

# revista de **e**EDUCACIÓN

Nº 383 ENERO-MARZO 2019



La evaluación de la competencia narrativa en la educación básica

The assessment of narrative competence during compulsory schooling

María-Jesús Fernández  
Manuel Montanero  
Manuel Lucero



# La evaluación de la competencia narrativa en la educación básica<sup>1</sup>

## The assessment of narrative competence during compulsory schooling

DOI: 10.4438/1988-592X-RE-2019-383-402

María-Jesús Fernández

Manuel Montanero

Manuel Lucero

*Universidad de Extremadura*

### Resumen

La redacción de relatos es una actividad fundamental para el desarrollo de la competencia escritora en los primeros años de escolaridad. Sin embargo, su evaluación en el contexto del aula se centra frecuentemente en aspectos superficiales, de carácter gramatical y ortográfico. Es notoria la escasez de instrumentos funcionales, con suficientes garantías de validez y fiabilidad, que faciliten la evaluación de aspectos semánticos y retóricos en la práctica del aula. En este sentido, el objetivo de esta investigación fue construir y validar una rúbrica para evaluar relatos (RER) redactados por estudiantes de entre 9 y 14 años. Para ello, se analizaron un total de 292 textos producidos por estudiantes escolarizados en estos niveles educativos. Los datos aportados apoyan la validez de la RER, lo que sugiere que tanto la estructura dimensional como la operativización de los criterios y niveles de logro son relevantes y suficientemente discriminativos para su aplicación en estos niveles educativos. También se obtuvieron datos positivos de la fiabilidad del instrumento, desde el punto de vista de su consistencia interna y de la objetividad de su aplicación por diversos

---

<sup>(1)</sup> Agradecimientos: A la Consejería de Economía, Competitividad e Innovación del Gobierno de Extremadura, y al Fondo Europeo de Desarrollo Regional por el apoyo financiero recibido (Referencia: GRU15153)

evaluadores. Los resultados del estudio-piloto muestran, sin embargo, pequeñas inconsistencias en algunos niveles de logro, como el nivel máximo del criterio “trama”. Esto se debe a la escasez de estudiantes de Secundaria que consiguen redactar tramas complejas, bien articuladas y con elementos realmente creativos. Finalmente, se discuten estos resultados y sus implicaciones para la evaluación y mejora de la competencia narrativa.

*Palabras-clave:* Escritura; Competencia narrativa; Relatos; Rúbrica; Educación Primaria; Educación Secundaria.

### **Abstract**

The writing of narrative texts is considered an important activity for the development of the writing skills in the first years of schooling. However, its assessment in the classroom context is frequently based on more superficial characteristics, such as grammar and spelling. There is a notable shortage of functional and standardized writing assessment tools, offering sufficient guarantees of validity and reliability, in order to facilitate the assessment of semantic and rhetorical aspects in classroom instruction. In this sense, the present research has two core aims: (1) to develop and validate a precise and easily interpretable rubric to assess narrative compositions written by students aged between 8 and 14; and (2) to provide more information about the evolution of narrative competence throughout the compulsory education. To this effect, 292 short stories produced by students from Elementary and Secondary education were analyzed. Related to the first objective, our findings support the content-related validity of the rubric. It suggests that the rubric’s constructs or dimensions developed, the assessment categories proposed, and the levels of execution defined within each category are relevant and sufficiently discriminative to be applied in these educational stages. Also, the reliability analysis revealed positive outcomes, as much from the point of view of its internal consistency as for its inter-rater reliability, in terms of how different instructors or raters award similar scores. Regarding the second objective, the low progression of rhetorical competence during the first years of Secondary Education should be noted, as can be seen in the shortage of students who are able to write complex and well-articulated plots, with truly creative elements. Finally, results and implications for the assessment and improvement of narrative competence are discussed.

*Key words:* Writing; Narrative competence; Short stories; Rubric; Primary Education; Secondary Education.

## Introducción

La composición de relatos es una actividad educativa fundamental para la adquisición de la competencia escritora en los primeros años de escolaridad. A partir de los 5 años los niños son ya capaces de construir verbalmente secuencias temporales sencillas, contextualizadas en marcos espacio-temporales, y que, poco a poco, van incorporando relaciones causales y motivacionales. A partir de los 7 años la mayoría de los niños son progresivamente capaces de componer por escrito dichos relatos. Los personajes se introducen por el nombre o desde la perspectiva del narrador, pero apenas se describen (Kernan, 1977); las acciones se articulan en episodios y los relatos adquieren una estructura cada vez más coherente, hasta el punto que el oyente puede predecir un final lógico durante la narración (Stadler y Ward, 2005). Las secuencias no muestran ya sólo relaciones causales entre ciertas acciones y las reacciones de los personajes, sino que se vertebran con cierta claridad en torno a un problema, una solución y en un elemento de mediación entre ambos (Iandolo, Esposito & Venuti, 2012). Entre los 8 y los 9 años se constata una mejora significativa en la cohesión y la complejidad de la estructura de la historia, que incorpora más elementos problemáticos y de mediación que retroalimentan el propio problema (Stein & Glenn, 1979). En torno a los 10 u 11 años las composiciones narrativas pueden considerarse ya relativamente maduras (Artiles & Jiménez, 2007; Yan *et al.*, 2012), con una estructura jerarquizada de episodios (primarios y secundarios), articulados en torno a un suceso inicial o un tema central (Botvin & Sutton-Smith, 1977; Ochoa-Angrino *et al.*, 2008).

Toda esta progresión se encuentra, no obstante, mediatizada por el nivel cognitivo, el dominio de la lengua y las emociones del individuo (Ryan, 2014). Así, por ejemplo, las redacciones de los niños que carecen de suficientes habilidades motoras relacionadas con la destreza escritora o que comenten muchas faltas de ortografía, manifiestan una peor organización del texto o una textualización más pobre (Berninger *et al.*, 1992; Cremin & Myhill, 2012). Cuando estos procesos se automatizan, las composiciones mejoran también en otras facetas (Graham, Harris & Fink-Chorzempa, 2002).

Desde hace décadas contamos, en definitiva, con un conocimiento relativamente amplio de la evolución de esta competencia narrativa de los niños. Sin embargo, el profesorado presta muy poca atención al

contenido semántico y a la estructura retórica de las narraciones cuando evalúa relatos (Fernández, Lucero & Montanero, 2016). La frecuencia de correcciones gramaticales y ortográficas que registran en los manuscritos de los alumnos constituye, por el contrario, el mejor predictor de la calificación global que los profesores otorgan a los textos (Lucero, Fernández & Montanero, 2018). Esta carencia podría explicar por qué los niños de Primaria mejoran significativamente estos últimos aspectos de sus borradores, pero no el contenido y la estructura retórica, cuando reciben la evaluación de sus profesores (Montanero, Fernández & Lucero, 2014).

A pesar de ser el género más frecuente en las tareas escolares de lectura y escritura, resulta también llamativo que casi ninguna de las pruebas estandarizadas más extendidas para evaluar el aprendizaje de la lectoescritura en la Educación Primaria examine específicamente la competencia retórica de los niños en tareas de composición de narraciones (véase Galve, 2005; García & Marbán 2001; Toro & Cervera, 1984; Toro, Cervera & Urío, 2002). En el contexto español el test de evaluación de procesos de escritura, PROESC (Cuetos, Ramos & Ruano, 2002) es prácticamente la única prueba estandarizada que permite evaluar específicamente dicha competencia. Su utilización en la práctica del aula es, no obstante, muy escasa, debido probablemente a su coste.

Cabe pensar que si el profesorado contara con instrumentos de evaluación fáciles de aplicar en la práctica del aula mejoraría la calidad de la retroalimentación que proporcionan a los estudiantes. En este sentido, se ha propuesto la utilización de instrumentos más sencillos, basados en escalas o rúbricas (Sundeen, 2014). Las rúbricas son un tipo particular de escalas de apreciación ordinal, en formato descriptivo, cuyos criterios aparecen operativizados en varios niveles de logro. Esta característica facilita que cada alumno conozca su situación con respecto a los objetivos que debe alcanzar en la realización de una tarea concreta (Gallavan & Kottler, 2009).

Así, por ejemplo, Benítez y García (2010) propusieron una escala de apreciación con 6 valores por criterio (desde insatisfactorio hasta excelente) que permite la evaluación de elementos retóricos de los relatos, tales como la identificación de los personajes o la construcción de la trama.

Yan *et al.* (2012) diseñaron una rúbrica de evaluación de la escritura, compuesta por 7 criterios divididos en 2 dimensiones (contenido y

organización), para alumnos de edades comprendidas entre 6 y 9 años. El instrumento valoraba la relevancia, la amplitud, el uso de conectores, la coherencia, la estructura del texto y la inteligibilidad del mismo.

Dunsmuir *et al.* (2015) propusieron otra rúbrica que permitía evaluar la composición de textos de niños con edades comprendidas entre 7 y 11 años; también contaba con 7 criterios y 4 niveles. Los aspectos que se tuvieron en cuenta, en este caso, eran la caligrafía, la ortografía, los signos de puntuación, la estructura textual, el vocabulario, la organización y la originalidad.

Se ha comprobado que este último tipo de instrumentos permiten aclarar a los estudiantes aquello que se tendrá en cuenta en la evaluación y facilitan más información que la corrección de redacciones que los profesores realizan habitualmente (Ferris & Hedgcock, 2014; Li & Lindsey, 2015). La necesidad de seleccionar uno de esos niveles en cada criterio de la rúbrica desencadena una interacción más rica en situaciones de co-evaluación, que tiene una influencia significativa en la calidad de composiciones narrativas (Montanero *et al.*, 2014; Ochoa-Angrino *et al.*, 2008), la autorregulación del aprendizaje y el sentimiento de autoeficacia (Panadero & Jonsson, 2013).

No obstante, otros autores han puesto también de manifiesto diversas debilidades, relacionadas, por un lado, con su validez y, por otro lado, con sus limitaciones de cara a la evaluación formativa de competencias (Spandel, 2006). En cuanto a la primera cuestión, Jonsson & Svingby (2007) hicieron una revisión sobre la validez y la fiabilidad de las rúbricas utilizadas en diversos estudios. Únicamente un tercio de las investigaciones tenidas en cuenta por los autores informaban de la validez de los instrumentos. Los indicadores de validez se basan principalmente en juicios de expertos, el análisis factorial confirmatorio (Baker *et al.*, 1995), o en correlaciones (superiores a 0,40) con otras medidas que se toman como criterio (Roblyer & Wiencke, 2003). En relación a la fiabilidad, las rúbricas revisadas por Jonsson & Svingby obtuvieron valores comprendidos entre 0,50 y 0,92 en la prueba Alfa de Cronbach; un grado de concordancia inter-jueces de entre el 55 y el 75%; así como índices kappa de Cohen entre 0,40 y 0,75. Tan solo uno de los trabajos revisados (Gearhart *et al.*, 1995) estaba centrado en analizar la validez de una rúbrica de evaluación de relatos.

En cuanto a la segunda cuestión, la utilización de la rúbrica en situaciones de evaluación formativa de la competencia escritora, algunos

autores han objetado el riesgo de que el instrumento encasille en cierta forma el supuesto desempeño escritor del alumnado (Fang & Wang, 2011). Diversos autores argumentan que, si bien facilita una evaluación rápida por parte del profesorado, no contribuye realmente a mejorar la capacidad de redacción del alumnado, debido a que los niveles de logro conllevan sugerencias genéricas (no personalizadas) y a menudo están formulados en un lenguaje difícil de comprender por el alumnado (Janssen *et al.*, 2015; Kohn, 2006; Panadero, Alonso-Tapia & Huertas, 2012; Schenck & Daly, 2012). Finalmente, se ha señalado la insuficiente formación de una parte del profesorado que la utiliza en el contexto del aula (East, 2009; Knoch, Read & Von Randow, 2007).

En definitiva, parece necesario generar más conocimiento sobre la validez y utilidad de este tipo de instrumentos en la práctica del aula, en particular como apoyo a la evaluación de la competencia escritora del alumnado de Educación Primaria y Secundaria. En este marco, el objetivo de este trabajo era diseñar un instrumento de evaluación de la escritura narrativa, que resulte útil y fácil de aplicar por el profesorado en la práctica del aula. En concreto, nos propusimos construir y validar una rúbrica “analítica” que permita la evaluación formativa, de manera rápida y sencilla, de textos narrativos redactados por estudiantes con edades comprendidas entre 9 y 14 años. En este sentido, es necesario aportar más evidencias de validez y fiabilidad de las que se han publicado en relación a las rúbricas anteriormente referenciadas. Analizaremos también las diferencias individuales obtenidas en la aplicación de la rúbrica a una muestra de estudiantes de diversas edades. De este modo, podemos comprobar si la rúbrica permite discriminar diferentes grados de competencia narrativa en las composiciones que redactan los estudiantes en función de su nivel educativo.

## Método

### Muestra

En el estudio participaron un total de 292 alumnos de 3º de Educación Primaria a 2º de Educación Secundaria Obligatoria (9-14 años). La muestra fue obtenida de 7 centros educativos de diferentes ciudades extremeñas, seleccionados por conveniencia en zonas de nivel socioeconómico medio

o medio-bajo. No participó el alumnado inmigrante con insuficiente conocimiento del español, ni los que presentaban necesidades educativas especiales. La tabla I refleja a distribución final de la muestra, en cuanto a edades, género y ciclo o nivel de escolarización.

