es el caso de PEG, otras se han centrado en entregar retroalimentación y evaluación formativa del escrito (Alikaniotis et al., 2016; Wilson y Czik, 2016; Cheng, 2017; Chu et al., 2018) y otras han emergido como tutores virtuales (Roscoe et al., 2014; Choi, 2016; Butterfuss et al., 2022). Las que han desarrollado una retroalimentación automatizada pueden tener dos objetivos: 1) apoyar el componente de identificación, diagnóstico y solución de la revisión y 2) complementar la enseñanza formal de la escritura académica en un género particular mediante material didáctico del género, corpus de consulta y retroalimentación automatizada (Cotos et al., 2015).

Por otro lado, los sistemas que desarrollan tutorías inteligentes tienen como objetivo apoyar el proceso de escritura de los estudiantes, permitiendo mejorar el dominio de la escritura (Roscoe et al., 2014). Por esto mismo, estas herramientas no pretenden sustituir la labor de los docentes, quienes son expertos en el área, sino ofrecer más oportunidades para desarrollar estrategias o prácticas de escritura en el aula (McNamara et al., 2015).

Como se ha observado en este apartado, los objetivos de la evaluación automática han ido variando con el paso del tiempo, el avance de la tecnología y los constructos teóricos (Figura 4).

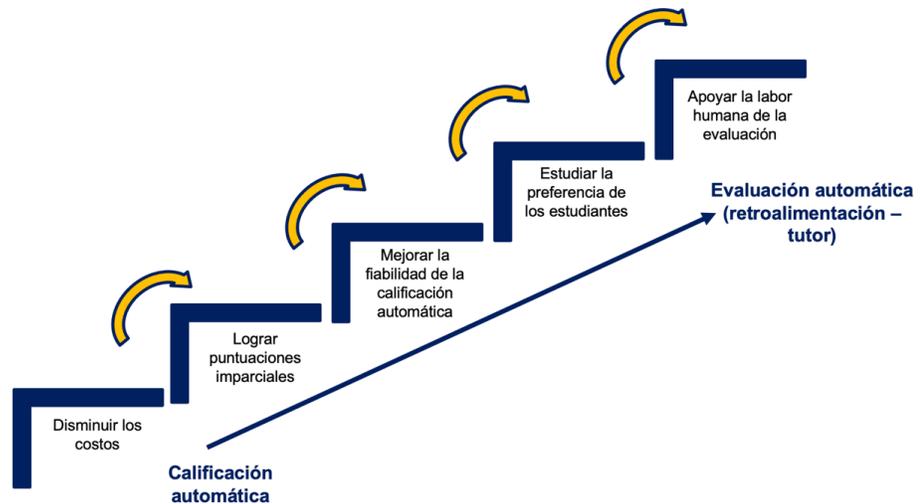


Figura 4: Evolución de los objetivos de la evaluación automática.

Fuente: Elaboración propia.

Tal como se observa, estos objetivos se han ido complementando y complejizando cada vez más con la evolución de la tecnología y los cambios de constructos. A su vez, los primeros objetivos se acercaban más a una calificación de los escritos (emulaban la puntuación de un humano), mientras que los últimos se acercan a una evaluación formativa, siempre con el fin de apoyar la labor del productor del texto y su revisor. Por esto, se debe reconsiderar la idea de que las herramientas han

sido creadas con el fin de reemplazar la evaluación humana, pues en realidad se plantean como un complemento a la retroalimentación brindada por el profesor y compañeros.

4.3. Tecnología empleada

Las propuestas de herramientas informatizadas presentadas en la Figura 3 han sido desarrolladas, en su gran mayoría, por ingenieros informáticos o desarrolladores de *software*, quienes, avalados por grandes empresas, especialmente editoriales y laboratorios, han impulsado y creado sistemas de evaluación automática. Debido a su poca relación con la evaluación y el estudio de la escritura, la mayoría de estos sistemas (43,3%) se han centrado en la eficacia y eficiencia del producto (Vajjala, 2018). Por su parte, un 23,2% de las herramientas se han desarrollado por lingüistas, psicólogos y sociólogos interesados en la evaluación automática de la escritura (Crossley et al., 2014).

Dentro de los desarrolladores de estas tecnologías también se encuentran expertos en educación (4,1%), quienes, desde el plano de la evaluación han desarrollado herramientas para la revisión de textos (Yang et al., 2022). Un 13,6% han sido diseñadas por expertos de diferentes áreas, no vinculadas a las ya mencionadas, quienes han contratado a informáticos para desarrollarlas y posteriormente comercializarlas (Bai y Hu, 2017). Finalmente, algunas propuestas (15,8%) se han formulado y desarrollado por grupos interdisciplinarios, lingüistas, psicólogos y educadores (Rapp y Kruse, 2020).

Respecto a los constructos teóricos que subyacen a estos sistemas, la mayoría (44,3%) emplea la teoría de la escritura para sustentar la propuesta. Así, algunos sistemas emplean el modelo sociocognitivo de escritura (Pittard y Martlew, 2000; Tolchinsky, 2000), otros, el enfoque por proceso (Flower y Hayes, 1981; Brookes y Grundy, 1998) y otros, la adquisición de segundas lenguas (Krashen y Terrel, 1983). También, existen trabajos (12,4%) que emplean teorías de evaluación en sus herramientas o bien la combinación de ambas (15,7%), es decir, teoría de la escritura y la evaluación (Rapp y Kauf, 2018). Por otra parte, algunos sistemas (9,8%) utilizan el análisis retórico-discursivo y la pedagogía basada en el género (Swales, 2004; Cotos et al., 2020). Por último, están los sistemas que no declaran el constructo teórico que los sustenta con un 17,5% (Saricaoglu, 2019).

Respecto a la tecnología detrás de las propuestas, una gran parte de ellas emplean motores de puntuación automatizados como *e-rater* e *Intellimetric*, entre otros (Powers et al., 2002). Estos motores, desarrollados a mediados de los sesenta y perfeccionados con los años, permiten emular las puntuaciones y algunas observaciones humanas a partir del PNL, la lingüística computacional y la inteligencia artificial. Si bien fueron diseñados inicialmente para el inglés, actualmente están disponibles para diferentes idiomas, como el chino, japonés, hebreo y malasio.

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

A su vez, varias de las herramientas emplean rasgos lingüístico-discursivos para predecir la calidad del texto, junto con otras herramientas. Así, seleccionan estos rasgos a partir de investigaciones empíricas (previas) en las que se ha atendido a un género discursivo, un nivel educacional y una comunidad discursiva. Como se ha mencionado previamente, la mayoría de las investigaciones han sido desarrolladas en el género ensayo, por lo que los rasgos más empleados corresponden a patrones que en este y otros géneros han permitido predecir la calidad de forma automatizada, a saber: longitud total del texto (mayor o menos), de palabras y de oraciones; densidad léxica y terminológica, ortografía puntual, acentual y literal; cohesión; complejidad sintáctica; organización del texto, contenido textual, entre otros (Crossley et al., 2014).

También, existen otros sistemas que utilizan herramientas del PNL, como, por ejemplo, lematizadores, analizadores sintácticos, analizadores retóricos y bases de datos, entre otros (McNamara y Graesser, 2012), para alimentar a algoritmos que evalúan la calidad de un texto en particular. A su vez, diversos prototipos e investigaciones desarrollados entre los años 2000 y 2010 emplearon el análisis semántico latente (ASL) con el fin de indagar y profundizar en nuevas formas de analizar y retroalimentar el contenido de los escritos en lengua inglesa.

En la actualidad, los desarrolladores se han inclinado por emplear Machine learning, Deep learning y redes neuronales para fortificar los sistemas ya existentes. Sumado a lo ya mencionado (PNL, algoritmos computacionales y otras herramientas), algunos autores han utilizado el análisis retórico-discursivo para evaluar la calidad de los segmentos funcionales de un texto (Cotos et al., 2020), demostrando que la evaluación automática progresa hacia la valoración de aspectos discursivos y de contenido textual.

A continuación, presentaremos investigaciones que emplean y prueban algunas técnicas y herramientas del PNL para evaluar la calidad de los escritos. Por temas de extensión, solo destacaremos las que utilizan otras tecnologías, se diseñan para un idioma diferente del inglés o proponen nuevas formas de realizar la evaluación.

Loraksa y Peachavanish (2007) emplearon ASL sumado a un motor de puntuación y una red neuronal para evaluar la calidad de los escritos en tailandés. Los autores realizaron dos experimentos en los que probaron diferentes combinaciones entre las herramientas empleadas, obteniendo resultados positivos. Venegas (2011) probó 3 métodos de evaluación con ASL para resúmenes en español. El autor demostró que existían correlaciones moderadas y altas entre las puntuaciones humanas y las alcanzadas con este método.

