

The effects of an experimental approach to writing instruction on monolingual and multilingual pupils in Italian primary schools

GABRIELE PALLOTTI^a & CLAUDIA BORGHETTI^b

^aUniversità di Modena and Reggio Emilia

^bUniversità di Bologna

Received 6 December 2018; received in revised form 24 May 2019; accepted 8 July 2019

ABSTRACT

EN This article reports on the results of an action-research project in Italian primary schools. Its main principles are formative evaluation, a process orientation to writing, cooperative learning, and an inclusive approach that does not separate pupils according to their linguistic or writing skills. Pupils aged eight to eleven years old participated in the study in seven experimental (N = 106) and seven control classes (N = 118). Their texts were examined both holistically, through scales of communicative effectiveness, and by means of analytical measurements. The comparison of monolingual and multilingual pupils in both experimental and control classes shows a number of similarities in terms of textual competence, as well as a few differences, especially regarding the more linguistic aspects of writing, such as verb tense cohesion. A remarkable result is that students' text quality seems to depend more on their participation to the program than on their monolingual or multilingual status.

Key words: MULTILINGUAL PUPILS, L1/L2 WRITING, WRITING DEVELOPMENT, WRITING INSTRUCTION, PRIMARY SCHOOL

ES Este artículo informa los resultados de un proyecto de investigación-acción en escuelas primarias italianas. Sus principios más importantes son la evaluación formativa, una orientación al proceso de escritura, aprendizaje cooperativo y un enfoque inclusivo que no separa a los estudiantes según sus habilidades lingüísticas o de escritura. Los alumnos de entre ocho y once años participaron en el estudio en siete clases experimentales (N = 106) y siete clases control (N = 118). Sus textos se examinaron de manera holística por medio de escalas de efectividad comunicativa y mediciones analíticas. La comparación de alumnos monolingües y multilingües en las clases experimentales y de control muestra un número de similitudes en términos de competencia textual al igual que algunas diferencias, especialmente con respecto a los aspectos más lingüísticos de la escritura, como la cohesión del tiempo verbal. Un resultado notable muestra que participar en el programa tuvo un mayor impacto en la calidad del texto de los estudiantes que sus condiciones iniciales mono- o multilingüe.

Palabras clave: ALUMNOS MULTILINGÜES, ESCRITURA L1/L2, DESARROLLO DE ESCRITURA, INSTRUCCIÓN DE ESCRITURA, ESCUELA PRIMARIA

IT Questo articolo dà conto dei risultati di un progetto di ricerca-azione in scuole primarie italiane. I principi centrali sono la valutazione della formazione, l'orientamento al processo della scrittura, l'apprendimento cooperativo e l'approccio inclusivo che non separa gli/le apprendenti in base alle loro capacità linguistiche o di scrittura. A questo studio hanno partecipato apprendenti di età compresa tra gli 8 e gli 11 anni, in sette classi sperimentali (N = 106) e in sette classi di controllo (N = 118). I loro testi sono stati esaminati sia olisticamente attraverso scale di efficacia comunicativa, sia mediante misurazioni analitiche. Il confronto tra alunni monolingui e multilingui, sia nelle classi sperimentali sia in quelle di controllo, mostra una serie di similarità in termini di competenza testuale e anche alcune differenze, in particolare per quanto riguarda gli aspetti più linguistici della scrittura, come la coesione verbale. Un risultato importante mostra che la partecipazione al programma ha avuto un impatto maggiore sulla qualità del testo degli/delle apprendenti rispetto alle loro condizioni iniziali di monolingui o multilingui.

Parole chiave: APPRENDENTI MULTILINGUI, SCRITTURA L1/L2, SVILUPPO DELLA SCRITTURA, INSEGNAMENTO DELLA SCRITTURA, SCUOLA PRIMARIA

✉ **Gabriele Pallotti**, Università di Modena e Reggio Emilia
gabriele.pallotti@unimore.it

1. Introduction

In the field of teacher education, literacy and second language acquisition (SLA) are often kept separate. In teaching practice too, actions aimed at improving writing skills and actions stimulating second (or additional) language acquisition are at times seen as complementary, but substantially distinct, based on different theoretical and methodological assumptions and approaches. This article reports the results of an action-research teacher-training project intended to help bridge these gaps, by adopting an approach to the teaching of writing that involves all students, regardless of whether they are using their first, second, or additional language. More specifically, the context of the study was a teacher-training initiative *Osservare l'interlingua* whose aim was to promote effective language education, with special emphasis on writing, at elementary and middle school levels.

