

Rojo sobre negro ¿Cómo evalúan los maestros las redacciones de sus estudiantes?

Red on black. How do teachers assess their students' narrative texts?

María-Jesús Fernández
Manuel Lucero
Manuel Montanero



IV CENTENARIO
DE LA MUERTE DE
CERVANTES



GOBIERNO
DE ESPAÑA

MINISTERIO
DE EDUCACIÓN, CULTURA
Y DEPORTE

revista de
eEDUCACIÓN

Nº 372 ABRIL-JUNIO 2016



IV CENTENARIO
CERVANTES

revista de **EDUCACIÓN**

Nº 372 Abril-Junio 2016

Revista trimestral

Fecha de inicio: 1952



MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE

SECRETARÍA DE ESTADO DE EDUCACIÓN, FORMACIÓN PROFESIONAL Y UNIVERSIDADES

Instituto Nacional de Evaluación Educativa

C/ San Fernando del Jarama, 14

28002 Madrid

Teléfono (+34) 91 745 92 39

revista@mecd.es

Edita

© SECRETARÍA GENERAL TÉCNICA

Subdirección General de Documentación y Publicaciones.

Catálogo de publicaciones del Ministerio: mecd.gob.es

Catálogo general de publicaciones oficiales: publicacionesoficiales.boe.es

Edición: 2016

NIPO línea: 030-15-016-X

NIPO ibd: 030-15-017-5

ISSN línea: 1988-592X 0034-8082

ISSN papel: 0034-8082

Depósito Legal: M.57/1958

Diseño de la portada: Dinarte S.L.

Maqueta: Solana e hijos, Artes Gráficas S.A.U.

Rojo sobre negro. ¿Cómo evalúan los maestros las redacciones de sus estudiantes?¹

Red on black. How do teachers assess their students' narrative texts?

DOI: 10.4438/1988-592X-RE-2015-372-315

María-Jesús Fernández

Manuel Lucero

Manuel Montanero

Universidad de Extremadura

Resumen

Uno de los recursos educativos más extendidos para la mejora de la competencia escritora consiste en evaluar y corregir los textos que los propios estudiantes redactan. Sin embargo, no contamos con estudios que analicen qué criterios tienen fundamentalmente en cuenta los maestros en dicha evaluación y qué tipo de retroalimentación proporcionan. En este sentido, en el presente estudio se describen los mensajes evaluativos que 21 maestros emitieron sobre las composiciones narrativas de 236 estudiantes de los últimos cursos de Educación Primaria, con el objetivo de ayudarles a mejorar su competencia escritora. En total, se analizaron 7851 registros verbales y gráficos de evaluación, en función de su presentación formal y su contenido evaluativo y meta-textual. Los resultados ponen de manifiesto un claro predominio de mensajes de tipo verbal, centrados principalmente en la corrección específica de aspectos ortográficos y gramaticales. La frecuencia de este tipo de correcciones se mostró como el mejor predictor de la calificación global que el profesorado otorgaba a las narraciones. Por el contrario, se constata una escasa presencia de mensajes que faciliten la comprensión y la autoevaluación de los errores cometidos, así como sugerencias de mejora que involucren aspectos macro-estructurales del

⁽¹⁾ **Agradecimientos:** Al Gobierno de Extremadura y al Fondo Social Europeo por la financiación recibida (PD12129).

texto. Finalmente, se discuten estos hallazgos y sus implicaciones en la enseñanza y evaluación de la competencia escritora. De estos resultados se derivan implicaciones relevantes para la enseñanza y evaluación de la escritura, que se discuten al final del artículo.

Palabras-clave: Evaluación del estudiante, Redacción, Ortografía, Gramática, Educación Básica.

Abstract

The correction of the texts written by students is one of the educative tool most frequently used for improving the writing skill. However, there is no research that analyses what criteria are to be considered by teachers in this kind of assessment and what kind of *feedback* is provided by them. In this sense, this study describes the improvement suggestions from 21 teachers about narrative texts written by 236 students in the last years of Primary Education, in order to help them improve their writing skills. A total number of 7851 verbal assessment records and charts were analysed, depending on their formal presentation and their assessment of meta-textual content. The results show a great predominance of verbal messages, focused mainly on spelling and grammar improvement. The frequency of such corrections could explain why teachers give higher scores to texts with better grammar and spelling use. On other hand, a low number of suggestions to try to develop the understanding and self-assessment of mistakes were also logged, as well as those related to improving macro-structural aspects of a text. Hence, the current study points out the implications for teaching and assessing writing, which are discussed in the conclusion of the article.

Keywords: Student evaluation, Writing, Spelling, Grammar, Basic Education.

Introducción

El desarrollo de la competencia escritora durante la escolarización obligatoria es una temática de interés para legisladores, investigadores y profesores, debido a que la escritura está ligada al éxito escolar de los estudiantes (Lennart, 2014). Los maestros dedican, de hecho, una gran cantidad de tiempo y esfuerzo a actividades relacionadas con la evaluación de la competencia escritora.

La práctica evaluativa de la escritura de narraciones en la Educación Primaria

Tradicionalmente el *feedback* proporcionado oralmente o por escrito ha sido considerado una parte importante de la instrucción y evaluación de la escritura (Parr y Timperley, 2008). Sin embargo, aunque se han investigado extensamente las revisiones y comentarios realizados en el contexto de actividades de autoevaluación y co-evaluación entre iguales (Cho y MacArthur, 2010; Fahimi y Rahimi, 2015; Lennart, 2014; Rouhi y Azizian, 2013), muy pocos estudios se han ocupado de analizar cómo los profesores evalúan los borradores de los alumnos en L1. Algunos trabajos han cuestionado la fiabilidad y validez de este tipo de evaluación, argumentando que, a pesar de su eficiencia para tener una impresión general de la producción escrita de los estudiantes, proporcionan una información diagnóstica limitada sobre los diversos componentes de la competencia escritora (Dunsmuir *et al.*, 2015). Otros, en cambio, ponen de manifiesto la relación entre la habilidad del profesor para proporcionar un *feedback* de calidad y el progreso del estudiante en el aprendizaje de la escritura de textos narrativos (Fife y O'Neill, 2001). La evaluación de los profesores suele incluir, de hecho, la corrección de fragmentos del texto original del estudiante (Peterson y McClay, 2010). Sin embargo, que la revisión conduzca a una mejora en la re-escritura del texto (en caso de que se diera esta opción) depende, no solo de la capacidad del escritor, sino también de la calidad de la retroalimentación que éste recibe (Tsui y Ng, 2000). En un estudio previo encontramos que el 67% de los cambios que una amplia muestra de estudiantes de Educación Primaria (de entre 10 y 12 años) introducía en una segunda versión de sus narraciones, tras recibir la evaluación de su profesor, se centraba en aspectos puramente gramaticales y ortográficos (Montanero, Lucero y Fernández, 2014). Otros estudios apuntan en la misma dirección al mostrar que la naturaleza de la evaluación del profesorado tiende a focalizarse en características muy superficiales del texto (Connors y Lunsford, 1993; Hargreaves y McCallum, 1998; Schwartz, 1984; Sommers, 2011). Aspectos retóricos, como la cohesión de las ideas, son, por el contrario, tradicionalmente ignorados, tanto en la evaluación como en la enseñanza de la competencia escritora (Struthers, Lapadat y MacMillan, 2013).

Algunos autores defienden que este tipo de prácticas evaluativas se relacionan con una escasa consciencia de los evaluadores sobre el conjunto de procesos que intervienen en la redacción de un texto (Cho, 2003; Dempsey, PytlikZillig y Bruning, 2009); así como sobre aquellos criterios que permiten juzgar su idoneidad, en función de sus objetivos

(Cassany, 2000; Morales, 2004). Una buena evaluación de los aspectos más profundos de una redacción requiere un complejo conocimiento sobre escritura, de la que incluso muchos de los escritores más competentes carecen: un meta-conocimiento explícito de la lengua, de las estructuras de los textos, de los procesos y estrategias de escritura; además del uso de métodos, actividades y recursos eficaces para ayudar a los alumnos con dificultades (Parr y Timperley, 2008). Los profesores que no dispongan de este compendio de conocimientos no sólo serán menos hábiles para formular sugerencias de mejora efectivas, sino que tendrán dificultades para anticipar los problemas de escritura que pueden tener los estudiantes (Jones y Moreland, 2005; Shepard et. al, 2005).

Además de los criterios y del contenido metalingüístico de las evaluaciones, otros aspectos relevantes de la evaluación, que no se han documentado suficientemente en esta etapa educativa, tienen que ver con el modo en que se expresan, «rojo sobre negro», en las propias redacciones. Los comentarios evaluativos pueden aparecer en forma de notas al margen, preguntas en las que el docente solicita información concreta, aclaraciones sobre aspectos globales y, sobre todo, correcciones (Ferris, 1997). Sabemos que los estudiantes de Educación Superior prefieren las preguntas abiertas y los comentarios que incluyen justificaciones y explicaciones de los que hay que mejorar (Straub, 1997). Sin embargo, no contamos con estudios que describan pormenorizadamente la práctica evaluativa, ni su impacto específico en los estudiantes de Primaria, en nuestro sistema educativo.

Modelos de análisis de los mensajes de evaluación de composiciones

La literatura recoge diversos sistemas de categorización de los mensajes de *feedback* que evaluadores (profesores o iguales) registran en las redacciones que revisan. Uno de los más citados es el que propuso Min (2003), basado a su vez en el clásico trabajo de Berg (1999). En cuanto a su naturaleza, los mensajes pueden clasificarse en *orientados* o *no orientados* a la revisión. Dentro de los primeros, cabe distinguir entre las *evaluaciones* o juicios que no conllevan sugerencias concretas de modificación del texto; las *clarificaciones*, que se expresan principalmente a través de preguntas sobre lo que el evaluador no entiende (aunque también se incluyen en esta categoría incitaciones a re-elaborar una intención comunicativa o el significado de un fragmento del texto); las *alteraciones*, o correcciones de supuestos errores; y las *sugerencias* de

información que ayuda al compañero a enriquecer o autocorregir el texto. Por último, en función del área, los mensajes pueden tener un carácter *global* (habitualmente referido al contenido) o *local*.

Cho y MacArthur (2010) utilizaron un doble sistema de categorías, más completo que el anterior. El análisis de los mensajes de *feedback* permite recoger otros matices, tales como la explicitud de la sugerencia de modificación (*directiva* vs. no *directiva*) o el sentido positivo (elogio) o crítico de la evaluación. Las modificaciones realizadas por los evaluadores se clasifican en tres niveles de complejidad: las *reparaciones simples*, en un nivel superficial, afectan a cuestiones, gramaticales u ortográficas; en un nivel microestructural, las reparaciones pueden ser *complejas* (borrando o cambiando contenido semántico de una oración o párrafo) o *extendidas* (añadiendo oraciones que precisan información); en un nivel macroestructural, pueden añadir nuevos *contenidos* o tópicos, o modificar la *organización* (títulos, movimientos de párrafos, frases conectivas), además de la adición o modificación de materiales (tablas, figuras...).

Por su parte, Liang (2010) propone un sistema de 5 categorías: *negociación* (cuya definición coincide aproximadamente con las clarificaciones de los trabajos anteriores); *discusión de contenido* (similares a las *sugerencias*, son expresiones de desacuerdo, que pueden ir seguidas de una justificación y/o propuesta); *corrección de errores* (tanto de contenidos, como gramaticales y ortográficos); *gestión* de la tarea y *comentarios* no relacionados con la revisión.

Estos antecedentes, no obstante, se centran principalmente en el aprendizaje de la escritura en el inglés como segunda lengua (L2) en niveles de Educación Secundaria o superior. Además, los datos que recogen estas investigaciones se basan en revisiones entre iguales, efectuadas por escrito, ya sea en papel o con el apoyo de nuevas tecnologías. Se echa en falta una aplicación de estos sistemas de análisis a las evaluaciones que realizan los propios profesores en los niveles educativos correspondientes a la Educación Primaria en el área de Lengua Castellana.