También resaltamos el trabajo de Da Cunha (2020), quien propone una herramienta que apoya y revisa la redacción de géneros discursivos ligados a la administración pública, el turismo, la medicina y el

ámbito académico (ArText). El sistema se asemeja a un editor de texto dividido en 3 módulos (estructura del género, forma y revisión ortográfica, y la adecuación del texto en cuanto a los rasgos y recomendaciones lingüísticas según el género en cuestión). Para llevar a cabo estas funciones, emplea herramientas del PNL como un analizador morfológico y segmentador discursivo.

4.4. Resultados obtenidos

A continuación, se expondrán los resultados obtenidos por las investigaciones al emplear una evaluación automatizada o semiautomatizada de la calidad de los textos escritos. En la Figura 5, presentamos una síntesis de las apreciaciones de los autores al respecto. Para ello, a partir del análisis realizado, se establecieron 7 categorías que condensan lo expresado en los 164 artículos revisados.

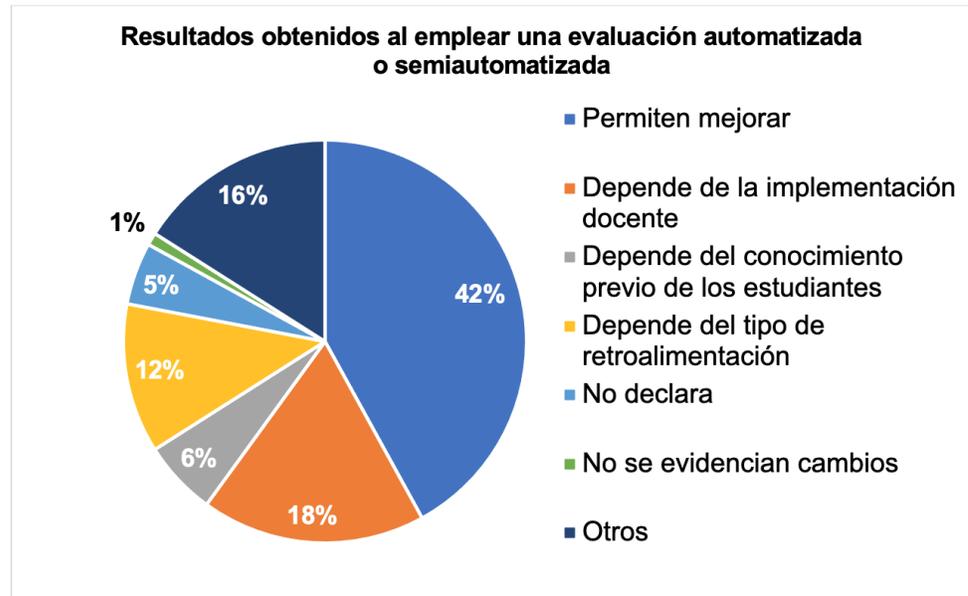


Figura 5: Resultados obtenidos al emplear la evaluación automatizada y semiautomatizada.

Fuente: Elaboración propia.

Tal como se evidencia en el Figura 5, la mayoría de los trabajos (42%) que han utilizado evaluación automatizada (motor de puntuación, sistema o tutor de escritura) han obtenido buenos resultados. Estos hallazgos positivos son atribuidos a diferentes factores, por un lado, se sostiene que empleando estos sistemas los estudiantes pueden concentrarse en revisar aspectos más profundos del texto, obviando otros más normativos (Palermo y Thomson, 2018), lo que mejoraría la calidad del escrito. Por otro lado, algunos trabajos exponen que la motivación de los estudiantes a emplear el sistema y la autogestión al utilizarlo impacta en los resultados obtenidos por la herramienta (Wilson y Czik, 2016).

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

En esta misma línea, algunos trabajos atribuyen el impacto positivo del instrumento a su diseño y, en especial, a las recomendaciones de sus diseñadores. Así, destacan que el consejo de emplear la herramienta al comienzo y al término del proceso de escritura resulta favorable para el escritor, pues permite avanzar y concluir con un sustento de que el trabajo se adecúa a los aspectos solicitados (Chen y Cheng, 2008). Otras investigaciones consideran que, la dosificación de la retroalimentación que poseen algunos sistemas resulta positiva para los usuarios, pues evita el agobio del productor del texto y les permite avanzar poco a poco, dependiendo del nivel que se desee revisar y mejorar (Cheng, 2017).

También existen artículos que exponen que el éxito o el fracaso del sistema dependerá de la implementación docente (18%). De esta manera, Liu et al. (2016) exponen que, si la retroalimentación automatizada se considera un apoyo para el docente, esta tendrá éxito, de lo contrario podría resultar un fracaso. Esto sobre todo porque la retroalimentación correctiva ha demostrado tener buenos resultados, pero requiere de un andamiaje y de la labor docente para extraer su máximo potencial.

Por otro lado, un 16% de los trabajos revisados mencionan que los resultados positivos o negativos de los sistemas automáticos dependen de múltiples factores. Entre ellos, se destaca la dificultad de la tarea de escritura (Wade-Stein y Kintsch, 2004). De esta manera, mientras más compleja resulte la tarea, más positiva será la opinión que posea el estudiante frente a la herramienta. A su vez, esta le permitirá mejorar su escrito, pues entregará recomendaciones de aspectos formales, normativos, estructurales e incluso de contenido (Cheng, 2017; Liu et al., 2016), lo que mejorará su calidad. En este mismo ítem, Dawson et al. (2018) declaran que el empleo de artefactos multimodales como parte del *feedback* automático es valorado positivamente por los estudiantes y que impacta positivamente en su desempeño, pues permite explicar de manera clara y sencilla.

Kellogg, Whiteford y Quinlan (2010) mencionan que el sistema de evaluación automatizada no impacta en la puntuación holística alcanzada. Sin embargo, al momento de revisar y comparar los escritos, los errores del producto final se reducen considerablemente. Por otro lado, algunas investigaciones mencionan el impacto positivo del sistema, pero no se refieren a la razón de la mejora.

También se observa en la Figura 5 que un 12% de los estudios atribuyen los resultados al tipo de retroalimentación que realizan y al tipo de error que corrigen (Ranalli et al., 2017). Así, relacionando el resultado de la herramienta con su precisión, exponen que los mejores resultados se producen en errores ortográficos y en aspectos formales y gramaticales. Sin embargo, la precisión y el desempeño baja cuando se emplean solo para retroalimentar el contenido. Por ello, la revisión debe ser integral e, idealmente, siempre con un docente como parte del proceso de retroalimentación (Bai y Hu, 2017; Ranalli et al., 2017).

Un 6% de los estudios declaran que los resultados obtenidos dependen de los conocimientos previos que posea el estudiante o productor. De esta manera, Franzke et al. (2005) demostraron que su herramienta tuvo buenos resultados para revisar el contenido, la organización y el estilo de la redacción. Sin embargo, estos resultados fueron beneficiosos solo para estudiantes con rendimiento medio o bajo, pues los avanzados no requerían las sugerencias que podía entregar el sistema, ya que no presentaban dificultades en esos niveles.

Finalmente, existe un 5% de los artículos en los que no se refieren a los resultados de sus sistemas. A su vez, un 1% expone que el uso del sistema no demostró ningún impacto positivo en el escrito. Así, Ware (2014) menciona que luego de la aplicación del sistema no se observan mejoras en los escritos, a diferencia de lo ocurrido con otros tipos de retroalimentación. Sin embargo, refiere a que esto se podría asociar a la impresión que tenía la docente del sistema, pues no solía emplearlo.

5. Comentarios de cierre

Como se puede evidenciar, la mayoría de los trabajos se han centrado en proponer herramientas automáticas y semiautomatizadas en el género ensayo y el idioma inglés. Si bien desde principio de los años 90 se refiere a la necesidad de realizar propuestas en otros idiomas y géneros, hasta la fecha más del 50% de los trabajos continúan efectuándose con los mismos criterios. A su vez, la mayoría de los desarrolladores de estos sistemas no son expertos de la lengua ni de la evaluación de aprendizajes, por lo que las consideraciones que emplean no son, necesariamente, las más idóneas para potenciar el aprendizaje de los estudiantes. A su vez, tampoco existe una preocupación por los rasgos lingüísticos que se empleen para determinar la calidad ni tampoco el tipo de evaluación que se utiliza, pues muchas veces el foco está en el desempeño de las herramientas más que en el constructo que las sustenta (Vajjala, 2018).