**TABLA I.** Distribución de la muestra (N) según el nivel de escolarización (EP: Educación Primaria; ESO: Educación Secundaria Obligatoria), la edad (en años) y el sexo (M: masculino; F: femenino)

Nivel	Curso	Edad	Sexo	N		
I	3° EP	8-9	M	14	40	92
			F	26		
	4° EP	9-10	M	17	52	
			F	35		
II	5° EP	10-11	M	18	49	102
			F	31		
	6° EP	11-12	M	25	53	
			F	28		
III	1° ESO	12-13	M	11	25	98
			F	14		
	2° ESO	13-14	M	30	73	
			F	49		

Fuente: Elaboración propia

## Instrumentos

### Hoja para la redacción

A cada estudiante se le proporcionó una hoja de papel en la que debían redactar un breve relato. En la parte superior se indicaba la temática (“un niño que se convierte en el héroe de su pueblo”) y la extensión (aproximadamente una cara de un folio).



## Rúbrica de evaluación de relatos (RER)

Se adaptó un instrumento de evaluación que había sido utilizado ya en investigaciones anteriores (Montanero, Lucero & Fernández, 2014), compuesto por 7 criterios de evaluación, distribuidos en dos grandes dimensiones. Por un lado, dichos criterios valoraban aspectos relativos a la organización y contenido de la narración: Marco, Tema, Trama y Creatividad. Por otro lado, se evaluaban aspectos gramaticales, léxicos y ortográficos: Oraciones, Vocabulario y Ortografía (Tabla II). Cada uno de estos criterios cuenta con 4 niveles de logro o ejecución, a los que podía asignarse las siguientes calificaciones numéricas: 0 puntos (nivel 1), 0,5 puntos (nivel 2), 1 punto (nivel 3) y 1,5 puntos (nivel 4).

## PROESC

El test de Procesos de Escritura (Cuetos, Ramos & Ruano, 2002) es una prueba estandarizada, comercializada por TEA, muy extendida en la evaluación de la escritura narrativa en la educación obligatoria. Se compone de 10 criterios de evaluación agrupados en dos grandes dimensiones: Contenido y Coherencia. La primera dimensión se concreta en los siguientes criterios: Dónde y cuándo, Personajes, Suceso con consecuencias, Desenlace coherente y Creatividad. La segunda dimensión se evalúa en función de los siguientes criterios: Continuidad lógica, Sentido unitario, Figuras literarias, Oraciones complejas y Vocabulario. Para alcanzar cada uno de estos criterios se definen varios requisitos.

De acuerdo con el manual, el instrumento tiene una consistencia interna de 0,82 (coeficiente Alfa de Cronbach). Como evidencia de validez de criterio, los resultados de la aplicación de la prueba a una muestra de 839 textos arrojaron una correlación de 0,46 con las calificaciones otorgadas por un grupo de profesores. En relación a la validez estructural, los 10 criterios de los que consta la prueba fueron sometidos a un análisis de componentes con rotación varimax para obtener como resultado la división de la prueba en 3 componentes que explicaban un 58% de la varianza total.

## Procedimiento

### Diseño de la rúbrica de evaluación de relatos (RER)

La rúbrica, objeto de este estudio, consistió en una escala de apreciación descriptiva-ordinal (con 4 niveles por criterio, que llevaban, además asignados puntuaciones numéricas), destinada a la evaluación de relatos redactados en L1 por estudiantes de entre 9 y 14 años. Aunque inicialmente el instrumento iba destinado al alumnado escolarizado en la educación obligatoria, tras un estudio preliminar (Fernández, 2017) se concluyó que no era suficientemente válido y discriminativo para los estudiantes escolarizados en los primeros cursos de Primaria (6-8 años), ni en los últimos de Secundaria (15-16 años).

Para la selección de las variables teóricas, en las que se fundamentaron los criterios de evaluación del instrumento, se revisaron diversos antecedentes de la literatura sobre evaluación de la competencia narrativa (Benítez & García; 2010; Cuetos *et al.*, 2002; Dunsmuir *et al.*, 2015; Montanero *et al.*, 2014; Sundeen, 2014; Yan *et al.*, 2012; Thorndyke, 1977).

A partir de esta revisión bibliográfica, se identificaron 7 criterios de calidad de una composición narrativa, que resumimos a continuación. Los 4 primeros se centran en la evaluación de la organización y contenido del relato; los siguientes se centra en los aspectos gramaticales, léxicos y ortográficos.

- **Marco.** Se corresponde con la presentación de un marco espacio-temporal más o menos definido e identificable, o bien ficticio, dentro del que se suelen presentar los personajes. Lo más importante, de hecho, es la riqueza de la descripción de los personajes (desde el punto de vista físico, psicológico, etc.).
- **Tema o suceso inicial.** Es el acontecimiento o suceso inicial que desencadena el conjunto de acciones posteriores. Un buen relato alude además a los sentimientos que hacen reaccionar a los personajes y a las intenciones que guían sus acciones.
- **Trama.** Se trata de la secuencia de eventos, posteriores al suceso inicial, que dan sentido a la historia. Es muy importante la coherencia y la claridad, pero también la riqueza de los eventos y acciones que la componen.

- *Creatividad e interés.* Se relaciona con la originalidad, la creatividad y la longitud de la historia. Se valora que el relato no sea una mera reproducción de un cuento conocido, que tenga una extensión adecuada y que el autor utilice ciertos recursos estilísticos, tales como metáforas, moraleja, el sentido del humor...
- *Oraciones.* Se corresponde con la estructura de las oraciones y el uso de signos de puntuación en el texto. Un relato adecuado se caracterizaría por la inteligibilidad de sus oraciones al poseer una estructura correcta y por el uso de los signos de puntuación de manera adecuada.
- *Vocabulario.* Se refiere a la variedad y adecuación de los vocablos que el autor incluye en su relato, sin ambigüedad ni excesivas repeticiones.
- *Ortografía.* Se refiere al correcto empleo de las reglas ortográficas (en la lengua materna).

En un segundo momento, se redactó un borrador de la rúbrica en el que se definieron los 4 niveles de logro que se esperaban para cada uno de los anteriores criterios para escritores de entre 10 y 12 años.

Finalmente, la RER fue aplicada a una treintena de composiciones de estudiantes de este nivel educativo en varios ciclos. Al acabar cada ciclo, los investigadores re-definieron la redacción de algunos niveles de logro, para mejorar su claridad y objetividad, así como su capacidad discriminativa (principalmente para evitar riesgos de efecto “suelo” en algunos de los criterios). La versión final del instrumento puede verse en la tabla II.

## Redacción y evaluación de relatos

La validación de la RER se basó en su aplicación a una muestra de composiciones narrativas de 292 estudiantes de 9 a 14 años.

El proceso de producción de textos se desarrolló en una sesión de una hora de duración. En una primera fase, durante unos 20 minutos, se realizó una breve introducción al concepto de narración, así como sobre las características y la estructura típica de un relato. Durante la segunda fase, cada alumno dispuso de 30 minutos para escribir una historia (inventada) cuya temática versaba sobre *un niño que se convierte en un héroe de su pueblo*, con una extensión entre una y dos caras de un folio.

Las narraciones producidas por los alumnos fueron analizadas desde tres perspectivas. Por un lado, se aplicó independientemente la rúbrica de evaluación de relatos y el PROESC. En ambos casos, cada composición recibió la puntuación correspondiente a cada criterio o nivel, únicamente cuando el texto cumplía con todos los requisitos enunciados en el mismo. En caso de faltar algún requisito, la puntuación obtenida fue la del nivel inmediatamente inferior.

Por otro lado, se seleccionaron intencionalmente cuatro maestros en activo (en centros educativos diferentes a los que se llevó a cabo el estudio) y con más de 5 años de experiencia profesional que calificaron las narraciones, utilizando una escala de 0 a 10, sin conocer las puntuaciones que dichas composiciones habían obtenido previamente con la RER y el PROESC.

## Resultados

A continuación, se exponen los resultados del análisis de la validez y la fiabilidad de la rúbrica construida para evaluar composiciones narrativas.

### Validez de contenido

La rúbrica fue sometida a un juicio de expertos para evaluar su validez de contenido. Se seleccionaron dos académicos, especialistas en escritura y un profesional con amplia experiencia en la Educación Primaria y Secundaria. Se les envió una hoja de validación en la que se les instaba a valorar, por un lado, la idoneidad de los criterios seleccionados y su importancia desde el punto de vista de la calidad de un relato; y, por otro lado, la claridad de la redacción de los respectivos niveles de logro.

El promedio obtenido en la valoración de la importancia (en una escala de 1 a 5) fue de 5 en todos los criterios, salvo en Creatividad (4,6), Oraciones (4,3) y Vocabulario (4,6). En cuanto a la claridad de la redacción, todos obtuvieron la máxima puntuación, salvo la trama (4,6). Los expertos hicieron, además, algunas observaciones cualitativas de carácter menor, que se tuvieron en cuenta para revisar la redacción final de la RER (tabla II).

**TABLA II.** Rúbrica de evaluación de relatos (RER)

<b>Criterio</b>	<b>Niveles de logro</b>	
1. Marco (Introducción)	1. No hay ninguna introducción a la historia	
	2. Sólo se nombran algunos personajes sin comentar nada de ellos.	
	3. Se cuenta cuándo y dónde ocurrió la historia y quiénes son los protagonistas pero no se describe cómo son.	
	4. Se cuenta cuándo y dónde ocurrió la historia. Se describe quiénes son los protagonistas y cómo son (físicamente, personalidad, etc.).	
2. Tema o suceso inicial	1. No se explica qué pasó al principio	
	2. Se explica algo que pasó al principio, pero no se entiende bien.	
	3. Se explica un poco lo que pasó al principio y cómo se sintieron los personajes.	
	4. Se describe con claridad todo lo que pasó al principio, cómo se sintieron los personajes y qué se propusieron hacer.	
3. Trama (episodios y resolución)	1. En la historia no pasa nada o no se entiende nada de lo que pasa.	
	2. La historia es un poco liosa o algunos hechos no tienen mucho sentido.	
	3. La historia es sencilla. Se explica con suficiente claridad lo que pasó después y al final.	
	4. Aunque la historia es larga, se explica muy bien todo lo que pasó después (en varios momentos bien ordenados), y cómo quedaron al final los personajes.	
4. Creatividad e interés	1. La historia está copiada o no se entiende.	
	2. La historia es original y se entiende, pero es demasiado corta o muy aburrida	
	3. La historia es original y entretenida	
	4. La historia es original y entretenida. Es bonita, tiene sentido del humor o nos enseña algo	
5. Oraciones	1. La mayoría de las frases no se entienden porque están mal construidas.	
	2. La mayoría de las frases se entienden bien, aunque algunas están mal hechas.	
	3. Las frases se entienden bien, pero hay pocos signos de puntuación (o sólo son comas)	
	4. Las frases se entienden bien y la mayoría están bien puntuadas, incluyendo los diálogos.	
6. Vocabulario	1. El vocabulario es muy pobre y con muchos errores (algunas palabras no son las más adecuadas para lo que se quiere decir).	
	2. Las palabras son correctas pero se repiten mucho.	
	3. Se utiliza un vocabulario correcto y variado (que no se repite demasiado).	
	4. Se utiliza un vocabulario rico, con algunas palabras poco conocidas.	
7. Ortografía	1. Hay muchas faltas de ortografía (una o más en cada oración).	
	2. Hay bastantes faltas de ortografía (una por cada dos oraciones aproximadamente)	
	3. Hay pocas faltas de ortografía	
	4. No hay ninguna falta de ortografía	

Fuente: adaptada de Montanero, Lucero y Fernández (2014)

## Validez de constructo: análisis de las diferencias individuales en función del nivel educativo

La versión final de la RER se aplicó a una muestra de 292 estudiantes de Educación Primaria y Secundaria. La tabla III refleja el promedio de los resultados de cada criterio en los tres niveles educativos que se compararon. Entre el primer y segundo nivel hay diferencias significativas en cuanto al marco ( $t=2,48$ ;  $p<0,05$ ), al tema ( $t=3,07$ ;  $p<0,01$ ), las oraciones ( $t=2,69$ ;  $p<0,01$ ), la ortografía ( $t=4,48$ ;  $p<0,01$ ); así como en cuanto al subtotal de la dimensión Organización y contenido ( $t = 3,18$ ;  $p<0,01$ ), al subtotal de la dimensión Aspectos gramaticales, léxicos y ortográficos ( $t = 4,25$ ;  $p<0,01$ ) y al total de la RER ( $t=4,42$ ;  $p<0,01$ ). Entre el segundo y el tercer nivel se detectan diferencias significativas, en cuanto a la creatividad ( $t=3,52$ ;  $p<0,01$ ) y la ortografía ( $t=2,58$ ;  $p<0,05$ ); así como al subtotal de Organización y contenido ( $t=3,18$ ;  $p<0,01$ ), al subtotal de Aspectos gramaticales ( $t=2,58$ ;  $p<0,05$ ) y al total de la rúbrica ( $t=3,50$ ;  $p<0,01$ ).