Objetivos de la investigación

En este marco, la presente investigación tiene dos objetivos: por un lado, describir las características de las calificaciones y comentarios evaluativos que los maestros de Primaria de nuestro sistema educativo registran sobre las narraciones que redactan sus alumnos; por otro lado, explorar la relación entre dichos comentarios evaluativos y las calificaciones globales

que se otorga a los relatos, así como los criterios específicos de evaluación que implícitamente traslucen.

Diseño y metodología

Participantes

En el estudio participaron 21 maestros-tutores de Primaria (15 mujeres y 6 hombres), de 8 colegios de distintas ciudades de Extremadura (5 públicos y 3 privados-concertados), seleccionados según la disponibilidad de los investigadores. Todos ellos contaban con más de 5 años de experiencia docente en los últimos cursos de Primaria. Cada maestro evaluó entre 5 y 20 textos de tipo narrativo, redactados por 236 alumnos de entre 10 y 12 años (tabla I).

TABLA I. Muestra

Centro	Evaluador/a	Nº de redacciones evaluadas
Centro 1	Maestro 1	5
	Maestra 2	12
	Maestra 3	7
	Maestro 4	5
Centro 2	Maestro 5	8
	Maestra 6	7
Centro 3	Maestra 7	12
	Maestra 8	14
Centro 4	Maestro 9	5
Centro 5	Maestra 10	5
	Maestra 11	5
	Maestra 12	5
	Maestra 13	5
Centro 6	Maestra 14	20
	Maestro 15	20
	Maestra 16	20
Centro 7	Maestra 17	19
	Maestra 18	9
	Maestra 19	20
	Maestro 20	20
Centro 8	Maestra 21	13

Fuente: Elaboración propia.

Procedimiento de recogida y segmentación de los registros evaluativos

La investigación se basó en el análisis documental de los registros evaluativos, expresado con bolígrafo por los profesores en los mismos papeles donde se habían escrito las narraciones.

Las narraciones fueron redactadas en unos 30 minutos por alumnos de 5º y 6º curso de Primaria. La temática sobre la que debían escribir, en aproximadamente un folio de extensión, fue la siguiente: “un niño que se convierte en héroe de su pueblo”. Los textos fueron repartidos aleatoriamente entre los 21 profesores, sin darles a conocer los propósitos de la investigación. Se les instó a que las corrigieran con un bolígrafo de otro color, como habitualmente hacían, registrando en el mismo manuscrito cualquier valoración que considerasen oportuna.

La segmentación de las unidades de análisis se realizó siguiendo un criterio semántico-pragmático. Cada unidad de análisis se conformó por un mensaje con un solo contenido meta-evaluativo (o bien, referido a un solo error o tipo de error), aunque estuviera insertado en un enunciado más amplio. Por ejemplo, la siguiente oración, empleada por el profesor 17, se segmentó en tres unidades de análisis: “Un poco sucio (1), está bien (2), cuidado con las faltas, piensa antes de escribir las palabras (3)”.

Procedimiento de categorización de los mensajes evaluativos

Para clasificar los mensajes utilizamos un sistema de categorías original (tabla II), elaborado por un procedimiento inductivo-deductivo, de *comparación permanente*. En primer lugar, a partir de la revisión de los modelos de análisis que hemos descrito en la introducción, se seleccionaron y concretaron, con igual o semejante denominación, las categorías empleadas para analizar el contenido evaluativo y meta-textual de los mensajes de *feedback*. En segundo lugar, tras un primer estudio de las más de 7000 unidades de análisis de nuestra muestra, se identificaron aquellas que no podían clasificarse con suficiente precisión y se añadieron nuevas categorías que no habían sido utilizadas explícitamente en investigaciones anteriores. Finalmente, se incorporó al sistema dos nuevas dimensiones de análisis, forma y ubicación de los mensajes, que tampoco habían sido contempladas en dichos antecedentes.

Para analizar los aspectos formales, se tuvo en cuenta, por un lado, la forma de expresión, verbal o no verbal, del registro evaluativo; y por otro

lado, la ubicación de dicho registro (más o menos encima del foco de la evaluación, en los márgenes, al principio o al final del texto).

Para el análisis de la amplitud y el contenido evaluativo de los mensajes identificamos una serie de categorías, basadas principalmente en el modelo de Min (2003). Así, en cuanto a la amplitud de los mensajes evaluativos, incluimos lo que Min y otros autores denominan *feedback global* y *local*. En cuanto a su contenido evaluativo, las «evaluaciones», sin contenido de mejora, se clasificaron en dos subtipos: *señalizaciones* y *calificaciones*. Siguiendo la propuesta de Cho y MacArthur (2010), dentro de estas últimas operativizamos un criterio para discriminar entre calificaciones positivas y negativas. Por su parte, dada su naturaleza interrogativa las «clarificaciones» se identifican en nuestro sistema como *preguntas*. Lo que Min calificaba indistintamente como «alteraciones» se han concretado en dos tipos de correcciones del texto: *genéricas* y *específicas*. Al término, un tanto ambiguo, de «sugerencia» le hemos añadido *de ampliación o alternativa*. Los mensajes «no orientados» los hemos incluido, finalmente, en la categoría *otros*. Nuevamente, el análisis inicial de los mensajes evaluativos nos permitió identificar un tipo de mensaje muy relevante, la *justificación*, que puede acompañar o no a una determinada corrección, y que no había sido considerada independientemente como categoría de análisis en las investigaciones previas que hemos revisado.

Para identificar las categorías de la dimensión meta-textual nos fundamentamos en la propuesta de Cho y MacArthur (2010). Las llamadas reparaciones superficiales o «simples» se han clasificado en dos tipos: gramaticales y ortográficas. En cambio, las complejas y extendidas, se han integrado en una sola categoría (contenido *semántico-léxico*), dado que el matiz que las diferencia queda ya recogido en la categoría *ampliación* de la dimensión anterior. Las reparaciones macro-estructurales de temas y organización se han identificado genéricamente como evaluaciones de contenido *organizativo*. Para clasificar otros contenidos evaluativos que no se ubican con claridad en ninguna de las anteriores categorías, como los comentarios sobre la caligrafía, la creatividad o cuestiones relativas a la gestión de la tarea (Liang, 2010), hemos incorporado las categorías de contenido *inespecífico* y *otros* (ver tabla II).

TABLA II. Sistema de categorías de *feedbacks* de la evaluación de textos

Dimensión	Categoría	Definición
A. Forma	1. Verbal	Expresiones verbales que contienen un conjunto de letras, palabra/s o frase/s. Incluye las tildes, los signos de puntuación y las siglas.
	2. No verbal	Expresiones numéricas (de uno o más dígitos) o gráficas, ya sean simples o complejas (tachado, subrayado lineal o marginal, flechas, óvalos).
	3. Mixta	<i>Feedbacks</i> que integran una expresión verbal y otra gráfica o numérica en la misma anotación.
B. Ubicación	1. Superpuesto	Anotación encima del contenido textual evaluado. Incluye el interlineado.
	2. Marginal	Glosa marginal o anotación superpuesta sobre un fragmento distinto del texto pero vinculado explícitamente a él.
	3. Desvinculado	Anotación registrada sin vinculación explícita con ningún fragmento del texto (por ejemplo al principio o final del texto) o en un papel diferente.
C. Amplitud	1. Global	<i>Feedbacks</i> sobre el conjunto del texto
	2. Local	<i>Feedbacks</i> sobre uno o varios fragmentos del texto
D. Contenido evaluativo	1. Señalización	Símbolos o expresiones verbales en los que únicamente se identifica un fragmento del texto, ya sea erróneo o acertado, sin aportar nada más (por ejemplo, se subrayan faltas de ortografía sin corregirlas)
	2. Calificación	Expresiones numéricas, gráficas (flecha hacia arriba, letras que representan valores de una escala ordinal) o verbales, que conllevan únicamente un juicio de calidad. Pueden tener una valencia positiva (como las numéricas superiores al 70% de la escala empleada), intermedia (como la letra R o las numéricas entre 50-70% de la escala) o negativa.
	3. Pregunta	Símbolos o expresiones verbales interrogativas de duda o en las que se solicita aclaración
	4. Corrección específica	Expresión gráfica o verbal que, además de señalar un error, proporciona una instrucción específica o una alternativa mejor (incluye la supresión o adición necesaria de un fragmento)
	5. Corrección genérica	Expresión verbal que identifica una carencia o error que se repite (sin señalarlo con precisión en el texto) y proporciona una instrucción genérica de mejora.

TABLA II. Continuación

Dimensión	Categoría	Definición
D. Contenido evaluativo	6. Ampliación/ alternativa	Comentario verbal en el que no se identifica un error pero que sugiere, ejemplifica o proporciona directamente una alternativa o ampliación no imprescindible del texto.
	7. Justificación	Comentario verbal en el que se explica o argumenta una alternativa (incluye los comentarios normativos).
	8. Otros comentarios	Otros comentarios verbales, como advertencias o medidas de refuerzo.
E. Contenido meta-textual	1. Organizativo	<i>Feedbacks</i> sobre la organización retórica del texto o sus párrafos. En las evaluaciones de textos narrativos cabe distinguir, más e específicamente entre marco, tema y trama.
	2. Semántico-léxico	<i>Feedbacks</i> sobre tópicos o ideas del texto. Incluye la evaluación de la coherencia o de vocabulario específico.
	3. Gramatical	<i>Feedbacks</i> sobre la construcción morfosintáctica de las oraciones. Incluye la alteración del orden de palabras, la evaluación de la cohesión o de los signos de puntuación.
	4. Ortográfico	<i>Feedbacks</i> sobre la construcción ortográfica de las palabras
	5. Otros	Comentarios sobre otros aspectos meta-textuales específicos, como los aspectos formales (caligrafía, margen) o la creatividad.
	6. Inespecíficos	<i>Feedbacks</i> globales, transversales o en los que no se expresa un contenido meta-textual específico. No incluye aquellas anotaciones que no tienen un carácter evaluativo.

Fuente: Elaboración propia.

La categorización de las unidades de análisis se realizó tras un proceso de entrenamiento en el que los investigadores afinaron el sistema de categorías y el grado de apertura de las mismas, a fin de asegurar la consistencia interna entre investigadores.

Para analizar la fiabilidad del proceso se calculó el grado de acuerdo entre dos evaluadores sobre una muestra de 7 textos, escogidos al azar, que contenían un total de 222 mensajes que categorizar. Como se aprecia en la siguiente tabla, los índices obtenidos mediante Kappa de Cohen son muy elevados en todas las dimensiones, salvo en cuanto a la forma y al contenido meta-textual, que pueden considerarse aceptables.

TABLA III. Datos de fiabilidad en la identificación de las categorías de cada una de las dimensiones de análisis ($p < 0,01$).

Dimensiones	Forma	Ubicación	Amplitud	Contenido evaluativo	Contenido meta-textual
Índice KC	0,85	0,97	0,97	0,96	0,82

Fuente: Elaboración propia.

Procedimiento de análisis cuantitativo y cualitativo

Para describir el perfil de evaluación de composiciones predominante en el profesorado de Primaria, se comparó la frecuencia y porcentaje de registros evaluativos clasificados en cada categoría.

Para analizar en qué medida el contenido meta-textual de los comentarios evaluativos predice la calificación global de las composiciones (que generalmente aparecían destacadas y localizadas en la esquina superior derecha del documento), se realizó un análisis de regresión entre dos variables: las *calificaciones* y la frecuencia de correcciones de cada categoría de la dimensión contenido meta-textual. Dado que el profesorado emplea diversas escalas de calificación, se equipararon todas en una misma escala de valores numéricos impares de 1 a 9, tal y como recoge la tabla IV.

