

Mayo Beltrán, A. M. (septiembre-diciembre, 2021). Recursos de evaluación de textos narrativos en educación primaria y análisis del feedback docente. *Revista Virtual Universidad Católica del Norte*, (64), 92-118.

<https://www.doi.org/10.35575/rvucn.n64a5>

Recursos de evaluación de textos narrativos en educación primaria y análisis del *feedback* docente

Resources to assess narrative texts in Primary Education and teacher feedback analysis

Alba María Mayo Beltrán

Doctoranda en Innovación en Formación del Profesorado
Departamento de Ciencias de la Educación, Universidad de Extremadura
Badajoz, España

amayob@unex.es

Orcid: <https://orcid.org/0000-0002-9003-388X>

Recibido: 5 de julio de 2020

Evaluado: 6 de septiembre de 2020

Aprobado: 6 de abril de 2021

Tipo de artículo: Investigación Científica y Tecnológica

Resumen

El presente estudio persigue un doble objetivo, explorar si la rúbrica y la lista de control tienen validez de criterio, pudiendo medir un mismo elemento, minimizando las diferencias entre las calificaciones, y analizar los comentarios evaluativos, realizados por los evaluadores, sobre la destreza escritora del alumnado de Educación Primaria. Para ello, 56 docentes en formación corrigieron dos textos narrativos diferentes, utilizando los instrumentos mencionados, y proporcionaron una serie de comentarios evaluativos, tomando como referencia los criterios considerados. Estos últimos fueron categorizados en función de la amplitud y del contenido evaluativo y meta-textual. Tras ello, se realizó un análisis factorial y descriptivo para el examen de los datos. Los resultados mostraron que las calificaciones globales, asignadas con la rúbrica y la lista de control a los textos por los evaluadores, tenían una relación significativa y que los docentes en formación proporcionan comentarios evaluativos empleando la rúbrica con contenido meta-textual

variado y completo. Así, se concluye la posibilidad de los docentes de utilizar cualquiera de las herramientas examinadas para proporcionar un juicio justo; así como los beneficios que comporta el uso de la rúbrica a la hora de generar mensajes de retroalimentación específicos, referidos a diversos aspectos y características de la lengua.

Palabras clave: Comentarios evaluativos; Lista de control; Rúbrica; Textos narrativos.

Abstract

The present investigation has a twin objective; exploring if both rubric and checklist have criterion validity, being able to assess same elements, minimizing differences between marks, and analyzing the *feedbacks*, given by the raters, to elementary students' writing competence. For that purpose, 56 future primary teachers assessed two different narrative text, using the two mentioned assessments instrument, and coming up with a *feedback* for each of the considered criteria. These latter feedbacks were categorized according to their amplitude and their evaluative and meta-textual content. Afterwards, a factorial and descriptive analysis were carried out to examine the data. Results show how global points, assigned by both instruments, rubric and checklist, are significantly related and how *feedbacks* given by future teachers using the rubric are meta-textually varied and complete. Thus, it could be concluded that both examined assessment instrument can be used by teachers to produce a fair judgement; as well as using rubric generates specific *feedbacks*, related to varied language aspects and characteristics.

Key words: Feedback; Checklist; Rubric; Narrative writing.

Introducción

La escritura y la lectura permiten a las personas alcanzar diferentes objetivos mediante su uso (Graham, 2006). Estas se consagran como herramientas que permiten mantener lazos de unión personales y profesionales, crear y contar historias o experiencias, compartir información, generar influencias sobre otros (Graham et al., 2013) o, incluso, beneficiar la salud mental, a través de la expresión de los sentimientos o de exploración de la propia persona (Smyth, 1998). La evidente importancia de ambas destrezas es corroborada por diferentes autores que señalan que las competencias lectoras

y escritoras suponen uno de los pilares sobre el que se asientan no solo aspectos académicos, sino también aquellos referidos a la vida cotidiana (Graham, 2006; Pérez, 2008).

Sin duda, será en el contexto escolar donde el alumnado desarrollará gran parte de estas habilidades mediante la enseñanza y guía de sus docentes. Así, es posible observar cómo el sistema educativo, que impera en la sociedad, se encuentra constituido por unas metodologías, unos estándares de aprendizaje y unos procedimientos, en los cuales el lenguaje y, dentro de este, los procesos lecto-escritores toman una gran relevancia. Los distintos exámenes de acceso a las diferentes enseñanzas, las clases guiadas por libros de texto o la realización de trabajos académicos son algunos ejemplos. De esta manera, se puede afirmar que el aprendizaje y el éxito escolar se verán influenciados, en gran medida, por la competencia escritora y lectora que desarrolle el alumnado (Fang & Wang, 2011; Jøll, 2014; Uzun et al., 2019), siendo necesario preguntarse cómo los docentes llevan a cabo la instrucción de dichas habilidades y cuál es el grado de adquisición y dominio del alumnado.

En este sentido, el último informe PISA (OCDE, 2018), pone de manifiesto la correlación existente entre ambas destrezas descritas y, debido a ello, únicamente evalúa la primera de ellas, la lectura. Consecuentemente a su estrecha relación, es posible observar cómo la valoración de las capacidades lectoras remite a aspectos necesarios para una adecuada competencia escritora, como la capacidad para comparar, contrastar o categorizar información. Los resultados de este informe muestran que solo el 53,5 % del alumnado evaluado en España superó el nivel tres en esta prueba, existiendo seis de ellos y siendo este último el superior, en el cual únicamente el 1,3 % de los estudiantes tuvieron éxito. El análisis puede dar a conocer que, aproximadamente, el 50 % del alumnado llega al nivel intermedio en lectura.

De acuerdo con estos datos, esta investigación se centrará de forma específica en las destrezas escritoras y en cómo futuros docentes llevan a cabo la evaluación de las composiciones de sus estudiantes, entendiendo esta como una parte esencial del proceso de enseñanza-aprendizaje que permite al docente analizar la práctica de su alumnado y proporcionarle un *feedback* para su necesaria mejora.

Referentes teóricos

La escritura es una de las bases sobre las que se asientan distintas disciplinas didácticas y aspectos de la vida cotidiana. Esto explicaría la transcendencia que posee para legisladores y docentes como objetivo curricular prioritario (Jølle, 2014). Así, resulta imprescindible que el alumnado adquiera “los recursos necesarios para participar, mediante el lenguaje, en las diferentes esferas de la vida” (Pérez, 2008, p. 16). Uribe-Enciso y Carrillo-García (2014) ratifican que ciertas tareas sencillas de la vida diaria y, especialmente, aquellas que tienen un carácter académico resultarán más complejas para aquellas personas con dificultades en la escritura y en su comprensión; esto es debido a que “la construcción de conocimiento se genera, en gran parte, a partir de la comprensión de la información escrita disponible en diversos medios” (p. 275), y dicho conocimiento es conservado en mayor grado cuando el alumnado escribe acerca de aquello que está leyendo (Graham & Hebert, 2010, 2011). De esta manera, existe una relación directa entre ambos procesos, lectores y escritores, que mejoran el rendimiento académico.

En relación con los procesos mencionados, Uribe-Enciso y Carrillo-García (2014) concluyen que la habilidad y la destreza en los procesos lectores permitirán a las personas desarrollar sus propias ideas, con aquella información que ya poseían y la que han adquirido gracias a la comprensión de textos. Sin embargo, la escritura posibilita la recogida y transmisión de dicha información, fomentando su examen y el establecimiento de conexiones entre ideas, encontrándose las mismas accesibles para ser revisadas y evaluadas (Applebee, 1984). Esta necesidad de llevar a cabo procesos de organización y planificación (Uzun et al., 2019), junto a la atención requerida en la construcción del significado, al mismo tiempo que se contemplan las convenciones de la propia escritura (Scull et al., 2020), convierte a esta en una de las habilidades más complejas dentro de la lengua, y de la cual depende el desarrollo del resto de disciplinas en la misma (Hyland, 2003). Así, se podría afirmar que el éxito escolar y las competencias lectoras y, especialmente, escritoras se encuentran íntimamente relacionados.