Writing is a complex activity, involving several sub-skills (e.g., sorting through one's memory of the events/facts, organizing them, selecting appropriate language resources), which may be more or less developed in different people. This complexity is greater in multilingual speakers with variable control of different linguistic codes and their sub-domains, which in turn may be more or less implicated in written communication. Teachers need to be able to take apart and carefully analyze all these factors, in order to allow for them in their daily practices. In most cases, these daily practices should be directed at the whole class, rather than specific sub-groups made up of monolingual or multilingual students, or first or second language learners, with higher or lower writing skills. This article reports on a program that attempts to meet these objectives, and whose effectiveness will be assessed using a quasi-experimental research design.

Before proceeding with the exposition, some terminological clarification is in order. The term *second language* (or *L2*) has been repeatedly called into question in recent years. Firstly, in many cases the language at stake is not the *second* a person is acquiring, but often a third, fourth, or further language. This is why *additional language acquisition* has been proposed as a more correct replacement for second language acquisition, and *LX* as a more general term than *L2* (Dewaele, 2017). As a consequence, many so-called *bilinguals* should actually be called *multilingual speakers*. This term *multilingual* also overcomes the limits of expressions such as *non-native speaker* or *L2 learner*, which imply some sort of deficit with respect to an idealized native speaker who is treated as the necessary benchmark of language performance (Cook, 1999). Secondly, the simple L1/L2 dichotomy does not apply to other groups of learners, such as so-called heritage speakers, i.e., children of immigrants who are either monolingual or dominant in the language of the family in early childhood, but progressively develop a balanced bilingualism or eventually become dominant in the majority language (Montrul, 2016).

This article adopts solely the distinction between monolingual and multilingual students (except when citing research that explicitly refers to "L2 acquisition" and "bilingualism"). This categorisation seems relatively descriptive and neutral: it does not require one to draw the problematic line between first and second language acquisition, or between native and non-native speakers, and it does not assume any deficit view of language learning and use, but rather describes a person's linguistic repertoire as being composed of one or more codes employed in everyday communication.¹

2. Writing development and instruction: an overview

2.1. Writing in L1 and LX

The practices of L1 and LX writing have frequently been compared, both in terms of processes (i.e., the cognitive activities employed when writing) and in terms of products (the texts). Analysing the interaction of four cognitive activities implied in writing (reading the assignment, planning, generating ideas, and formulating), van Weijen (2009) concludes that the distribution of each activity is similar in both L1 and L2 writing. These results confirm many previous studies (Cumming, Rebuffot & Ledwell, 1989; Uzawa, 1996; for a review: Cumming, 2001). However, a consistent number of differences between L1 and LX cognitive activities have also been highlighted in text planning and revision, and in writing fluency (Jones & Tetroe, 1987; Hall, 1990; Stevenson, Schoonen, & de Gloppe, 2006; van Waes & Leijten, 2015). This overall empirical evidence (e.g., Leki, Cumming, & Silva, 2006; Silva, 1993) tends to converge into a general agreement that writing processes in L1 and LX are similar, even if "L2 composing is more constrained, more difficult, and less

¹ The *multilingual* category does not include those children who only know a smattering of a foreign language taught at school in addition to their first language.

effective” (Silva, 1993, p. 668), mainly because LX writers need to concentrate on language forms as well as on content.

Contrasting conclusions have also been drawn from research which compares the linguistic properties of texts produced by L1 and LX writers. For example, Silva (1993) claims that L2 writers’ texts contain fewer words, while Cameron and Besser (2004) point out that text length does not differ significantly, even if L2 writers tend to make lesser use of adverbials, which in turn reduces the amount of information they provide about the story (e.g., in terms of time, place and manner of the events).