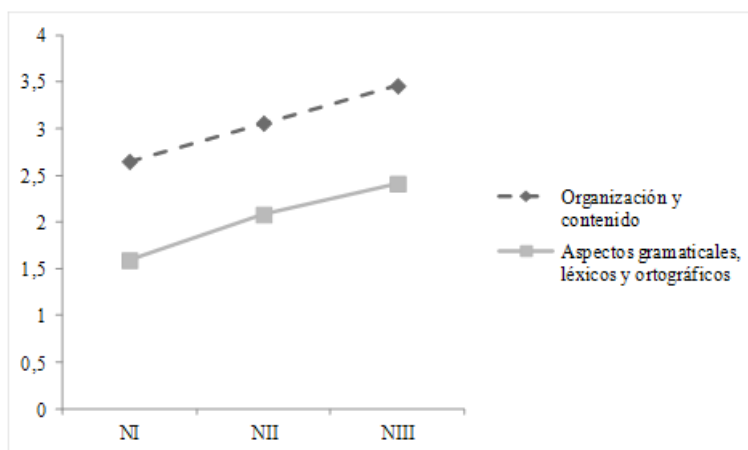
**TABLA III.** Diferencias en los resultados de la rúbrica de evaluación de relatos (RER), en función del nivel de escolarización (NI: 3º y 4º de Educación Primaria; NII: 5º y 6º de Educación Primaria; NIII: 1º y 2º de Educación Secundaria Obligatoria)

RER	Media (M) y desviación típica (DT)						Diferencia de medias (DM)		
	Nivel I		Nivel II		Nivel III		NI-II	NI-III	NII-III
	M	DT	M	DT	M	DT	DM	DM	DM
Marco	0,79	0,4	0,92	0,35	0,99	0,38	-0,13*	-0,2**	-0,07
Tema	0,63	0,31	0,77	0,34	0,87	0,37	-0,14*	-0,24**	-0,1
Trama	0,77	0,32	0,83	0,37	0,89	0,28	-0,06	-0,12*	-0,06
Creatividad	0,47	0,32	0,54	0,38	0,7	0,28	-0,07	-0,23**	-0,16*
Subtotal Organización y contenido	2,65	0,87	3,06	0,95	3,46	0,8	-0,41**	-0,81**	-0,4**
Oraciones	0,67	0,33	0,81	0,39	0,91	0,41	-0,14*	-0,24**	-0,1
Vocabulario	0,63	0,35	0,71	0,35	0,8	0,45	-0,08	-0,17**	-0,09
Ortografía	0,28	0,45	0,55	0,40	0,7	0,4	-0,27**	-0,42**	-0,15*
Subtotal Aspectos gramaticales, léxicos y ortográficos	1,59	0,75	2,08	0,85	2,41	0,95	-0,49**	-0,82**	-0,33*
Total	4,23	1,35	5,14	1,50	5,87	1,43	-0,91**	-1,64**	-0,73**

Fuente: Elaboración propia. (\*)  $p<0,05$ ; (\*\*)  $p<0,01$

Un indicador de la validez de un instrumento es su capacidad de discriminar entre diferentes niveles de desarrollo de la competencia que supuestamente evalúa. En el siguiente gráfico se observa una relación aumento progresivo y lineal de los promedios de la rúbrica a medida que aumenta el nivel en el que están escolarizados los estudiantes que redactaron los textos.

**GRÁFICO I.** Evolución de los criterios globales en función del nivel de escolarización (NI: 3º y 4º de Educación Primaria; NII: 5º y 6º de Educación Primaria; NIII: 1º y 2º de Educación Secundaria Obligatoria).



Fuente: Elaboración propia

Más específicamente, la tabla IV recoge la distribución de los sujetos en cada uno de los niveles de logro de los criterios de la RER.

**TABLA IV.** Porcentaje de estudiantes localizados en cada nivel de logro (NL) de la rúbrica de evaluación de relatos (RER) en función del nivel de escolarización (NI: 3° y 4° de Educación Primaria; NII: 5° y 6° de Educación Primaria; NIII: 1° y 2° de Educación Secundaria Obligatoria)

<b>Criterio</b>	<b>NL</b>	<b>NI</b>	<b>NII</b>	<b>NIII</b>
1. Marco	1	1,1	1	0
	2	58,7	30,4	28,6
	3	21,7	52	43,9
	4	18,5	16,7	27,6
2. Tema o suceso inicial	1	5,4	1	0
	2	68,5	53,9	42,9
	3	21,7	35,3	39,8
	4	4,3	9,8	17,3
3. Trama	1	2,2	4,9	0
	2	47,8	35,3	29,6
	3	44,6	48	63,3
	4	5,4	11,8	7,1
4. Creatividad	1	22,8	21,6	3,1
	2	62	52	53,1
	3	14,1	23,5	43,9
	4	1,1	2,9	0
5. Oraciones	1	7,6	6,9	2
	2	52,2	35,3	38,8
	3	38	46,1	34,7
	4	2,2	11,8	24,5
6. Vocabulario	1	12	5,9	10,2
	2	52,2	52	36,7
	3	33,7	36,3	35,7
	4	2,2	5,9	17,3
7. Ortografía	1	68,4	25,5	15,3
	2	8,7	39,2	33,7
	3	20,7	34,3	46,9
	4	2,2	1	4,1

Fuente: Elaboración propia



Como cabía esperar, el *contenido* de los relatos se va haciendo más complejo a medida que el alumno va evolucionando por los niveles del sistema educativo. Así, por ejemplo, el alumnado de tercer curso no suele referir al marco *espacio-temporal* en el que se desarrolla el relato. En los últimos cursos de Primaria, por el contrario, se menciona frecuentemente el nombre del lugar y se utilizan fórmulas específicas del género para aludir al contexto temporal (por ejemplo, “érase una vez” o “había una vez”). En Primaria en general la narración gira casi siempre alrededor de un único personaje, que apenas se describe o sobre el que se señala alguna característica física. En ocasiones, se alude a personajes que se desconocía que existían o cuyo papel en la trama no queda del todo claro. En Secundaria, en cambio, la historia se contextualiza y se describe a los personajes con mayor detalle, incluyendo rasgos psicológicos o psicosociales (“Luis no era buen estudiante, pero se caracterizaba por ser solidario y tranquilo”). Además, aparecen casi siempre *personajes secundarios* que participan en eventos relevantes.

La RER pone también de manifiesto cómo la *trama* se va haciendo más compleja a lo largo de los niveles educativos. En bastantes textos de 3º de Educación Primaria prácticamente no ocurre nada (son meras descripciones sin sucesos problemáticos que desencadenen acciones posteriores); mientras que en otros la trama resulta muy confusa, con disrupciones en el hilo conductor y desenlaces poco coherentes con el inicio de la historia. En este nivel muchos niños no componen una historia original, sino que adaptan argumentos de cuentos conocidos. A partir de 5º curso se aprecia una mejora significativa en la coherencia. Se alude ya explícitamente a los sentimientos de los personajes y lo que éstos se propusieron hacer una vez delimitado el problema inicial, aunque todavía en bastantes textos las acciones se yuxtaponen o carecen de una lógica causal o motivacional. Los relatos son más extensos y creativos e incorporan frecuentemente diálogos y moralejas. En los cursos de Secundaria la trama se extiende, al incorporar más episodios y al aparecer, en ocasiones, temas multi-problemáticos que generan ciclos de desenlace.

Sin embargo, sorprendentemente el porcentaje de estudiantes que alcanzan el máximo nivel de logro en la trama es ligeramente inferior a los que lo consiguen en los últimos cursos de Primaria. Los textos de Secundaria son más extensos, pero no mucho más creativos que

en el nivel anterior y curiosamente suelen carecer de diálogos y otros elementos que enriquecen el relato.

En cuanto a los aspectos *formales y gramaticales*, los resultados de la RER son también coherentes con la evolución que se esperaba en función del nivel educativo del alumnado. En los textos de 3º y 4º predominan las oraciones simples y las compuestas coordinadas de un modo muy pobre (con un abuso de la conjunción “y”, así como de anáforas fieles, que resultan excesivamente repetitivas). Un 70% de los estudiantes de 3º y 4º tienen más de 10 faltas en el texto (sin contar tildes); mientras que en 5º y 6º la mayoría tiene más de 5. Hasta Secundaria prácticamente no aparecen oraciones subordinadas sustantivas y adjetivas y un empleo algo más preciso de los signos de puntuación (aunque el abuso de la coma, donde debería aparecer un punto o un punto y coma, es una constante). El vocabulario y los recursos anafóricos, no obstante, siguen resultando poco variados repetitivos en la mayoría de los alumnos de esta etapa. Casi todos los textos de Secundaria presentan todavía errores ortográficos, aunque la mayoría contienen menos de 5 faltas.

## Normalidad

En la siguiente tabla se observa que las puntuaciones obtenidas en las composiciones de los diferentes niveles educativos en cada una de las dimensiones se adecuan suficientemente a los modelos de una distribución normal, teniendo en cuenta la bondad de ajuste, la asimetría y la curtosis obtenida en cada caso.

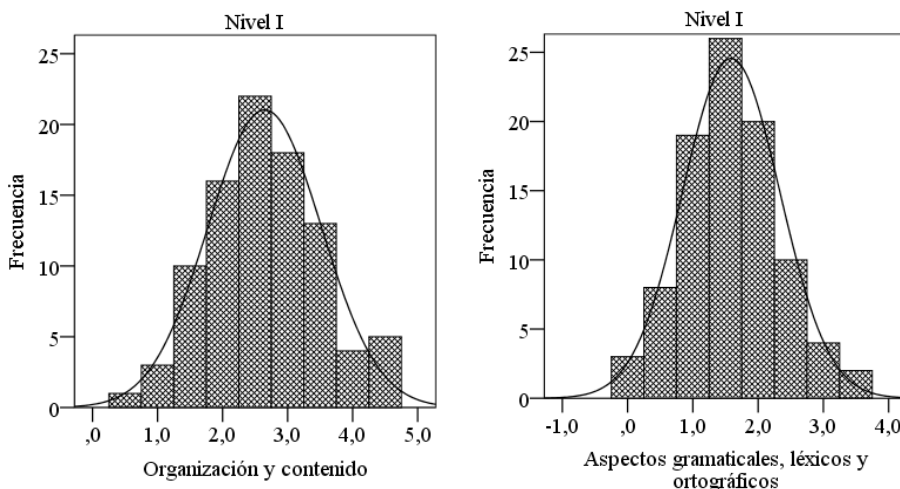
**TABLA V.** Distribuciones de las puntuaciones de la rúbrica de evaluación de relatos (RER) por dimensiones en cada nivel de escolarización (NI: 3º y 4º de Educación Primaria; NII: 5º y 6º de Educación Primaria; NIII: 1º y 2º de Educación Secundaria Obligatoria).

Dimensiones	Nivel I	Nivel II	Nivel III
<i>Organización y contenido</i>	Media: 2,65 DT: 0,87 Asimetría: 0,13 Curtosis: -0,20 Z de K-S: 1,27	Media: 3,06 DT: 0,95 Asimetría: 0,14 Curtosis: -0,59 Z de K-S: 1,48	Media: 3,46 DT: 0,80 Asimetría: 0,06 Curtosis: -0,67 Z de K-S: 1,43
<i>Aspectos gramaticales, léxicos y ortográficos</i>	Media: 1,59 DT: 0,75 Asimetría: 0,22 Curtosis: 0,01 Z de K-S: 1,49	Media: 2,08 DT: 0,85 Asimetría: 0,10 Curtosis: 0,03 Z de K-S: 1,36	Media: 2,41 DT: 0,95 Asimetría: 0,05 Curtosis: -0,51 Z de K-S: 1,24

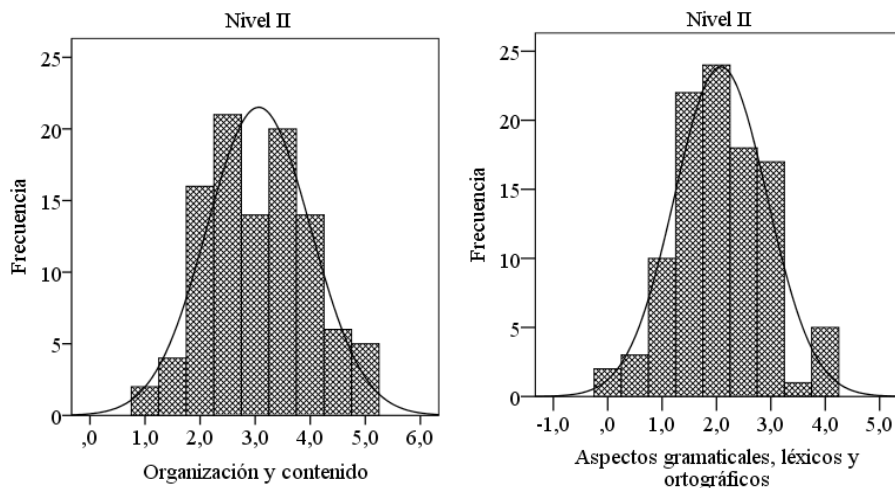
Fuente: Elaboración propia

En las gráficas II, III, IV se puede observar un mejor ajuste en los niveles educativos I y II. En la dimensión *Organización y contenido* la curva resulta en todos los niveles ligeramente platicúrtica, especialmente en el nivel III.