TABLA IV. Equiparación de las diversas escalas de calificación en una escala estándar

Estándar	Cualitativa 1	Cualitativa 2	Númerica
1	Muy mal	Insuficiente	1-2
3	Mal	Suficiente	3-4
5	Regular	Bien	5-5
7	Bien	Notable	7-8
9	Muy bien	Sobresaliente	9-10

Fuente: Elaboración propia

Finalmente, con objeto de profundizar en el análisis de los mensajes evaluativos que abordaban aspectos macroestructurales se analizaron cualitativamente los registros evaluativos, particularmente los de tipo organizativo.

Resultados

En las 236 narraciones evaluadas por los maestros participantes en la investigación, se identificaron un total de 7851 unidades de análisis. La tabla V refleja los resultados de la categorización de dichos mensajes de evaluación en función de las diferentes dimensiones consideradas en el estudio.

TABLA V. Frecuencia y naturaleza de los comentarios evaluativos

Dimensión	Categoría	Registros	Profesores
Forma	Verbal	5956 (75,9%)	20 (95,2%)
	No verbal	1120 (14,3%)	20 (95,2%)
	Mixta	775 (9,8%)	20 (95,2%)
Ubicación	Superpuesto	7096 (90,4%)	20 (95,2%)
	Marginal	123 (1,6%)	10 (47,6%)
	Desvinculado	632 (8,0%)	21 (100%)
Amplitud	Global	600 (7,6%)	21 (100%)
	Local	7251 (92,4%)	20 (95,2%)
Contenido evaluativo	Señalización	219 (2,8%)	13 (61,9%)
	Calificación	243 (3,1%)	21 (100%)
	Pregunta	54 (0,7%)	14 (66,7%)
	Corrección específica	6946 (88,5%)	20 (95,2%)
	Corrección genérica	293 (3,7%)	19 (90,5%)
	Ampliación/alternativa	24 (0,3%)	10 (47,6%)
	Justificación	31 (0,4%)	9 (42,9%)
	Otros comentarios	41 (0,5%)	11 (52,4%)
Contenido meta-textual	Organizativo	91 (1,2%)	12 (57,1%)
	Semántico-léxico	976 (12,4%)	20 (95,2%)
	Gramatical	2260 (28,8%)	20 (95,2%)
	Ortográfico	4082 (52,0 %)	20 (95,2 %)
	Otros	230 (2,9 %)	18 (85,7 %)
	Inespecífico	212 (2,7 %)	20 (95,2 %)

Fuente: Elaboración propia.

Presentación formal de los mensajes evaluativos

La tabla V refleja que casi todos los profesores han empleado expresiones *verbales*, *no verbales* y *mixtas* para proporcionar sus *feedbacks*, aunque la mayoría de los mensajes evaluativos fueron *verbales* (75,9%). El más frecuente fue la corrección en el propio texto de la ausencia de tildes: “En un pueblo a poco tiempo de Madrid vivía Jose, el era un niño sencillo.” [Profesor 9]. Entre las *no verbales*, los símbolos más utilizados son las tachaduras y los subrayados o círculos. Por ejemplo, el profesor 21 rodea la palabra “entonces” en varias ocasiones para indicar que se repite con demasiada frecuencia. Los mensajes *mixtos* consistieron en combinaciones de símbolos gráficos, como flechas o tachaduras, con una corrección de tipo verbal. Por ejemplo, el profesor 8 sobre el fragmento “a el”, el profesor 8 tacha “el” y escribe una “l” pegada a la “a”.

En relación a la *ubicación*, la inmensa mayoría fueron los *superpuestos* (90,4%), habitualmente usados para corregir aspectos locales (el maestro 20, por ejemplo, escribe entre las líneas del texto “error de concordancia” para señalar que hay dos palabras que no concuerdan en número). En cambio, los mensajes *desvinculados* parecen reservarse para cuestiones globales de carácter calificativo. El profesor 21, por ejemplo, hace un comentario evaluativo global al final del texto: “*Muy bien diferenciadas las partes del cuento. Buena redacción. Repites: oyó*”. Resulta sorprendente la escasez de anotaciones marginales, probablemente debido al poco espacio que los estudiantes dejaban en los márgenes del texto. De hecho, más de la mitad de los profesores no las emplearon en ninguna ocasión. El profesor 20, por ejemplo, realiza una sugerencia al margen del texto para hacer una recomendación sobre los sentimientos de los personajes. “Este es un buen momento para que cuentes como se sentía la madre”. Ante la ausencia de guiones en el diálogo, el profesor 18 escribe en el margen derecho lo siguiente: “¡Los guiones del diálogo!”.

Finalmente, en cuanto a la amplitud de los mensajes, más del 90% estaban referidos aspectos *locales* del texto. Son bastante frecuentes los comentarios en el que se evalúan positivamente los fragmentos correspondientes a la trama del relato (“Muy bien la trama” [Profesor 8]); si bien, la mayoría consisten en correcciones ortográficas. Las pocas valoraciones de carácter más o menos *global* se centraron principalmente en calificaciones de la calidad percibida del conjunto de la redacción; o bien, errores superficiales, pero que se repiten demasiado a lo largo del texto, como la indicación del profesor 5 acerca de la presentación del texto: «Un poco sucio».

Contenido evaluativo de los mensajes

La inmensa mayoría de los mensajes analizados (88,5%) consistieron en correcciones de errores específicos. Un caso típico es la confusión de “ahí” con “hay”. El maestro número 14 pone entre paréntesis el adverbio mal empleado (“ahí”) y encima escribe un verbo («hay»). Otra situación habitual es la ausencia de tildes. El profesor número 21 corrige una falta de ortografía al escribir una tilde encima de la “i” de la forma verbal “tenía”. También son frecuentes las incorporaciones de signos de puntuación (comas y puntos) y la eliminación y sustitución de unas letras por otras. Pueden servir a modo de ejemplo: la corrección de un profesor que escribe una “ll” encima de la “y” que ha utilizado el autor del relato para escribir la palabra “medaya” [profesor 21] o la realizada por el maestro número 17 al eliminar la “h” en la expresión “ha correr”.

El resto de categorías de mensajes evaluativos no alcanzan el 5%. El porcentaje de correcciones generales (o de fragmentos amplios del texto) y las calificaciones están cercanas a este dato. Por ejemplo, ante la ausencia de signos de puntuación a lo largo del relato, el profesor 8 escribe la siguiente corrección general: “Utiliza puntos, comas...”. Otra corrección de índole general es la que realiza un maestro que felicita al autor del texto que ha corregido por la calidad caligráfica del mismo; para ello, escribe en la parte inferior de la hoja: “Buena letra” [profesor 18].

Frecuentemente este tipo de comentarios o instrucciones acompañan a alguna forma de calificación. Todos los docentes incluyeron entre sus mensajes evaluativos alguno que expresaba una calificación del texto, ya fuera global o local. Por ejemplo, en la parte superior de uno de los textos el profesor 14 escribe: “Suficiente”. Ante la ausencia de faltas de ortografía, el autor de otra historia es felicitado: “Muy bien la ortografía” [Profesor 13]. En total se registraron 243 mensajes calificativos. La mayor parte se expresaron de modo verbal (45,9%) o mixto (38,8%). Por ejemplo, muchos maestros escribieron en la parte superior del papel un adjetivo calificativo que luego subrayaban. Las calificaciones numéricas representan sólo un 15,3% del total (un 73% de las cuales, son números rodeados, y el resto subrayados). En cuanto a la valencia de las calificaciones, sólo un 16,1% fueron negativos; el resto puede considerarse como positivos o de un nivel intermedio (como «bien»).

Cabe destacar la escasez de comentarios que señalicen, justifiquen, soliciten aclaración o amplíen la corrección de un error. El profesor 7, por ejemplo, subraya la palabra «triavesuras» para señalar el error

ortográfico, sin llegar a corregirlo. El maestro número 4 rodea la palabra “moustro” para llamar la atención del autor del relato, aunque no realiza ninguna corrección sobre la misma. Sólo en ocasiones puntuales, algunos profesores escriben sus comentarios interrogativamente (por ejemplo, el profesor 3 formula varias preguntas junto a un fragmento introductorio del relato: “¿Dónde? ¿Cuándo?”); hacen una sugerencia no correctiva (“Podrías utilizar en más ocasiones la palabra héroe” [Profesor 17]); o justifican sus correcciones (El profesor 20, por ejemplo, ante un error en el que el alumno escribe “de el...”, proporciona la regla a tener en cuenta en estos casos: “de + el = del”). De hecho, más de la mitad de los maestros nunca justificaron sus apreciaciones ni realizaron sugerencias para ampliar o re-escribir determinados fragmentos del texto.

Finalmente, se registraron algunos comentarios de gestión, refuerzo, etc. que no podían clasificarse en ninguna de las categorías anteriores. Por ejemplo el maestro 14 escribe en una de las composiciones: “las tildes también son faltas de ortografía”. Ante el gran número de errores ortográficos que ha cometido el autor del texto, el mismo maestro escribe lo siguiente al final de otra de las composiciones: “En el cuaderno de lengua, las palabras rodeadas las copias 10 veces correctamente”.

Contenido meta-textual

La mayor parte de los comentarios evaluativos (casi un 80%) se refieren a aspectos gramaticales y ortográficos. En cuanto a los primeros cabe destacar la incorporación de puntos y comas a lo largo del texto y los referidos a la falta de concordancia gramatical entre vocablos; por ejemplo, el profesor número 20 escribe “error de concordancia” para señalar que hay una palabra en plural que no concuerda con el número del artículo que la precede (singular). En relación a los comentarios evaluativos de índole ortográfica puede servir a modo de ejemplo la corrección del profesor número 14 que indica: “Faltan muchas tildes”.

En cambio, el porcentaje de comentarios semántico-léxicos es muy bajo (12,4%). Veamos tres ejemplos: “No puedes repetir tanto la misma palabra” [Profesor 5]; “¿Lo de budistas aporta algo o tiene algo que ver?” [Profesor 17]; “Debes evitar repeticiones” [Profesor 18].

Los comentarios *organizativos*, por su parte, son muy pocos (1,2%). El porcentaje de profesores que realiza alguno de estos comentarios sobre la macro-estructura del texto, apenas la mitad de la muestra, es mucho

menor que en el resto de criterios de la dimensión *contenido meta-textual*. En las 236 narraciones evaluadas se identificaron un total de 91 registros evaluativos referidos a aspectos organizativos. Un 73,6% de los registros de contenido organizativo estaban expresados de manera verbal, frente a casi un 20% que, además de texto, incluían algún elemento gráfico para señalar el fragmento de texto en el que se encontraba el error. Un 71,4% se presentaba desvinculado del contenido al que se refería, mientras que un 7,7 % de los comentarios evaluativos aparecían superpuestos y un 20,9% tenía una ubicación marginal. En su mayoría se trataba de registros que se referían a la totalidad del relato (63,7% eran globales); si bien, se expresaban como correcciones específicas (46,2%) o genéricas (23,1%).

En cuanto a los componentes de la estructura narrativa objeto de evaluación, encontramos que 24% de los registros aluden al marco del cuento (descripción de los personajes, espacios y tiempo). Así, el profesor 7 sugiere mejoras ante la ausencia de algunos elementos del marco: “Faltan datos del niño, del lugar y del tiempo en que ocurrió”. Ante la ausencia de la descripción del personaje de la historia, el maestro número 18 escribe “no nos presentas al personaje de la historia”.