Some research has also highlighted that LX writers employ a restricted repertoire of features (especially conjunctions and demonstrative pronouns) to give cohesion to their texts (Hinkel, 2001; see also: Oi, 1984 and Reid, 1988, both cited in Silva, 1993). However, less is known about how L1 and LX writers manage reference to entities, that is, which linguistic and discourse features they use to make the characters of their stories identifiable by the reader. While it has been claimed that L1 children learn to systematically use appropriate referent introductions from seven years onwards (Hickmann, 1995), little or no research has been conducted, to the best of our knowledge, on the developmental patterns of young LX writers, and one of the aims of the present article is to start filling this gap. Hendriks’s (1998) study, comparing the written productions of adult LX learners with those of L1 children, may in fact be more informative in terms of the impact of age—rather than “native” status—on cohesion skills and/or preferences.

2.2. Research-based writing instruction

A considerable body of research has accumulated on the effectiveness of instruction on writing quality (e.g., Graham, Harris, & Santangelo, 2015; Rietdijk, Janssen, van Weijen, van den Bergh, & Rijlaarsdam, 2017), and meta-analyses are also available (e.g., Graham, Harris, & Chambers, 2016), some of which specifically focus on primary school children (Graham, McKeown, Kiuahara, & Harris, 2012; Koster, Tribushinina, de Jong, & van den Bergh, 2015). These last authors identified a number of instructional practices that have systematically produced positive results with younger learners. These include:

- Teaching writing strategies for planning, drafting, or revising one’s text, e.g., teaching pupils to organize ideas in writing plans and to use/upgrade the plan when writing (Tracy, Reid, & Graham, 2009).
- Assigning specific goals before and during the writing process, like asking students to add information about a specific fact/character, to divide the text into paragraphs, or, as Ferretti, Lewis, and Andrews-Weckerly (2009) have shown, to follow some given prompts when organising their argumentative discourse.
- Promoting group work and peer collaboration, as pupils plan, draft, and revise their text; see, for example, Brakel Olson (1990) about peer-to-peer support in revision.
- Teaching text structure explicitly, e.g., how narrative texts work in terms of story constituents and event sequences (Fitzgerald & Teasley, 1986).
- Providing teacher and peer-to-peer feedback during and after writing (e.g., Schunk & Swartz, 1993).

Not surprisingly, these research-based best practices in L1 writing instruction are also often mentioned in relation to LX writing. For example, Olson, Scarcella, and Matuchniak (2013) suggest that, in order to limit the effect of LX writers’ *cognitive constraints*, educators should focus on the teaching of strategies which help students set intermediate goals and thus manage the complexity of the writing process. Moreover, LX collaborative writing and peer-to-peer feedback are often advocated (Storch, 2013), as well as genre pedagogy (Hyland, 2007).

However, research to date on the (differential) effects of pedagogical interventions on monolingual and multilingual pupils’ writing remains scant. This is one of the motivations behind this study, whose pedagogical approach largely draws on the works briefly reviewed in this section.

3. The study

3.1. Research questions

The following research questions were formulated in order to investigate the outcomes of an inclusive writing program on monolingual and multilingual learners:

- RQ1. In a primary school context, what are the effects on the whole class of a long-term pedagogical intervention based on the process approach to writing, formative assessment, and cooperative learning?
- RQ2. If there are any effects, do they differ for monolingual and multilingual pupils?
- RQ3. What are the differences between monolingual and multilingual pupils in control classes as regards writing skills?

3.2. The project

The project *Osservare l'interlingua*, promoted by the Reggio Emilia municipality together with the University of Modena and Reggio Emilia (interlingua.comune.re.it; Pallotti, 2010, 2017a, 2017b; Pallotti & Rosi, 2017) started in Reggio Emilia (Italy) in 2007. It involves 12 elementary and 3 middle schools, 28 teachers and 35 classes, for a total of about 650 pupils aged between 6 and 14 (Grades 1 to 8). While, at the beginning, it was meant to address issues of LX Italian learning, today it is directed toward all the pupils in class, with an inclusive approach involving monolingual and multilingual students with varying proficiency levels.