Romero et al. (2002) continúan con esta idea y señalan, como aspecto común a diferentes problemas educativos, las dificultades en los procesos matemáticos, de lectura y escritura, e incluso exponen ciertas habilidades relacionadas con estos procesos, como

pueden ser la comprensión y manejo del lenguaje, o de conceptos numéricos, como prerrequisito para el aprendizaje.

La síntesis de lo señalado hasta el momento bien podría encontrarse en lo expuesto por Grové y Hauptfleish (1982, como se citó en Uribe-Enciso y Carrillo-García, 2014), los cuales afirman que “un individuo que lea y escriba bien se desempeña con mayor eficiencia en los distintos entornos de su vida” (p. 275). Es decir, las habilidades lectoras y escritoras, alcanzadas por las personas en un nivel de competencia adecuado, les permitirán desarrollarse de forma satisfactoria en distintos ámbitos de su vida, personal, profesional y social (Graham, 2006).

La evidente importancia de la escritura es, por tanto, puesta de manifiesto. Así, en el contexto escolar, el alumnado requiere de una serie de competencias escritoras que pueda trasladar a diferentes áreas de aprendizaje y le permitan construir su propio conocimiento, y de esta manera conseguir éxito educativo. Para ello, desde los primeros años de la Educación Primaria el tiempo y esfuerzo empleado por docentes en actividades relacionadas con la escritura y su evaluación es cuantioso y constante (Fernández, Lucero et al., 2016), siendo estos conscientes de que los estudiantes han de emplear años para adquirir una competencia escritora adecuada y efectiva en todas sus dimensiones, gramatical, semántica, organizativa y de adecuación a diferentes contextos (Bazerman, 2013; Gallego y Muñoz 2020; Graham et al., 2013), requiriendo este proceso de unas “condiciones físicas y cognitivas” adecuadas, además de “ambientes favorecedores y estimulantes” (Gallego y Muñoz 2020). En definitiva, “el aprendizaje sobre las habilidades escritoras del alumnado en educación juega un importante rol en la planificación y la gestión de los procesos educativos, y el desarrollo efectivo de las habilidades escritoras individuales” (Uzun et al., 2019, p.169).

En ese transcurso del desarrollo de la escritura, por parte del alumnado, la evaluación y el *feedback*, que el docente proporciona al estudiante y que guiará al mismo en su desempeño, cobran especial relevancia, ya que permitirá planificar su enseñanza e instrucción. Esta evaluación ha de ser concebida como un proceso en el cual se han de considerar diferentes aspectos que lo dirijan hacia su eficacia. De tal modo, será fundamental valorar el trabajo del estudiante en diferentes fases encadenadas, que permitan comprender sus necesidades y adecuar la enseñanza a las mismas, estimar si dicha instrucción está siendo efectiva y examinar los resultados finales, para comenzar de nuevo el ciclo con los objetivos apropiados. Asimismo, resulta necesario distinguir

criterios de los que partir, en los que sentar las bases para el éxito en la evaluación y que proporcionen conocimiento en relación con los diferentes niveles de desempeño de los estudiantes en el aula, explorando las debilidades y fortalezas que estos manifiesten, con el objetivo de satisfacer las necesidades individuales (Fang & Wang, 2011). En este sentido, existen antecedentes que demuestran los beneficios que proporciona el uso de instrumentos de evaluación para tal fin. Un ejemplo de ello es el estudio llevado a cabo por Ghaffar et al. (2020), el cual expuso que el uso de la rúbrica en la evaluación de textos expositivos permitió al docente y a los estudiantes identificar con cierta exactitud los errores existentes en sus escritos, y proporcionó, a estos últimos, conocimiento en relación con los objetivos se esperaba que cumplieren y capacidad para generar diferentes formas de mejorarlo. En definitiva, será esencial tener en cuenta los diferentes tipos de evaluación existentes, así como conocer los múltiples instrumentos ideados para tal fin, de manera que el proceso se ajuste, lo máximo posible, a cada individuo.

Por otro lado, analizados los diferentes aspectos que caracterizan el proceso de una evaluación efectiva de la escritura, será necesario preguntarse cómo los docentes llevan a cabo dicha evaluación, cuáles son las dificultades a la que se enfrentan y qué tipo de problemáticas se originan en consecuencia en el procedimiento. Así, Parr & Timperley (2008, como citó en Fernández, Lucero et al. 2016,) señalan que una evaluación de calidad de la escritura requiere de un conocimiento profundo de la misma, “un meta-conocimiento explícito de la lengua, de las estructuras de los textos, de los procesos y estrategias de la escritura” (p. 66). De acuerdo con esto, Jones y Moreland (2005) y Shepard et. al (2005), como se citó en Fernández, Lucero et al. (2016), exponen que la falta de dicha competencia propia será muy en parte responsable de las dificultades del profesorado para proporcionar a su alumnado una retroalimentación efectiva que le permita mejorar sus escritos y anticipar las dificultades que pueda tener en el proceso. Asimismo, será esencial que los docentes posean un conocimiento de los procedimientos de aprendizaje y enseñanza de la composición de un escrito, que les permita organizar y planificar dicho proceso (Sandiford & Macken-Horarik, 2020; Scull et al., 2020).

No obstante, la formación docente no suele concebir un entrenamiento profundo en las características discursivas de la escritura (Schleppegrell, 2004), por lo que, en consonancia con lo señalado anteriormente, puede deducirse la falta de preparación y conocimiento para llevar a cabo y proporcionar una evaluación de calidad de la escritura del alumnado, por parte de estos. Así, diversos estudios han puesto de manifiesto la

tendencia del profesorado a centrarse en aspectos superficiales de la escritura, gramaticales y ortográficos, en su corrección, tales como la puntuación, el uso de mayúscula o las concordancias sujeto-verbo o género-número (Montanero et al., 2014; Schleppegrell, 2004). Asimismo, un estudio liderado por Otnes & Solheim (2019) determinó que la mayoría de los comentarios proporcionados por los docentes, al evaluar los textos de su alumnado, tenían la característica de ser directivos, y de forma inusual se llevaban a cabo de una forma dialógica, conllevando una interacción profesor-alumno unidireccional. A este respecto, sin embargo, Tsui y Ng (2000, como se citó en Fernández, Lucero et al., 2016) revelan que, junto a la capacidad del alumnado, la retroalimentación dada por el docente a este será esencial en las revisiones de sus textos y su mejora. Por tanto, acorde a lo expuesto, se sustenta la necesidad de proporcionar una evaluación eficaz, que vaya más allá de los elementos superfluos y considere aquellos transcendentales que permitan al alumnado optimizar su competencia escritora mediante un *feedback* de calidad.