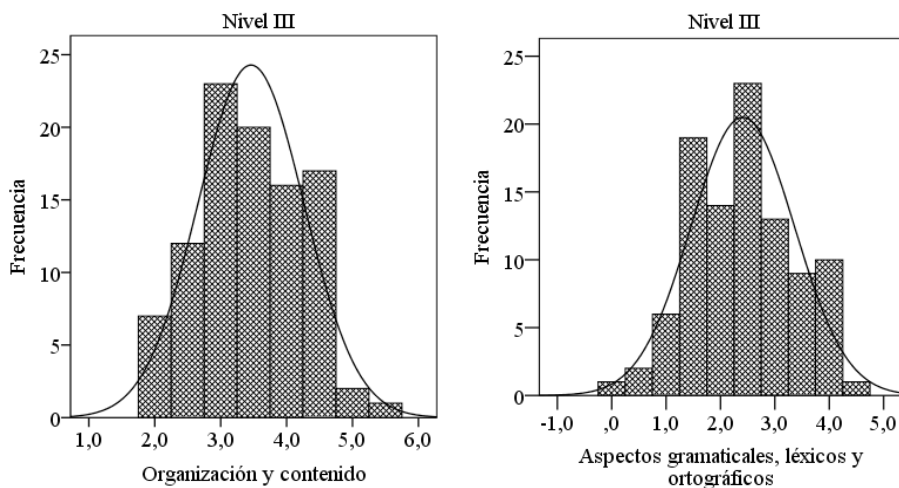
**GRÁFICO II.** Distribución de las puntuaciones en Nivel I (3º y 4º de Primaria)



**GRÁFICO III.** Distribución de las puntuaciones en Nivel II (5° y 6° de Primaria)



**GRÁFICO IV.** Distribución de las puntuaciones en Nivel III (1° y 2° de ESO)



## Validez estructural

Para valorar la estructura de las dimensiones y criterio de la RER se realizó un análisis factorial confirmatorio de componentes principales de la rúbrica. Previamente se aplicaron dos pruebas que evidenciaron que el conjunto de datos era adecuado para ser factorizado: la medida de la adecuación muestral de Kaiser-Meyer-Olkin (0,75) y la prueba de esfericidad de Barlett (Chi-cuadrado=161,45;  $p < 0,01$ ).

En la tabla VI se presenta la matriz factorial de componentes rotados, según el procedimiento varimax, en la que aparecen los 2 componentes extraídos y la saturación de las variables en cada uno de ellos.

TABLA VI. Matriz de componentes rotados según varimax

Criterios	Componentes	
	1	2
Marco	0,206	0,421
Tema	<b>0,715</b>	0,064
Trama	<b>0,711</b>	0,174
Creatividad	<b>0,739</b>	0,200
Oraciones	0,356	<b>0,600</b>
Vocabulario	0,012	<b>0,755</b>
Ortografía	0,083	<b>0,697</b>

Fuente: Elaboración propia

Se observan en negrita los valores de cada criterio que se corresponden con una saturación mayor en cada uno de los componentes. Para vincular un factor con un componente se aconseja que el primero de ellos tenga un peso factorial significativo (superior a 0,50). Aplicando un criterio más laxo (superior a 0,2) el criterio Marco podría considerarse realmente en cualquiera de los dos componentes. Se ha incluido en el primero porque resulta más coherente con su validez teórica de constructo (a pesar de contar con un peso factorial algo menor).

Los 7 factores quedan, por tanto, agrupados en 2 componentes que en su conjunto explican el 49% de la varianza total, teniendo mayor peso el primer componente que explica un 34%. De acuerdo con la estructura

del instrumento, los componentes parecen por tanto coherentes con las dimensiones Contenido y organización (componente 1), así como con la dimensión relativa a los aspectos gramaticales, léxicos y ortográficos (Componente 2).

## Validez de criterio

La correlación entre el sumatorio de las puntuaciones obtenidas en la RER y en el test estandarizado PROESC, a partir de la evaluación de cada uno de los textos, resultó muy elevada  $r=0,79$  ( $p<0,01$ ), aunque no tanto con las calificaciones que el profesorado en activo en estos mismos niveles educativos otorgó a los textos ( $r=0,41$ ;  $p<0,01$ ). No obstante, esta última correlación fue superior a la que obtuvimos entre dichas calificaciones y los resultados del PROESC ( $r=0,30$ ;  $p<0,01$ ).

Más específicamente, en la tabla VII se muestra la correlación con cada uno de los criterios del PROESC que evalúan aspectos similares (no se ha incluido el criterio “ortografía” de la rúbrica por que no existe ninguno similar en el PROESC).

TABLA VII. Correlación de Pearson ( $r$ ) entre los criterios específicos de los dos instrumentos

RER	PROESC	$r$
Marco	Dónde y cuándo Personajes	0,8* 0,27*
Tema o suceso inicial	Suceso con consecuencias	0,28*
Trama	Suceso con consecuencias Desenlace coherente Continuidad lógica Sentido unitario	0,29* 0,57* 0,43* 0,46*
Creatividad	Creatividad Figuras literarias	0,34* 0,14
Oraciones	Oraciones complejas	0,28*
Vocabulario	Vocabulario	0,48*
Ortografía	-	-

Nota: Correlaciones estadísticamente significativas: (\*)  $p < 0,01$ .

Fuente: Elaboración propia

## Consistencia interna

Al analizar la homogeneidad de los 4 elementos de los que consta la dimensión *Organización y contenido* del instrumento se obtuvo un Alfa de Cronbach de 0,57, mientras que en la dimensión *Aspectos gramaticales, léxicos y ortográficos* fue de 0,60. Se trata de un valor satisfactorio y superior al que hemos obtenido con el PROESC (0,37 en ambas dimensiones). Se observa también que si se eliminase algún criterio de la rúbrica el valor de Alfa no mejoraría.

## Fiabilidad interjueces

Para calcular la fiabilidad en la aplicación de la RER se analizó el grado de acuerdo entre 3 evaluadores sobre una muestra de 30 narraciones, escogidas al azar, tras varias sesiones de entrenamiento en su aplicación. Los índices de fiabilidad obtenidos por el método kappa de Cohen, fueron superiores a 0,85 ( $p < 0,01$ ) en todos los criterios. Nuevamente, los resultados fueron ligeramente superiores a los obtenidos con el PROESC.

## Conclusiones

La evaluación formativa de la escritura es una práctica habitual del aula que se encuentra mediatizada por las opiniones subjetivas de los maestros sobre los aspectos que reflejan una buena escritura (Barton, Hamilton, & Ivancic, 2000). La revisión de la bibliografía que hemos realizado en la introducción de este trabajo muestra la escasez de instrumentos objetivos, con suficientes garantías de validez y fiabilidad, que sean accesibles para el profesorado y permitan, al mismo tiempo una utilización funcional en el aula. Las escalas ordinales de apreciación en formato descriptivo (o rúbricas) se adecuan muy bien a ese objetivo: por un lado, por que ofrecen indicaciones concretas sobre el nivel de logro alcanzado por el estudiante en los criterios específicos de evaluación de una competencia o tarea, así como las siguientes mejoras o avances que puede realizar en la progresión de su aprendizaje; por otro lado, porque su diseño funcional e intuitivo (normalmente en una sola hoja), facilita

su utilización en la práctica del aula, no sólo por el profesorado, sino también por los propios estudiantes.

En este sentido, el principal objetivo de este trabajo era diseñar y validar una rúbrica para la evaluación de relatos (RER) en los primeros años de la escolaridad obligatoria.

Los datos aportados en primer lugar en el apartado de Resultados apoyan la validez de contenido de la rúbrica, lo que sugiere que tanto la estructura dimensional como la operativización de los criterios y niveles de logro son relevantes y representativos del constructo para su evaluación en los tres niveles escolares seleccionados. Los expertos (investigadores y profesionales de la educación) que han tenido la oportunidad de revisar y utilizar la RER han valorado muy positivamente, tanto su construcción como su utilidad práctica. El análisis factorial de los datos obtenidos tras su aplicación a una amplia muestra de relatos ha confirmado la adecuación de la estructura de dos factores o dimensiones que agrupan los diversos criterios de evaluación. Al aplicar la rúbrica a la evaluación de relatos redactados por estudiantes de diferentes edades, se ha comprobado también que el instrumento discrimina aceptablemente la progresiva evolución en la madurez en la competencia narrativa que se espera en alumnos de 9 a 14 años. Además, la correlación global de 0,79, con una prueba estandarizada (el PROESC), puede considerarse más que suficiente como indicador de *validez criterial* (Roblyer & Wiencke, 2003). Otro importante indicador de validez es la correlación con las calificaciones que el profesorado en activo en estos mismos niveles educativos otorgó a los textos (Gearhart *et al.*, 1995), que resultó incluso superior a la que hemos registrado con el PROESC.

En segundo lugar, hemos obtenido también datos positivos de la fiabilidad del instrumento, desde el punto de vista de su consistencia interna y de la objetividad de su aplicación por diversos evaluadores. En cuanto a la consistencia (medida con la prueba Alfa de Cronbach), los resultados se encuentran comprendidos dentro de los parámetros que diversos expertos recomiendan en su revisión de investigaciones sobre rúbricas (Graham & Perin, 2007; Jonsson & Svingby, 2007; Knoch, 2009). Los resultados de la aplicación del índice kappa de Cohen a nuestra rúbrica (superiores a 0,85 en todos los criterios) son también mejores que en la mayoría de dichos trabajos.

En definitiva, la rúbrica estudiada parece un instrumento sencillo, válido y fiable para la evaluación de relatos redactados por estudiantes



de 9 a 14 años. Dadas sus características y nuestra experiencia en trabajos previos (Montanero *et al.*, 2014), cabe suponer además su potencialidad para apoyar actividades de evaluación y co-evaluación formativa en el contexto del aula.

Por último, las diferencias encontradas entre la calidad de los relatos en diferentes niveles educativos es, en general, coherente con lo que se esperaba. Cabe señalar, sin embargo, la escasez de estudiantes de los últimos cursos que consiguen redactar tramas complejas, bien articuladas y con elementos realmente creativos. En particular, se observa una débil progresión en cuanto a la competencia retórica de la mayoría del alumnado de los primeros cursos de Secundaria, a la que ya aludían otras investigaciones previas en nuestro sistema educativo (Artiles & Jiménez, 2007; Salvador, 2000). Una posible explicación para este resultado es que la RER responde a una estructura clásica de relato sencillo y lineal. No identifica una diversidad de recursos (como las analepsias, el contraste de personajes en diferentes tiempos, etc.) que escritores más avanzados pueden utilizar para enriquecer el relato; lo que limita su capacidad discriminativa en un rango tan amplio de edad. En futuros estudios pretendemos adecuar y validar diferentes versiones de la RER para rangos de edad de sólo 2 años.

En cualquier caso, la incorporación al aula de instrumentos que faciliten una evaluación formativa de esta competencia puede contribuir a potenciarla. Para confirmar la utilidad de este instrumento en futuros trabajos pretendemos también analizar longitudinalmente la evolución de la calidad de los relatos de estudiantes que participen en actividades de evaluación con esta rúbrica, en comparación con otros que no lo hagan.

## Referencias bibliográficas

Artiles, C., & Jiménez, J. E. (2007). *Proescrí Primaria. Prueba de evaluación de procesos cognitivos en la escritura*. Canarias: Consejería de Educación, Cultura y Deportes del Gobierno de Canarias. Dirección General de Ordenación e Innovación Educativa.

- Baker, E.L., Abedi, J.; Linn, R. L., & Niemi, D. (1995). Dimensionality and generalizability of domain-independent performance assessments. *Journal of Educational Research* 89 (4), 197–205.
- Barton, D., Hamilton, M., & Ivanic, R. (2000). *Situated Literacies: reading and writing in context*. Routledge: London.
- Benítez, R., & García, G. (2010). El razonamiento analógico verbal: una habilidad cognitiva esencial de la producción escrita. *Onomázein*, 22 (2), 165-194.-
- Berninger, V.W, Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing: An interdisciplinary Journal*, 4, 257-280.
- Botvin, G. J., & Sutton-Smith, B. (1977). The development of structural complexity in children's fantasy narratives. *Developmental Psychology*, 13 (4), 377-388.
- Cremin, T., & Myhill, D. (2012). *Writing Voices: Creating communities of writers*. London: Routledge.
- Cuetos, F., Ramos, J. L., & Ruano, E. (2002). *PROESC: Bateria de evaluación de los procesos de escritura*. Madrid: TEA Ediciones.
- Dunsmuir, S. Kyriacou, M. Batuwitige, S. Hinson, E. Ingram, V., & O'Sullivan, S. (2015). An evaluation of the Writing Assessment Measure (WAM) for children's narrative writing. *Assessing Writing*, 23 (2015) 1–18.
- East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, 14, 88-115. DOI: 10.1016/j.asw.2009.04.001
- Fang, Z., & Wang, Z. (2011). Beyond rubrics: Using functional language analysis to evaluate students writing. *Australian Journal of Language and Literacy*, 34 (2), 147-165.
- Fernández, M. J., Lucero, M., & Montanero, M. (2016). Rojo sobre negro. ¿Cómo evalúan los maestros las redacciones de sus estudiantes? *Revista de educación*, 372 (abril-junio), 63-82.
- Fernández, M. J. (2017). Co-evaluación de textos narrativos en la Educación Obligatoria. Universidad de Extremadura, Badajoz, España.
- Ferris, D. R., & Hedgcock, J. (2014). *Teaching ESL composition: Purposes, process, and practice* (3<sup>rd</sup> ed.). Mahwah, NJ: Lawrence Erlbaum.
- Gallavan, N., G. & Kottler, E. (2009). Constructing rubrics and assessing progress collaboratively with social studies students, social studies students. *The Social Studies*, 100 (4), 154-158.