The project is both a teacher training and action-research initiative, with teachers actively involved in several ways. Its main tenet is that effective teaching practices should be grounded in the careful observation of students' needs, competences, and learning strategies. Thus, teachers are first of all trained to conduct systematic analyses of their pupils' oral and written productions, in line with the interlanguage approach (Pallotti, 2017b; Selinker, 1972). This is a kind of formative assessment that aims to overcome the traditional stance of "hunting for errors" and to allow teachers to develop the activities that best suit the needs of their class. It also promotes the learners' autonomy, since they are not seen as passive recipients of knowledge that has been elaborated elsewhere, but rather as active creators of their own learning. This orientation, and the ability to analyze children's texts accordingly, takes some time to develop. It begins with pre-service training for some of the younger teachers, while others often acknowledge that it took them some years to get used to looking at pupils' productions in positive terms through recognizing their strategies and developments, rather than just spotting errors. This is why text-analysis sessions are periodically held (about once a year) as part of the teachers' permanent training; in these sessions, more and less experienced participants assist each other in building their formative analytic skills.

Teachers also actively take part in planning pedagogical actions. At the beginning of each school year, the texts produced by pupils are analysed by the group of teachers and researchers. Considering the properties of the texts and thus the needs of each class, the group of teachers defines a common theme (e.g., narrative, expository, descriptive texts) to be pursued by all participants in their classes and they agree on common core activities. Then, each teacher introduces the due adaptations according to age and overall proficiency level of their pupils, receiving guidance from the researchers. The group of teachers and university researchers meets approximately every two months throughout the entire school year, so that planned activities can be monitored, discussed, and possibly revised.

The project as a whole promotes a process approach to teaching writing (e.g., White & Arndt, 1991), according to which teaching activities guide the pupils to experience and progressively acquire the stages implied in text production: generating and organizing ideas, writing the initial draft, revising, and editing. Most of these activities are carried out in a cooperative manner, in small and large groups (following Ede & Lunsford, 1990; Storch, 2013), and, in line with socio-cultural approaches, they encourage the semiotic mediation of thought by making the writing process more concrete through the use of artefacts such as boxes, envelopes, paper strips, and posters (following Englert, Mariage, & Dunsmore, 2006). Feedback, too, is primarily provided by the students, as the program particularly encourages peer-to-peer response, both within the members of the group and among groups (following Andrade & Evans, 2013). In line with the interlanguage approach, this feedback is not summative (no grades are assigned, by either teachers or students), but formative, as it focuses on pupils' strengths and weaknesses, analyses their strategies, and suggests ways of developing their skills.

Teachers are actively involved in data collection, too, which is usually carried out at the beginning and end of every school year, in relatively controlled and standardized conditions (that are nonetheless absolutely natural for a classroom setting—no laboratory experiments, in other words), so as to allow for comparisons over time and across pupils and classes. This data is kept by schools as a portfolio of pupils' linguistic development over the years, qualitatively analyzed by the teachers, and also used by researchers

(and some pre-service-training students as part of their master's thesis) for subsequent quantitative analyses, such as the kind reported in this article.

3.3. The pedagogical intervention (experimental classes)

In the course of the 2013-2014 school year, when data collection for this study took place, the pedagogical intervention carried out in most classes was aimed at developing the pupils' writing skills as applied to narrative texts.

Although the seven classes had slightly different needs and were thus engaged with partially diverse tasks, all of them were involved in the following core activities:²