Por otro lado, se ha de tener en cuenta que este tipo de evaluación directa, en la que el docente evalúa un texto escrito del estudiante, se ve afectada por muy diferentes aspectos en su puesta en marcha, influyendo estos en la variabilidad de su calificación (Uzun et al., 2019). Por tanto, las dificultades a la que se enfrenta el docente a la hora de evaluar ya no solo residen en el conocimiento de este para realizarla y, por consiguiente, en aquello que evalúa, sino también en los múltiples aspectos que intervienen en la fiabilidad de la evaluación que ha realizado. De este modo, Uzun et al. (2019) apuntan al propio evaluador como uno de los elementos más decisivos que atañen a la variabilidad en la evaluación. A modo de ejemplo puede citarse la disparidad de opiniones entre docentes, encontrada por Skar & Jølle (2017) en relación con las características significativas que debían poseer los tipos de textos evaluados en dicha investigación, narrativo y expositivo. Así, un texto puede ser considerado y calificado de forma diferente, en función del docente que realice dicha evaluación, llevando esta variabilidad inter-jueces a equívocos en el dictamen de las competencias escritas del alumnado (Popham, 1990). En relación con esta problemática, Heldsinger & Humphry (2010) declaran que la producción, por parte del docente, de una evaluación de la escritura fiable pasará por el uso de una “metodología rigurosa y apropiada”. De esta manera, el uso de escalas e instrumentos de evaluación es una forma de reducir la inconsistencia en la evaluación y aumentar la fiabilidad de la misma, distanciándose de la subjetividad y las

impresiones generales, y generando un examen justo y efectivo (Kayapinar, 2014; Uzun et al., 2019).

Estudios previos han comparado el uso que realiza el profesorado de diferentes métodos y estrategias evaluativas en la escritura. Entre ellos, el llevado a cabo por Kayapinar (2014), quien concluye que el uso de escalas de apreciación numéricas y listas de control se presenta con una mayor fiabilidad y consistencia que el examen de un texto mediante una impresión generalizada, la cual no cumple estas condiciones en ningún caso. Entre ambos instrumentos comparados, escala de apreciación y lista de control, se resuelve la primera de estas como la más fiable, aunque con resultados muy similares. Asimismo, la investigación llevada a cabo por Uzun et al. (2019) examinó las diferencias entre la utilización para la evaluación de la escritura de la lista de control y la rúbrica, mostrando la segunda de ellas como la más fiable y generalizable. Este último estudio señaló, además, algunas de las fuentes de error más comunes en el manejo de estos instrumentos, como son la insuficiente formación de los docentes en el uso de los mismos o la necesidad de descansos frecuentes en el proceso.

En síntesis, la evaluación que realizan los docentes de los textos del alumnado podría decirse que se encuentra marcada por múltiples y diversos aspectos, por lo que es posible señalar con certeza que “se requiere de una estrategia evaluativa que dé cuenta de la capacidad del estudiante para escribir, que lo ayude a analizar y mejorar su escritura mediante estándares de ejecución específicos, así como que proporcione un lenguaje común” (Contreras et al., 2009, p. 520). Los instrumentos de evaluación de texto, como las rúbricas, las listas de control o las escalas de apreciación, proporcionan esos estándares o criterios fundamentales que orientan y guían el proceso, y en los cuales el docente basa su juicio y provee con un *feedback* al alumnado. Sin embargo, es necesario conocer con profundidad los entresijos que conllevan su uso, siendo conscientes de sus limitaciones, ventajas y, en general, aspectos a tener en cuenta en su utilización. Para ello, la investigación y estudio de sus características y de su empleo en la práctica docente serán fundamentales, siendo en líneas generales el objetivo del estudio que subyace a continuación.

Así, se seleccionaron para su análisis dos tipos de herramientas concretas para la evaluación de un tipo de texto determinado, el narrativo. Los instrumentos elegidos fueron la lista de control y la rúbrica, cuyo estudio tiene diversos fines en esta investigación, en relación con las problemáticas explicitadas hasta el momento. De esta

manera, por un lado, se pretende indagar si ambos instrumentos tienen validez de criterio, pudiendo afirmar la capacidad de estos para medir un mismo elemento y minimizando las diferencias entre las calificaciones que se pudiesen obtener con uno u otro; y, por otro lado, se quiere registrar los comentarios evaluativos realizados por los evaluadores, futuros docentes de Educación Primaria, y advertir si el uso de la rúbrica permite generar un *feedback* de calidad, teniendo en cuenta las dificultades existentes a este respecto.

Metodología

A continuación, se presenta la metodología llevada a cabo para la realización de esta investigación, señalando para ello la población a la que va dirigida y las características de la muestra, así como las técnicas e instrumentos utilizados para la recogida de la información.

Población y muestra

La presente investigación estuvo dirigida a la población universitaria, perteneciente al primer curso del grado de Educación Primaria, de la Facultad de Educación de la Universidad de Extremadura. En total, participó un número de 56 futuros docentes, los cuales analizaron dos textos narrativos de dos estudiantes de primaria de 5º curso, con dos instrumentos diferentes: rúbrica y lista de control.

La asignación del texto a corregir por cada sujeto fue por conveniencia, teniendo en cuenta el grupo-clase al que pertenecían y adjudicando los textos uno y dos, y los grupos uno y dos, respectivamente. Finalmente, se obtuvieron 35 correcciones del texto uno, 31 con cada uno de los instrumentos; y 21 correcciones del texto 2, 14 con rúbrica y 16 con lista de control. La diferencia perceptible en el número de veces de utilización de los diferentes instrumentos, para el examen de cada uno de los textos, se debe a la falta de tiempo de algunos de los evaluadores para proceder a la evaluación con alguna de las herramientas.

A lo largo de la evaluación de los textos se pidió a los evaluadores que proporcionasen un *feedback* escrito para cada uno de los criterios de uno de los instrumentos, la rúbrica. Estos fueron analizados en el estudio, tal y como se verá más adelante, generándose una muestra de 381 mensajes en total a examinar.

Técnicas e instrumentos de recogida y análisis de la información

La recogida de la muestra se llevó a cabo de forma presencial en el aula. A los evaluadores se les entregó una serie de documentos, los cuales contenían el texto seleccionado para su corrección y una copia de cada uno de los instrumentos a utilizar. Asimismo, al inicio de la sesión de evaluación se les proporcionó una serie de instrucciones en relación con el uso de la rúbrica y la lista de control. La misma tuvo una duración aproximada de 20 minutos.

La rúbrica se define como un instrumento de evaluación constituido por una escala descriptivo-ordinal, la cual permite medir una habilidad o constructo, teniendo en cuenta una serie de categorías, para las cuales se precisan niveles o rangos. Los mismos describen de forma precisa y específica las características de actuación a evaluar en cada uno de estos criterios (Contreras et al., 2009; Ministry of Education of Ontario, 2006; Montanero et al., 2014). En cuanto a la lista de control (en inglés *checklist*), esta se constituye como un instrumento de evaluación, determinado por una serie de conductas, características o habilidades, que aparecen en forma de lista y en relación con la cuales el evaluador registra su existencia o ausencia en el texto que está evaluando (Montanero, 2019; Uzun et al., 2019). En concreto, las empleadas en el actual estudio fueron, por un lado, una rúbrica de evaluación específica para textos narrativos, de los autores Montanero et al. (2014) y, por otro lado, una lista de control constituida mediante una adaptación de esta última. De este modo, ambos instrumentos están conformados por siete criterios (marco, tema, trama, creatividad, oraciones, vocabulario y ortografía), proporcionando la rúbrica una escala descriptivo-ordinal con una serie de rangos para cada uno de los criterios, y permitiendo puntuar cada uno de ellos del uno al cuatro, siendo el uno el nivel más bajo y el cuatro el más alto. Igualmente, a cada nivel se le asignó una puntuación que permitió obtener una nota final de evaluación del texto, partiendo del 0 en el nivel uno y sumando 0,5 a cada nivel posterior, hasta llegar a 1,5 puntos, correspondientes al nivel cuatro. Por su parte, la lista de control fue conformada teniendo en cuenta la descripción de los niveles superiores de cada criterio del instrumento anterior, para cuyos aspectos el evaluador debía de registrar su existencia o no en el texto. Al igual que en la rúbrica, se determinó un sistema de puntuación para esta herramienta que consistió en la

adjudicación de 1,5 puntos a la presencia de los aspectos descritos y 0 a su ausencia. De esta manera, ambos instrumentos tuvieron un sistema de puntuación en la misma escala que permitió su posterior comparación.