Algo más de un tercio de los mensajes (35%) se centraron más bien en la trama de la historia. Algunos mensajes se refieren explícitamente al tema inicial o al «nudo» de la historia. Por ejemplo, el profesor 18 solicita una ampliación al autor del cuento con respecto a los sentimientos que experimenta uno de los personajes: “Este es un buen momento para que cuentes cómo se sentía la madre”. Otros aluden al encadenamiento (a veces poco claro) de los eventos de la trama. El maestro 5 realiza un comentario para resaltar que la historia respeta un orden temporal lógico y no hay saltos en la narración: “Se entiende bien, está ordenado”. El profesor 7 realiza un comentario para resolver una duda que le surge con respecto al momento temporal en el que surge una acción enclavada espacialmente en el nudo de la historia: “Antes o después de haber cogido la comida”. La mayoría, no obstante se refieren al desenlace de la historia o al final del texto. Por ejemplo, el maestro 17 advierte que el texto carece de un final que sirva de cierre, e indica lo siguiente: “Falta el final”. El profesor número 18 señala que el relato que ha corregido cuenta con un desenlace muy escueto y considera necesario ampliarlo: “El desenlace de la historia debe ser más desarrollado”

Otros maestros, proponen sugerencias organizativas relacionadas con aspectos más formales, como la extensión de los párrafos del texto. El

profesor 3 escribe en dos ocasiones “puedes hacer párrafos más cortos”, “separa las partes del texto en párrafos”. Igualmente se detectan bastantes comentarios evaluativos poco específicos: “Está bien estructurado” [maestro 19]; “Acuérdate de las partes de la narración” [maestro 5].

Finalmente, se registraron algunos comentarios cuyo contenido meta-textual no podía clasificarse en ninguna de las anteriores categorías, centrados sobre todo en aspectos caligráficos o relacionados con la presentación visual del texto. Por ejemplo, ante la ausencia de márgenes en la composición, el profesor 5 refleja la siguiente sugerencia: “Hay que dejar un margen”. El profesor 17 escribe en una de las composiciones: “Debes mejorar la letra”; mientras que en otra sugiere: “Podrías añadir diálogos”. [Profesor 17]

Relación entre comentarios evaluativos y calificaciones

Un último análisis cuantitativo se centró en la relación entre el contenido meta-textual de los mensajes evaluativos y las calificaciones globales que los maestros otorgaban a las composiciones. Concretamente se analizó en qué medida la presencia de cada una de estas categorías podía predecir la valoración global que los maestros hacían de cada narración. En la tabla VI se observan los valores de la regresión lineal obtenidos entre dos aspectos: la calificación que el profesorado ha asignado a las narraciones de los estudiantes (una vez transformadas a la escala estandarizada que aparece en la tabla IV) y la frecuencia de mensajes evaluativos, en función de su contenido meta-textual.

TABLA VI. Regresión lineal y correlación entre la calificación y el número de correcciones según el criterio de evaluación.

Calificaciones	Organizativo	Semántico-léxico	Gramatical	Ortográfico	Otros	Inespecífico
Coefficiente de regresión (F)	2,0	0,3	-2,7**	-4,8**	-0,9	-0,6
Correlación de Pearson	0,1*	-0,1*	-0,3**	-0,4**	-0,2*	-0,1*
Media	0,3	3,4	8,9	17,0	1,1	1,1
Desviación típica	0,7	3,9	7,3	14,2	1,1	0,3

(*) $p < 0,05$; (**) $p < 0,01$

Fuente: Elaboración propia.

Podemos apreciar que existe una regresión significativa entre la nota con la que los profesores califican las narraciones y los criterios gramatical y ortográfico. Además, se observa una correlación de Pearson negativa y significativa en ambos criterios ($p < 0,01$). Se podría decir que cuantos más errores ortográficos y gramaticales corrige el profesorado, la calificación global registrada es menor. En cambio, el número de correcciones y comentarios evaluativos sobre los aspectos semánticos y organizativos resulta muy poco relevante para predecir dicha calificación.

Discusión y conclusiones

En este estudio se analiza la práctica evaluativa de los maestros de Primaria aplicada a textos narrativos producidos por el alumnado de los últimos cursos de dicha etapa educativa. Más específicamente hemos tratado de documentar el tipo de *feedback* que proporcionan a sus alumnos cuando revisan sus textos narrativos.

Los resultados sugieren que cuando se evalúan las composiciones narrativas de los estudiantes, los maestros se centran principalmente en corregir errores específicos de tipo ortográfico y gramatical, que son sustituidos (encima o en el interlineado) por las letras o palabras adecuadas. Entre el 75% y el 90% de los registros evaluativos analizados y en torno al 95% del profesorado coincide con este perfil. Por el contrario, si exceptuamos las calificaciones, los maestros participantes expresaron muy pocos mensajes evaluativos no meramente correctivos (apenas un 4,5% del total de mensajes). Casi la mitad de los profesores no proporcionó ningún *feedback* acerca de aspectos macro-estructurales del texto. Sólo un 60% de los ya de por sí escasos mensajes centrados en mejorar la organización del texto abordaban aspectos específicos de la contextualización, la trama y el desenlace de la historia.

Los errores ortográficos y gramaticales son más visibles y fáciles de evaluar que los de carácter retórico (Morales, 2004). Cabría argumentar, sin embargo, que esta abrumadora comparación entre la frecuencia de correcciones gramaticales y ortográficas, frente a otros aspectos semánticos y estructurales de la composición, se debe simplemente al proceso natural de adquisición de la competencia escritora en esta etapa educativa. Lógicamente, cualquier texto tiene muchas más palabras que ideas, y un niño de entre 10 y 12 años, a diferencia de los estudiantes maduros, comete una mayor cantidad de errores de este tipo que de cualquier otro.

Sin embargo, la frecuencia de mensajes de evaluación referidos a aspectos ortográficos y gramaticales se ha mostrado también como el mejor predictor de las calificaciones globales que los maestros hacían de los estudiantes. El resto de mensajes evaluativos, no son sólo escasos, sino que no ayudan a predecir la valoración global que el profesorado realiza de la redacción. Aunque la mayor parte de los estudiantes que cometen muchos errores ortográficos y gramaticales suelen también organizar y articular peor las ideas en los textos, esto no es siempre así. Y cuando lo fuera, no explicaría por qué los evaluadores prestaron tan poca atención a este último aspecto.

Esta conclusión coincide con estudios previos, contextualizados en otros países, en los que se constató también que la evaluación del profesorado tiende focalizarse en características muy superficiales del texto (Connors y Lunsford, 1993; Hargreaves y McCallum, 1998; Schwartz, 1984; Sommers, 2011; Struthers, Lapadat y MacMillan, 2013). También es coherente con otros en los que se ha comprobado que los estudiantes de Primaria mejoran mucho más los aspectos superficiales que los macro-estructurales en sus composiciones, tras recibir la evaluación de sus profesores (Matsumara, Patthye-Chávez, Valdés, y Garnier, 2002; Montanero *et al.*, 2014; Searle & Dillon, 1980).

Por otro lado, el hecho de que la inmensa mayoría de los mensajes evaluativos tengan un carácter explícitamente correctivo, con muy pocas señalizaciones, preguntas, ampliaciones o justificaciones, que estimulen comportamientos autorregulatorios (relacionados con la comprensión y autoevaluación del error), podría estar relacionado con el propio contexto de la evaluación. Posiblemente, si la muestra de evaluaciones hubiera sido recogida en el marco de una evaluación más formativa, en la que los estudiantes tuvieran que mejorar posteriormente sus propias narraciones, este tipo de mensajes serían más frecuentes. En este mismo sentido, es posible que, si los profesores tuvieran la posibilidad de comentar públicamente los errores más distintivos o repetidos de su evaluación, habrían añadido oralmente algunas preguntas o justificaciones, e incluso hubieran ejemplificado algunas ampliaciones.

Volviendo al análisis de las calificaciones, se ha constatado una clara tendencia a calificar globalmente (85,7%) las composiciones, con una escala de tipo cualitativo ordinal (61,9%). Dichas calificaciones se acompañaban mayoritariamente de valoraciones verbales negativas o de desaprobación de aspectos particulares de la narración (86,3%), focalizados sobre todo, nuevamente, en los aspectos gramaticales (63,6%).

No hay acuerdo en la literatura sobre la idoneidad del uso de comentarios positivos y negativos en la evaluación de redacciones. Algunos autores han puesto de manifiesto el riesgo que comportan las alabanzas excesivas para resaltar aspectos positivos (DeGroff, 1992; Smith, 1997); mientras que otros argumentan que una evaluación centrada en la identificación de errores fomenta actitudes negativas en los estudiantes (Connors y Lunsford, 1993; Lee, 2003). Se ha observado que, aunque los sistemas aditivos de corrección pueden tener efectos psicológicos más positivos en los estudiantes que los sustractivos, resulta pedagógicamente beneficioso que los estudiantes desarrollen la conciencia del error (Morales, 2004).

En cualquier caso parece claro que el juicio de valor, (ya sea en la forma de calificación numérica, o a través de expresiones tales como “muy bien la trama” o “no me convence”) debería acompañarse de consejos, sugerencias o procedimientos para la revisión y corrección por parte del alumno. Casi la totalidad de las calificaciones globales o locales registradas en nuestro estudio estaban acompañadas de una corrección u orientación de mejora. No obstante, la mayoría de estos comentarios, o bien son muy generales (sin especificar dónde se localiza exactamente el problema o qué normas o reglas se deben seguir para la resolución del mismo), o bien consisten simplemente en la corrección de una letra o palabra. Sólo en un 9,4% de las correcciones de este último tipo el profesor proporciona la norma gramatical u ortográfica.

Es posible que esta circunstancia esté relacionada con la falta de tiempo para corregir exhaustivamente los aspectos gramaticales y ortográficos de grupos relativamente amplios de estudiantes de Primaria, haciendo al mismo tiempo comentarios explicativos o sugerencias de mejora. En cualquier caso, además de restarle validez y fiabilidad (Espin, Weissenburger, y Benson, 2004; Hayes, Hatch, y Silk, 2000; Miller y Crocker, 1990), la mera identificación de errores ortográficos y gramaticales parece un tipo de evaluación insuficiente para establecer un diagnóstico de la habilidades y limitaciones reales del alumno (Dunsmuir *et al.*, 2015).

La claridad de los comentarios evaluativos que el estudiante recibe por este medio es fundamental para que pueda interpretarlos y mejorar su competencia escritora. Se ha constatado que, en demasiadas ocasiones, el mensaje evaluativo se encuentra muy desvinculado del fragmento de texto al que se refiere. En otras ocasiones, el contenido del mensaje del profesor es demasiado impreciso e/o inespecífico; por ejemplo, cuando el maestro indica “separa el texto en párrafos” podría aclarar que cada

una de las tres partes que componen el argumento de un cuento debería formar cada párrafo. Otros casos recurrentes de mensajes que no facilitan su interpretación son el empleo de símbolo de interrogación o de líneas separando distintas zonas del texto (sin ninguna nota aclaratoria al respecto).

Limitaciones y prospectiva

Si asumimos que la evaluación que propone el maestro es un buen indicador de lo que previamente ha enseñado, cabe finalmente suponer que las conclusiones anteriores tienen implicaciones más allá de la práctica evaluativa en las aulas de Primaria. La competencia escritora sigue sujeta en nuestras escuelas a estilos tradicionales de instrucción y evaluación que hacen que los alumnos sigan poniendo atención en aspectos superficiales del código (reglas gramaticales, ortografía, limpieza...), en contraste con las habilidades autorregulatorias del contenido; lo que convierte las tareas de escritura en un aprendizaje “mecánico” y aburrido (Akkaya y Susar Kırımı, 2010).

Estas conclusiones deben considerarse, no obstante, con prudencia, dada las limitaciones inherentes al presente estudio. Los docentes que participaron en esta investigación no eran profesores del alumnado que escribió las narraciones. Tampoco esperaban que los estudiantes reescribieran y mejoraran el texto posteriormente, como consecuencia de su evaluación. Esta circunstancia podría haber influido en la orientación y precisión de sus mensajes evaluativos. En futuros trabajos sería conveniente, por tanto, replicar estos análisis con una muestra más amplia de profesores, seleccionados aleatoriamente, en contextos de evaluación formativa con sus propios estudiantes.