- Galve, J. (2005). BECOLE. *Batería de Evaluación Cognitiva de la Lectura y Escritura*. Instituto de Orientación Psicológica. Madrid: Editorial EOS.
- García, J. N., & Marbán, J. M. (2001): Instrumento de Evaluación Individual de los Procesos Cognitivos de la Escritura (IEPCE). En J. N. García, *Dificultades de aprendizaje e intervención psicopedagógica* (pp. 141-149). Barcelona: Ariel.
- Gearhart, M., Herman, J. L., Novak, J. R., & Wolf, S. A. (1995). Toward the instructional utility of large-scale writing assessment: Validation of a new narrative rubric. *Assessing Writing*, 2, 207–242.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99 (3), 445–476.
- Graham, S., Harris, K.R., & Fink-Chorzempa, B. (2002). Contributions of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology*, 94, 669-686.
- Iandolo, G., Esposito, G., & Venuti, P. (2012). Cohesión, micro-organización, estructura narrativa y competencias verbales entre tres y once años: el desarrollo narrativo formal. *Estudios de Psicología*, 34(2), 141-160.
- Janssen, G., Meier, V., & Trace, J. (2015). Building a better rubric: Mixed methods rubric revision. *Assessing Writing*, 26, 51-66. DOI: 10.1016/j.asw.2015.07.002
- Jonsson A., & Svingby, G. (2007). The Use Of Scoring Rubrics: Reliability, Validity And Educational Consequences, *Educational Research Review*, 2, 130–144. DOI: 10.1016/j.edurev.2007.05.002
- Kernan, Keith T. (1977). Semantic and expressive elaboration in children's narratives. *Child discourse*, 91-102.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275-304. DOI: 10.1177/0265532208101008
- Knoch, U., Read, J., & Von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, 12(1), 26–43. DOI: 10.1016/j.asw.2007.04.001
- Kohn, A. (2006). The trouble with rubrics. *English Journal*, 95(4), 12–15.
- Li, J., & Lindsey, P. (2015). Understanding variations between student and teacher application of rubrics. *Assessing Writing*, 26, 67–79. DOI: 10.1016/j.asw.2015.07.003

- Lucero, M., Fernández, M. J., & Montanero, M (2018). Assessment of short stories in Elementary and Secondary Education. A study of teachers' responses to students' texts. *Studies in Educational Evaluation*, 59, 158-167.
- Montanero, M., Lucero, M., & 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 (1), 184-220. DOI: 10.1080/02103702.2014.881653
- Ochoa-Angrino, S., Aragón, L., Correa, M., & Mosquera, S. (2008). Funcionamiento metacognitivo de niños escolares en la escritura de un texto narrativo antes y después de una pauta de corrección conjunta [Versión electrónica], *Acta Colombiana de Psicología*, 11 (2), 77-88.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(0), 129-144. DOI: <http://dx.doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences* (0). DOI: 10.1016/j.lindif. 2012.04.007
- Roblyer, M. D., & Wiencke, W. R. (2003). Design and use of a rubric to assess and encourage interactive qualities in distance courses. *American Journal of Distance Education*, 17, 77-99.
- Ryan, M. (2014). Reflexive writers; Re-thinking writing development and assessment in schools. *Assessing Writing*, 22, 60-74. DOI:10.1016/j.asw.2014.08.002
- Salvador, F. (2000). Habilidades narrativas de alumnos de Educación Primaria en la producción de textos escritos. *Enseñanza*, 17-18, 145-163.
- Schenck, A. D., & Daly, E. (2012). Building a better mousetrap: Replacing subjective writing rubrics with more empirically-sound alternatives for EFL learners. *Creative Education*, 3 (8), 1310-1325. DOI: 10.4236/ce.2012.38193
- Spandel, V. (2006). Speaking my mind: In defense of rubrics. *English Journal*, 96 (1), 19-22.
- Stadler, M. A., & Ward, G. C. (2005). Supporting the narrative development of young children. *Early Childhood Education Journal*, 33, 2, 73-80.

- Stein, N., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. En R. O. Freedle (Ed.), *Advances in discourse processing*. Vol. 2. Norwood, Nueva Jersey: Ablex.
- Sundeen, T. H. (2014). Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, 21, 74-88. DOI: 10.1016/j.asw.2014.03.003
- Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9, 77-110.
- Toro, J., & Cervera, M. (1984). *TALE. Test de análisis de la lectoescritura*. Madrid: Aprendizaje Visor.
- Toro, J., Cervera, M., & Urío, C. (2002). *Escalas Magallanes de Lectura y Escritura. TALE-2000*. Vizcaya: Albor-COHS.
- Yan, C. M. W., McBride-Chang, C., Wagner, R. K., Zhang, J., Wong, A. M. Y., & Shu, H. (2012). Writing quality in Chinese children: speed and fluency matter. *Reading and Writing*, 25 (7), 1499-1521.

**Dirección de contacto:** María Jesús Fernández Sánchez, Universidad de Extremadura, Facultad de Educación, Departamento de Ciencias de la Educación, Avda. de Elvas s/n. 06006 Badajoz, España. E-mail: mafernandezs@unex.es

# The assessment of narrative competence during compulsory schooling<sup>1</sup>

## La evaluación de la competencia narrativa en la educación básica

DOI: 10.4438/1988-592X-RE-2019-383-402

María-Jesús Fernández

Manuel Montanero

Manuel Lucero

*Universidad de Extremadura*

### Summary

Story writing is a fundamental activity toward the development of writing skills in the first years of school education. In a classroom context however, its assessment is often focused on superficial aspects of grammar and spelling. It is widely known that there is a lack of instruments with demonstrated validity and reliability to enable the assessment of semantics and rhetoric in this context. Therefore, the objective of this research was to build and validate a story assessment rubric (RER) to assess the writing tasks of school children of 8-14 years of age. We analyzed 292 texts written by students at those levels of education. The data provided support the validity of the story assessment rubric (RER). This suggests that the dimensional structure, as well as the operationalization of criteria and levels of achievement, are relevant and discriminative enough to be implemented at these levels of the education system. The reliability of the instrument also yielded positive results, in terms of internal consistency and objectivity of implementation. However, the results of the pilot study show some small inconsistencies in a few levels of achievements, such as in the highest

---

<sup>(1)</sup> Acknowledgements: To the Consejería de Economía, Competitividad e Innovación of the Regional Government of Extremadura, and to the European Fund of Regional Development for financial support received (Reference: GRU15153).

level of the “plot” criterion. This is due to the insufficient number of secondary education students capable of writing complex and articulate plots with truly creative elements. Finally, these results are discussed, as well as their implications toward the assessment and improvement of narrative skills.

*Key words:* Writing; Narrative Skills; Stories; Rubric; Elementary Education; Secondary Education.

### **Resumen**

La redacción de relatos es una actividad fundamental para el desarrollo de la competencia escritora en los primeros años de escolaridad. Sin embargo, su evaluación en el contexto del aula se centra frecuentemente en aspectos superficiales, de carácter gramatical y ortográfico. Es notoria la escasez de instrumentos funcionales, con suficientes garantías de validez y fiabilidad, que faciliten la evaluación de aspectos semánticos y retóricos en la práctica del aula. En este sentido, el objetivo de esta investigación fue construir y validar una rúbrica para evaluar relatos (RER) redactados por estudiantes de entre 8 y 14 años. Para ello, se analizaron un total de 292 textos producidos por estudiantes escolarizados en estos niveles educativos. Los datos aportados apoyan la validez de la RER, lo que sugiere que tanto la estructura dimensional como la operativización de los criterios y niveles de logro son relevantes y suficientemente discriminativos para su aplicación en estos niveles educativos. También se obtuvieron datos positivos de la fiabilidad del instrumento, desde el punto de vista de su consistencia interna y de la objetividad de su aplicación por diversos evaluadores. Los resultados del estudio-piloto muestran, sin embargo, pequeñas inconsistencias en algunos niveles de logro, como el nivel máximo del criterio “trama”. Esto se debe a la escasez de estudiantes de Secundaria que consiguen redactar tramas complejas, bien articuladas y con elementos realmente creativos. Finalmente, se discuten estos resultados y sus implicaciones para la evaluación y mejora de la competencia narrativa.

*Palabras-clave:* Escritura; Competencia narrativa; Relatos; Rúbrica; Educación Primaria; Educación Secundaria.

## **Introduction**

Story writing is a fundamental educational activity toward the acquisition of writing skills during the first years of schooling. From the age of 5, children are already able to build simple temporal sequences verbally,

contextualizing them in space-time frames, and then progressively adding causal and motivational relations. From the age of 7, most children are gradually able to put those stories in writing. The characters are introduced by their names or from the narrator's perspective, but are scarcely described (Kernan, 1977); the actions are articulated in episodes and the stories become progressively more coherent, to the extent that the listener can predict a logical ending during the narration (Stadler & Ward, 2005). The sequences are no longer only showing causal relations between certain actions and the characters' reactions, but they are rather clearly articulated around a problem, a solution, and a mediating element between them (Iandolo, Esposito & Venuti, 2012). Between the ages of 8 and 9, there is a significant improvement in cohesion and complexity of the story's structure, incorporating more problematic and mediation elements that reinforce the problem itself (Stein & Glen, 1979). Around the ages of 10 or 11, the narrative essays can be considered relatively mature (Artiles & Jiménez, 2007; Yan *et al.*, 2012), with a hierarchical structure of episodes (primary and secondary) articulated around an inciting incident or a central topic (Botvin & Sutton-Smith, 1977; Ochoa-Angrino *et al.*, 2008).

This progression is determined by the cognitive level, language proficiency and the individual's emotions (Ryan, 2014). For example, the compositions of children with some deficiency in motor skills required for the writing skills, or those who make many spelling mistakes, demonstrate a worse organization of the text or textualization (Berninger *et al.*, 1992; Cremin & Myhill, 2012). When those processes become more automatic, the compositions also improve in other aspects (Graham, Harris & Fink-Chorzempa, 2002).

For decades, we have had a relatively broad knowledge of the evolution of the children's narrative skills. However, when assessing stories, teachers usually neglect the semantic content and rhetorical structure of the narration (Fernández, Lucero & Montanero, 2016). Rather, the frequency of grammar and spelling corrections recorded in students' manuscripts has proven to be the best predictor of the overall grade that teachers give to the texts (Lucero, Fernández & Montanero, 2018). This could be the reason why elementary school children substantially improve those last aspects in their drafts after receiving their teachers' assessment, although not the content and rhetorical structure (Montanero, Fernández & Lucero, 2014).



Although the genre of narrative texts is the most frequent in reading and writing assignments in elementary education, it is striking that very few of the most popular standardized tests for its assessment take rhetorical skills into consideration (see Galve, 2005; García & Marbán 2001; Toro & Cervera, 1984; Toro, Cervera & Urío, 2002). In Spain, the assessment test of writing processes (PROESC) (Cuertos, Ramos & Ruano, 2002) is practically the only standardized test that allows for the assessment of these skills. Its use in the classroom, however, is marginal, probably due to its cost.

It is reasonable to think that if teachers had practical assessment tools in the classroom, the quality of feedback offered to students would improve. Thus, the use simpler instruments, based on scales or rubrics, has been proposed (Sundeen, 2014). A rubric is a particular type of ordinal measurement, in a descriptive format. Its criteria are operationalized in various levels of achievements. This characteristic allows the students to know their situation regarding the objectives they must reach within a specific task.

For example, Benítez & García (2010) proposed an assessment scale of six values per criteria (from unsatisfactory to excellent), allowing the assessment of rhetorical elements in the stories, such as the identification of characters or construction of the plot.

Yan *et al.* (2012) devised a writing assessment rubric, composed of seven criteria divided in two dimensions (content and organization), for students of 6-9 years of age. The instrument valued relevance, range, the use of connectors, coherence, text structure, and intelligibility.

Dunsmuir *et al.* (2015) proposed another rubric allowing for the assessment of written texts in children between 7-11 years of age; this rubric had seven criteria and four levels. The aspects taken into account in this case were the calligraphy, spelling, punctuation marks, structure of the text, vocabulary, organization and originality.

It has been proven that these last instruments enable teachers to explain to the students which aspects will be considered in the assessment, and provide more information than the usual correction of compositions (Ferris & Hedgcock, 2014; Li & Lindsey, 2015). The necessity to select one of those levels for each criteria of the rubric triggers a richer interaction in situations of co-assessment, which in turn has a significant influence in the quality of narrative compositions (Montanero *et al.*, 2014; Ochoa-

Angrino *et al.*, 2008), the self-regulation of learning and the feeling of self-efficacy (Panadero & Jonsson, 2013).