- 1) Students wrote individual texts after watching, twice, a video clip of a few minutes taken from *Modern Times* by Charlie Chaplin. This served to collect baseline data to assess students' progress over the school year and to fine-tune teaching activities based on their competences and needs.
- 2) In groups, the students viewed the initial video clip again, identified the main sequences of the story (primary information units), and labelled them with short descriptions written on paper envelopes or cardboard boxes.
- 3) The groups watched the video once more and gathered additional information about characters, actions, etc. (secondary information units), which they noted on paper strips to be placed in the envelopes/boxes.
- 4) In their groups, pupils glued the strips on big sheets of paper, deciding on the order to best convey the story of the video clip. This procedure led to the text outline. Since a main objective of the program was to promote a better organization of texts, special emphasis was placed on the need to cluster similar strips of information together, in order to facilitate the following segmentation of the text into paragraphs.
- 5) The groups produced the first versions of their texts, based on the video clip they had worked on in the first four steps.
- 6) The texts were then circulated among groups, who had been previously instructed on how to make a peer-to-peer revision based on explicit and shared criteria (e.g., clarity, lexical choices, spelling). It is worth noting that a number of aspects of writing, such as verb tense cohesion and punctuation, were addressed within this phase merely by asking pupils to take them into consideration in the reviewing process. These topics did not receive significant attention within the other activities.
- 7) Each group reviewed their own text, taking into account the comments of the other groups. They then rewrote it, after having to decide which suggestions for improvement they wanted to accept.
- 8) Teachers guided a class debriefing on the experience and stimulated meta-cognitive reflections on how to produce good narrative texts.
- 9) At the end of the school year, teachers used an additional video stimulus taken from *World of Comedy* by Harold Lloyd to collect a second round of individual texts. This video was different from the first one, in order to assess the transferability of skills across tasks, although the two inputs had some similar characteristics (e.g., number of main characters, plot complexity, duration) and were administered in identical conditions.

In terms of White & Arndt's (1991) process approach to writing, Steps 2 and 3 correspond to the stages of *Generating ideas* and *Focusing*; Step 4 corresponds to *Text Structuring*, and step 5 to *Drafting*. The processes of *Evaluating* and *Re-viewing* were covered in Steps 6 and 7.

The activities above were just a part of the experimental classes' syllabus, accounting for approximately two hours a week of class time over the entire school year. The project involved all classes, from Grades 1 to 5. Obviously, activities for the first two grades were simplified and involved much less writing, but they still followed the general orientation of engaging with different phases of the writing process one by one. Given that the data discussed in the following pages come from Grades 3-5, no further details on the first two grade levels are offered here: for more information about the teaching phases and

² All texts are translated from Italian. Originals are available on the author's website: <http://www.gabrielepallotti.it>. Concerning translations, when the morpho-syntactic and orthographic mistakes made by the pupils did not have any straightforward equivalent in English, similar types of errors were introduced in order to give a better sense of the overall writing quality.

methods and the variations introduced in different classes and age levels, see Pallotti (2017a, 2017b), Pallotti & Rosi (2017) and the project website (interlingua.comune.re.it).

3.4. Data collection

To evaluate the effectiveness of the training program, individual texts were collected in seven experimental and seven control classes, Grades 3-5 (ages 8-11), at the beginning and end of the school year. Most experimental classes had been participating in the project since Grade 1, so that at the time of data collection they had already been performing these types of activities for two, three, or four years. Pupils in control classes at the beginning and end of the school year wrote their texts under exactly the same conditions as their peers in the experimental classes (i.e., They watched the two video clips by Chaplin and Lloyd twice) and were given the same generic assignment (i.e., Describe the video to a teacher who has not seen it), with no directions about text planning and writing and no time limit to accomplish the task. Control classes followed a rather regular language education curriculum (for Italy), in which writing activities mainly consisted in composing short essays or narrative texts, training in orthography and grammar, or answering comprehension questions after reading. Little or no attention was devoted to individual phases of the writing process, no artefacts were used as semiotic mediators for higher-order cognitive functions like text planning and revision, and most writing activities had no real communicative intent but were conceived as exercises to be assessed by teachers, who were the only feedback providers.

Although great care was taken to ensure that control classes paralleled experimental classes as much as possible, the correspondence was not perfect, as is often the case with quasi-experimental studies. Control classes came from the same schools or areas of the city as the experimental ones and they had a comparable number of pupils per class. However, their demographic composition differed from multilingual classes, as shown in Table 1. Experimental classes had more multilingual children, over one third of whom were not born in Italy and had spent an average of just two years in the country. Control classes had fewer multilingual pupils, less than one fifth of them were not born in Italy and the average length of residence in Italy for these children was longer than in experimental classes.