Por otro lado, los *feedbacks* dados por los evaluadores fueron categorizados en función de la clasificación generada para tal fin por Fernández (2017). En la tabla 1 se muestra dicha categorización.

Tabla 1

Sistema de categorías de feedbacks de la evaluación de textos

Dimensión	Categoría	Definición
Amplitud	1. Global	<i>Feedbacks</i> sobre el conjunto del texto
	2. Local	<i>Feedbacks</i> sobre uno o varios fragmentos del texto
Contenido Evaluativo	1. Señalización	Símbolos o expresiones verbales en los que únicamente se identifica un fragmento del texto, ya sea erróneo o acertado, sin aportar nada más (por ejemplo, se subrayan faltas de ortografía sin corregirlas)
	2. Calificación	Expresiones numéricas, gráficas (flecha hacia arriba, letras que representan valores de una escala ordinal) o verbales, que conllevan únicamente un juicio de calidad. Pueden tener una valencia positiva (como las numéricas superiores al 70% de la escala empleada), intermedia (como la letra R o las numéricas entre 50-70% de la escala) o negativa
	3. Pregunta	Símbolos o expresiones verbales interrogativas de duda o en las que se solicita aclaración
	4. Corrección específica	Expresión gráfica o verbal que, además de señalar un error, proporciona una instrucción específica o una alternativa mejor (incluye la supresión o adición necesaria de un fragmento)
	5. Corrección genérica	Expresión verbal que identifica una carencia o error que se repite (sin señalarlo con precisión en el texto) y proporciona una instrucción genérica de mejora.
	6. Ampliación/ alternativa	Comentario verbal en el que no se identifica un error pero que sugiere, ejemplifica o proporciona directamente una alternativa o ampliación no imprescindible del texto.
	7. Justificación	Comentario verbal en el que se explica o argumenta una alternativa (incluye los comentarios normativos).
	8. Otros comentarios	Otros comentarios verbales, como advertencias o medidas de refuerzo.
Contenido Meta-textual	1. Organizativo	<i>Feedbacks</i> sobre la organización retórica del texto o sus párrafos. En las evaluaciones de textos narrativos cabe distinguir, más específicamente entre marco, tema y trama.
	2. Semántico-léxico	<i>Feedbacks</i> sobre tópicos o ideas del texto. Incluye la evaluación de la coherencia o de vocabulario específico.
	3. Gramatical	<i>Feedbacks</i> sobre la construcción morfosintáctica de las oraciones. Incluye la alteración del orden de palabras, la evaluación de la cohesión o de los signos de puntuación

4. Ortográfico	<i>Feedbacks</i> sobre la construcción ortográfica de las palabras
5. Otros	Comentarios sobre otros aspectos meta-textuales específicos, como los aspectos formales (caligrafía, margen) o la creatividad.

Nota: Adaptado de Fernández (2017).

Una vez recogida la muestra y categorizados los *feedbacks* de los evaluadores se procedió a su análisis mediante a los programas estadísticos SPSS y Excel. De esta manera, se utilizó el primero de ellos para conocer la validez de criterio de los instrumentos estudiados, estableciendo una relación entre las notas globales obtenidas con cada una de las herramientas en los textos, mediante una correlación de Pearson y el coeficiente de regresión. Por otro lado, mediante el programa Excel se llevó a cabo un análisis descriptivo, en el cual se calcularon los porcentajes de utilización de cada categoría de comentarios.

Resultados

Los resultados obtenidos en relación con los análisis realizados para cada uno de los objetivos del estudio son los que se encuentran a continuación.

De esta manera, para conocer la validez de criterio de los instrumentos utilizados, se estableció una relación entre las notas globales obtenidas en cada uno de ellos en los diferentes textos, utilizando para ello una correlación de Pearson y el coeficiente de regresión.

En primer lugar, los resultados de las correlaciones entre la rúbrica y la lista de control se muestran en la tabla 2.

Tabla 2

Resultado de correlaciones de Pearson entre rúbrica y lista de control

		Correlaciones	
		Rúbrica	Lista de control
Rúbrica	Correlación de Pearson	1	,620**
	Sig. (bilateral)		,000
	N	36	36
Lista de Control	Correlación de Pearson	,620**	1
	Sig. (bilateral)	,000	
	N	36	36

Nota: Elaboración propia

** La correlación es significativa en el nivel 0,01 (bilateral).

De acuerdo con los resultados obtenidos se puede concluir que las puntuaciones de los textos corregidos, según los criterios de la rúbrica, correlacionan con los de la lista de control de manera significativa (correlación de Pearson = 0,620**, $p < 0,01$).

En segundo lugar, el análisis de regresión lineal simple para los diferentes instrumentos analizados se muestra en la tabla 3, que se expone a continuación.

Tabla 3

Resultados análisis de varianzas entre rúbrica y lista de control

ANOVA ^a						
Modelo		Suma de cuadrados	de gl	Media cuadrática	F	Sig.
1	Regresión	36,011	1	36,011	21,208	,000 ^b
	Residuo	57,732	34	1,698		
	Total	93,743	35			

Nota: Elaboración propia

a. Variable dependiente: rúbrica

b. Predictores: (Constante), checklist

La tabla ANOVA permite conocer si existe o no relación significativa entre las variables, es decir entre las puntuaciones dadas utilizando la rúbrica y la lista de control. En este caso, dado que el nivel crítico (sig) asociado al estadístico F (21,21) es igual 0,00, es decir $p < 0,01$, puede confirmarse que los instrumentos de evaluación de textos narrativos analizados están linealmente relacionados.

En relación con el segundo de los objetivos de esta investigación, se llevó a cabo un análisis descriptivo, en el cual se calcularon los porcentajes de utilización de cada categoría de comentario y se elaboraron una serie de gráficos.

En la tabla 4 pueden observarse las frecuencias y porcentajes de cada una de las categorías de las tres dimensiones de categorización de los *feedbacks*, proporcionados por los evaluadores en cada uno de los textos.

Tabla 4

Porcentajes y frecuencias de los comentarios evaluativos

Dimensión	Categoría	Texto 1	Texto 2
Amplitud	Global	131 (50,4%)	69 (57,5%)
	Local	129 (49,6%)	51 (42,5%)
Contenido Evaluativo	Señalización	109 (41,9%)	39 (32,5%)
	Calificación	11 (4,2%)	18 (15,0%)
	Específica	80 (30,8%)	35 (29,2%)
	Genérica	34 (13,1%)	10 (8,3%)
	Ampliación/alternativa	11 (4,2%)	6 (5,0%)
	Justificación	10 (3,8%)	0 (0%)
	Otros comentarios	5 (1,9%)	12 (10,0%)
	Organizativo	123 (47,3%)	48 (40,0%)
Contenido meta-textual	Semántico	35 (13,5%)	15 (12,5%)
	Gramatical	39 (15,0%)	21 (17,5%)
	Ortográfico	28 (10,8%)	20 (16,7%)
	Otros	30 (11,5%)	15 (12,5%)
	Inespecífico	5 (1,9%)	1 (0,8%)

Nota: Elaboración propia.

En las siguientes líneas se procede a describir los resultados expuestos en la tabla 4, para cada una de las dimensiones de categorización.