A pesar de este vacío, en la actualidad son cada vez más los educadores, lingüistas y psicólogos que se interesan por indagar en el campo computacional. Por ello, desde el 2006 en adelante, se han comenzado a desarrollar nuevas herramientas, esta vez por expertos en ámbitos ligados a la evaluación automática. Así, los nuevos sistemas han incluido aspectos que antes no se consideraban, como, por ejemplo, mayor énfasis en el tipo de retroalimentación que se entrega, los rasgos que se emplean para esto, el rol del docente, entre otros. Estos puntos son relevantes, pues por mucho tiempo se estuvo realizando un mal uso de los sistemas, planteándolos como sustitutos de la instrucción humana (Cotos, 2016). Sin embargo, con los trabajos interdisciplinarios actuales y la inclusión de expertos en el tema, la situación se ha modificado.

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

En líneas similares a lo planteado, la revisión ha dejado de manifiesto que existen pocas herramientas y sistemas que basen sus propuestas en investigaciones empíricas. Si bien destacan algunas realizadas por lingüistas y psicólogos ligados a la evaluación, aún falta que esto se extrapole a otros campos en los que se está incursionando en la evaluación automática.

Un punto a destacar es que aún se observan escasas propuestas desarrolladas para el español (ArText, Estilector, STILUS y LinguaKit). Este tema resulta preocupante, pues la mayoría de las herramientas y sistemas diseñados para el inglés no pueden aplicarse al español, ya que resultan específicas para un idioma en particular. No obstante, existen algunos avances emergentes para la evaluación y retroalimentación, tales como WriteWise (<https://web.writewise.io/writewise-spanish/?nab=2>) para el género artículo científico o PEUMO (www.redilegra/peumo) para informes de ingeniería.

Otro punto interesante que se ha constatado con la revisión realizada es que, efectivamente, el desconocimiento de los nuevos paradigmas, de los constructos teóricos que subyacen a las herramientas y de la competencia de sus desarrolladores ha llevado a tener un rechazo a la evaluación automática (Shermis, 2020). Así, uno de los grandes motivos de este rechazo es la supuesta suplantación humana, la que relegaría a un segundo plano el rol del docente (Palermo y Wilson, 2020). Sin embargo, los nuevos paradigmas y propuestas no sugieren esto, sino más bien todo lo contrario, como se ha constatado en los objetivos declarados por los desarrolladores de las herramientas y en las evaluaciones de usuarios.

A modo de conclusión, es posible establecer que el campo de la evaluación automatizada de textos seguirá expandiéndose y mejorando la calidad de sus aportes, dada la mayor capacidad computacional existente, la rápida aparición de nuevos y mejores modelos de lenguaje basados en aprendizaje profundo, la ampliación de cobertura de las redes digitales (red 5G), la amplia disponibilidad de datos *online* y el compromiso de trabajo interdisciplinario. Así, lejos de poner en riesgo el rol docente, estos avances permiten complementar su labor, permitiendo una mayor focalización en el contenido y en las estrategias de aprendizaje. A los estudiantes les permitirá mejorar sus textos, acorde con su nivel lingüístico y discursivo, enfatizando la función epistémica de la escritura.

6. Referencias

- Acosta, N. y Nazar, R. (21-23 de octubre de 2020). *Estilector. com: herramienta de ayuda a la redacción en castellano* [ponencia]. Congreso Internacional de Lingüística Computacional y de Corpus - CILCC 2020 y V Workshop en Procesamiento Automatizado de Textos y Corpus - WoPATeC 2020, Medellín, Colombia.
- Ai, H. (2017). Providing graduated corrective feedback in an intelligent computer-assisted language learning environment. *ReCALL*, 29(3), 313–334.
- Andueza, A. (2019). Evaluación de la escritura académica: construcción y validación de un instrumento integrado basado en tareas para evaluar habilidades específicas de escritura. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, 25(2), 1–20.
- Ajay, H., Tillet, P. y Page, E. (1973). *Analysis of essays by computer (AEC-II)*. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education, National Center for Educational Research and Development.
- Alikaniotis, D., Yannakoudakis, H. y Rei, M. (7-12 de agosto de 2016). *Automatic text scoring using neural networks* [conferencia]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlín, Alemania.
- Bai, L. y Hu, G. (2017). In the face of fallible AWE feedback: How do students respond? *Educational Psychology*, 37(1), 67–81. <https://doi.org/10.1080/01443410.2016.1223275>
- Belcadhi, L. C. (2016). Personalized feedback for self assessment in lifelong learning environments based on semantic web. *Computers in Human Behavior*, 55, 562–570.
- Butterfuss, R., Roscoe, R., Allen, L., McCarthy, K. y McNamara, D. (2022). Strategy Uptake in Writing Pal: Adaptive Feedback and Instruction. *Journal of Educational Computing Research*, 60(3), 696-721.
- Braun, H. I., Bennett, R. E., Frye, D. y Soloway, E. (1990). Scoring constructed responses using expert systems. *Journal of Educational Measurement*, 27, 93–108.
- Brookes, A. y Grundy, P. (1998). *Beginnig to write: writing activities for elementary and intermediate learners*. Cambridge University Press.
- Burstein, J. (2003). The e-rater® scoring engine: Automated essay scoring with natural language processing. En M. Shermis y J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective*. (pp. 113-123). Lawrence Erlbaum Associates.
- Caro, E. F., Cid, H. A., Rodríguez, S., Venegas, R., Pavez, J., Palma, W. y VanCott, A. (2020). *WriteWise: software that guides scientific writing* [ponencia]. Congreso Internacional de Lingüística Computacional y de Corpus - CILCC 2020 y V Workshop en Procesamiento Automatizado de Textos y Corpus - WoPATeC 2020, Medellín, Colombia.
- Chen, C. y Cheng, W. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2), 94–112.

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- Cheng, G. (2017). The impact of online automated feedback on students' reflective journal writing in an EFL course. *The Internet and Higher Education*, 34, 18–27.
- Choi, I. (2016). Efficacy of an icall tutoring system and process-oriented corrective feedback. *Computer Assisted Language Learning*, 29(2), 334–364.
- Chu, T., Taelle, P. y Hammond, T. (8-11 de mayo de 2018). *Supporting Chinese character educational interfaces with richer assessment feedback through sketch recognition* [conferencia]. Graphics Interface (GI 2018), Toronto, Canadá.
- Cowie, B. y Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles, Policy & Practice*, 6(1), 101-116.
- Contreras, L., González, M. y Urías, E. (2009). Evaluación de la escritura mediante rúbrica en la educación primaria en México. *Revista Interamericana de Psicología/Interamerican Journal of Psychology*, 43(3), 518-531.
- Cotos, E., Huffman, S. y Link, S. (2015). Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for EAP pedagogy. *Journal of English for Academic Purposes*, 19, 52-72.
- Cotos, E. (2016). Computer-assisted research writing in the disciplines. En S. A. Crossley, y D. S. McNamara (Eds.), *Adaptive educational technologies for literacy instruction* (pp. 225-242). Routledge.
- Cotos, E., Huffman, S. y Link, S. (2020). Understanding graduate writers' interaction with and impact of the Research Writing Tutor during revision. *Journal of Writing Research*, 12(1), 187-232.
- Crossley, S., Allen, L., Kyle, K. y McNamara, D. (2014). Analyzing discourse processing using a simple natural language processing tool. *Discourse Processes*, 51(5-6), 511-534.
- Da Cunha, I. (2020). Una herramienta TIC para la redacción del Trabajo de Fin de Grado (TFG). *ELUA*, 34, 39-72.
- Davidson, L. Y. J., Richardson, M. y Jones, D. (2014). Teachers' perspective on using technology as an instructional tool. *Research in Higher Education*, 24, 1–25. <https://eric.ed.gov/?id=EJ1064110>
- Dawson, P., Henderson, M., Ryan, T., Mahoney, P., Boud, D., Phillips, M. y Molloy, E. (2018). Technology and feedback design. En *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*. M. J. Spector (eds) 1–45. Springer International Publishing.
- Díaz López, M. (2018). Impacto de la retroalimentación y la evaluación formativa en la enseñanza-aprendizaje de Biociencias. *Educación Médica Superior*, 32(3), 147-156.
- Dikli, S. (2006). An overview of automated scoring of essays. *The Journal Technology, Learning and Assessment*, 5(1),1-35.
- Dwyer, H. y Sullivan, H. (1993). Student preferences for teacher and computer composition marking. *The Journal of Educational Research*, 86(3), 137–141.

