Rojo sobre negro ¿Cómo evalúan los maestros las redacciones de sus estudiantes?
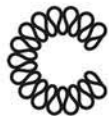
Red on black. How do teachers assess their students' narrative texts?

María-Jesús Fernández
Manuel Lucero
Manuel Montanero

revista de EDUCACIÓN

IV CENTENARIO
CERVANTES

revista de EDUCACIÓN

Nº 372 Abril-Junio 2016

Revista trimestral
Fecha de inicio: 1952

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

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

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.

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 a 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 o 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 massage 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 where 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	*Feedback* 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	*Feedback* on the topic or ideas of the text. Includes assessment on coherence or specific vocabulary use.
	3. Grammar	*Feedback* about the morph syntactical construct of sentences. Includes the modification of the word order, cohesion assessment cohesion or punctuation.
	4. Spelling	*Feedback* 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	*Global Feedback*, 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.

TABLE V. 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 vivia 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 (DeGroff, 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 Kırmızı, 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