Amplitud de los comentarios

En relación con la amplitud, los resultados son muy similares para ambos textos. De esta manera, en los textos uno y dos los comentarios globales y locales aparecen en un porcentaje muy equitativo para cada categoría y tipo de texto (50 %).

De igual forma, se advirtió que tanto los comentarios globales como los locales estaban en su mayoría referidos a señalizaciones (globales: 50 %; locales: 26,7 %), como por ejemplo “*tienes varias faltas de acentuación*”; y a correcciones genéricas (globales: 16,5 %; locales: 53,3 %), como puede ser “*hay palabras que repites y para solucionarlo puedes buscar sinónimos*”.

Sobre el contenido meta-textual, se muestra relevante cómo el 90 % de las correcciones realizadas de forma local se refieren a la organización con *feedbacks*, como “*en la introducción debes explicar quiénes y como son los protagonistas, y eso no lo has hecho*” o “*empezaría explicando de forma más detallada cómo es el pueblo y cuáles y cómo son los personajes*”. Sin embargo, en la categoría de comentarios globales los porcentajes se reparten con escasa diferencia, a excepción de las categorías “organizativo” e “inespecífico”, que ocupan el 4,5 % y 3 %, respectivamente.

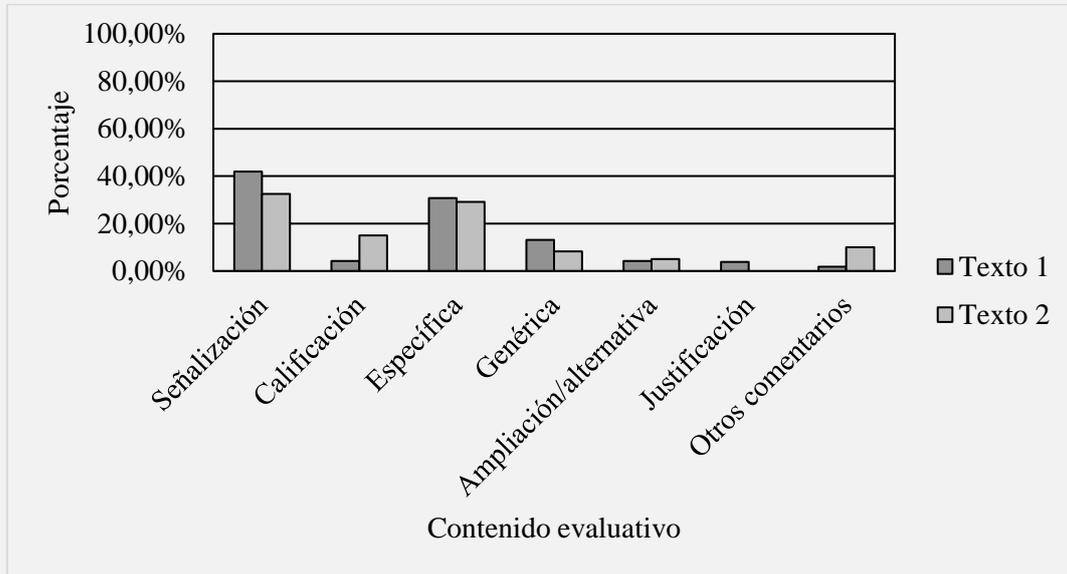
En cuanto a la parte de los criterios de la rúbrica a la que se refieren los comentarios globales y locales se ve una clara diferencia entre ambos. De esta manera, los globales se ocupan principalmente de los criterios “creatividad” (25 %), “oraciones” (30 %), “vocabulario” (23,5 %) y “ortografía” (20 %), y los locales de los criterios “marco” (31,1 %), “tema” (28,3 %) y “trama” (31,7 %).

Contenido evaluativo

En la figura 1 se muestra que todas las categorías adscritas a la dimensión de contenido evaluativo han sido utilizadas en mayor o menor medida. La más usada en la evaluación de ambos textos ha sido la señalización, con un 41,9 % y 32,5 % para los textos uno y dos, respectivamente, seguida de las correcciones específicas que abarcan un 30,8 % y 29,2 % de las retroalimentaciones. En los resultados hallados cabe destacar la baja utilización de justificaciones en el texto 2, y la no utilización de la categoría pregunta.

Figura 1

Contenido evaluativo de los comentarios utilizados



Nota: Elaboración propia.

Los comentarios referidos a señalizaciones tuvieron una menor presencia en el marco y tema, con solo un 4,7 % y 8,8 %, y en los contenidos meta-textuales relacionados con la organización (12,2 %) y los inespecíficos (1,4 %).

En cuanto a las calificaciones, estas fueron referidas en mayor medida a nivel global (62,1 %), como puede ser el ejemplo *“pero en general está bien”*. En lo que se refiere al contenido meta-textual y los criterios de la rúbrica, las calificaciones se reparten de forma aproximadamente igualada en las diferentes categorías, a excepción de la ortografía en ambas dimensiones y de la trama como criterio de la rúbrica, coincidiendo las tres categorías en un porcentaje de 6,9 %.

Las correcciones específicas se centraron especialmente en comentarios locales (83,5 %), lo que podría resultar comprensible y evidente. Asimismo, este tipo de correcciones se encuentra relacionado mayormente con la categoría “organización” en el contenido meta-textual (76,5 %), con comentarios como *“añadiría la apariencia de los protagonistas y sobre todo cómo son sentimentalmente”*. Por otro lado, este tipo de comentarios se concentran en gran medida en el marco (33,9 %), el tema (22,6 %) y la trama (19,1 %).

En relación con las correcciones de carácter genérico, el contenido evaluativo con el que se conectan es el organizativo (27,3 %), aunque se reparten de forma muy igualada en todas las categorías de esta dimensión, excluyendo la inespecífica con solo un 2,3 %. De forma adversa a lo ocurrido en las correcciones específicas, las genéricas se reúnen en los criterios de la rúbrica “creatividad” (15,9 %), “oraciones” (22,7 %), “vocabulario” (20,5 %) y “ortografía” (15,19 %).

Las retroalimentaciones de ampliación o alternativa se dan de manera muy similar a nivel local (52,9 %) y global (47,1 %), con comentarios como “*contaría una historia algo más larga*” y “*falta quizá poner dónde estaba al principio de la historia, dónde está corriendo, aunque se puede sobreentender*”, para cada criterio respectivamente. Este último comentario, puesto como ejemplo, se refiere, además, a aspectos organizativos, categoría con el que se relacionan mayoritariamente y con gran diferencia, suponiendo un 41,2 %. Asimismo, este tipo de comentarios se encuentran referidos en mayor cantidad a la trama y a la creatividad, coincidiendo ambos en un 29,4 %. Un ejemplo de este último podría ser “*para mejorar podría hacer alusión a los hechos que pueden enseñar algo*”.

En lo que se refiere a las justificaciones, los escasos comentarios se reparten de manera muy similar en las categorías global (60 %) y local (40 %). En cuanto al contenido meta-textual, estos *feedbacks* se acumulan en los criterios “organización” y “gramatical”, con un porcentaje del 40 % en ambos. Por otro lado, las justificaciones se refieren especialmente al marco (30 %) y a las oraciones (40 %). Entre los ejemplos de *feedback* justificativo, referido al marco, se encuentran “*para que el lector pueda empatizar más con la historia*”, refiriéndose con anterioridad a que se debería proporcionar más detalles sobre los personajes; y entre los concernientes a las oraciones, “*para que tu cuento tenga más sentido y se pueda entender mejor*”, señalándose de forma preliminar que se ha de dar más conexión a las oraciones.