Referencias bibliográficas

Akkaya, N., & Susar Kırımı F. (2010). Relationship between attitudes to reading and time allotted to writing in primary education. *Procedia Social and Behavioral Sciences*, 2, 4742–4746

- Allan, E., & Driscoll, D. (2014). The three-fold benefit of reflective writing: Improving program assessment, student learning, and faculty professional development. *Assessing Writing*, 21, 37-55
- Berg, E. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing*, 8, 215-241.
- Cassany, D. (2000). *Reparar la escritura*. Barcelona: Graó.
- Cho, K. & MacArthur, C (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20, 328-338.
- Cho, Y. (2003). Assessing writing: Are we bound by only one method? *Assessing Writing*, 8 (3), 165-191.
- Connors, R. & Lunsford, A. (1993). Teachers' rhetorical comments on student papers. *College Composition and Communication*, 44 (2), 200-223.
- DeGroff, L. (1992). Process-writing teachers' responses to fourth-grade writers' first drafts. *The Elementary School Journal*, 93 (2), 131-144.
- Dempsey, M. S., PytlikZillig, L. M., & Bruning, R. H. (2009). Helping preservice teachers learn to assess writing: Practice and feedback in a Web-based environment. *Assessing Writing*, 14, 38-61.
- 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
- Espin, C. A., Weissenburger, J. W., & Benson, B. J. (2004). Assessing the writing performance of students in special education. *Exceptionality*, 12, 55-66.
- Fahimi, Z., & Rahimi, A. (2015) On the Impact of Self-assessment Practice on Writing Skill. *Procedia - Social and Behavioral Sciences*, 192, 730 - 736.
- Ferris, D. R. (1997). The Influence of Teacher Commentary on Student Revision. *TESOL Quarterly*, 31, 315-339.
- Fife, J. M., & O'Neill, P. (2001). Moving beyond the written comment: narrowing the gap between response practice and research. *College Composition and Communication*, 53 (2), 300-321.
- Glaser, C., & Brunstein, J. (2007). Improving fourth-grade students' composition skills: effects of strategy instruction and self-regulation procedures. *Journal of Educational Psychology*, 99(2), 297-310.
- Graham, S., Harris, K., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: the effects of self-regulated strategy development. *Contemporary Educational Psychology*, 30(2), 207-241.

- Hargreaves, E., & McCallum, B. (1998). *Written feedback to children from teachers*, ERSC Project 'Teaching, learning and feedback strategies'. Project Paper No. 7, London: University of London, Institute of Education.
- Hayes, J. R., Hatch, J. A., & Silk, C. M. (2000). Does holistic assessment predict writing performance. *Written Communication*, 17, 3-26
- Jones, A., & Moreland, J. (2005). The importance of pedagogical content knowledge in assessment for learning practices: A case-study of a whole-school approach. *Curriculum Journal*, 16(2), 193-206.
- Lee, I. (2003). L2 writing teachers' perspectives practices, and problems regarding error feedback. *Assessing Writing*, 8 (3), 216-237. DOI: 10.1016/j.asw.2003.08.002
- Lennart, J. (2014). Pair assessment of pupil writing: A dialogic approach for studying the development of rater competence. *Assessing Writing*, 20, 37-52. DOI: 10.1016/j.asw.2014.01.002
- Liang, M-Y. (2010). Using synchronous online peer response groups in EFL writing: revision-related discourse. *Language Learning & Technology*, 14(1), 45-64.
- Matsumara, L. C., Patthye-Chavez, G. G., Valdes, R., & Garnier, H. (2002). Feedback, writing assignment quality, and third-grade students' revisions in lower- and higher-achieving urban schools. *The Elementary School Journal*, 103 (1), 3-25.
- Miller, H. D., & Crocker, L. (1990). Validation methods for direct writing assessment. *Applied Measurement in Education*, 3, 285-296
- Min, H. T. (2003). Why peer comments fail? *English Teaching and Learning*, 27(3), 85-103.
- Montanero, M., Lucero, M. & Fernández, M. J. (2014). Iterative co-evaluation with a rubric of narrative texts in Primary Education / Coevaluación iterativa con rúbrica de textos narrativos en la Educación Primaria. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, 37 (1), 184-220. DOI: 10.1080/02103702.2014. 881653
- Morales, F. (2004). Evaluar la escritura, sí... Pero ¿Qué y cómo evaluar?. *Acción Pedagógica*, 1 (13), pp. 38-48.
- Parr, J., & Timperley, H. (2008). Teachers, schools and using evidence: Considerations of preparedness. *Assessment in Education: Principles, Policy and Practice*, 15(1), 57-71.
- Peterson, S. S. & McClay, J. (2010). Assessing and providing feedback for student writing in Canadian classrooms. *Assessing Writing*, 15 (2), 86-99. DOI: 10.1016/j.asw.2010.05.003

- Rouhi, A., & Azizian, E. (2013). Peer review: Is giving corrective feedback better than receiving it in L2 writing? *Procedia - Social and Behavioral Sciences*, 93, 1349-1354.
- Schwartz, M. (1984). Response to writing: a college-wide perspective. *College English*, 46(1), 55-62.
- Searle, D., & Dillon, D. (1980). The message of marking: Teacher written responses to student writing at intermediate grade levels. *Research in the Teaching of English*, 14, 233-242.
- Shepard, L., Hammerness, K., Darling-Hammond, L., Rust, F., Baratz-Snowdon, J., Gordon, E., et al. (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275-326). San Francisco: Jossey-Bass.
- Slomp, D. H. (2012). Challenges in assessing the development of writing ability: Theories, constructs and methods. *Assessing Writing*, 17, 81-91.
- Smith, S. (1997). The genre of the end comment: Conventions in teacher responses to student writing. *College Composition and Communication*, 48 (2), 249-268.
- Sommers, J. (2011). Reflection revisited: Using the class collage. *Journal of Basic Writing*, 30(1), 99-129.
- Straub, R. (1997). Students' reactions to teacher comments: An exploratory study. *Research in the Teaching of English*, 31, 91-119.
- Struthers, L., Lapadat, C. y MacMillan, D. (2013). Assessing cohesion in children's writing: Development of a checklist. *Assessing Writing*, 18 (3), 187-201.
- Struthers, L., Lapadat, C., & MacMillan, D. (2013). Assessing cohesion in children's writing: Development of a checklist. *Assessing Writing*, 18 (3), 187-201.
- Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9 (2), 147-170.

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

Red on black. How do teachers assess their students' narrative texts?¹

Rojo sobre negro. ¿Cómo evalúan los maestros las redacciones de sus estudiantes?

DOI: 10.4438/1988-592X-RE-2015-372-315

María-Jesús Fernández

Manuel Lucero

Manuel Montanero

University of Extremadura

Abstract

The correction of the texts written by students is one of the educative tool most frequently used for improving the writing skill. However, there is no research that analyses what criteria are to be considered by teachers in this kind of assessment and what kind of *feedback* is provided by them. In this sense, this study describes the improvement suggestions from 21 teachers about narrative texts written by 236 students in the last years of Primary Education, in order to help them improve their writing skills. A total number of 7851 verbal assessment records and charts were analysed, depending on their formal presentation and their assessment of meta-textual content. The results show a great predominance of verbal messages, focused mainly on spelling and grammar improvement. The frequency of such corrections could explain why teachers give higher scores to texts with better grammar and spelling use. On other hand, a low number of suggestions to try to develop the understanding and self-assessment of mistakes were also logged, as well as those related to improving macro-structural aspects of a text. Hence, the current study points out the implications for teaching and assessing writing, which are discussed in the conclusion of the article.

Keywords: Student evaluation, Writing, Spelling, Grammar, Basic Education.

⁽¹⁾ **Acknowledgements:** Extremadura Regional Government and the European Social Fund for the financial support (PD12129).

Resumen

Uno de los recursos educativos más extendidos para la mejora de la competencia escritora consiste en evaluar y corregir los textos que los propios estudiantes redactan. Sin embargo, no contamos con estudios que analicen qué criterios tienen fundamentalmente en cuenta los maestros en dicha evaluación y qué tipo de retroalimentación proporcionan. En este sentido, en el presente estudio se describen los mensajes evaluativos que 21 maestros emitieron sobre las composiciones narrativas de 236 estudiantes de los últimos cursos de Educación Primaria, con el objetivo de ayudarles a mejorar su competencia escritora. En total, se analizaron 7851 registros verbales y gráficos de evaluación, en función de su presentación formal y su contenido evaluativo y meta-textual. Los resultados ponen de manifiesto un claro predominio de mensajes de tipo verbal, centrados principalmente en la corrección específica de aspectos ortográficos y gramaticales. La frecuencia de este tipo de correcciones se mostró como el mejor predictor de la calificación global que el profesorado otorgaba a las narraciones. Por el contrario, se constata una escasa presencia de mensajes que faciliten la comprensión y la autoevaluación de los errores cometidos, así como sugerencias de mejora que involucren aspectos macro-estructurales del texto. Finalmente, se discuten estos hallazgos y sus implicaciones en la enseñanza y evaluación de la competencia escritora. De estos resultados se derivan implicaciones relevantes para la enseñanza y evaluación de la escritura, que se discuten al final del artículo.

Palabras-clave: Evaluación del estudiante, Redacción, Ortografía, Gramática, Educación Básica.

Introduction

The development of written competence during compulsory schooling is a topic of interest to legislators, researchers and teachers, because writing is linked to academic achievement of students (Lennart, 2014). In fact, Teachers dedicate much time and effort in regards to the evaluation of written competence activities.

The evaluation practice of story-writing in Primary Education

Traditionally *feedback* provided orally or written has been considered an important part of writing instruction and assessment (Parr and Timperley,

2008). However, even though there has been extensive research, revisions and commentaries made in contexts of self and peer-evaluation activities (Cho and MacArthur, 2010; Fahimi and Rahimi, 2015; Lennart, 2014; Rouhi and Azizian, 2013), very few studies have focused on analysing the way teachers evaluate students writing drafts in their native language (L1). Some papers have questioned the reliability and the validity of this type of assessment, arguing that despite of their efficiency to attain the general impression of students written productions, they provide diagnostic information on the various components of written competence (Dunsmuir *et al.*, 2015). Others, instead, highlight the relation between a teacher's ability to provide quality *feedback* and students' progress in learning to write narrative texts (Fife and O'Neill, 2001). The assessment of teachers normally includes, in fact, the correction of original fragments of text from students (Peterson and McClay, 2010). Although, the revision drives towards an improvement in the re-writing of the text (when a re-write option is available) it depends, not only on the writer's capacity, but also on the quality of the *feedback* received (Tsui and Ng, 2000). In a previous study researchers found that 67% of modifications that an ample sample of Primary School students (between 10 and 12 years old) were introduced in a second version of their stories, after receiving assessment of their teachers, were centred around purely grammatical and spelling aspects (Montanero, Lucero and Fernández, 2014). Other studies point in the same direction showing that the nature of teacher assessment is generally based on the more superficial characteristics of the text (Connors and Lunsford, 1993; Hargreaves and McCallum, 1998; Schwartz, 1984; Sommers, 2011). Rhetoric aspects, such as the cohesion of ideas, are, on the contrary, traditionally ignored, in both written competence assessment and teaching (Struthers, Lapadat and MacMillan, 2013).

Some authors argue that this type of evaluative practises are related to a lack of consciousness by evaluators about the different processes that intervene in the writing of a text (Cho, 2003; Dempsey, PytlikZillig and Bruning, 2009); as well as those criteria that permit for an ideal judgment, in regards to the marked objectives. (Cassany, 2000; Morales, 2004). A good evaluation of the deeper aspects of a writing requires complex knowledge about writing, of which many competent writers are lacking: an explicit meta-knowledge of the language, of text structures, of writing process and strategies; in addition to the use of methods,

activities and efficient resources that help students with difficulties (Parr and Timperley, 2008). Teaching staff that do not have grasp of this knowledge will not only be less able to formulate suggestions for improvement, but will also have difficulty in anticipating writing problems that their students may encounter (Jones and Moreland, 2005; Shepard et. al, 2005).