Table 1
Group composition

	Control	Experimental
Sample size	118	106
% multilingual pupils in entire class	24%	36%
% multilingual pupils not born in Italy among all multilingual pupils	18%	38%
Mean period of residency in Italy for multilingual pupils born abroad	3.6 yrs	2.1 yrs

A second difference concerns the scores at the beginning of the school year, which, for several dimensions discussed in this article, were higher in experimental classes. While this may be seen as unfortunate from a strictly scientific point of view, from a pedagogical perspective this difference is quite positive, since it seems to indicate that classes taking part in the experimental program from first grade gradually accrue an advantage that sets them apart from those not involved in the program. This assumption is substantiated by the fact that, for most measurements, the initial gap between experimental and control classes was lower in the third grade and higher in the fourth and fifth grades—that is, the gap grew larger in parallel to a longer participation in the project. Furthermore, the experimental group's higher scores at the beginning of the school year cannot be ascribed to the fact that these pupils came from privileged socio-cultural environments; on the contrary, most experimental classes were located in somewhat disadvantaged areas of the city.

Given this difference in initial scores, this article focuses on final results only, as a way of comparing the program's effects accrued after three, four, or five years. In other words, it seems more appropriate to assess the cumulative effects of the experimental program (which began in first grade for all experimental classes), rather than discussing the gains obtained by the pupils from the beginning to the end of just one school year.

3.5. Data analysis

Handwritten texts were first transcribed and assigned arbitrary codes that did not allow raters and coders to know whether they came from experimental or control classes, or whether they were the work of monolingual or multilingual students, to avoid any possible bias in the procedures requiring a qualitative evaluation. Pupils diagnosed with cognitive or linguistic impairment, or who did not attend the initial or final data collection sessions, were excluded from the analysis, which was thus based on 106 participants in experimental classes and 118 in control classes. Finally, scripts were examined both holistically and by means of analytic measurements.

Holistic examination was carried out using scales to rate overall communicative effectiveness. In particular, functional adequacy was rated with three scales slightly adapted from Kuiken and Vedder (2017), while a scale taken from the Common European Framework (Council of Europe, 2001) was used for assessing text coherence and cohesion. The scales are reported in Appendix A. In order to assess inter-rater reliability, 10% of the scripts were rated on the four scales by two independent researchers, who showed a substantial level of agreement: the average Intraclass correlation coefficient (two-way mixed effects model) was .85. A series of features was then considered for the analytical measurement of text properties, covering the dimensions outlined below.

3.5.1. Content

This dimension was assessed by looking at text length, measured by the number of words produced and by counting the number of primary and secondary idea units. Similar to other studies (e.g., Boscolo, Arfé, & Quarisa, 2007), these were agreed upon by the whole research team, who produced a list of five main information units, representing the story's key sequences, and 43 secondary units, representing specific events in the story plot.

3.5.2. Text organization

This dimension refers to the pupils' ability to effectively divide and organize their text into thematic and syntactic blocks. It is evaluated firstly by counting the number of paragraphs, and their correlated measure, the number of words per paragraph. Paragraphing is normally not taught in Italian schools, so that pupils often end up writing texts in just one single long block. Our pedagogic intervention explicitly focused on this skill, because it was considered closely related to planning the text in several thematic units.

At a different level of textual organization, the comma serves to divide syntactic units, and its appropriate use is rarely taught in Italian schools, where it is often simplistically explained as a way of marking pauses and intonation or to separate items in a list. Pupils are introduced to this punctuation mark between Grades 2 and 3 and they soon start using it, initially in a few prototypical contexts, such as lists. Gradually, its usage is expanded to express more sophisticated meanings, like differentiating levels of text organization or signalling discourse structure, but this comes with a higher risk of making mistakes, which persist throughout middle and high school with pupils rarely being completely aware of their nature (Solarino, 2009). Thus, a higher number of commas may be taken as an indicator of pupils' willingness to experiment with this important but challenging and often misunderstood punctuation mark. However, a greater use of the comma should not be automatically interpreted as an indicator of better text quality. This is why the proportion of inappropriate commas to all commas used was measured, too. Omissions were intentionally not counted, since establishing when a comma is "missing" is often a matter of discretion. Another indicator of a good command of punctuation is the ratio of inappropriate uses of the period. However, the number of periods per 100 words was not considered to be a clearly interpretable measure: on the one hand, it may signal a desirable feature, that is a higher degree of text segmentation and avoidance of run-on sentences; on the other, it may imply a very paratactic style, with lots of simple, telegraphic sentences, which do not necessarily contribute to text quality.