Finalmente, en esta dimensión de contenido evaluativo, aquellos comentarios que no se pudieron categorizar en los criterios anteriores tuvieron un carácter generalmente global (88,2 %), se refirieron especialmente a la gramática (23,5 %) y a la ortografía (41,2%), y, en consonancia, se realizaron en los apartados de la rúbrica a las oraciones (23,5 %), el vocabulario (17,6 %) y la ortografía (35,3 %). Entre los comentarios de carácter global y ortográfico, y clasificados en “otros comentarios” se pueden encontrar algunos como “*recomendaría que repasara más las faltas de ortografía y así hacer de su escritura algo más bueno y vistoso*” o “*antes de entregar la actividad al profesor haría*

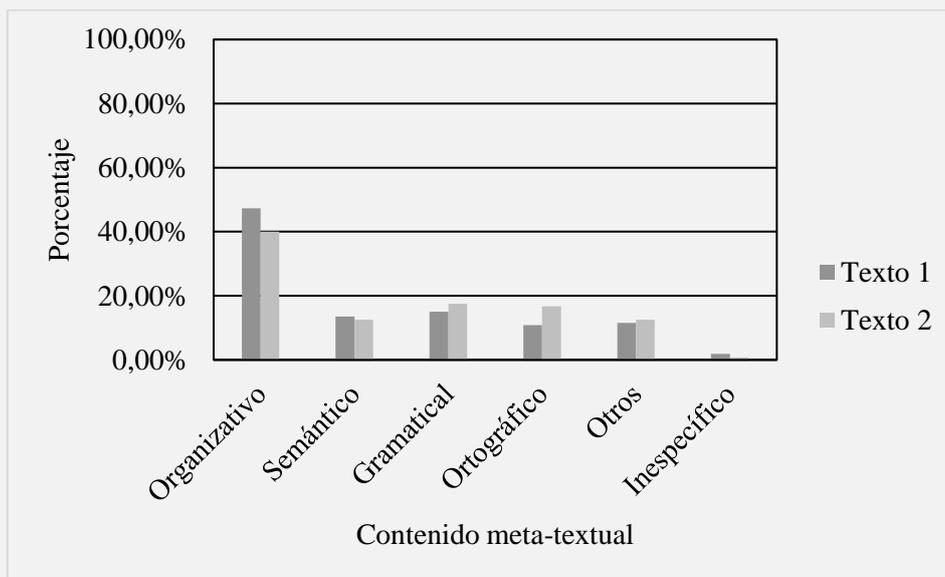
un repaso leyendo la historia y viendo los errores, en este caso". Otro ejemplo sería *"cuidado las reglas de puntuación"*, referido este a la gramática y las oraciones.

Contenido meta-textual

En cuanto al contenido meta-textual, la figura 2 muestra con claridad la preponderancia de los comentarios de tipo organizativo, respecto al resto de categorías en ambos textos analizados. A pesar de ello, se puede afirmar la manifestación de retroalimentaciones en diferentes ámbitos lingüísticos.

Figura 2

Contenido meta-textual de los comentarios utilizados



Nota: Elaboración propia.

Los comentarios de tipo organizativo, tal y cómo ya se ha expuesto anteriormente, se concentran en aquellos referidos a aspectos locales en el texto y a correcciones específicas, seguidas de las señalizaciones (27,5 %). Estos *feedbacks* se presentan con práctica totalidad en las partes de los textos, referidas al marco (32,7 %), el tema (29,2 %) y la trama (33,3 %).

En relación con las retroalimentaciones dadas en aspectos semántico-léxicos del texto, cabe destacar el predominio de comentarios de su carácter global (94 %). Dichos

feedbacks se dan mediante señalizaciones (36 %), correcciones específicas (22 %) y correcciones genéricas (20 %), principalmente.

A nivel gramatical, las correcciones se llevan a cabo fundamentalmente a la globalidad del texto (88,3 %), y prácticamente la mitad de las mismas se realizan mediante señalizaciones (51,7 %); algunos ejemplos de ellos son: “*faltan varias comas y algún que otro punto seguido*” o “*hay algunas partes de algunas oraciones que no tienen sentido*”.

En cuanto a la ortografía, los resultados son muy similares a los hallados en los aspectos gramaticales; las correcciones se ejecutan a nivel global (87,5 %) y a través de señalizaciones (56,3 %). Un ejemplo de ello sería “*tiene varias faltas de acentuación*” o “*si nos fijamos bien, apreciamos prácticamente una falta de ortografía por oración*”.

En lo que se refiere a los comentarios que no pueden ser clasificados en los criterios anteriores y a los cuales se denominan como “otros”, estos se concentran prácticamente en su totalidad en los aspectos creativos (95,6 %), y se realizan a la globalidad del escrito en su mayoría (95,6 %), mediante señalizaciones en prácticamente la mitad de los casos (51,1 %). Un comentario registrado con estas categorías sería “*cuenta una historia interesante*” o “*la historia es original, pero al no estar bien narrada no se llega a entender fácilmente, lo que la hace aburrida*”.

Por último, los comentarios determinados como “inespecíficos”, los que no reseñan un contenido meta-textual concreto, son muy escasos (texto uno: 1,9 %; texto dos: 0,8 %). Todos ellos estuvieron referidos al escrito en su globalidad y repartidos entre las partes que aluden a la trama (16,7 %), el vocabulario (16,7 %), la creatividad (33,3 %) y las oraciones (33,3 %) en el texto.

Discusión

Múltiples son los autores que señalan las habilidades y destrezas en lectura y escritura como fundamentales en el desarrollo eficaz de la vida académica, personal y profesional (Romero et al., 2002; Uribe-Enciso y Carrillo-García, 2014), poniéndose de manifiesto la importancia de una adecuada adquisición de dichas competencias, con el objetivo de formar personas que se desenvuelvan en todos los ámbitos de su vida de forma competente.

La escuela será el principal agente de formación en relación con estas destrezas, y será en los años de la educación primaria en la que se asienten las bases para un desarrollo adecuado de ellas. Así, será esencial que los docentes conozcan cómo guiar este aprendizaje de forma eficaz.

Sin embargo, una de las problemáticas principales en relación con la enseñanza que proporcionan los docentes, respecto a las habilidades escritoras, es que su capacidad para evaluar textos dependerá en gran medida del conocimiento profundo y exhaustivo que ellos mismos tengan de la lengua (Parr & Timperley, 2008), y en torno al cual no suelen recibir una formación específica (Schleppegrell, 2004). Esto provoca que la retroalimentación dada por los docentes se centre en aspectos superfluos de la escritura, no pudiéndose provocar una mejora significativa en estas destrezas en el alumnado.

De acuerdo con esto y debido a la suma importancia que tiene la escritura en los diversos aspectos vitales, se requiere de métodos y estrategias de evaluación de estas habilidades que permitan conocer en qué medida el alumnado tiene desarrolladas las destrezas escritoras y cómo potenciar su perfeccionamiento. Para ello, se considera que el uso de instrumentos y herramientas de evaluación que guíen y orienten el proceso favorecerá dicho conocimiento por parte del docente y le permitirá proveer al alumnado con *feedbacks* que le faciliten alcanzar una mejora.

No obstante, es necesario considerar el estudio de las herramientas de evaluación propuestas y utilizadas con el fin de conocer su validez y eficacia para el fin que pretenden cumplir.

De esta manera, en el presente estudio se procedió al análisis de dos herramientas seleccionadas: la rúbrica para textos narrativos de Montanero et al. (2014), y una lista de control, creada a partir de dicha rúbrica. En la investigación se juzgó si ambos instrumentos tenían la capacidad de medir los textos propuestos sin existir grandes discrepancias entre las calificaciones obtenidas. De igual forma, se examinaron los mensajes evaluativos que docentes en formación producían al emplear una de las herramientas: la rúbrica.