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- Eid, S. y Wanas, N. (2017). Automated essay scoring linguistic feature: Comparative study. *En Intl Conf on Advanced Control Circuits Systems (ACCS) Systems & 2017 Intl Conf on New Paradigms in Electronics & Information Technology (PEIT)*, Alexandria, 212-217.
- Elliot, S. (1999). *Construct validity of IntelliMetric with international assessment*. Yardley, PA: Vantage Technologies.
- Feng, H., Saricaoglu, A. y Chukharev-Hudilainen, E. (2016). Automated error detection for developing grammar proficiency of ESL learners. *Calico Journal*, 33(1), 49-70.
- Fernández, S. (2017). Evaluación y aprendizaje. *Marco ELE: Revista de Didáctica Español Lengua Extranjera*, 24, 1-43.
- Flower, L. y Hayes, J. (1981). A cognitive process theory of writing. *College composition and communication*, 32(4), 365-387.
- Franzke, M., Kintsch, E., Caccamise, D., Johnson, N. y Dooley, S. (2005). Summary Street®: Soporte informático para la comprensión y escritura. *Revista de Investigación en Informática Educativa*, 33(1), 53-80.
- Granic, A. y Marangunic, N. (2019). Technology acceptance model in educational context: A systematic literature review. *British Journal of Educational Technology*, 50(5), 2572–2593.
- Harvey, L. (2002). Evaluation for what?. *Teaching in higher education*, 7(3), 245-263.
- Huang, Y. y Wilson, J. (2021). Using automated feedback to develop writing proficiency. *Computers and Composition*, 62, 1-27. <https://doi.org/10.1016/j.compcom.2021.102675>
- Krashen, S. y Terrel, T. (1983). *The Natural Approach: Language Acquisition in the Classroom*. Pergamon.
- Kellogg, R., Whiteford, A. y Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research*, 42(2), 173–196.
- Liu, M., Li, Y., Xu, W. y Liu, L. (2016). Automated essay feedback generation and its impact on revision. *IEEE Transactions on Learning Technologies*, 10(4), 502–513.
- Lopez, X., Valenzuela, J., Nussbaum, M. y Tsai, C. (2015). Some recommendations for the reporting of quantitative studies. *Computers & Education*, 91, 106-110.
- Loraksa, C. y Peachavanish, R. (2007). Automatic Thai-language essay scoring using neural network and latent semantic analysis. *En First Asia International Conference on Modelling & Simulation (AMS'07)* (pp. 400-402). IEEE.
- McNamara, D. S. y Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. McCarthy C. Boonthum-Denecke (Eds.), *Applied natural language processing and contentanalysis: Identification, investigation, and resolution* (pp. 188–205). IGI Global.
- McNamara, D., Crossley, S., Roscoe, R., Allen, L. y Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing*, 23, 35–59.
- Morales, F. (2004). Evaluar la escritura, sí... Pero ¿Qué y cómo evaluar?. *Acción pedagógica*, 13(1), 38-48.

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- Muñoz, C. y Pérez, B (2021). Elaboración de una prueba diagnóstico para medir habilidades de escritura académica. *Educatio Siglo XXI*, 39(1), 131-146.
- Navarro, F., Ávila-Reyes, N. y Gómez Vera, G. (2019). Validez y justicia: hacia una evaluación significativa en pruebas estandarizadas de escritura. Meta: *Avaliacao*, 11(31), 1-35.
- Palermo, C. y Thomson, M. M. (2018). Teacher implementation of self-regulated strategy development with an automated writing evaluation system: Effects on the argumentative writing performance of middle school students. *Contemporary Educational Psychology*, 54, 255–270.
- Palermo, C. y Wilson, J. (2020). Implementing automated writing evaluation in different instructional contexts: A mixed-methods study. *Journal of Writing Research*, 12(1), 63–108.
- Page, E. (1966). The imminence of...grading essays by computer. *Phi Delta Kappan*, 47(5), 238–243.
- Page, E. (1968). The use of the computer in analyzing student essays. *International Review of Education*, 14, 210–225.
- Page, E. (2003). Project essay grade: PEG. En M. Shermis y J. Burstein (eds), *Automated essay scoring: A cross-disciplinary perspective*, 43– 54. Lawrence Erlbaum Associates.
- Parra G, y Calero S. (2019). Automated writing evaluation tools in the improvement of the writing skill. *International Journal of Instruction*, 12(2), 209-226.
- Phandi, P., Chai, K. y Ng, H. (19- 21 de septiembre de 2015). *Flexible domain adaptation for automated essay scoring using correlated linear regression* [conferencia]. Conference on Empirical Methods in Natural Language Processing, Lisboa, Portugal.
- Pittard, V. y Martlew, M. (2000). Socially-situated cognition and metalinguistic activity. En A. Camps y M. Millán (Eds.). *Metalinguistic activity in learning to write* (pp. 79-102). Amsterdam University Press.
- Perdomo, B. (2021). Publicaciones científicas de países latinoamericanos sobre educación ante el COVID-19: Revisión sistemática de la literatura. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, 28, 344-351.
- Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E. y Kukich, K. (2002). Stumping e-rater: Challenging the validity of automated essay scoring. *Computers in Human Behavior*, 18, 103–134. [https://doi.org/10.1016/S0747-5632\(01\)00052-8](https://doi.org/10.1016/S0747-5632(01)00052-8)
- Ranalli, J., Link, S. y Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educ. Psychol.* 37(1), 8–25.
- Rapp, C. y Kauf, P. (2018). Scaling Academic Writing Instruction: Evaluation of a Scaffolding Tool (Thesis Writer). *International Journal of Artificial Intelligence in Education*, 28, 590-615.
- Rapp, C. y Kruse, O. (2020). Thesis writer 2.0: a system supporting academic writing, its instruction and supervision. En *GMW Tagung 2020, Winterthur, 24.-26. August 2020* (pp. 235-240).

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- Reilly, E., Stafford, R., Williams, K. y Corliss, S. (2014). Evaluating the validity and applicability of automated essay scoring in two massive open online courses. *International Review of Research in Open and Distributed Learning*, 15(5), 83-98.
- Roscoe, R. D., Allen, L. K., Weston, J. L., Crossley, S. A. y McNamara, D. S. (2014). The Writing Pal intelligent tutoring system: Usability testing and development. *Computers and Composition*, 34, 39-59.
- Ruz, I. (2018). Evaluación para el aprendizaje. *Revista Educación las Américas*, 6, 13-28.
- Saricaoglu, A. (2019). The impact of automated feedback on L2 learners' written causal explanations. *ReCALL*, 31(2), 189-203. <https://doi.org/10.1017/S095834401800006X>
- Shermis, M. y Burstein, J. (2003). *Automated essay scoring: A cross-disciplinary perspective*. Lawrence Erlbaum Associates.
- Shermis, M., Burstein, J. y Bursky, S. (2013). Introduction to automated essay evaluation. En *Handbook of automated essay evaluation* (pp. 23-37). Routledge.
- Shermis, M. (2020). International Applications of Automated Scoring. En D. Yan., Rupp, A. y Foltz, P. (eds). *Handbook of Automated Scoring* (pp. 113-132). Chapman and Hall/CRC.
- Stevenson, M. y Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19, 51-65.
- Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A. y Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers & education*, 131, 33-48.
- Swales, J. (2004). *Research genres: Explorations and applications*. Cambridge University Press.
- Talanquer, V. (2015). La importancia de la evaluación formativa. *Educación química*, 26(3), 177-179.
- Tolchinsky L. (2000). Distintas perspectivas acerca del objeto y propósito del trabajo y la reflexión metalingüística en la escritura académica. En M. Miliam y A. Camps (Eds.). *El papel de la actividad metalingüística en el aprendizaje de la escritura*. Homo Sapiens.
- Urrútia, G. y Bonfill, X. (2010). Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Medicina clínica*, 135(11), 507-511.
- Vajjala, S. (2018). Automated Assessment of Non-Native Learner Essays: Investigating the Role of Linguistic Features. *International Journal of Artificial Intelligence in Education* 28(1), 79-105.
- Venegas, R. (2011). Evaluación de resúmenes en español con Análisis Semántico Latente: Una implementación posible. *Revista signos*, 44(75), 85-102.
- Wade-Stein, D. y Kintsch, E. (2004). Summary street: Interactive computer support for writing. *Cognition and Instruction*, 22(3), 333-362.
- Wang, E., Matsumura, L., Correnti, R., Litman, D., Zhang, H., Howe, E., Magooda, A. y Quintana, R. (2020). eRevis(ing): Students' revision of text evidence use in an automated writing evaluation system. *Assessing Writing*, 44, 100449.
- Ware, P. (2014). Feedback for adolescent writers in the english classroom. *Writing & Pedagogy*, 6(2), 223-249. <https://doi.org/10.1558/wap.v6i2.223>

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- Wilson, J. y Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers and Education, 100*, 94–109.
- Wilson, J. y Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research, 58*(1), 87–125. <https://doi.org/10.1177/0735633119830764>
- Yang, L. P., Xin, T., Luo, F., Zhang, S. y Tian, X. T. (2022). Automated evaluation of the quality of ideas in compositions based on concept maps. *Natural Language Engineering, 28*(4), 449-486.