However, other studies have highlighted certain weaknesses related on the one hand, to its validity, and on the other, to its limitations in the formative assessment of competences (Spandel, 2006). About the former, Jonsson & Svingby (2007) did a review of the validity and reliability of various rubrics in different studies. Only a third of these gave some information as to the validity of the instruments. Most validity indicators are based on expert opinions, confirmatory factor analysis (Baker *et al.*, 1995), or on correlations (greater than 0,40) with other measurements taken as criteria (Roblyer & Wiencke, 2003). As regards reliability, the rubrics reviewed by Jonsson & Svingby yielded values between 0,50 and 0,92 in the Cronbach's Alpha test; a 55%-75% degree of agreement between judges; and a Cohen's kappa coefficient between 0,4 and 0,75. Only one of the reviewed studies (Gearhart *et al.*, 1995) was focused on analyzing the validity of a story assessment rubric.

Regarding the second question, the use of the rubric in situations of formative assessment of writing skills, some authors pointed out the risk that the instrument might hinder the students' writing performance (Fang & Wang, 2011). Many authors argue that although they enable teachers to do a quick assessment, they do not really contribute to improve students' writing skills. This is due to the fact that the achievement levels involve generic (not personalized) suggestions, and they are often formulated in a language that is difficult for students to understand (Janssen *et al.*, 2015; Kohn, 2006; Panadero, Alonso-Tapia & Huertas, 2012; Schenck & Daly, 2012). Finally, they also pointed out the lack of training of some teachers who are using it in a classroom context (East, 2009; Knoch, Read & von Randow, 2007).

Ultimately, it seems necessary to shed more light on the validity and usefulness of this type of classroom instruments, particularly as a support for the assessment of writing skills in elementary and secondary education. Within this framework, the objective of this study was to devise a narrative writing assessment tool that would be useful and practical for teachers in the classroom. We built and validated an "analytical" rubric enabling a quick and simple formative assessment of narrative texts written by students of 8-14 years of age. To this effect, we will provide more evidence of validity and reliability than what has been published so far in relation to the abovementioned rubrics. We will also analyze the

individual differences observed when using the rubric with a sample of students of different ages. In doing so, we will be able to verify if the rubric allows us to distinguish between different levels of narrative skills in students' compositions in relation to their level of education.

## Methods

### Sample

A total of 292 students participated in the study, ranging from the third year of elementary education to the second year of compulsory secondary education (8-14 years old). The sample was obtained from seven schools of different cities in Extremadura, selected for convenience in areas of low or average to low socio-economic level. Immigrant students with an insufficient level of Spanish did not participate, nor students with special educational needs. Table 1 shows the final distribution of the sample by age, gender, and level of schooling.

**TABLE I.** Sample distribution (N) according to the level of education (EP: elementary school; ESO: compulsory secondary education), age (in years) and gender (M: male; F: female)

Level	School year	Age	Gender	N		
I	3° EP	8-9	M	14	40	92
			F	26		
	4° EP	9-10	M	17	52	
			F	35		
II	5° EP	10-11	M	18	49	102
			F	31		
	6° EP	11-12	M	25	53	
			F	28		
III	1° ESO	12-13	M	11	25	98
			F	14		
	2° ESO	13-14	M	30	73	
			F	49		

Prepared by the authors

## Instruments

### Writing sheet

Each student was provided with a writing sheet, on which they had to write a short story. The topic was indicated at the top (“a child who becomes the hero of his village”), as well at the required length (approximately one page).

### Story assessment rubric

We adapted an instrument which had already been used in previous studies (Montanero, Lucero & Fernández, 2014), comprising seven assessment criteria, distributed in two dimensions. On the one side, those criteria evaluated aspects related to content and narrative organization: Frame, Topic, Plot, and Creativity. On the other side, grammar, lexical and orthographic aspects were evaluated: Sentences, Vocabulary, and Spelling (Table 2). Each of those criteria had four levels of accomplishment or execution, which could be rated as such: 0 points (level 1), 0,5 points (level 2), 1 point (level 3) and 1,5 points (level 4).

## PROESC

The test of writing processes (PROESC) (Cuetos, Ramos & Ruano, 2002) is a standardized test, commercialized by TEA, which is widely used in the narrative writing assessment in compulsory education. It is composed of ten assessment criteria grouped into two main dimensions: *content* and *coherence*. The first dimension focuses on the following criteria: Where and When, Characters, Events and Consequences, Coherent Ending, and Creativity. The second dimension is assessed according to the following criteria: Logical Continuity, Sense of Unity, Literary Figures, Complex Sentences, and Vocabulary. To achieve each of those criteria, many requisites are defined.

According to the manual, the instrument has an internal consistency of 0,82 (Cronbach's Alpha coefficient). As evidence of criterion validity,

the application of the test to a sample of 839 texts yielded a correlation of 0,46 with the grades given by a group of teachers. Regarding the structural validity, the ten criteria which constitute the test were submitted to a components analysis with varimax rotation. The result was the division of the test in three components which explained 58% of the total variance.

## Procedure

### Design of the story assessment rubric

The object of this study, the rubric, consists of a descriptive-ordinal assessment scale (with four levels per criteria, which were also given a numerical score), with the purpose of assessing stories written in L1 by students of 9-14 years of age. Although the instrument was initially intended for all compulsory education students, a preliminary study (Fernández, 2017) concluded that it was not valid nor discriminative enough for the early elementary school students (6-8 years old), or for the last years of secondary school (15-16 years).

The assessment criteria of the instrument were based on theoretical variables chosen after reviewing previous studies on the assessment of narrative skills (Benítez & García; 2010; Cuetos *et al.*, 2002; Dunsmuir *et al.*, 2015; Montanero *et al.*, 2014; Sundeen, 2014; Yan *et al.*, 2012; Thorndyke, 1977).

From this bibliography review, seven quality criteria for a narrative composition were identified. The first four are focused on the assessment of the story's organization and content; the others focus on the grammar, lexical and orthographic aspects.

- *Frame*. Corresponds to the presentation of a space-time frame, which can be more or less defined and identifiable, or fictitious, and in which the characters are usually presented. The most important aspect is the richness of character descriptions (from a physical or psychological viewpoint).
- *Topic or inciting incident*. This is the inciting incident which triggers all subsequent actions. A good story refers to the emotions

which make the characters react, as well as the intentions guiding their actions.

- *Creativity and interest.* It is related to originality, creativity and the length of the story. It is highly valued when the story is no mere reproduction of a known tale, if it has an adequate length, and whether the author uses certain stylistic resources such as metaphors, moral of the story, sense of humor, etc.
- *Sentences.* It corresponds to sentence structure and the use of punctuation in the text. An adequate story is characterized by the intelligibility of its sentences, by having an appropriate structure, and by a suitable use of punctuation marks.
- *Vocabulary.* Refers to the variety and correctness of the words the authors include in their story, without ambiguity or excessive repetitions.
- *Spelling.* Refers to the correct use of the grammar rules (in their mother tongue).

Subsequently, a draft of the rubric was drawn up, in which we defined the four levels of achievement for each of the previously mentioned criteria expected from writers of 10-12 years of age.

Finally, the story assessment rubric was applied in various cycles to about 30 compositions of students of this level of education. At the end of each cycle, the researchers re-defined the formulation of some levels of achievement, to improve clarity and objectivity, as well as its capacity of discrimination (mostly to avoid the risk of floor effect in some criteria). Table 2 shows the final version of the instrument.

## Story writing and assessment

The validation of the story assessment rubric (RER) was based on its application to a sample of narrative texts of 292 students of 8-14 years of age.

The text writing process happened during one hour-long session. In a first phase of 20 minutes, the concept of narration was briefly introduced, as well as the characteristics and typical structure of a story. In the second phase, each student had 30 minutes to write an invented story, with the

topic *a child who becomes the hero of his village*, of a length of one or two pages.

The narrative texts produced by the children were analyzed from three perspectives. On the one hand, we applied independently the story assessment rubric (RER) and the PROESC. In both cases, each composition received the score corresponding to a criteria or level when the text met all the requirements. If any of those were missing, the next lower score was obtained.

On the other hand, four teachers with more than five years of professional experience were selected (in other schools than those where the study was carried out). These teachers graded the compositions using a scale between 0 and 10, without knowledge of the score achieved previously by the story assessment rubric and the PROESC.

## Results

Next, we will present the results of the validity and reliability analysis for the rubric we devised to assess narrative texts.

### Content validity

The rubric was submitted to a group of experts to evaluate its content validity. Two scholars were selected, specialists in writing, with an experienced professional in elementary and secondary education. They were sent a validation sheet in which they were asked to assess, on the one hand, the suitability of the selected criteria and their importance regarding the quality of a story; on the other hand, the clarity of formulation of the achievement levels.

The average obtained in the evaluation of importance (on a scale from 1 to 5) was 5 in all criteria, except in Creativity (4,6), Sentences (4,3), and Vocabulary (4,6). In terms of clarity of formulation, only the highest score was obtained, except for the Topic (4,6). Furthermore, the experts made some minor observations, which were taken into account for the revision of the final draft of the story assessment rubric (Table 2).

TABLE II. Story Assessment Rubric (RER)

Criteria	Levels of Achievement	
1. Frame (Introduction)	1. There is no introduction to the story	
	2. Only a few characters are named, without mentioning anything else about them.	
	3. The story tells where and when the action happened and who are the protagonists, but they are not described any further.	
	4. The story tells where and when the action happened and who are the protagonists; their physical aspect and personality is also described.	
2. Topic or Inciting Incident	1. What happened at the beginning is not explained.	
	2. What happened at the beginning is partially explained, but it is not comprehensible.	
	3. What happened at the beginning is partially explained, as well as how the characters felt.	
	4. What happened at the beginning is clearly explained, as well as how the characters felt and what they planned to do.	
3. Plot (episodes and resolution)	1. Nothing happens in the story or what happens is not comprehensible.	
	2. The story is a bit confusing or some facts do not make much sense.	
	3. The story is simple. What happened next and at the end is explained with enough clarity.	
	4. Although the story is long, what happened is very clear (in well-organized segments), as well as what happened to the characters at the end.	
4. Creativity and Interest	1. The story is copied or cannot be understood.	
	2. The story is original and can be understood, but it is too short or very boring.	
	3. The story is original and entertaining.	
	4. The story is original and entertaining. It is nice, shows a sense of humor, or teaches us something.	
5. Sentences	1. Most sentences cannot be understood because they are badly built.	
	2. Most sentences can be understood, but some are badly built.	
	3. The sentences are easily understood, but there are too few punctuation marks (or there are only commas).	
	4. The sentences are easily understood, most of the punctuation marks are in the right places, even in the dialogues.	
6. Vocabulary	1. The vocabulary is very poor and with many mistakes (some sentences are not correctly representing what they are trying to say).	
	2. The words are correct, but there are a lot of repetitions.	
	3. The vocabulary used is correct and varied (not repeated too often).	
	4. The vocabulary used is rich, with some little-known words.	
7. Spelling	1. There are a lot of spelling mistakes (at least one in each sentence).	
	2. There are quite a lot of spelling mistakes (at least one in every other sentence).	
	3. There are few spelling mistakes.	
	4. There are no spelling mistakes.	

Adapted from Montanero, Lucero & Fernández (2014)



## Construct validity: analysis of the individual differences according the level of education

The final version of the story assessment rubric was applied to a sample of 292 students of elementary and secondary education. Table 3 reflects the average results for each criterion in the three levels of education that were compared. Between the first and second level, there were no significant differences in terms of the frame ( $t=2,48$ ;  $p<0,05$ ), the topic ( $t=3,07$ ;  $p<0,01$ ), the sentences ( $t=2,69$ ;  $p<0,01$ ), the spelling ( $t=4,48$ ;  $p<0,01$ ); as well as in terms of the subtotal of the Organization and Content dimension ( $t = 3,18$ ;  $p<0,01$ ), the subtotal of the grammar, lexical and orthographic aspects dimension ( $t = 4,25$ ;  $p<0,01$ ), and for the total of the story assessment rubric ( $t=4,42$ ;  $p<0,01$ ). Significant differences were detected between the second and third level, in terms of creativity ( $t=3,52$ ;  $p<0,01$ ) and spelling ( $t=2,58$ ;  $p<0,05$ ); as well as in the subtotal of Organization and Content ( $t=3,18$ ;  $p<0,01$ ), the subtotal of Grammar Aspects ( $t=2,58$ ;  $p<0,05$ ), and the total for the rubric ( $t=3,50$ ;  $p<0,01$ ).

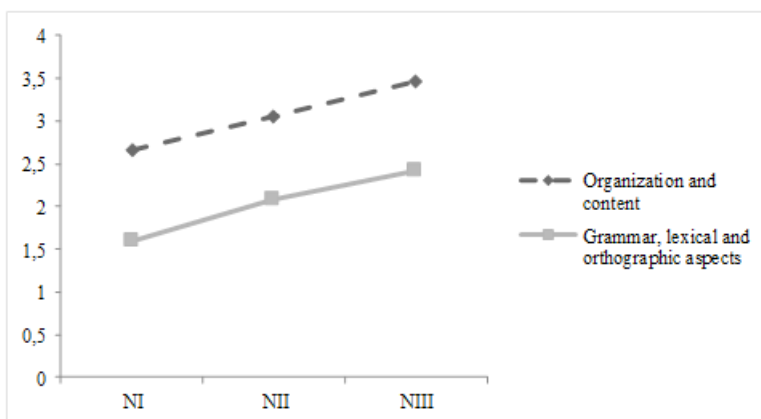
**TABLE 3.** Differences in results of the story assessment rubric (RER), according to the level of schooling (LI: third and fourth year of elementary education; LII: fifth and sixth year of elementary education; LIII: first and second year of compulsory secondary education).