En cuanto a la correlación entre el sumatorio de las puntuaciones obtenidas al corregir con la rúbrica y en la lista de control, el coeficiente de correlación de Pearson y el análisis de regresión lineal ratificaron una relación significativa entre las puntuaciones otorgadas a los textos uno y dos, utilizando ambas herramientas. De esta manera, la rúbrica para texto narrativos de Fernández, Montanero et al. (2019) demuestra su

capacidad de evaluar textos de este género con la misma validez que el PROESC (Cuetos, Ramos & Ruano, 2002, como se citó en Fernández, 2017) y la lista de control usada en esta investigación.

Respecto a los comentarios evaluativos generados empleando la rúbrica, los análisis descriptivos permitieron obtener las conclusiones que se exponen a continuación.

En primer lugar, mientras que la retroalimentación proporcionada a los estudiantes en sus textos narrativos sin el uso de ningún instrumento es fundamentalmente a nivel local (Fernández, 2017), el empleo de la rúbrica generó, en una proporción de aproximadamente el 50 %, comentarios referidos tanto al conjunto del texto (globales) como a uno o varios fragmentos (locales).

En segundo lugar, en relación con el contenido evaluativo, el uso del instrumento de evaluación examinado en la investigación plasmó la utilización de los diferentes criterios considerados, al igual que ocurría en el caso de la corrección no guiada, a excepción de la pregunta, la cual no fue usada en ningún momento para proporcionar retroalimentación al emplear la herramienta de evaluación. De la misma manera, en el presente estudio se puso de relieve la preponderancia de las señalizaciones (38,9 %) y las correcciones específicas (30,3 %), mientras que en el estudio previo existía un predominio de éstas últimas con un porcentaje del 88,5 % (Fernández, 2017).

Finalmente, en cuanto al contenido meta-textual y como ya se adelantaba, la evaluación espontánea registraba un análisis de los textos centrados fundamentalmente en aspectos ortográficos y gramaticales, con unos porcentajes del 52 % y el 28,8 % (Fernández, 2017). El uso de la rúbrica, sin embargo, originó comentarios evaluativos principalmente en los aspectos organizativos (45 %), quedando el resto de las categorías repartidas con porcentajes similares (semántico, 13,2 %; gramatical, 15,8 %; ortográfico, 12,6 % y otros, 11,8 %), a excepción de los aspectos inespecíficos que solo ocuparon el 1,6 %. De esta manera, puede concluirse que el uso de la rúbrica provoca una mayor variedad en las correcciones en el contenido meta-textual, procediéndose a un análisis del texto más profundo. Fernández, Lucero et al. (2016) señalan como aspecto habitual y lógico la corrección principalmente de aspectos gramaticales y ortográficos por parte de los docentes, dada la propia naturaleza del desarrollo en el aprendizaje de la escritura, y existiendo en consecuencia una gran cantidad de errores de este tipo en las redacciones del alumnado de Educación Primaria que llaman la atención del examinador. Sin embargo, y como se advertía con anterioridad, la redacción de un texto conlleva otros

aspectos más profundos como la organización y planificación que no pueden pasar desapercibidos en su corrección para una enseñanza eficaz y una mejora de las habilidades de escritura.

En consecuencia con lo expuesto hasta el momento, la presente investigación pone de manifiesto los comentarios evaluativos que los docentes proporcionan al evaluar los textos con una herramienta específica para ello: la rúbrica, frente a otros estudios realizados que categorizaron mensajes evaluativos de docentes en la corrección de textos de forma espontánea, sin uso de ningún instrumento de evaluación. Ejemplo de ellos son los llevado a cabo por Fernández (2017), o el expuesto por Otnes & Solheim (2019). Asimismo, a diferencia de las investigaciones señaladas, las cuales asignaron la evaluación de multitud de textos diferentes a distintos docentes, este estudio propuso la evaluación de dos textos narrativos y específicos a múltiples docentes en formación, pudiendo examinar y analizar los comentarios expuestos y las calificaciones otorgadas por diferentes evaluadores a un mismo texto.

Por último, a estos resultados subyacen una serie de implicaciones educativas prácticas como son la posibilidad de que los docentes utilicen cualquiera de las herramientas examinadas, rúbrica y lista de control, y generar un juicio justo y sin diferencias significativas; así como los beneficios que comporta el uso de la rúbrica a la hora de generar mensajes de retroalimentación específicos, referidos a diversos aspectos y características de la lengua y en diferentes formatos, los cuales favorezcan una mejora real de la capacidad escritora del alumnado.

Conclusiones

Este estudio explora el carácter de la relación entre el sumatorio de puntuaciones obtenidas con la lista de control y la rúbrica, a partir de la evaluación de diferentes textos. Así, se quería conocer si ambos instrumentos proporcionaban calificaciones globales semejantes al ser empleadas, siendo conscientes de que los criterios a considerar eran los mismos, difiriendo en su escala. De igual forma, se deseaba analizar las características de los comentarios evaluativos del profesorado en formación al corregir narraciones mediante la rúbrica.

Los resultados obtenidos concluyeron que la correlación entre el sumatorio de las puntuaciones otorgadas con la rúbrica y con la lista de control a los textos era elevada,

demostrando que la rúbrica y la lista de control, usadas en esta investigación, pueden ser empleadas para la evaluación de textos narrativos, sin perjuicio de hallar grandes controversias en las puntuaciones obtenidas para ellos y disminuyendo de forma indirecta la subjetividad y los aspectos que puedan causar dichas diferencias.

Asimismo, se pudo confirmar que los docentes en formación proporcionan comentarios evaluativos empleando la rúbrica con contenido meta-textual variado y completo, más allá de los aspectos gramaticales y ortográficos, y utilizando múltiples fórmulas para ello. A este respecto, se encontró que la retroalimentación proporcionada a los estudiantes en sus textos narrativos estaba referida tanto al conjunto del texto (globales) como a uno o varios fragmentos (locales), en una proporción de aproximadamente el 50 %; que los comentarios originados se referían principalmente a aspectos organizativos, quedando el resto de categorías repartidas con porcentajes similares, a excepción de los aspectos inespecíficos que solo ocuparon el 1,6 %; y que estos *feedbacks* se producían especialmente mediante señalizaciones y las correcciones específicas.

En cuanto a las limitaciones, la primera de ella está relacionada con la muestra, pues es posible que su selección pueda esconder sesgos al ser llevada a cabo mediante un procedimiento no probabilístico; dicho aspecto influiría de manera decisiva en la validez interna y externa de la investigación desarrollada. Así, se considera que una futura réplica de esta investigación sería conveniente, llevando a cabo procesos probabilísticos de selección de la muestra y equiparando la misma en relación con la evaluación de los textos evaluados y al empleo de cada uno de los instrumentos para ello.

En cuanto a este último aspecto, los escritos a analizar fueron únicamente dos y del mismo curso, lo cual podría afectar a la posibilidad de generalización de los resultados. De esta manera, se recomendaría el uso de una variedad más extensa de los mismos.

Finalmente, cabe destacar que el uso de la rúbrica y la lista de control, y el manejo de ambas pudo estar condicionado por la falta de entrenamiento previo de los evaluadores.