In addition to the criteria and the metalinguistic content of the evaluations, other relevant assessment aspects, which have not been sufficiently documented in this educational stage, are related to the way that they are expressed “red on black” on students writing. The assessment comments can be found as margin annotations, questions asking for concrete or additional information, clarifications about general aspects and, mostly, corrections (Ferris, 1997). Research has found that Higher Education students prefer open questions and commentaries that include arguments and explanations on where to improve (Straub, 1997). However, there are no studies that can describe detailed evaluation practices, nor its specific impact on Primary School students in our educational system.

Analysis models of assessment messages of compositions

The literature provides a series of categorization systems for *feedback* messages that evaluators (teachers or peers) register in the works that they revise. One of the most cited is the proposal made by Min (2003) based upon the classical works of Berg (1999). In regards to the nature of the message they can be classified into *oriented* or *not oriented* on the revision. In the first group, we can differentiate between *assessment* and judgement that don't behold concrete suggestions to modify the text; *clarifications*, that are normally expressed via questions, about what the evaluator has not understood (however, we can also include in this category incitements to re-develop communicative intent or meaning of a text fragment); *alterations*, or corrections of supposed mistakes; and the information *suggestions* that help a peer to enrich or self-correct a text. Finally, in regards to area, messages can be *global* (normally referring to content) or *local*.

Cho and MacArthur (2010) used a dual system of categories that are more complete than the previous system. The analysis of *feedback*

messages permits the collection of other matrixes, such as the explicitness of the modification suggestion (*directive* vs. *non-directive*) or in the positive sense (eulogy) or evaluation critique. The modifications made by evaluators can be classified into three levels of complexity: the *simple fixes*, on a superficial level, affecting spelling or grammatical questions; on a microstructural level, where the corrections can be either *complex* (deleting or changing information) or *extended* (adding sentences that provide more information); on a macrostructural level, in which students can add new *content* or topics, or modify the *organization* (titles, paragraph movement, connecting phrases), as well as the addition or modification of materials (e.g. tables, figures...).

On her part, Liang (2010) suggests a five-category system: *negotiation* (whose definition approximately coincides with *clarifications* of previous works); *content discussion* (similar to *suggestions*, they are expressions of disagreement, that can be followed by an argument and/or proposal); *error correction* (either for content, grammar or spelling); *management task* and *commentaries* not related to the revision.

This literature background, however, is centred mainly on learning to write in English as a foreign language (L2) in Secondary and Higher Education Levels. Moreover, the collected research data is based upon written peer revision, either on paper or through the use of new technologies. There is a lack of application of these assessment analysis systems that are made by teachers in the corresponding levels of Primary Education in the Spanish Literacy subject.

Aims of the research

In this framework, the present research has two core objectives: on one hand, to describe the characteristics of assessment grading and commentaries registered by primary school teachers from our educational system of the redacted narrations provided by students. On the other hand, to explore the relation between assessment commentaries and the general grades given to the stories, in addition to the implicitly made specific assessment commentaries.

Design and Method

Participants

In this study 21 teacher-tutors (15 female and 6 male) participated from 8 schools of different cities of Extremadura (5 public and 3 public-funded private schools). They were selected due to the availability of the researchers. All of the participants have more than 5 years teaching experience in the last years of Primary Education. Each teacher assessed between 5 and 20 narrative type texts written by 236 students aged between 10 and 12 (Table I).

TABLE I. Sample

Centre	Assessor	N° of texts assessed
Centre 1	Male Teacher 1	5
	Female Teacher 2	12
	Female Teacher 3	7
	Male Teacher 4	5
Centre 2	Male Teacher 5	8
	Female Teacher 6	7
Centre 3	Female Teacher 7	12
	Female Teacher 8	14
Centre 4	Male Teacher 9	5
Centre 5	Female Teacher 10	5
	Female Teacher 11	5
	Female Teacher 12	5
	Female Teacher 13	5
Centre 6	Female Teacher 14	20
	Male Teacher 15	20
	Female Teacher 16	20
Centre 7	Female Teacher 17	19
	Female Teacher 18	9
	Female Teacher 19	20
	Male Teacher 20	20
Centre 8	Female Teacher 21	13

Source: Table of own design.

Collection and segmentation procedures of the evaluation records

The research was based on the document analysis of assessment registers, handwritten expressions made by teachers on the hand-in stories of students.

Narrations were written in thirty minutes by year 5 and year 6 students of Primary Education. Students had to write, in approximately one page, about the following topic: "A child becomes a town hero". The finished papers were randomly distributed between the 21 teachers, who were unaware of the research aims. They were asked to correct with a different coloured pen, as usual, registering in the same paper any relevant annotation and/or judgement.

The segmentation of the analysis units was done following pragmatic and semantic criteria. Each analysis unit was composed by a single meta-evaluative message (referring to an error or type of error), even though it was inserted into an extended sentence. For example, the following sentence, used by teacher 17, was segmented into three units of analysis: "A Little dirty (1), its good (2), be careful with spelling, think before writing the words (3)".

Categorization procedures of assessment messages

In order to classify messages researchers used an original system of categories (table II), created from an inductive-deductive procedure of *continuous comparison*. In first place, building upon the review of analysis models that was described in the introduction, investigators then selected and defined, with the same or similar denomination, the categories used in this study to analyse the assessment and meta-textual content of *feedback* messages. In second place, after a first study of more than 7000 analysis units from the sample, researchers identified those messages that were not able to be classified with sufficient precision and added new categories that were not explicitly used in previous studies. Finally, two more dimensions that were not included in the other works (form and location of messages) were incorporated into the system.

To analyse the formal aspects, investigators took into account, on the one hand, the form of expression (e.g. verbal or non-verbal) of the assessment record and on the other hand, the location of the register (e.g.

roughly above the evaluation focus point, in the margins, at the beginning or at the end of the text).

To examine the amplitude and the assessment content of the messages researchers identified a series of categories primarily based upon the model by Min (2003). As to the extent of the evaluation messages, the investigation team included to what Min and other authors defined as *global* and *local feedback*. In terms of evaluative content, the “assessments” that did not include content for improvement were classified into two subtypes: *signals* and *grades*. As proposed by Cho and MacArthur (2010), within the latter researchers defined a criteria to discriminate between positive and negative ratings. Given the interrogative nature of the “clarifications” they are identified in the system as a *question*. What Min interchangeably defines as “modifications” have been defined into two types of text correction: *generic* and *specific*. The term “suggestion” is somewhat ambiguous so academics included *extension* or *modification*. The “non-targeted” messages have been categorised into the *other(s)* category. Again, the initial analysis of the assessment messages brought about the identification of a very relevant message type, *justification* that may or may not accompany a particular correction, which was not considered as an independent category in previous studies.

As to identify category the meta-textual dimension researchers used the proposal made by Cho and MacArthur (2010). Superficial or simple repairs or corrections were classified, also, into two types: grammatical and spelling. However, the complex and extended messages was classified into a single category (*semantic-lexical* content), since the matrix difference is already included in the *extension* category of the previous dimension. The macro-structural repairs and organization issues were generally identified as assessments of *organizational content*. To classify other evaluation content that do not fall clearly into any of the above categories, as comments on handwriting, creativity or issues relating to the management of the task (Liang, 2010), we have incorporated the categories of *non-specific* content and *others* (see table II).

TABLE II. Category Systems of *feedback* from the text assessment

Dimension	Category	Definition
A. Form	1. Verbal	Verbal expressions that contain a group of letters, words or phrases. Including accent marks, punctuation and acronyms.
	2. Non-verbal	Numerical expressions (one or more digits) or graphics, either simple or complex (scoring, underlining or margined lines, arrows, circles).
	3. Mixed	Feedback that integrate a verbal expression and another graphic or numerical annotation.
B. Location	1. Overlay	Annotations overlaying the content of the assessed text. (Includes between lines).
	2. Margin	Marginal space or superimposed annotation on a different part of the text, but still explicitly linked to it.
	3. Separated	A registered annotation without an explicit link to any fragment of the text (e.g. at the beginning or at the end of the work) or in a different piece of paper.
C. Amplitude	1. Global	Feedback on the whole of the text
	2. Local	Feedback on a fragment or specific part of the text
D. Assessment Content	1. Signals	Symbols or verbal expressions in which the assessor exclusively identifies a single fragment of the text, this being a mistake or a correct use, without giving any more information. (E.g. underlining spelling mistakes without correction).
	2. Grade	Numerical expressions, graphics (upwards arrow, letters representing an ordinal scale) o verbal expressions that emit a judgement about the quality of the work. They can have a positive valence (such as numbers reaching 70% or higher on the selected scale), intermediate (e.g. the letter R or numbers between 50-70% of the scale) or negative.
	3. Question	Interrogative symbols or verbal expressions that require a clarification.
	4. Specific Correction	Graphic or verbal expression that, in addition to signalling an error; provides a specific instruction or a better suited alternative (includes the deletion or addition of a fragment)

TABLE II. Continuation

Dimension	Category	Definition
D. Assessment Content	5. Generic correction	Verbal expression that identifies an element that is lacking or a repeating error (without signalling it in the text) and provides a generic improvement instruction.
	6. Extension/Improvement	Verbal commentary that does not identify an error but suggests, exemplifies or directly provides an alternative or extension not essential to the text.
	7. Justification	Verbal commentary that explains or provides an argument for an alternative or improvement. (Includes regulatory comments).
	8. Other Comments	Other verbal comments, such as warnings or reinforcement messages.
E. Meta-textual content	1. Organization	<i>Feedback</i> about the rhetoric of the text or paragraphs. In the assessment of narrative texts the assessor can distinguish elements, specifically in this framework, about topic and plot.
	2. Lexical and semantic	<i>Feedback</i> on the topic or ideas of the text. Includes assessment on coherence or specific vocabulary use.
	3. Grammar	<i>Feedback</i> about the morph syntactical construct of sentences. Includes the modification of the word order, cohesion assessment cohesion or punctuation.
	4. Spelling	<i>Feedback</i> on spelling or word construction.
	5. Other(s)	Comments on specific meta-textual aspects, such as form (e.g. handwriting, margin use) or creativity.
	6. Non-specific	<i>Global Feedback</i> , transversal or in which no specific meta-textual content is expressed. Excludes those entries that are not evaluative.

Source: Table of own design

The unit analysis categorization was made after a training process in which researchers refined the category system and the permissive grade of said categories, as to assure the internal consistency between researchers.

In order to analyse the reliability of the process researchers calculated the degree of agreement between two evaluators on a sample of 7 randomly chosen texts that contained a total of 222 categorized messages.

As shown in the following table, the rates obtained from Cohen's kappa coefficient are very high in all dimensions, except form and meta-textual content which can be considered as acceptable.

TABLE III. Reliability data on the category identification of the dimension analysis. ($p < 0, 01$)

Dimensions	Form	Location	Amplitude	Assessment Content	Meta-textual content
KC index	0,85	0,97	0,97	0,96	0,82

Source: table of own design.

Quantitate and Qualitative analysis procedure

To describe the assessment profile of predominant compositions in Primary school teachers, researchers compared the frequency and percentile of assessment registers classified into each category.

In order to analyse to what extent the meta-textual content of the assessment commentaries predict the overall rating of the compositions (which usually appear highlighted and located in the top right-hand corner), researchers made a regression analyse between two variables: the *grading* and the correction frequency of each category of the meta-textual content dimension. Since teachers use various rating scales, all were matched into a same scale of odd numeric values from 1 to 9, as stated in table IV.