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

Anexo 1: Tabla con fuentes empleadas en el desarrollo de la revisión sistemática

Fuente utilizada en la revisión sistemática	
1	Ai, H. (2017). Providing graduated corrective feedback in an intelligent computer-assisted language learning environment. <i>ReCALL</i> , 29(3), 313–334. https://doi.org/10.1017/S095834401700012X
2	Ajay, H., Tillet, P. y Page, E. (1973). Analysis of essays by computer (AEC-II). Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education, National Center for Educational Research and Development.
3	Akçapınar, G. (2015). How automated feedback through text mining changes plagiaristic behavior in online assignments. <i>Computers & Education</i> , 87, 123–130. https://doi.org/10.1016/j.compedu.2015.04.007
4	Aluthman, E. S. (2016). The effect of using automated essay evaluation on ESL undergraduate students' writing skill. <i>International Journal of English Linguistics</i> , 6(5), 54–67. https://doi.org/10.5539/ijel.v6n5p54 .
5	Alikianiotis, D., Yannakoudakis, H. y Rei, M. (7-12 de agosto de 2016). <i>Automatic text scoring using neural networks</i> [conferencia]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlín, Alemania.
6	Attali, Y., Bridgeman, B. y Trapani, C. (2010). Performance of a generic approach in automated essay scoring. <i>The Journal of Technology, Learning, and Assessment</i> , 10(3), 1–17.
7	Attali, Y. y Burstein, J. (2005). Automated essay scoring with e-rater v.2.0 (Research Report No. RR-04-45). Princeton, NJ: Educational Testing Service. doi:10.1002/j.2333-8504.2004.tb01972.x
8	Acosta, N. y Nazar, R. (21-23 de octubre de 2020). <i>Estilector. com: herramienta de ayuda a la redacción en castellano</i> [ponencia]. Congreso Internacional de Lingüística Computacional y de Corpus - CILCC 2020 y V Workshop en Procesamiento Automatizado de Textos y Corpus - WoPATeC 2020, Medellín, Colombia.
9	Ashrafganjoe, M., Rezai, M. y Elhambakhsh, S (2022). Proporcionar retroalimentación basada en computadora a través de Grammarly® en clases de escritura. <i>Revista de Lengua y Traducción</i> , 12 (2), 163-176.
10	Bai, L. y Hu, G. (2017). In the face of fallible AWE feedback: How do students respond? <i>Educational Psychology</i> , 37(1), 67–81. https://doi.org/10.1080/01443410.2016.1223275
11	Bárcena, E. y Read, T. (2007). El diseño y arquitectura de herramientas para la redacción asistida de textos en inglés. <i>Porta Linguarum</i> , 8, 225-244.
12	Barrot, J. (2021). Using automated written corrective feedback in the writing class- rooms: Effects on L2 writing accuracy. <i>Computer Assisted Language Learning</i> . https://doi.org/10.1080/09588221.2021.1936071
13	Belcadhi, L. (2016). Personalized feedback for self assessment in lifelong learning environments based on semantic web. <i>Computers in Human Behavior</i> , 55, 562–570.
14	Braun, H. I., Bennett, R. E., Frye, D. y Soloway, E. (1990). Scoring constructed responses using expert systems. <i>Journal of Educational Measurement</i> , 27, 93–108. doi:10.1111/j.1745-3984.1990.tb00736.x
15	Bridgeman, B., Trapani, C. y Attali, Y. (2012). Comparison of human and machine scoring of essays: Differences by gender, ethnicity, and country. <i>Applied Measurement in Education</i> , 25, 27–40. doi:10.1080/08957347.2012.635502
16	Bond, M. y Pennebaker, J. W. (2012). Automated computer-based feedback in expressive writing. <i>Computers in Human Behavior</i> , 28(3), 1014–1018. https://doi.org/10.1016/j.chb.2012.01.003
17	Burstein, J., Chodorow, M. y Leacock, C. (2004). Automated essay evaluation: The Criterion online writing service. <i>Ai magazine</i> , 25(3), 27-27. https://doi.org/10.1609/aimag.v25i3.1774
18	Burstein, J., Elliot, N. y Molloy, H. (2016). Informing automated writing evaluation using the lens of genre: Two studies. <i>CALICO Journal</i> , 3, 117–141. doi:10.1558/ cj.v33i1.26374
19	Butterfuss, R., Roscoe, R. D., Allen, L. K., McCarthy, K. S. y McNamara, D. S. (2022). Strategy Uptake in Writing Pal: Adaptive Feedback and Instruction. <i>Journal of Educational Computing Research</i> , 60(3), 696-721.
20	Buzick, H., Oliveri, M. E., Attali, Y. y Flor, M. (2016). Comparing human and automated essay scoring for prospective graduate students with learning disabilities and/or ADHD. <i>Applied Measurement in Education</i> , 29, 161–172. doi:10.1080/08957347.2016.1171765

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- 21 Chapelle, C. A., Cotos, E. y Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, 32(3), 385–405. <https://doi.org/10.1177/0265532214565386>
- 22 Calvo, R. y Ellis, R. (2010). Students' conceptions of tutor and automated feedback in professional writing. *Journal of Engineering Education*, 99(4), 427–438. <https://doi.org/10.1002/j.2168-9830.2010.tb01072.x>
- 23 Caro, E. F., Cid, H. A., Rodríguez, S., Venegas, R., Pavez, J., Palma, W. y VanCott, A. (2020). *WriteWise: software that guides scientific writing* [ponencia]. Congreso Internacional de Lingüística Computacional y de Corpus - CILCC 2020 y V Workshop en Procesamiento Automatizado de Textos y Corpus - WoPATeC 2020, Medellín, Colombia.
- 24 Chen, C. y Cheng, W. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2), 94–112.
- 25 Chen, M. H., Huang, S. T., Chang, J. S. y Liou, H. C. (2015). Developing a corpus-based paraphrase tool to improve EFL learners' writing skills. *Computer Assisted Language Learning*, 28(1), 22–40. <https://doi.org/10.1080/09588221.2013.783873>
- 26 Cheng, G. (2017). The impact of online automated feedback on students' reflective journal writing in an EFL course. *The Internet and Higher Education*, 34, 18–27. <https://doi.org/10.1016/j.iheduc.2017.04.002>
- 27 Chodorow, M., Gamon, M. y Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing*, 27(3), 419–436. <https://doi.org/10.1177/0265532210364391>
- 28 Choi, I.-C. (2016). Efficacy of an icall tutoring system and process-oriented corrective feedback. *Computer Assisted Language Learning*, 29(2), 334–364.
- 29 Chu, T., Taelle, P. y Hammond, T. (8-11 de mayo de 2018). *Supporting Chinese character educational interfaces with richer assessment feedback through sketch recognition* [conferencia]. Graphics Interface (GI 2018), Toronto, Canadá.
- 30 Chung, Z., Nagai, H. y Rodrigo, M. M. T. (9-13 de junio de 2013). *An intelligent tutoring system for Japanese language particles with user assessment and feedback* [conferencia]. International Conference on Artificial Intelligence in Education, Memphis, USA.
- 31 Cotos, E. (2011). Potential of automated writing evaluation feedback. *Calico Journal*, 28(2), 420–459. <https://doi.org/10.11139/cj.28.2.420-459>
- 32 Cotos, E., Link, S. y Huffman, S. (2017). Effects of DDL technology on genre learning. *Language Learning & Technology*, 21(3), 104–130.
- 33 Cotos, E., Huffman, S. y Link, S. (2020). Understanding graduate writers' interaction with and impact of the Research Writing Tutor during revision. *Journal of Writing Research*, 12(1), 187-232.
- 34 Da Cunha, I. (2020). Una herramienta TIC para la redacción del Trabajo de Fin de Grado (TFG). *ELUA*, 34, págs. 39-72.
- 35 Da Cunha, I., Montané, M. y Hysa, L. (3-7 de abril de 2017). *The arText prototype: an automatic system for writing specialized texts* [conferencia]. 15th Conference of the European Chapter of the Association for Computational Linguistics. Proceedings of the Software Demonstrations, Valencia, España.
- 36 Davidson, L. Y. J., Richardson, M. y Jones, D. (2014). Teachers' perspective on using technology as an instructional tool. *Research in Higher Education*, 24, 1–25. <https://eric.ed.gov/?id=EJ1064110>
- 37 Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, 18(1), 7–24. <https://doi.org/10.1016/j.asw.2012.10.002>
- 38 Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, 5(1), 1–36. <https://ejournals.bc.edu/index.php/jtla/article/view/1640>
- 39 Dikli, S. (2010). The nature of automated essay scoring feedback. *CALICO Journal*, 28(1), 99–134. <https://doi.org/10.11139/cj.28.1.99-134>
- 40 Dikli, S. y Bleyl, S. (2014). Automated Essay Scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, 22, 1–17. <https://doi.org/10.1016/j.asw.2014.03.006>
- 41 Di Eugenio, B., Fossati, D., Haller, S., Yu, D. y Glass, M. (2008). Be brief, and they shall learn: Generating concise language feedback for a computer tutor. *International Journal of Artificial Intelligence in Education*, 18(4), 317–345.