Por otro lado, esta investigación deja abierta algunas líneas de investigación que resultaría interesante estudiar en el futuro. De esta forma, dado que se han analizado los *feedbacks* únicamente proporcionados con la guía y orientación de la rúbrica, un estudio futuro podría examinar aquellos que se generan con la utilización de la lista de control, y llevar a cabo un examen comparativo entre ambos instrumentos y dichos comentarios

evaluativos. En este estudio no fue posible debido a la utilización de ambas herramientas por cada uno de los evaluadores, pudiendo haber sido condicionados los mensajes, dados en la lista de control, por el uso previo de la rúbrica, dado que esta última resulta mucho más detallada y específica.

Por último, sería interesante llevar a cabo este estudio con docentes en activo en un contexto natural de aula, y cotejar los resultados con los obtenidos por docentes en formación.

Referencias

- Applebee, A. (1984). Writing and reasoning. *Review of Educational Research*, 54(4), 577–596. <https://www.jstor.org/stable/1170176?seq=1>
- Bazerman, C. (2013). Understanding the lifelong journey of writing development. *Infancia y Aprendizaje*, 36(4), 421-441. <https://doi.org/10.1174/021037013808200320>
- Contreras, L. A., González, M., y Urías, E. (2009). Evaluación de la escritura mediante rúbrica en la educación primaria en México. *Interamerican Journal of Psychology*, 43(3), 518-531. <https://www.redalyc.org/pdf/284/28412903011.pdf>
- Fang, Z., & Wang, Z. (2011). Beyond rubrics: Using functional language analysis to evaluate student writing. *Australian Journal of Language and Literacy*, 34(2), 147-165. <https://search.informit.org/doi/10.3316/ielapa.112929734219836>
- Fernández, M. J. (2017). *Co-evaluación de textos narrativos en la Educación Obligatoria* (Tesis de doctorado). Universidad de Extremadura.
- Fernández, M. J., Lucero, M., y Montanero, M. (2016). Rojo sobre negro. ¿Cómo evalúan los maestros las redacciones de sus estudiantes? *Revista de Educación*, (372), 63-82. <https://dialnet-unirioja.es/ezproxy.nebrija.es/servlet/articulo?codigo=5399659>
- Fernández, M. J., Montanero, M., y Lucero, M. (2019). La evaluación de la competencia narrativa en la educación básica. The assessment of narrative competence during compulsory schooling. *Revista de Educación*, (383), 85-112. 4. <https://www.doi.org/10.4438/1988-592X-RE-2019-383-402>

- Gallego, T. M., y Muñoz, M. T. (enero-abril, 2020). Construcción de la lengua escrita en preescolar, a través de la variedad textual, en una zona rural de Antioquia. *Revista Virtual Universidad Católica del Norte*, (59), 111-132. <https://doi.org/10.35575/rvucn.n59a7>
- Ghaffar, M. A., Khairallah, M., & Salloum, S. (2020). Co-constructed rubrics and assessment for learning: The impact on middle school students' attitudes and writing skills. *Assessing Writing*, 45, 1-12. <https://doi.org/10.1016/j.asw.2020.100468>
- Graham, S. (2006). Writing. In P. Alexander y P. Winne (Eds.), *Handbook of educational psychology* (pp. 457–478). Erlbaum.
- Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing*, 26(1), 1-15. <https://doi.org/10.1007/s11145-012-9395-2>
- Graham, S., & Hebert, M. (2010). *Writing to Read*. Alliance for Excellent Education.
- Graham, S., & Hebert, M. (2011). Writing-to-read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, 81(4), 710–744. <https://doi.org/10.17763/haer.81.4.t2k0m13756113566>
- Heldsinger, S., & Humphry, S. (2010). Using the method of pairwise comparison to obtain reliable teacher assessments. *The Australian Educational Researcher*, 37(2), 1-19. <https://doi.org/10.1007/BF03216919>
- Hyland, K. (2003). *Second Language Writing*. Cambridge University Press.
- Jølle, L. (2014). Pair assessment of pupil writing: A dialogic approach for studying the development of rater competence. *Assessing Writing*, 20, 37–52. <https://doi.org/10.1016/j.asw.2014.01.002>
- Kayapinar, U. (2014). Measuring essay assessment: Intra-rater and inter-rater reliability. *Eurasian Journal of Educational Research*, 14(57), 113-136. <https://doi.org/10.14689/ejer.2014.57.2>
- Ministry of Education of Ontario. (2006). *A guide to effective literacy instruction. Volume two. Assessment*. http://www.eworkshop.on.ca/edu/resources/guides/Guide_Lit_456_Vol_2_Assessment.pdf
- Montanero, M. (2019). *Didáctica General*. Universidad de Extremadura.

- Montanero, M., Lucero, M., y Fernández, M. J. (2014). Iterative co-evaluation with a rubric of narrative texts in primary education. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, 37, 184–220. <https://doi.org/10.1080/02103702.2014.881653>
- OCDE. (2018). *Programa para la evaluación internacional de estudiantes (PISA)*. https://www.oecd.org/pisa/publications/PISA2018_CN_esp_ESP.pdf
- Otnes, H., & Solheim, R. (2019). Acts of responding. Teachers' written comments and students' text revisions. *Assessment in Education: Principles, Policy & Practice*, 26(6), 700-720. <https://doi.org/10.1080/0969594X.2019.1595524>
- Parr, J., & Timperley, H. (2008). Teachers, schools and using evidence: Considerations of preparedness. *Assessment in Education: Principles, Policy and Practice*, 15(1), 57-71. <https://doi.org/10.1080/09695940701876151>
- Pérez, P. (2008). La comprensión lectora y la competencia en comunicación lingüística en el nuevo marco curricular: algunas claves para su desarrollo. *Participación Educativa*, 8, 41-56. <https://revistas-um-es.ezproxy.nebrija.es/educatio/article/view/71081>
- Popham, W. J. (1990). *Modern Educational Measurement: A Practitioner's Perspective* (2nd Ed.). Prentice-Hall.
- Romero, M. A., Edna, L., y Silva, A. (2002). Evaluación de las aptitudes para el aprendizaje escolar. En L. Edna y A. Silva, *Evaluación psicológicas en el área educativa* (pp. 37-81). Pax México.
- Sandiford, C., & Macken-Horarik, M. (2020). Changing stories: Linguistically-informed assessment of development in narrative writing. *Assessing Writing*, 45, 1-15. <https://doi.org/10.1016/j.asw.2020.100471>
- Schleppegrell, M. (2004). *The language of schooling: A functional linguistics perspective*. Erlbaum.
- Scull, J., Mackenzie, N. M., & Bowles, T. (2020). Assessing early writing: a six-factor model to inform assessment and teaching. *Educational Research for Policy and Practice*, 1-21. <https://doi.org/10.1007/s10671-020-09257-7>
- Skar, G. B., & Jølle, L. J. (2017). Teachers as raters: Investigation of a long term writing assessment program. *L1 Educational Studies in Language and Literature*, 17, 1-30. <https://doi.org/10.17239/L1ESLL-2017.17.01.06>

- Smyth, J. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. *Journal of Consulting and Clinical Psychology*, 66(1), 174–184. <https://doi.org/10.1037/0022-006X.66.1.174>
- Uribe-Enciso, O. L., y Carrillo-García, S. (2014). Relación entre la lecto-escritura, el desempeño académico y la deserción estudiantil. *Entramado*, 10(2), 272-285. <https://revistas.unilibre.edu.co/index.php/entramado/article/view/3509>
- Uzun, N. B., Alici, D., & Aktas, M. (2019). Reliability of the Analytic Rubric and Checklist for the Assessment of Story Writing Skills: G and Decision Study in Generalizability Theory. *European Journal of Educational Research*, 8(1), 169-180. <https://doi.org/10.12973/eu-jer.8.1.169>