TABLE IV. Comparative of different scales of grading in a standard scale

Standard	Qualitative 1	Qualitative 2	Numeric
1	Very Bad	Insufficient	1-2
3	Bad	Sufficient	3-4
5	Regular	Acceptable	5-5
7	Good	Remarkable	7-8
9	Wey Good	Outstanding	9-10

Source: table of own design

Finally, in order to deepen the analysis of the assessment messages that address macrostructural aspects, the assessment registers were qualitatively analysed, particularly those related to organization.

Results

Researchers identified 7851 units of analysis from the 236 stories evaluated by the participant-teachers in the research. Table V reflects the categorization results of the assessment messages in regards to the dimensions considered in the study.

TABLEV. Frequency and nature of assessment commentaries

Dimension	Category	Register	Teachers
Form	Verbal	5956 (75,9%)	20 (95,2%)
	Non-verbal	1120 (14,3%)	20 (95,2%)
	Mixed	775 (9,8%)	20 (95,2%)
Location	Overlay	7096 (90,4%)	20 (95,2%)
	Margin	123 (1,6%)	10 (47,6%)
	Separated	632 (8,0%)	21 (100%)
Amplitude	Global	600 (7,6%)	21 (100%)
	Local	7251 (92,4%)	20 (95,2%)
Assessment Content	Signals	219 (2,8%)	13 (61,9%)
	Grade	243 (3,1%)	21 (100%)
	Question	54 (0,7%)	14 (66,7%)
	Specific Correction	6946 (88,5%)	20 (95,2%)
	Generic correction	293 (3,7%)	19 (90,5%)
	Extension/Improvement	24 (0,3%)	10 (47,6%)
	Justification	31 (0,4%)	9 (42,9%)
	Other Comments	41 (0,5%)	11 (52,4%)
Meta-textual content	Organization	91 (1,2%)	12 (57,1%)
	Lexical and semantic	976 (12,4%)	20 (95,2%)
	Grammar	2260 (28,8%)	20 (95,2%)
	Spelling	4082 (52,0 %)	20 (95,2 %)
	Other(s)	230 (2,9 %)	18 (85,7 %)
	Non-specific	212 (2,7 %)	20 (95,2 %)

Source: Table of own design.

Formal presentation of assessment messages

The previous table (table V) shows that all teachers employed *verbal*, *non-verbal* or *mixed* expressions to provide their *feedback*, although the majority of which were *verbal* (75.9%). The lack of accent marks was the most frequent corrective message: “En un pueblo a poco tiempo de Madrid vivía Jose, el era un niño sencillo.” [Teacher 9]. In regards to *non-verbal* symbols, the most used were scoring and underlining or circling. For example, teacher 21 circled the word “entonces” in several occasions to indicate its frequent use. *Mixed* messages consisted of a combination of graphic symbols, like arrows or scoring, with a verbal correction. Take for instance, teacher 8 who writes on top of “a el”, a line through “el” and then adds an “l” to the remaining “a”.

In regards to the *location*, the bulk of comments and annotations were *overlaid* (90.4%), normally correcting local mistakes (teacher 20, for example, writes “concordance error” between the lines of the text to signal that the number of two words did not match). Instead, *unrelated* messages seemed to be reserved for general grading questions. Teacher 21, for example, made a final global assessment comment stated that: “*The parts of the story are distinguishable. Good composition. You repeat: heard*”. The lack of margin annotations is surprising, however this is probably due to the absence of space left by students. In fact, more than half of the teachers did not use them. Teacher 20, for instance, wrote a margin annotation to make a recommendation about the characters feelings. “This is a good moment to talk about how the mother feels”. Before the lack of dialogue scripts, teacher 18 wrote in the right margin the following: “Dialogue scripts!”

Finally, in terms of message amplitude, more than 90% were referred *local* aspects of the text. The most frequent positive assessment fragment was in regards to the plot of the story (“very good plot” [teacher 8]); although, most consist of spelling corrections. The few more or less *global* assessments focussed mainly on the quality perception of the whole text; or on repetitious superficial mistakes, as indicated by teacher 5 in regards to the presentation of the paper: “A little dirty”.

Assessment content of messages

The immense majority of analysed messages (88.5%) were related to specific error correction.

A usual case is mistaking the Spanish word “ahí” with “hay”. Teacher number 14 draws a parenthesis around the misused adverb (“ahí”) and writes the correct verb form (“hay”) above. Another frequent situation is the lack of accent marks. Teacher number 21 corrects a spelling mistake by including the accent mark on top of the “i” from the verb form “tenía”. Furthermore, there is another trend of adding punctuation marks (e.g. semi-colons) and the deletion and modification of some letters for others. As means of an example: a teachers correction of the letter “ll” instead of “y”, which was used by a storywriter in the word “medaya” [teacher 21] or that made by teacher 17 deleting the “h” from the expression “ha correr”.

The rest of the assessment message categories do not reach 5%. The general percentage of corrections (of ample fragments of text) and grading are close to this result. For example, when faced with missing punctuation marks in a text, teacher 8 writes the following general correction: “Use full stops and commas...”. Another general correction is that of a teacher praising the author of an evaluated text for the quality of his handwriting; as a means of doing so, he writes at the bottom of the page: “Good handwriting” [Teacher 18].

Normally these type of commentaries or instructions are accompanied by some form of grade. All of the participant teacher included in their assessment messages a global or local grade. For example, teacher 14 writes in the top section of one of the evaluated text: “Sufficient”. Moreover, Teacher 13 congratulates a student for the lack of spelling mistakes by writing: “Very good spelling” [teacher 13]. In total researches registered 243 grading messages. The majority of them where verbal (45.9%) or mixed (38.8%). For instance, many teachers wrote and underlined a qualifying adjective in the superior part of the text. Numerical grading represents 15.3% of the total (73% of which, were circles and the others underlined). In regards to grading valence, only 16.1% were negative; the rest can be considered as a positive or intermediate (good) level.

It is important to highlight the lack of commentaries that signal, argument, ask for clarification or extension, of the correction of a mistake. Teacher 7, for example, underlines the word “triavesuras” to signal that it is a spelling mistake, without correcting it. Teacher 4 circles the word “moustro” to draw the author’s attention, however does not make any error correction. Only in a few cases, did some teachers write an interrogative comment (take for instance, teacher 3 who formulates a

series of questions adjacent to the introductory fragment of the text: "Why? When?"); Teacher 17 makes a non-corrective suggestion by stating: "The student could use the word hero more" and Teacher 20, when correcting a student who has wrote "de el...", writes an argument to her correction by providing the misused grammatical rule "de + el = del". In fact, more than half of the teachers never justified their comments or made suggestions to extend or re-write fragments of text.

Finally, some management and reinforcement comments that could not be categorized into any of the previous categories, were also registered. For instance, teacher 14 writes on one of the compositions that: "Accent marks also form part of spelling mistakes". Having found many spelling mistakes the same teacher writes at the end of another paper: "Copy the circled spelling mistakes correctly, 10 times in you literacy notebook".

Meta-textual content

The main part of assessment commentaries (nearly 80%) refer to grammatical and spelling aspects. In regards to the former, teacher 20 writes: "concordance error" in order to signal that there is a word in plural that does not correspond to the article that precedes (which is in singular form). This is an example of a correction of grammatical concordance, the following example highlights an assessment commentary of a spelling correction. Teacher 14 indicates that: "There are a lot of accent marks missing".

In contrast, the percentage of semantic-lexical comments is very low (12.4%). Take the following three examples: «You must not repeat the same word» [Teacher 5] «Does the part about Buddhist bring something or have anything to do with the composition?» [Teacher 17]; «You must avoid repetitions» [Teacher 18].

As well as the previous commentaries, the ones referred to *management* are also very few (1.2%). The percentage of teachers who provide comments on the macro-structure of the text are just half of the sample. In comparison, it is much smaller than in the other criteria of the *meta-textual content* dimension. Among the 236 evaluated texts, we identified a total of 91 assessment records regarding management aspects, of which 73.6% where expressed verbally. Whereas nearly 20%, that in

addition to text, included a graphic element to signal the mistake in the fragment of text. 71.4 % appeared unrelated to the concerned, while 7.7% of the evaluative comments appeared superimposed and 20.9 % had a marginal location. In general terms the registers that referred to the whole text (63.7 % were global) mostly; although corrections were expressed as specific (46.2%) or generic (23.1%).

As for the components of the narrative structure under evaluation, we found that 24 % of the records refer to the frame of the story (description of the characters, space and time). Thus teacher 7 suggests improvements in the absence of some elements of the framework: «Missing information about the child, the place and the time it happened.» In the absence of the description of the character in the story, teacher 18 writes: «You do not present the characters of the story».

A little more than a third of the messages (35%) where related to the plot of the story. Some messages refer explicitly to the opening theme or «the middle» of the story. For example, teacher 18 requested an extension on the feelings that one of the characters may experience: “This is a good time for you to tell us how the mother felt”. Other allude to the flow of events (sometimes unclear) in the plot. Teacher 5 makes a comment to highlight that the story follows a logical and temporal sequence and that there are no narrative interruptions: “Understandable and well organized”. Teacher 7 comments in order to clarify a doubt in regards to the temporal aspects of the middle of the story, asking: “Is this before or after grabbing the food”. Most, however refer to the outcome of the story or the end of the text. For example, teacher 17 warns that the text lacks a final closure, and states: «The end is missing». Teacher 18 indicated that the story has an abrupt ending and is in need of an extension: «The outcome of the story should be developed».

Other teachers propose organizational suggestions regarding formal aspects, such as paragraph extension. Teacher 3 writes twice: «You can make shorter paragraphs», “You can separate the parts of the text into paragraphs». Also quite unspecific evaluative comments were detected, such as: «It is well structured» [teacher 19]; «Remember the parts of the narrative» [teacher 5].

Finally, researches recorded some meta-textual content comments could not be classified in any of the previously mentioned categories, they primarily focused on handwriting or visual presentation related aspects of the text. For example, in the absence of margins in the

composition, teacher 5 makes the following suggestion: «We must leave a margin». Teacher 17 writes in one of the compositions: «You must improve the letter» while in another suggested: «You might want to add a dialogue» [teacher 17].

Relationship between assessment commentaries and grading

A final quantitative analysis focused on the relationship between the meta-textual messages of evaluative content and overall ratings that teachers awarded to the compositions. Specifically, the analysis determined to what extent the presence of each of these categories could predict the overall assessment teachers made for each story. The following table VI shows the rates of lineal regression obtained from the two aspects: the grade that teachers assigned to the stories of students (once transformed to a standard scale shown in Table IV) and the frequency of assessment messages based on their meta-textual content.

TABLE VI. Lineal Regression and correlation between grading and the number of corrections according to assessment criteria

Description	Organization	Semántical and Lexical	Grammar	Spelling	Other(s)	Non-specific
Regression Coefficient (F)	2,0	0,3	-2,7**	-4,8**	-0,9	-0,6
Pearson Correlation	0,1*	-0,1*	-0,3**	-0,4**	-0,2*	-0,1*
Average	0,3	3,4	8,9	17,0	1,1	1,1
Typical deviation	0,7	3,9	7,3	14,2	1,1	0,3

(*) $p < 0,05$; (**) $p < 0,01$

Source: Table of own design.

Researchers can determine that there is a significant regression between the score that teachers awarded the students stories and the grammar and spelling criteria. In addition, the data shows a negative and significant Pearson correlation in both criteria ($p < 0.01$). Researchers could say that when more spelling and grammatical errors are registered and are corrected by teachers the overall grade is lower. However, the

number of corrections and evaluative comments on semantic and organizational aspects has little relevance to predict the grading.

Discussion and Conclusions

In this study we analyse the evaluation practice of primary school teachers applied to narrative texts produced by students in their final years of this educational stage. Specifically, we have tried to document the type of *feedback* they provide to their students when they review their written work.