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- 42 Dodigovic, M. (2007). Artificial intelligence and second language learning: An efficient approach to error remediation. *Language Awareness, 16*(2), 99–113.
- 43 Dwyer, H. J. y Sullivan, H. J. (1993). Student preferences for teacher and computer composition marking. *The Journal of Educational Research, 86*(3), 137–141. <https://doi.org/10.1080/00220671.1993.9941152>
- 44 El Ebyary, K. y Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies, 10*(2), 121-142.
- 45 El Ebyary, K. y Windeatt, S. (2019). Eye tracking analysis of EAP students' regions of interest in computer-based feedback on grammar, usage, mechanics, style and organization and development. *System, 83*, 36–49. <https://doi.org/10.1016/j.system.2019.03.007>
- 46 Fang, Y. (2010). Perceptions of the computer-assisted writing program among EFL college learners. *Journal of Educational Technology & Society, 13*(3), 246–256.
- 47 Ferrero, C. L., Renau, I., Nazar, R. y Torner, S. (2014). Computer-assisted revision in Spanish academic texts: peer-assessment. *Procedia-Social and Behavioral Sciences, 141*, 470-483.
- 48 Franzke, M., Kintsch, E., Caccamise, D., Johnson, N. y Dooley, S. (2005). Summary street®: Computer support for comprehension and writing. *Journal of Educational Computing Research, 33*(1), 53–80. <https://doi.org/10.2190/DH8F-QJWM-J457-FQVB>
- 49 Garcia-Gorrostieta, J. M., Lopez-Lopez, A. y Gonzalez-Lopez, S. (2018). Automatic argument assessment of final project reports of computer engineering students. *Computer Applications in Engineering Education, 26*(5), 1217–1226. <https://doi.org/10.1002/cae.21996>
- 50 Gao, J. W. y Ma, S. (2019). The effect of two forms of computer-automated metalinguistic corrective feedback. *Language Learning & Technology, 23*(2), 65–83.
- 51 Gao, J. W. y Ma, S. (2020). Instructor feedback on free writing and automated corrective feedback in drills: Intensity and efficacy. *Language Teaching Research, https://doi.org/10.1177/136216882091533.7*
- 52 Gamallo, P., Garcia, M., Pineiro, C., Martinez-Castano, R. y Pichel, J. C. (15-18 de octubre de 2018). *LinguaKit: a Big Data-based multilingual tool for linguistic analysis and information extraction* [conferencia]. Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), Valencia, España.
- 53 Gerard, L. F. y Linn, M. C. (2017). Using automated scores of student essays to support instructor guidance in classroom inquiry. *Journal of Science Instructor Education, 27*(1), 111–129. [doi:10.1007/s10972-016-9455-6](https://doi.org/10.1007/s10972-016-9455-6)
- 54 Graesser, A. C. (2016). Conversations with AutoTutor help students learn. *International Journal of Artificial Intelligence in Education, 26*, 124–132. [doi:10.1007/s40593-015-0086-4](https://doi.org/10.1007/s40593-015-0086-4)
- 55 Graesser, A. C., McNamara, D. S., Louwerse, M. M. y Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers, 36*(2), 193-202.
- 56 Grimes, D. y Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment, 8*(6), 1–44. <https://ejournals.bc.edu/index.php/jtla/article/view/1625>
- 57 Guo, Q., Feng, R. L. y Hua, Y. F. (2021). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning, 35*(9), 2312-2331. <https://doi.org/10.1080/09588221.2021.1879161>
- 58 Gutierrez, F. y Atkinson, J. (2011). Adaptive feedback selection for intelligent tutoring systems. *Expert Systems with Applications, 38*(5), 6146–6152.
- 59 Han, Y., Zhao, S. y Ng, L. L. (2021). How technology tools impact writing performance, lexical complexity, and perceived self-regulated learning strategies in EFL academic writing: A comparative study. *Frontiers in Psychology, 12*, 1–18. <https://doi.org/10.3389/fpsyg.2021.752793>.
- 60 Hassanzadeh, M. y Fotoohnejad, S. (2021). Implementing an automated feedback program for a foreign language writing course: A learner-centric study implementing an AWE tool in a L2 class. *Journal of Computer Assisted Learning, 37*(5), 1494–1507. <https://doi.org/10.1111/jcal.12587>
- 61 Hazelton, L., Nastal, J., Elliot, N., Burstein, J. y McCaffrey, D. (2021). Formative automated writing evaluation: A standpoint theory of action. *Journal of Response to Writing, 7*(1), 37–91.
- 62 Heift, T. y Hegelheimer, V. (2017). Computer-assisted corrective feedback and language learning. *Corrective feedback in second language teaching and learning, 51-65*.
- 63 Hegelheimer, V., Dursun, A. y Li, Z. (2016). Automated writing evaluation in language teaching: Theory, development, and application. *Calico Journal, 33*(1), 1-5.

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- 64 Hockly, N. (2019). Automated writing evaluation. *ELT Journal*, 73(1), 82– 88. <https://doi.org/10.1093/elt/ccy044>
- 65 Hoang, G. T. L. y Kunnan, A. J. (2016). Automated essay evaluation for English language learners: A case study of MY Access. *Language Assessment Quarterly*, 13(4), 359–376. <https://doi.org/10.1080/15434303.2016.1230121>
- 66 Huang, Y. y Wilson, J. (2021). Using automated feedback to develop writing proficiency. *Computers and Composition*, 62, 1-27. <https://doi.org/10.1016/j.compcom.2021.102675>
- 67 Huang, S. y Renandya, W. A. (2020). Exploring the integration of automated feedback among lower-proficiency EFL learners. *Innovation in language learning and teaching*, 14(1), 15-26.
- 68 Jiang, L. y Yu, S. (2022). Appropriating automated feedback in L2 writing: Experiences of Chinese EFL student writers. *Computer Assisted Language Learning*, 35(7), 1329-1353. <https://doi.org/10.1080/09588221.2020.1799824>
- 69 Kellogg, R. T., Whiteford, A. P. y Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research*, 42(2), 173–196. <https://doi.org/10.2190/EC.42.2.c>
- 70 Kim, K., Clarianay, R. y Kim, Y (2019). Automatic representation of knowledge structure: enhancing learning through knowledge structure reflection in an online course. *Education Tech Research*, 67, 105–122. <https://doi.org/10.1007/s11423-018-9626-6>
- 71 Koh, W. Y. (2017). Effective applications of automated writing feedback in process-based writing instruction. *English Teaching*, 72(3), 91-118.
- 72 Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44, 1-12.
- 73 Lachner, A., Burkhart, C. y Nuckles, M. (2017). Mind the gap! Automated concept map feedback supports students in writing cohesive explanations. *Journal of Experimental Psychology. Applied*, 23(1), 29–46. <https://doi.org/10.1037/xap0000111>
- 74 Lai, Y. H. (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology*, 41(3), 432–454. <https://doi.org/10.1111/j.1467-8535.2009.00959.x>
- 75 Landauer, T. y Psootka, J. (2000). Simulating text understanding for educational applications with latent semantic analysis: Introduction to LSA. *Interactive Learning Environments*, 8(2), 73–86. [https://doi.org/10.1076/1049-4820\(200008\)8:2;1-b;ft073](https://doi.org/10.1076/1049-4820(200008)8:2;1-b;ft073)
- 76 Landauer, T., Lochbaum, K. y Dooley, S. (2009). A new formative assessment technology for reading and writing. *Theory into Practice*, 48(1), 44–52. <https://doi.org/10.1080/00405840802577593>
- 77 Lavolette, E., Polio, C. y Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language, Learning & Technology*, 19(2), 50-68.
- 78 Lawley, J. (2016). Spelling: Computerised feedback for self-correction. *Computer Assisted Language Learning*, 29(5), 868–880.
- 79 Lee, C. (2020). A study of adolescent English learners' cognitive engagement in writing while using an automated content feedback system. *Computer Assisted Language Learning*, 33(1–2), 26–57. <https://doi.org/10.1080/09588221.2018.1544152>
- 80 Lee, Y.-J. (2020). The long-term effect of automated writing evaluation feedback on writing development. *English Teaching*, 75(1), 67–92. <https://doi.org/10.15858/engtea.75.1.202003.67>
- 81 Lee, C., Wong, K. C. K., Cheung, W. K. y Lee, F. S. L. (2009). Web-based essay critiquing system and EFL students' writing: A quantitative and qualitative investigation. *Computer Assisted Language Learning*, 22(1), 57–72. <https://doi.org/10.1080/09588220802613807>
- 82 Leontjev, D. (2014). The effect of automated adaptive corrective feedback: L2 English questions. *Apples-Journal of Applied Language Studies*, 8(2), 43–66.
- 83 Li, Z., Feng, H. H. y Saricaoglu, A. (2017). The short-term and long-term effects of AWE feedback on ESL students' development of grammatical accuracy. *CALICO journal*, 34(3), 355-375.
- 84 Li, J. R., Link, S. y Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27, 1–18. <https://doi.org/10.1016/j.jslw.2014.10.004>
- 85 Liao, H. C. (2016). Using automated writing evaluation to reduce grammar errors in writing. *Elt Journal*, 70(3), 308–319. <https://doi.org/10.1093/elt/ccv058>
- 86 Liao, H.-C. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System*, 62, 77–92. <https://doi.org/10.1016/j.system.2016.02.007>