3.5.3. Reference to entities

Managing reference to entities in discourse is a notoriously difficult task for young learners, and sometimes for adult speakers and writers as well (Giuliano, 2007; Hendriks, 1998; Hickmann, 1995). An entity can be *introduced*, upon its first mention, and then *maintained* in the following clauses. This referential chain can be interrupted with the introduction of new entities, but the original entity can be later on *reintroduced* in the text. In the following Fictive Text (1), introductions are indicated by (I), maintenances by

(M), and reintroductions by (R). Coding and analysis were conducted on animate entities only, as they are a key element in plot progression and thus play a central role in narrative texts.

Fictive Text (1)

There was a boy(I) who went to the park. He(M) was very shy and Ø(M) had very few friends. An old man(I) came to the park and Ø(M) took out a worn-out kite. He(M) looked a bit lonesome, too. The boy(R) asked him(M) if they(M) could play together

Normally, entities are introduced by lexical noun phrases with indefinite determiners (*a boy, an old man*), they are maintained by pronouns or zero anaphora (*he, Ø, him*) and are re-introduced by lexical noun phrases with definite determiners (*the boy*). A common feature of young children's speech and writing is that entities are often introduced with definite determiners, so that, in the example above, the first sentence could have been something like, *The boy went to the park*. These introductions may be called inappropriate, as they indicate that their author improperly assumes the referent is accessible to the hearer/reader. However, although they may sound childish or slightly egocentric, they do not compromise communication and the meaning remains clear. In this article, however, these phenomena will not be discussed (for further discussion on this aspect, see Pallotti, Borghetti, & Rosi, in preparation).

What will be focused on here are ambiguous references that prevent the receiver from correctly interpreting their meaning. If the text above had begun, *He went to the park*, this would have been counted as an ambiguous introduction. Likewise, if the last sentence had been, *He asked him if they could play together*, it would have been impossible to establish who was asking whom. Only by knowing the intended meaning (which is possible, for instance, when pupils are retelling a film or a picture story) can one score the first *he* as an ambiguous reintroduction and the immediately following *him* in this sentence as an ambiguous maintenance.

3.5.4. Verb tense cohesion

Another common feature of young children's language is the frequent shifting of verb tenses in the same narrative (Kersten, 2009), as in sentences like: *Then he starts skating and she went to bed*. A verb shift is defined as a change of verb tense that is not justified for textual reasons, such as the move from the description of background states to the narration of events in the plot, or from indirect to direct speech. The operational definition of all these categories was set out in a detailed coding manual, which provided specific examples for each phenomenon, including possible counterexamples and borderline cases.

As regards the statistical procedures employed in the analysis, given that the study was a quasi-experimental investigation of intact classes and that initial levels were not homogenous for several dimensions, Welch t-tests for independent samples were employed to compare the averages of experimental and control classes at the end of the school year. According to Fagerland and Sandvik (2009), this test is best suited for samples like ours, which consist in about 100 cases, normally distributed or moderately skewed, with similar but not identical variances. For holistic assessment data, whose distributions were substantially normal, t-tests without Welch correction were used, since they are more appropriate for Likert-like data, with many tied ranks (de Winter & Dodou, 2010).

4. Results

This section reports on the main results regarding program effectiveness at the end of the school year. Tables include the mean and standard deviation, the 95% confidence interval for the estimation of the difference between the means of experimental and control classes, Cohen's *d* effect size, calculated as the ratio between the difference between means and the standard deviation of the control group, and the t-test's two-tailed *p* values. Results will be reported firstly for entire classes, then for monolingual and multilingual sub-groups.

4.1. Holistic assessment

At the end of the school year, text assessment through rating scales showed very positive results for the experimental classes as a whole. Their compositions were perceived as being richer in terms of content and more coherent and cohesive, both considering Kuiken and Vedder's (2017) scale and for the Common

European Framework scale (Council of Europe, 2001) with rather large effect sizes of between 0.61 and 0.78. They were also more comprehensible, with a smaller but still statistically significant difference ($d = 0.36$).