The results suggest that when students' narrative compositions are evaluated, teachers mainly focus on correcting specific errors of spelling and grammar, which are replaced (either above or in the spacing) by the appropriate letters or words. Between 75 % and 90 % of the analysed assessment records and around 95 % of teachers matches this profile. On the contrary, except for grades, participating teachers provide very few assessment messages that are not merely corrective messages (just 4.5% of message total). Almost half of the teachers did not provide any *feedback* about macro-structural aspects of the text. Only 60 % of the already scarce messages focused on improving the organization of the text addressed specific aspects of contextualization, plot and outcome of the story.

Spelling and grammar errors are more visible and easier to evaluate than rhetorical mistakes (Morales, 2004). It could be argued, however, that this overwhelming comparison between the frequency of grammatical and spelling corrections, compared to other semantic and structural aspects of the composition, is simply due to the natural process of acquisition of writer competence in this educational stage. Logically, any text has many more words than ideas, and a child between 10 and 12 of age, unlike mature students, makes a larger amount of errors of this type than any other.

Although, the frequency of evaluation messages related to spelling and grammar has also shown to be the best predictor of overall ratings towards students. The rest of the evaluative messages are not only rare, but they do not help predict the overall assessment of the compositions. Although most of the students who make many spelling and grammatical errors are usually also the worst in organizing and articulating ideas in

the text, this is not always the case. And when it was, it would not explain why the evaluators paid little attention to this aspect.

These findings are consistent with previous studies, contextualized in other countries, where it was also found that the evaluation of teachers tends to focus on very superficial characteristics of the text (Connors and Lunsford, 1993; Hargreaves and McCallum, 1998; Schwartz, 1984; Sommers, 2011; Struthers, Lapadat and MacMillan, 2013). It is also consistent with others that have found that primary school students improve superficial aspects much more than macro-structural elements of their, after receiving the evaluation of teachers (Matsumara, Patthye-Chávez, Valdés, and Garnier, 2002; Montanero et al., 2014; Searle & Dillon, 1980).

On the other hand, the fact is that the vast majority of evaluation messages have an explicitly corrective nature, with few signs, questions, additions or justifications, which encourage self-regulatory behaviour (related to understanding and self-assessment error). This may be related to the assessment context itself. Possibly, if the sample of assessments had been collected in the framework of a formative evaluation, in which students had to subsequently improve their own narratives, such messages would be more frequent. In this sense, it is possible that if teachers had the opportunity to publicly comment on the most distinctive or repeated evaluation errors, they would have orally added some questions or justifications, and even have exemplified some extensions.

Returning to the grading analysis, there has been a clear tendency to globally grade (85.7 %) compositions with a qualitative ordinal scale (61.9 %). These ratings were accompanied mostly with negative reviews or verbal disapproval of particular aspects of the narrative (86.3 %), mainly focused, again, on grammar (63.6 %). There is no consensus in the literature on the appropriateness of the use of positive and negative comments in the evaluation of essays. Some authors have highlighted the risk that is involved with excessively praising positive aspects (DeGross, 1992; Smith, 1997); while others argue that a focus on identifying errors fosters negative attitudes in students (Lunsford and Connors, 1993; Lee, 2003). It has been observed that although the additive correction systems may have more positive psychological effects on students than subtractive systems, it is pedagogically beneficial for students to develop error awareness (Morales, 2004).

In any case it seems clear that judgment value, (whether in the form of a numerical rating, or through expressions such as “great plot” or “it

does not convince me») should be accompanied by advice, suggestions or review procedures and correction by the student. Almost all global or local qualifications registered in our study were accompanied by a correction or guidance for improvement. However, most of these comments were either very general (without specifying where exactly is your problem or what rules or regulations should be followed for the thereof, resolution), or comments simply consist in correcting a letter or word. Only in 9.4% of the latter type of corrections did the teacher provided the grammatical or spelling rule. It is possible that this circumstance is related to the lack of time to fully correct the grammar and spelling aspects of relatively large groups of Primary School students, while making explanatory comments or suggestions for improvement. In any case, in addition to subtracting validity and reliability (Espin, Weissenburger, and Benson, 2004; Hayes, Hatch, and Silk, 2000; Miller and Crocker, 1990), the mere identification of spelling and grammatical errors seems an insufficient kind of evaluation to establish a diagnosis of the skills and actual limitations of students (Dunsmuir et al., 2015).

The clarity of the evaluative comments that the student receive, in this way, is essential for a teacher to interpret and improve their writer competence. It was found that, too often, the evaluation message is very detached from the corresponding fragment of text. In other cases, the message content is too vague or non-specific; for example, when the teacher says: "Separate into paragraphs" he or she could clarify that each of the three parts of a story that make up the plot should be separated into those three paragraphs. Other recurring instances of messages that do not provide an easy interpretation are the use of question mark or separation lines (without any related explanatory note) in different parts of the text.

Limitations and prospective

Assuming that the proposed evaluation by the teacher is a good indicator of what has been previously taught, it can be assumed that previous findings have implications beyond the assessment practice in primary classrooms. Writer competence in our schools is still subject to traditional styles of instruction and assessment that make students continue paying attention to superficial aspects of code (such as: rules of grammar,

spelling , presentation...), in contrast to the self-regulatory skills of the content; it makes writing assignments be «mechanical» learning and boring (Akkaya and Susar Kırmızı, 2010).

Having said that these conclusions should be considered with caution, given the inherent limitations of this study. Teachers who participated in this research were not the student's teacher. Neither had they expected the students to re-write and improve the text as a result of the assessment afterwards. This could have influenced the accuracy of their orientation and evaluative messages. In future researches it would be desirable, therefore, to replicate these analyses with a larger selected randomly sample of teachers, in a formative assessment context with their own students.

References

- Akkaya, N., & Susar KİRMİZİ, F. (2010). Relationship between attitudes to reading and time allotted to writing in primary education. *Procedia Social and Behavioural Sciences*, 2, 4742–4746.
- Allan, E., & Driscoll, D. (2014). The three-fold benefit of reflective writing: Improving program assessment, student learning, and faculty professional development. *Assessing Writing*, 21, 37–55.
- Berg, E. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing*, 8, 215-241.
- Cassany, D. (2000). *Reparar la escritura*. Barcelona: Graó.
- Cho, K. & MacArthur, C (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20, 328-338.
- Cho, Y. (2003). Assessing writing: Are we bound by only one method? *Assessing Writing*, 8 (3), 165-191.
- Connors, R. & Lunsford, A. (1993). Teachers' Rhetorical comments on student papers. *College Composition and Communication*, 44 (2), 200-223.
- DeGroff, L. (1992). Process-writing teachers' responses to fourth-grade writers' first drafts. *The Elementary School Journal*, 93 (2), 131–144.

- Dempsey, M. S., PytlikZillig, L. M., & Bruning, R. H. (2009). Helping preservice teachers learn to assess writing: Practice and *feedback* in a Web-based environment. *Assessing Writing*, 14, 38-61.
- Dunsmuir, S. Kyriacou, M. Batuwitage, 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.
- Espin, C. A., Weissenburger, J. W., & Benson, B. J. (2004). Assessing the writing performance of students in special education. *Exceptionality*, 12, 55–66.
- Fahimi, Z., & Rahimi, A. (2015) On the Impact of Self-assessment Practice on Writing Skill. *Procedia-Social and Behavioural Sciences*, 192, 730 – 736.
- Ferris, D. R. (1997). The Influence of Teacher Commentary on Student Revision. *TESOL Quarterly*, 31, 315–339.
- Fife, J. M., & O'Neill, P. (2001). Moving beyond the written comment: narrowing the gap between response practice and research. *College Composition and Communication*, 53 (2), 300-321.
- Glaser, C., & Brunstein, J. (2007). Improving fourth-grade students' composition skills: effects of strategy instruction and self-regulation procedures. *Journal of Educational Psychology*, 99(2), 297-310.
- Graham, S., Harris, K., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: the effects of self-regulated strategy development. *Contemporary Educational Psychology*, 30(2), 207-241.
- Hargreaves, E., & McCallum, B. (1998). *Written feedback to children from teachers*, ERSC Project 'Teaching, learning and *feedback* strategies'. Project Paper No. 7, London: University of London, Institute of Education.
- Hayes, J. R., Hatch, J. A., & Silk, C. M. (2000). Does holistic assessment predict writing performance? *Written Communication*, 17, 3–26.
- Jones, A., & Moreland, J. (2005). The importance of pedagogical content knowledge in assessment for learning practices: A case-study of a whole-school approach. *Curriculum Journal*, 16(2), 193-206.
- Lee, I. (2003). L2 writing teachers' perspectives practices, and problems regarding error *feedback*. *Assessing Writing*, 8 (3), 216–237. DOI: 10.1016/j.asw.2003.08.002

- Lennart, J. (2014). Pair assessment of pupil writing: A dialogic approach for studying the development of writer competence. *Assessing Writing*, 20, 37-52. DOI: 10.1016/j.asw.2014.01.002
- Liang, M-Y. (2010). using synchronous online peer response groups in EFL writing: revision-related discourse. *Language Learning & Technology*, 14(1), 45-64.
- Matsumara, L. C., Patthye-Chavez, G. G., Valdes, R., & Garnier, H. (2002). *Feedback*, writing assignment quality, and third-grade students' revisions in lower- and higher-achieving urban schools. *The Elementary School Journal*, 103 (1), 3-25.
- Miller, H. D., & Crocker, L. (1990). Validation methods for direct writing assessment. *Applied Measurement in Education*, 3, 285-296.
- Min, H. T. (2003). Why peer comments fail? *English Teaching and Learning*, 27(3), 85-103.
- Montanero, M., Lucero, M. & Fernández, M. J. (2014). Iterative co-evaluation with a rubric of narrative texts in Primary Education / Coevaluación iterativa con rúbrica de textos narrativos en la Educación Primaria. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, 37 (1), 184-220. DOI: 10.1080/02103702.2014.881653
- Morales, F. (2004). Evaluar la escritura, sí...Pero ¿Qué y cómo evaluar?. *Acción Pedagógica*, 1 (13), pp. 38-48.
- Parr, J., & Timperley, H. (2008). Teachers, schools and using evidence: Considerations of preparedness. *Assessment in Education: Principles, Policy and Practice*, 15(1), 57-71.
- Peterson, S. S. & McClay, J. (2010). Assessing and providing *feedback* for student writing in Canadian classrooms. *Assessing Writing*, 15 (2), 86-99. DOI: 10.1016/j.asw.2010.05.003
- Rouhi, A., & Azizian, E. (2013). Peer review: Is giving corrective *feedback* better than receiving it in L2 writing? *Procedia-Social and Behavioral Sciences*, 93, 1349-1354.
- Schwartz, M. (1984). Response to writing: a college-wide perspective. *College English*, 46(1), 55-62.
- Searle, D., & Dillon, D. (1980). The message of marking: Teacher written responses to student writing at intermediate grade levels. *Research in the Teaching of English*, 14, 233-242.
- Shepard, L., Hammerness, K., Darling-Hammond, L., Rust, F., Baratz-Snowdon, J., Gordon, E., et al. (2005). Assessment. In L. Darling-

- Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275-326). San Francisco: Jossey-Bass.
- Slomp, D. H. (2012). Challenges in assessing the development of writing ability: Theories, constructs and methods. *Assessing Writing*, 17, 81–91.
- Smith, S. (1997). The genre of the end comment: Conventions in teacher responses to student writing. *College Composition and Communication*, 48 (2), 249–268.
- Sommers, J. (2011). Reflection revisited: Using the class collage. *Journal of Basic Writing*, 30(1), 99–129.
- Straub, R. (1997). Students' reactions to teacher comments: An exploratory study. *Research in the Teaching of English*, 31, 91–119.
- Struthers, L., Lapadat, C. y MacMillan, D. (2013). Assessing cohesion in children's writing: Development of a checklist. *Assessing Writing*, 18 (3), 187-201.
- Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9 (2), 147–170.

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