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- 87 Link, S., Mehrzad, M. y Rahimi, M. (2020). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 33, 1–30.
- 88 Liu, M., Li, Y., Xu, W. y Liu, L. (2016). Automated essay feedback generation and its impact on revision. *IEEE Transactions on Learning Technologies*, 10(4), 502–513. <https://doi.org/10.1109/TLT.2016.2612659>
- 89 Liu, S. y Kunna, A. J. (2017). Investigating the application of automated writing evaluation to Chinese undergraduate English majors: A case study of WriteToLearn. *CALICO Journal*, 33(1), 71–91. <https://doi.org/10.1558/cj.v33i1.26380>
- 90 Luo, Y. y Liu, Y. (2017). Comparison between peer feedback and automated feedback in college English writing: A case study. *Open Journal of Modern Linguistics*, 7(4), 197–215. <https://doi.org/10.4236/ojml.2017.74015>
- 91 Lu, X., Di Eugenio, B., Ohlsson, S. y Fossati, D. (8-11 de junio de 2008). *Simple but effective feedback generation to tutor abstract problem solving* [conferencia]. Proceedings of the Fifth International Natural Language Generation Conference, Ohio, USA.
- 92 Ma, H. y Slater, T. (2015). Using the developmental path of cause to bridge the gap between AWE scores and writing teachers' evaluations. *Writing & Pedagogy*, 7(2), 395–422. <https://doi.org/10.1558/wap.v7i2-3.26376>.
- 93 MacArthur, C. A., Jennings, A. y Philippakos, Z. A. (2018). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction? *Reading and Writing: An Interdisciplinary Journal*, 32, 1553–1574.
- 94 Matthews, S. y Matthews, B. (2021). Reconceptualising feedback: Designing educational tangible technologies to be a creative material. *International Journal of Child-Computer Interaction*, 29, 1-12. <https://doi.org/10.1016/j.ijcci.2021.100278>
- 95 McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K. y Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing*, 23, 35–59. <https://doi.org/10.1016/j.asw.2014.09.002>.
- 96 Mirzaeian, V. R., Kohzadi, H. y Azizmohammadi, F. (2016). Learning Persian grammar with the aid of an intelligent feedback generator. *Engineering Applications of Artificial Intelligence*, 49, 167–175
- 97 Morch, A. I., Engeness, I., Cheng, V. C., Cheung, W. K. y Wong, K. C. (2017). EssayCritic: Writing to learn with a knowledge-based design critiquing system. *Educational Technology & Society*, 20(2), 213–223.
- 98 Na, Z. y Ma, X. (2021). Automated writing evaluation (AWE) feedback: A systematic investigation of college students' acceptance. *Computer Assisted Language Learning*, 35(9), 2817–2842.
- 99 O'Neill, R. y Russell, A. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, 35(1), 42–46. <https://doi.org/10.14742/ajet.3795>
- 100 Page, E. B. (1966). The imminence of ... grading essays by computer. *Phi Delta Kappan*, 47(5), 238–243.
- 101 Page, E. B. (1968). The use of the computer in analyzing student essays. *International Review of Education*, 14, 210–225. doi:10.1007/BF01419938
- 102 Pallant, A. (2018). Validation of automated scoring for a formative assessment that employs scientific argumentation. *Educational Assessment*, 23(2), 121–138. doi:10.1080/10627197.2018.1427570
- 103 Palermo, C. y Thomson, M. M. (2018). Teacher implementation of self-regulated strategy development with an automated writing evaluation system: Effects on the argumentative writing performance of middle school students. *Contemporary Educational Psychology*, 54, 255–270. <https://doi.org/10.1016/j.cedpsych.2018.07.002>
- 104 Palermo, C. y Wilson, J. (2020). Implementing automated writing evaluation in different instructional contexts: A mixed-methods study. *Journal of Writing Research*, 12(1), 63–108. <https://doi.org/10.17239/jowr-2020.12.01.04>
- 105 Parra, L. y Calero, X. (2019). Automated writing evaluation tools in the improvement of the writing skill. *International Journal of Instruction*, 12(2), 209–226. <https://doi.org/10.29333/iji.2019.12214a>

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- 106 Perin, D. y Lauterbach, M. (2018). Assessing text-based writing of low-skilled college students. *International Journal of Artificial Intelligence in Education*, 28(1), 56–78. <https://doi.org/10.1007/s40593-016-0122-z>.
- 107 Persico, D., Passarelli, M., Manganello, F., Pozzi, F., Dagnino, F. M., Ceregini, A. y Caruso, G. (2020). Automatic feedback, self-regulated learning and social comparison: A case study. *Qwerty-Open and Interdisciplinary Journal of Technology, Culture and Education*, 15(2), 27-44.
- 108 Peterson, E. K. (2017). *The impact of computer-generated feedback on student perceptions of revision process*. https://sophia.stkate.edu/maed/247?utm_source=sophia.stkate.edu%2Fmaed%2F247&utm_medium=PDF&utm_campaign=PDFCoverPages
- 109 Phandi, P., Chai, K. y Ng, H. (19- 21 de septiembre de 2015). *Flexible domain adaptation for automated essay scoring using correlated linear regression* [conferencia]. Conference on Empirical Methods in Natural Language Processing, Lisboa, Portugal.
- 110 Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E. y Kukich, K. (2002). Stumping e-rater: Challenging the validity of automated essay scoring. *Computers in Human Behavior*, 18, 103–134. [https://doi.org/10.1016/S0747-5632\(01\)00052-8](https://doi.org/10.1016/S0747-5632(01)00052-8)
- 111 Powers, D. E., Escoffery, D. S. y Duchnowski, M. P. (2015). Validating automated essay scoring: A (modest) refinement of the “gold standard”. *Applied Measurement in Education*, 28, 130–142. doi:10.1080/08957347.2014.1002920
- 112 Proske, A., Narciss, S. y McNamara, D. S. (2012). Computer-based scaffolding to facilitate students’ development of expertise in academic writing. *Journal of Research in Reading*, 35(2), 136–152. <https://doi.org/10.1111/j.1467-9817.2010.01450.x>
- 113 Ramineni, C. y Williamson, D. M. (2013). Automated essay scoring: Psychometric guidelines and practices. *Assessing Writing*, 18, 25–39. <https://doi.org/10.1016/j.asw.2012.10.004>
- 114 Ranalli, J., Link, S. y Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educ. Psychol.* 37(1), 8–25.
- 115 Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it?. *Computer Assisted Language Learning*, 31(7), 653-674
- 116 Rapp, C. y Kauf, P. (2018). Scaling Academic Writing Instruction: Evaluation of a Scaffolding Tool (Thesis Writer). *International Journal of Artificial Intelligence in Education*, 28, 590-615.
- 117 Rapp, C. y Kruse, O. (24-27 de agosto de 2020). *Thesis writer 2.0: a system supporting academic writing, its instruction and supervision* [conferencia]. GMW Tagung 2020, Winterthur, Alemania.
- 118 Reynolds, B. L., Kao, C. W. y Huang, Y. Y. (2021). Investigating the effects of perceived feedback source on second language writing performance: A quasi-experimental study. *The Asia-Pacific Education Researcher*, 30(6), 585–595. <https://doi.org/10.1007/s40299-021-00597-3>
- 119 Reilly, E. D., Stafford, R. E., Williams, K. M. y Corliss, S. B. (2014). Evaluating the validity and applicability of automated essay scoring in two massive open online courses. *International Review of Research in Open and Distributed Learning*, 15(5), 83-98.
- 120 Roscoe, R. D., Wilson, J., Johnson, A. C. y Mayra, C. R. (2017). Presentation, expectations, and experience: Sources of student perceptions of automated writing evaluation. *Computers in Human Behavior*, 70, 207–221. <https://doi.org/10.1016/j.chb.2016.12.076>
- 121 Roscoe, R. D., Allen, L. K., Weston, J. L., Crossley, S. A. y McNamara, D. S. (2014). The Writing Pal intelligent tutoring system: Usability testing and development. *Computers and Composition*, 34, 39-59.
- 122 Roscoe, R. D. y McNamara, D. S. (2013). Writing pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, 105(4), 1010–1025. <https://doi.org/10.1037/a0032340>
- 123 Rupp, A. (2018). Designing, evaluating, and deploying automated scoring systems with validity in mind: Methodological design decisions. *Applied Measurement in Education*, 31, 191–214. doi:10.1080/08957347.2018.1464448
- 124 Saricaoglu, A. (2019). The impact of automated feedback on L2 learners’ written causal explanations. *ReCALL*, 31(2), 189–203. <https://doi.org/10.1017/S095834401800006X>
- 125 San Mateo, A. (2016). Un corpus de bigramas utilizado como corrector ortográfico y gramatical destinado a hablantes nativos de español. *Revista signos*, 49(90), 94-118.