RER	Mean (M) y standard deviation (SD)						Difference of means (DM)		
	Level I		Level II		Level III		LI-II	LI-III	LII-III
	M	SD	M	SD	M	SD	DM	DM	DM
Frame	0,79	0,4	0,92	0,35	0,99	0,38	-0,13*	-0,2**	-0,07
Topic	0,63	0,31	0,77	0,34	0,87	0,37	-0,14*	-0,24**	-0,1
Plot	0,77	0,32	0,83	0,37	0,89	0,28	-0,06	-0,12*	-0,06
Creativity	0,47	0,32	0,54	0,38	0,7	0,28	-0,07	-0,23**	-0,16*
Subtotal Organization y Content	2,65	0,87	3,06	0,95	3,46	0,8	-0,41**	-0,81**	-0,4**
Sentences	0,67	0,33	0,81	0,39	0,91	0,41	-0,14*	-0,24**	-0,1
Vocabulary	0,63	0,35	0,71	0,35	0,8	0,45	-0,08	-0,17**	-0,09
Grammar	0,28	0,45	0,55	0,40	0,7	0,4	-0,27**	-0,42**	-0,15*
Subtotal grammar, lexical, and orthographic aspects	1,59	0,75	2,08	0,85	2,41	0,95	-0,49**	-0,82**	-0,33*
Total	4,23	1,35	5,14	1,50	5,87	1,43	-0,91**	-1,64**	-0,73**

Prepared by the authors. (\*)  $p<0,05$ ; (\*\*)  $p<0,01$

The validity indicator of an instrument is its capacity to discriminate between the different levels of development of the competence it is supposed to assess. The next graphic shows a progressive and linear increase of the rubric's average as the students' level of schooling progresses.

**GRAPH 1.** Evolution of global criteria in terms of level of schooling (LI: third and fourth year of elementary education; LII: fifth and sixth year of elementary education; LIII: first and second year of compulsory secondary education).



In a more specific way, Table 4 shows the distribution of subjects for each level of achievement, according to the criteria of the story assessment rubric (RER).

As predicted, the *content* of stories becomes more complex as students advance through the different levels of the education system. For example, a third grade student does not usually refer to the *space-time* frame of the text. In the last years of elementary education, on the other hand, the name of the place is often mentioned, and they use specific genre formulas to refer to the temporal context (e.g. “Once upon a time”). In elementary education, in general, the narration almost always revolves around a single character who is scarcely described, or if there is a description, it is only to point out some physical characteristic. Sometimes, there is a mention of a character which was unknown until then, or whose role in the plot is not entirely clear. In secondary education, in contrast, a context is given to the story and the characters

are described in more details, including psychological and psycho-social traits (“Luis wasn’t a good student, but he was quiet and he was a team player”). Additionally, *secondary characters* start appearing and participate in relevant events.

**TABLE 4.** Percentage of students in each level of achievement (LoA) of the story assessment rubric (RER) according to the level of schooling (LI: third and fourth year of elementary education; LII: fifth and sixth year of elementary education; LIII: first and second year of compulsory secondary education).

Criterion	LoA	LI	LII	LIII
1. Frame	1	1,1	1	0
	2	58,7	30,4	28,6
	3	21,7	52	43,9
	4	18,5	16,7	27,6
2. Topic or inciting incident	1	5,4	1	0
	2	68,5	53,9	42,9
	3	21,7	35,3	39,8
	4	4,3	9,8	17,3
3. Plot	1	2,2	4,9	0
	2	47,8	35,3	29,6
	3	44,6	48	63,3
	4	5,4	11,8	7,1
4. Creativity	1	22,8	21,6	3,1
	2	62	52	53,1
	3	14,1	23,5	43,9
	4	1,1	2,9	0
5. Sentences	1	7,6	6,9	2
	2	52,2	35,3	38,8
	3	38	46,1	34,7
	4	2,2	11,8	24,5
6. Vocabulary	1	12	5,9	10,2
	2	52,2	52	36,7
	3	33,7	36,3	35,7
	4	2,2	5,9	17,3
7. Spelling	1	68,4	25,5	15,3
	2	8,7	39,2	33,7
	3	20,7	34,3	46,9
	4	2,2	1	4,1

Prepared by the authors.

The story assessment rubric also shows how the *plot* becomes more complex throughout the different levels of education. In many third grade texts, almost nothing happens (they are simple descriptions without problematic events to trigger subsequent actions); while in others, the plot is confusing, with disruptions in the common thread, and the ending is often not coherent with the beginning of the story. At this level, many children do not write an original story, but rather adapt elements of well-known tales. A significant improvement in coherence can be observed from the fifth grade onward. The characters' feelings are now explicitly mentioned, as well as what they plan to do once the initial problem has been delimited. In many texts though, the actions are still juxtaposed, and there can be a lack of causal or motivational logic. The stories are longer and more creative, and often incorporate dialogues and morals. In the first years of secondary school, the plot is extended, incorporating more episodes, and there are sometimes multi-problem topics which generate cycles of denouement.

Surprisingly, the percentage of those secondary school students reaching the highest level of achievement is slightly lower than the students in their last years of elementary education. The texts from secondary school students are longer, but not much more creative than in the previous level of education, and moreover they usually lack dialogues and other enriching elements.

Regarding the *formal* and *grammar* aspects, the results of the story assessment rubric are in line with the expected progress according to the students' level of education. In the texts of third and fourth grade students, the sentences are mostly simple, or the compound sentences are badly coordinated (with an excessive use of the anaphora, as well as the conjunction "and"). 70% of third and fourth grade students made more than ten mistakes per text (not including the Spanish *tilde*); while most fifth and sixth grade students made more than five. Substantive and adjective subordinate clauses practically do not exist until secondary school, and the punctuation marks are used poorly (although there is a systematic abuse of commas where there should be a period or a semi-colon). The vocabulary and the anaphoric references are still limited and repetitive for most students at this level of education. Almost all texts in secondary education still contain spelling mistakes, although for most of them, less than five.

## Normality

The following table shows the scores obtained by the compositions for the different levels of education, and for each of the dimensions. They are sufficiently adapted to the normal distribution models, taking into account in each case the goodness of fit, asymmetry, and kurtosis.

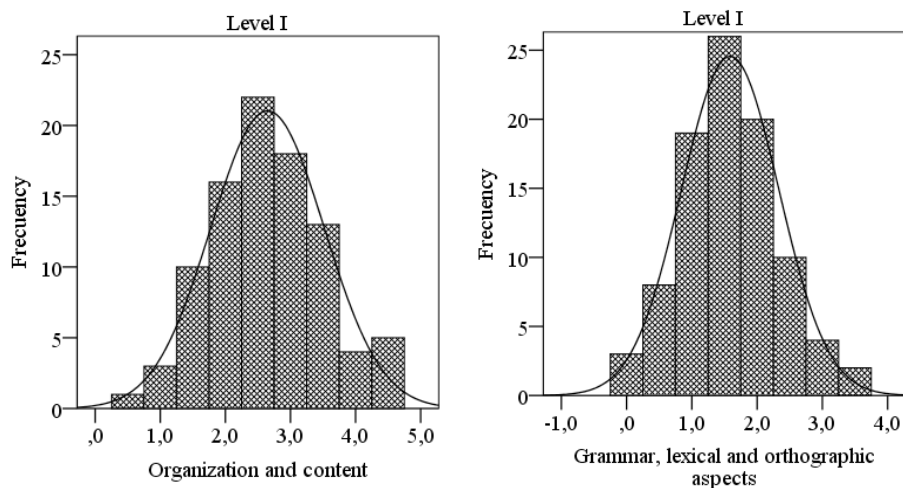
**TABLE 5.** Score distribution for the story assessment rubric (RER) per dimension for each level of schooling (LI: third and fourth year of elementary education; LII: fifth and sixth year of elementary education; LIII: first and second year of compulsory secondary education).

Dimensions	Level I	Level II	Level III
<i>Organization and Content</i>	Mean: 2,65 SD: 0,87 Asymmetry: 0,13 Kurtosis: -0,20 Z of K-S: 1,27	Mean: 3,06 SD: 0,95 Asymmetry: 0,14 Kurtosis: -0,59 Z of K-S: 1,48	Mean: 3,46 SD: 0,80 Asymmetry: 0,06 Kurtosis: -0,67 Z of K-S: 1,43
<i>Grammar, lexical y orthographic aspects</i>	Media: 1,59 DT: 0,75 Asymmetry: 0,22 Kurtosis: 0,01 Z of K-S: 1,49	Media: 2,08 DT: 0,85 Asymmetry: 0,10 Kurtosis: 0,03 Z of K-S: 1,36	Media: 2,41 DT: 0,95 Asymmetry: 0,05 Kurtosis: -0,51 Z of K-S: 1,24

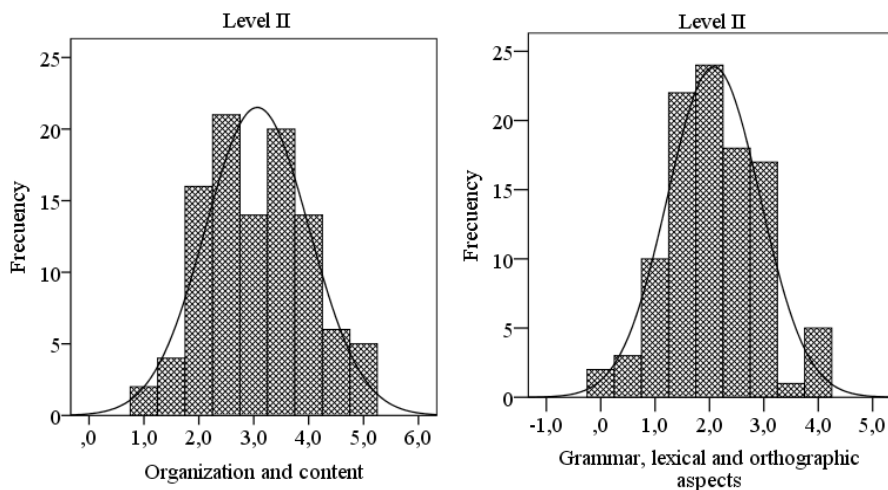
Prepared by the authors

Graphs 2, 3, and 4 show a better adjustment in levels of education I and II. In the Organization and Content dimension, the curve is slightly platykurtic in all levels, but especially in level III.

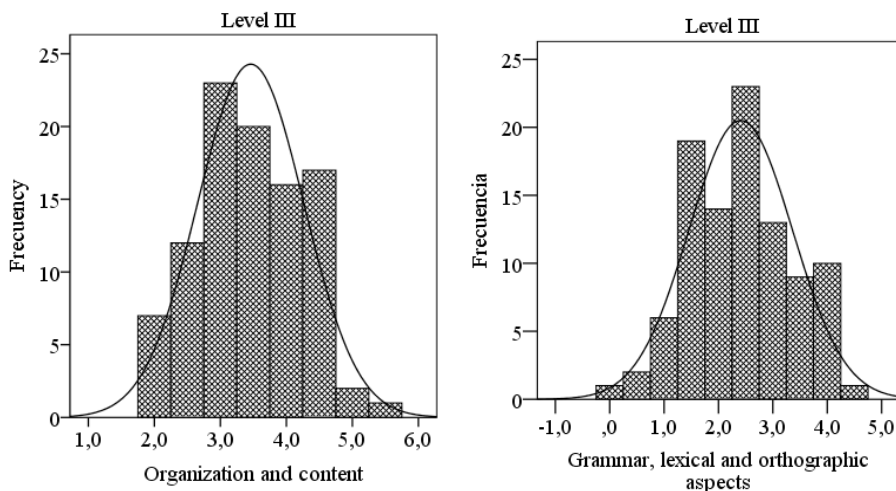
GRAPH 2. Score distribution in Level I (third and fourth grade)



GRAPH 3. Score distribution in Level 2 (fifth and sixth grade)



**GRAPH 4.** Score distribution in Level III (first and second year of secondary education)



### Structural validity

A confirmatory factor analysis of the main components of the rubric was carried out in order to evaluate the dimensions and criteria of the story assessment rubric. Two tests were previously performed, which confirmed that the data set could be factorized: the Kaiser-Meyer-Olkin measure of sampling adequacy (0,75) and Bartlett's test for sphericity (Chi-squared=161,45;  $p < 0,01$ ).

Table 6 shows the factor matrix of rotated components, according to the varimax procedure. Both extracted components can be observed, as well as the saturation of variables for each of them.

**TABLE 6.** Matrix of rotated components, according to varimax.

Criterion	Components	
	1	2
Frame	0,206	0,421
Topic	<b>0,715</b>	0,064
Plot	<b>0,711</b>	0,174
Creativity	<b>0,739</b>	0,200
Sentences	0,356	<b>0,600</b>
Vocabulary	0,012	<b>0,755</b>
Spelling	0,083	<b>0,697</b>

Prepared by the authors

The values in bold for each criterion correspond to a higher saturation in each of the components. In order to connect a factor with a component, it should be advisable for the first of them to have a significant factorial weight (higher than 0,50). If applying a laxer criterion (higher than 0,2), the Frame criterion could be considered for both components. Only the first was included, since it was more coherent with its construct's theoretical validity (despite having a slightly smaller weight factor).