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- 126 Shaalan, K., Magdy, M. y Fahmy, A. (2015). Analysis and feedback of erroneous Arabic verbs. *Natural Language Engineering*, 21(2), 271–323.
- 127 Shang, H. (2017). An exploration of asynchronous and synchronous feedback modes in EFL writing. *Journal of Computing in Higher Education*, 29(3), 496–513.
- 128 Shang, H. (2019). Exploring online peer feedback and automated corrective feedback on EFL writing performance. *Interactive Learning Environments*, 30(1):4–16.
- 129 Sherafati, N., Largani, F. M. y Amini, S. (2020). Exploring the effect of computer-mediated teacher feedback on the writing achievement of Iranian EFL learners: Does motivation count? *Education and Information Technologies*, 25(5), 4591–4613. <https://doi.org/10.1007/s10639-020-10177-5>
- 130 Shermis, M. D. (2014). State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing*, 20, 53–76. doi:10.1016/j.asw.2013.04.001
- 131 Shermis, M. D. (2018). Establishing a crosswalk between the Common European Framework for Languages (CEFR) and writing domains scored by automated essay scoring. *Applied Measurement in Education*, 31, 177–190. doi:10.1080/08957347.2018.1464451
- 132 Sinharay, S., Zhang, M. y Deane, P. (2019). Prediction of essay scores from writing process and product features using data mining methods. *Applied Measurement in Education*, 32, 116–137. doi:10.1080/08957347.2019.1577245
- 133 Stevenson, M. (2016). A critical interpretative synthesis: The integration of automated writing evaluation into classroom writing instruction. *Computers and Composition*, 42, 1-16
- 134 Stevenson, M. y Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19, 51–65. <https://doi.org/10.1016/j.asw.2013.11.007>
- 135 Sung, Y. T., Liao, C. N., Chang, T. H., Chen, C. L. y Chang, K. E. (2016). The effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. *Computers & Education*, 95, 1–18. <https://doi.org/10.1016/j.compedu.2015.12.003>
- 136 Tang, J. y Rich, C. S. (2017). Automated writing evaluation in an EFL setting: Lessons from China. *JALT CALL Journal*, 13(2), 117–143. <https://doi.org/10.29140/jaltcall.v13n2.215>
- 137 Tian, L. L. y Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System*, 91, 102247. <https://doi.org/10.1016/j.system.2020.102247>
- 138 Villena, J., González, B., González, B. y Muriel, M. (2002). STILUS: Sistema de revisión lingüística de textos en castellano. *Procesamiento del lenguaje natural*, 29, 305-306.
- 139 Ware, P. (2014). Feedback for adolescent writers in the english classroom. *Writing & Pedagogy*, 6(2), 223–249. <https://doi.org/10.1558/wap.v6i2.223>
- 140 Warschauer, M. y Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language teaching research*, 10(2), 157-180.
- 141 Warschauer, M. y Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 3(1), 22–36. <https://doi.org/10.1080/15544800701771580>
- 142 Wade-Stein, D. y Kintsch, E. (2004). Summary street: Interactive computer support for writing. *Cognition and Instruction*, 22(3), 333–362. https://doi.org/10.1207/s1532690xci2203_3
- 143 Wang, P.-L. (2015). Effects of an automated writing evaluation program: Student experiences and perceptions. *Electronic Journal of Foreign Language Teaching*, 12(1), 79–100.
- 144 Wang, E. L., Matsumura, L. C., Correnti, R., Litman, D., Zhang, H., Howe, E., Magooda, A. y Quintana, R. (2020). Wang, E. L., Matsumura, L. C., Correnti, R., Litman, D., Zhang, H., Howe, E., Magooda, A. y Quintana, R. (2020). eRevis(ing): Students' revision of text evidence use in an automated writing evaluation system. *Assessing Writing*, 44, 1-19. <https://doi.org/10.1016/j.asw.2020.100449>
- 145 Wang, Z. y Han, F. (2022). The Effects of Teacher Feedback and Automated Feedback on Cognitive and Psychological Aspects of Foreign Language Writing: A Mixed-Methods Research. *Frontiers in Psychology*, 13, 1-12.
- 146 Wang, Y., Harrington, M. y White, P. (2012). Detecting breakdowns in local coherence in the writing of Chinese English learners. *Journal of Computer Assisted Learning*, 28(4), 396–410. <https://doi.org/10.1111/j.1365-2729.2011.00475.x>
- 147 Wang, Y. J., Shang, H. F. y Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning*, 26(3), 234–257. <https://doi.org/10.1080/09588221.2012.655300>

CHARACTERIZATION AUTOMATED AND SEMI-AUTOMATED ASSESSMENT OF THE QUALITY OF WRITTEN TEXTS: A SYSTEMATIC REVIEW

- 148 Williamson, M. M. (2003). Validity of automated scoring: Prologue for a continuing discussion of machine scoring student writing. *Journal of Writing Assessment*, 1(2), 85–104.
- 149 Williamson, D. M., Xi, X. y Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice*, 31(1), 2–13. doi:10.1111/j.1745-3992.2011.00223.x
- 150 Wilson, J. (2017). Associated effects of automated essay evaluation software on growth in writing quality for students with and without disabilities. *Reading and Writing: An Interdisciplinary Journal*, 30(4), 691–718. <https://doi.org/10.1007/s11145-016-9695-z>
- 151 Wilson, J., Olinghouse, N. G. y Andrada, G. N. (2014). Does automated feedback improve writing quality? *Learning Disabilities: A Contemporary Journal*, 12(1), 93–118.
- 152 Wilson, J. y Czik, A. (2016). Automated essay evaluation software in English language arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, 100, 94–109. <https://doi.org/10.1016/j.compedu.2016.05.004>
- 153 Wilson, J. y Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, 58(1), 87–125. <https://doi.org/10.1177/0735633119830764>
- 154 Wilson, J., Huang, Y., Palermo, C., Beard, G. y MacArthur, C. A. (2021). Automated feedback and automated scoring in the elementary grades: Usage, attitudes, and associations with writing outcomes in a districtwide implementation of MI Write. *International Journal of Artificial Intelligence in Education*, 31(2), 234-276.
- 155 Wood, J. (2021). A dialogic technology-mediated model of feedback uptake and literacy. *Assessment & Evaluation in Higher Education*, 46(8), 1173-1190.
- 156 Yang, L. P., Xin, T., Luo, F., Zhang, S. y Tian, X. T. (2022). Automated evaluation of the quality of ideas in compositions based on concept maps. *Natural Language Engineering*, 28(4), 449-486.
- 157 Zaini, A. y Mazdayasna, G. (2015). The impact of computer-based instruction on the development of EFL learners' writing skills. *Journal of Computer Assisted Learning*, 31(6), 516-528.
- 158 Zaini, A. (2018). Word processors as monarchs: Computer-generated feedback can exercise power over and influence EAL learners' identity representations. *Computers & Education*, 120, 112–126. <https://doi.org/10.1016/j.compedu.2018.01.014>
- 159 Zhang, M. (2013). Contrasting automated and human scoring of essays. *R&D Connections*, 21, 1–11.
- 160 Zhang, Z. (2017). Student engagement with computer-generated feedback: A case study. *Elt Journal*, 71(3), 317–328. <https://doi.org/10.1093/elt/ccw089>
- 161 Zhang, Z. V. y Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90–102. <https://doi.org/10.1016/j.asw.2018.02.004>
- 162 Zhang, Z. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, 43, 78–91. <https://doi.org/10.1016/j.asw.2019.100439>
- 163 Zhu, M. X., Lee, H. S., Wang, T., Liu, O. L., Belur, V. y Pallant, A. (2017). Investigating the impact of automated feedback on students' scientific argumentation. *International Journal of Science Education*, 39(12), 1648–1668. <https://doi.org/10.1080/09500693.2017.1347303>
- 164 Zhu, M. X., Liu, O. L. y Lee, H. S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education*, 143, 103668. <https://doi.org/10.1016/j.compedu.2019.103668>