The seven remaining factors, grouped in two components, explain 49% of the total variance, the first of those having more weight, explaining 34%. According to the structure of the instrument, the components seem therefore coherent with the Organization and Content dimension (component 1), as well as the dimension related to the grammar, lexical and orthographic aspects (component 2).

### Criterion validity

The correlation between the sum of the scores obtained from the assessment of each text in the story assessment rubric (RER), and those for the standardized test PROESC, was very high  $r=0,79$  ( $p<0,01$ ). However, the teachers gave lower scores to the same texts ( $r=0,41$ ;  $p<0,01$ ). Nonetheless, this last correlation was higher than what we obtained between those scores and the results from the PROESC ( $r=0,30$ ;  $p<0,01$ ).



Specifically, table 7 shows the correlation between each of those criteria from the PROESC assessing similar aspects (the criterion “spelling” from the rubric was not included, since there are no similar criteria in the PROESC).

**TABLE 7.** Pearson correlation (r) between the specific criteria of both instruments.

RER	PROESC	r
Frame	Where and when Characters	0,8* 0,27*
Topic or initial event	Event with consequences	0,28*
Plot	Event with consequences	0,29*
	Coherent ending	0,57*
	Logical continuity	0,43*
	Sense of unity	0,46*
Creativity	Creativity	0,34*
	Literary figures	0,14
Sentences	Complex sentences	0,28*
Vocabulary	Vocabulary	0,48*
Spelling	-	-

Note: Statistically significant correlations: (\*)  $p < 0,01$ . Prepared by the authors.

### Internal consistency

The four elements comprising the *Organization and Content* dimension were analyzed for homogeneity, and a score of 0,57 was obtained in Cronbach’s Alpha, while for the *grammar, lexical and orthographic aspects* dimension, the result was 0,60. It is a satisfactory value, higher than what we obtained with the PROESC (0,37 in both dimensions). We also observed that if we eliminate a criterion from the rubric, the Alpha value does not improve.

## Interjudge reliability

To calculate the reliability in the implementation of the rubric, the degree of agreement between three evaluators was analyzed over a sample of 30 randomly chosen compositions. This was done after various training sessions for its implementation. The reliability indices obtained from Cohen's kappa method were higher than 0,85 ( $p < 0,01$ ) for all criteria. Again, the results were slightly higher than those obtained with the PROESC.

## Conclusions

The formative assessment of writing is a common classroom practice. It relies largely on the subjective opinions of teachers to determine which aspects demonstrate good writing (Barton, Hamilton, & Ivanic, 2000). The literature review we did in the introduction of the present study shows the lack of objective instruments with sufficient validity and reliability. These instruments should also be practical and easily available for teachers. The ordinal scales of measurement in descriptive format (i.e. rubric) are well adapted to this objective: on one side because they offer concrete indications about the students' level of achievement within the specific criteria of a skill or task assessment, as well as what they can improve next in their learning progression. On the other side its functional and intuitive design (usually on a single sheet) facilitates its use in the classroom, not only for teachers, but for students as well.

To this effect, the main objective of this study was to design and validate a story assessment rubric (RER) for the first years of compulsory education.

The data provided in the Results section support the rubric's content validity, which suggests that the dimensional structure, operationalization of criteria, and levels of achievements are relevant and representative of the construct for its evaluation in the three levels of schooling selected. The experts (researchers and education professionals) who had the opportunity of reviewing and using the story assessment rubric (RER), valued its construction and usefulness very positively. The factor analysis of the data gathered from its application to a large sample of texts, confirmed the adequacy of the structure with two factors or dimensions

containing all the assessment criteria. The application of the rubric to the assessment of texts written by students of different ages also allowed us to confirm that the instrument satisfactorily discriminates the progressive maturation in narrative skills, expected in students of 9 to 14 years of age. In addition, the global correlation of 0,79 with a standardized test (the PROESC) can be considered as an adequate *criterion validity* indicator (Roblyer & Wiencke, 2003). Another important validity indicator is the correlation with the grades that the teachers gave to these texts (Gearhart *et al.*, 1995), which was even higher than what was registered with the PROESC. We also obtained positive results from various evaluators regarding the instrument's reliability, in terms of internal consistency and the objectivity of its application. As regards consistency (measured by the Cronbach's Alpha test), the results are within the parameters recommended by many researchers in their review of research on rubrics (Graham & Perin, 2007; Jonsson & Svingby, 2007; Knoch, 2009). When applying Cohen's kappa coefficient to our rubric (higher than 0,85 for all criteria), the results were also better than for most of those studies.

In conclusion, the rubric in this study is a simple, valid, and reliable instrument for the assessment of texts written by students of 9-14 years of age. Given its characteristics and our previous experience (Montanero *et al.*, 2014), we can assume it has the potential to support formative assessment and co-assessment in the classroom.

In general, the difference of quality found between the texts of different levels of education was in line with our expectations. However, we must point out the lack of students in the last years of compulsory education who are able to write complex and well-articulated plots with truly creative elements. In particular, there is a weak progression in the rhetorical skills for most students in the first years of secondary education, a fact that was pointed out by previous studies in our education system (Artiles & Jiménez, 2007; Salvador, 2000). A possible explanation for this result is that the story assessment rubric (RER) responds to the classic structure of a simple and linear story. It does not identify a variety of resources (e.g. the analepsis, the contrast of characters in different times, etc.) that more advanced writers might use to enrich the story; this limits its discriminative capacity for such a wide range of ages. In future studies, we would like to adapt and validate different versions of the RER for an age range of only two years.

In any event, the incorporation of instruments facilitating the formative assessment of this competence in the classroom could contribute to its improvement. To confirm the usefulness of this instrument, we plan to design a longitudinal research in order to analyze the evolution of quality in the texts of students participating in assessment activities with this rubric, compared with others who do not.

## Bibliographical references

- Artiles, C., & Jiménez, J. E. (2007). *Proescri Primaria. Prueba de evaluación de procesos cognitivos en la escritura*. Canarias: Consejería de Educación, Cultura y Deportes del Gobierno de Canarias. Dirección General de Ordenación e Innovación Educativa.
- Baker, E.L., Abedi, J.; Linn, R. L., & Niemi, D. (1995). Dimensionality and generalizability of domain-independent performance assessments. *Journal of Educational Research* 89 (4), 197–205.
- Barton, D., Hamilton, M., & Ivanic, R. (2000). *Situated Literacies: reading and writing in context*. Routledge: London.
- Benítez, R., & García, G. (2010). El razonamiento analógico verbal: una habilidad cognitiva esencial de la producción escrita. *Onomázein*, 22 (2), 165-194.-
- Berninger, V.W, Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing: An interdisciplinary Journal*, 4, 257-280.
- Botvin, G. J., & Sutton-Smith, B. (1977). The development of structural complexity in children's fantasy narratives. *Developmental Psychology*, 13 (4), 377-388.
- Cremin, T., & Myhill, D. (2012). *Writing Voices: Creating communities of writers*. London: Routledge.
- Cuetos, F., Ramos, J. L., & Ruano, E. (2002). *PROESC: Batería de evaluación de los procesos de escritura*. Madrid: TEA Ediciones.
- Dunsmuir, S. Kyriacou, M. Batuwitige, S. Hinson, E. Ingram, V., & O'Sullivan, S. (2015). An evaluation of the Writing Assessment Measure (WAM) for children's narrative writing. *Assessing Writing*, 23 (2015) 1–18.

- East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, 14, 88-115. DOI: 10.1016/j.asw.2009.04.001
- Fang, Z., & Wang, Z. (2011). Beyond rubrics: Using functional language analysis to evaluate students writing. *Australian Journal of Language and Literacy*, 34 (2), 147-165.
- Fernández, M. J., Lucero, M., & Montanero, M. (2016). Rojo sobre negro. ¿Cómo evalúan los maestros las redacciones de sus estudiantes? *Revista de educación*, 372 (abril-junio), 63-82.
- Fernández, M. J. (2017). Co-evaluación de textos narrativos en la Educación Obligatoria. Universidad de Extremadura, Badajoz, España.
- Ferris, D. R., & Hedgcock, J. (2014). *Teaching ESL composition: Purposes, process, and practice* (3<sup>rd</sup> ed.). Mahwah, NJ: Lawrence Erlbaum.
- Gallavan, N., G. & Kottler, E. (2009). Constructing rubrics and assessing progress collaboratively with social studies students, social studies students. *The Social Studies*, 100 (4), 154-158.
- Galve, J. (2005). BECOLE. *Batería de Evaluación Cognitiva de la Lectura y Escritura*. Instituto de Orientación Psicológica. Madrid: Editorial EOS.
- García, J. N., & Marbán, J. M. (2001): Instrumento de Evaluación Individual de los Procesos Cognitivos de la Escritura (IEPCE). En J. N. García, *Dificultades de aprendizaje e intervención psicopedagógica* (pp. 141-149). Barcelona: Ariel.
- Gearhart, M., Herman, J. L., Novak, J. R., & Wolf, S. A. (1995). Toward the instructional utility of large-scale writing assessment: Validation of a new narrative rubric. *Assessing Writing*, 2, 207-242.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99 (3), 445-476.
- Graham, S., Harris, K.R., & Fink-Chorzempa, B. (2002). Contributions of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology*, 94, 669-686.
- Iandolo, G., Esposito, G., & Venuti, P. (2012). Cohesión, micro-organización, estructura narrativa y competencias verbales entre tres y once años: el desarrollo narrativo formal. *Estudios de Psicología*, 34(2), 141-160.
- Janssen, G., Meier, V., & Trace, J. (2015). Building a better rubric: Mixed methods rubric revision. *Assessing Writing*, 26, 51-66. DOI: 10.1016/j.asw.2015.07.002

- Jonsson A., & Svingby, G. (2007). The Use Of Scoring Rubrics: Reliability, Validity And Educational Consequences, *Educational Research Review*, 2, 130–144. DOI: 10.1016/j.edurev.2007.05.002
- Kernan, Keith T. (1977). Semantic and expressive elaboration in children's narratives. *Child discourse*, 91-102.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275-304. DOI: 10.1177/0265532208101008
- Knoch, U., Read, J., & Von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, 12(1), 26–43. DOI: 10.1016/j.asw.2007.04.001
- Kohn, A. (2006). The trouble with rubrics. *English Journal*, 95(4), 12–15.
- Li, J., & Lindsey, P. (2015). Understanding variations between student and teacher application of rubrics. *Assessing Writing*, 26, 67–79. DOI: 10.1016/j.asw.2015.07.003
- Lucero, M., Fernández, M. J., & Montanero, M (2018). Assessment of short stories in Elementary and Secondary Education. A study of teachers' responses to students' texts. *Studies in Educational Evaluation*, 59, 158-167.
- Montanero, M., Lucero, M., & 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 (1), 184-220. DOI: 10.1080/02103702.2014.881653
- Ochoa-Angrino, S., Aragón, L., Correa, M., & Mosquera, S. (2008). Funcionamiento metacognitivo de niños escolares en la escritura de un texto narrativo antes y después de una pauta de corrección conjunta [Versión electrónica], *Acta Colombiana de Psicología*, 11 (2), 77-88.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(0), 129-144. DOI: <http://dx.doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences* (0). DOI: 10.1016/j.lindif. 2012.04.007
- Roblyer, M. D., & Wiencke, W. R. (2003). Design and use of a rubric to assess and encourage interactive qualities in distance courses. *American Journal of Distance Education*, 17, 77–99.

- Ryan, M. (2014). Reflexive writers; Re-thinking writing development and assessment in schools. *Assessing Writing*, 22, 60-74. DOI:10.1016/j.asw.2014.08.002
- Salvador, F. (2000). Habilidades narrativas de alumnos de Educación Primaria en la producción de textos escritos. *Enseñanza*, 17-18, 145-163.
- Schenck, A. D., & Daly, E. (2012). Building a better mousetrap: Replacing subjective writing rubrics with more empirically-sound alternatives for EFL learners. *Creative Education*, 3 (8), 1310-1325. DOI: 10.4236/ce.2012.38193
- Spandel, V. (2006). Speaking my mind: In defense of rubrics. *English Journal*, 96 (1), 19-22.
- Stadler, M. A., & Ward, G. C. (2005). Supporting the narrative development of young children. *Early Childhood Education Journal*, 33, 2, 73-80.
- Stein, N., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. En R. O. Freedle (Ed.), *Advances in discourse processing*. Vol. 2. Norwood, Nueva Jersey: Ablex.
- Sundeen, T. H. (2014). Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, 21, 74-88. DOI: 10.1016/j.asw.2014.03.003
- Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9, 77-110.
- Toro, J., & Cervera, M. (1984). *TALE. Test de análisis de la lectoescritura*. Madrid: Aprendizaje Visor.
- Toro, J., Cervera, M., & Urío, C. (2002). *Escalas Magallanes de Lectura y Escritura. TALE-2000*. Vizcaya: Albor-COHS.
- Yan, C. M. W., McBride-Chang, C., Wagner, R. K., Zhang, J., Wong, A. M. Y., & Shu, H. (2012). Writing quality in Chinese children: speed and fluency matter. *Reading and Writing*, 25 (7), 1499-1521.

**Contact address:** María Jesús Fernández Sánchez, Universidad de Extremadura, Facultad de Educación, Departamento de Ciencias de la Educación, Avda. de Elvas s/n. 06006 Badajoz, España. E-mail: mafernandezs@unex.es