This trend is also confirmed for multilingual pupils, who wrote more complete, coherent, and cohesive texts than their control peers, although differences were less marked (d between 0.39 and 0.53) and even smaller when compared to monolingual students in control classes (d between 0.02 and 0.18). Multilingual pupils in experimental classes scored slightly lower ($M = 3.24$) than multilingual and monolingual peers ($M = 3.41$) in control classes only with regards to text comprehensibility however the difference is not significant. Monolingual pupils in experimental classes largely outperformed their monolingual peers in control classes in all dimensions. These overall results are shown in Table 2.

Table 2
Overview of results from holistic assessment

	Control Group Mean (SD)	Experimental Group Mean (SD)	95% CI of the difference between means	Cohen's d	p
<i>Content (1-6)</i>	3.39 (1.13)	4.09 (1.06)	0.39 - 0.99	0.61	< .0001
Monolingual pupils	3.46 (1.19)	4.32 (1.07)	0.50 - 1.22	0.72	< .0001
Multilingual pupils	3.19 (0.92)	3.68 (1.08)	-0.99 - 0.01	0.53	(ns)
<i>Comprehensibility (1-6)</i>	3.53 (0.93)	3.86 (1.01)	0.07 - 0.60	0.36	< .0001
Monolingual pupils	3.56 (0.95)	4.21 (0.87)	0.36 - 0.94	0.68	< .0001
Multilingual pupils	3.41 (0.89)	3.24 (0.98)	-0.30 - 0.64	-0.18	ns
<i>Coherence-Cohesion KV (1-6)</i>	2.89 (0.81)	3.47 (1.04)	0.33 - 0.82	0.72	< .0001
Monolingual pupils	2.95 (0.82)	3.74 (0.92)	0.50 - 1.07	0.96	< .0001
Multilingual pupils	2.65 (0.73)	2.97 (1.01)	-0.74 - 0.13	0.42	ns
<i>Coherence-Cohesion CEF (1-8)</i>	3.25 (1.32)	4.27 (1.40)	0.66 - 1.39	0.78	< .0001
Monolingual pupils	3.33 (1.31)	4.71 (1.25)	0.97 - 1.79	1.05	< .0001
Multilingual pupils	2.96 (1.34)	3.49 (1.39)	-1.21 - 0.17	0.39	ns

Note. KV = rating scale by Kuiken and Vedder (2017); CEF = rating scale of the Common European Framework (Council of Europe, 2001, 2017).

4.2. Content

We now turn to analytical measures of text quality. The texts produced in experimental classes are much longer than those of control classes, with an average of 243.84 words versus 177.93; they also contain more secondary information units (22.56 vs. 17.81) and thus provide a more detailed and rich content, with considerable effect sizes (0.83 and 0.78, respectively). The difference is smaller, but still statistically significant, as regards main idea units (4.79 vs 4.60), since most pupils, in both experimental and control classes, were able to mention all or almost all of them.

The trends identified at the entire-class level are also confirmed by multilingual pupils in experimental classes, who write longer texts than multilingual students in the control group. Interestingly, their productions are just slightly shorter than those of their monolingual peers in experimental classes, and considerably longer than the texts written by monolingual pupils in the control group.

This greater text length reflects the number of primary and secondary information units. The number of units is higher among the multilingual students in the experimental group than in both monolingual and multilingual controls. Similar to the results discussed in the previous section, the pupils in the experimental groups who demonstrate the greatest advantage over their counterparts in the control groups are the monolingual students (see Table 3).

Table 3
Content (analytic measures)

	Control Group Mean (SD)	Experimental Group Mean (SD)	95% CI of the difference between means	Cohen's <i>d</i>	<i>p</i>
Words	177.93 (79.12)	243.84 (96.67)	41.82 - 90.00	0.83	< .0001
<i>Monolingual pupils</i>	170.09 (69.04)	258.95 (97.00)	61.03 - 116.70	1.29	< .0001
<i>Multilingual pupils</i>	203.19 (102.83)	216.89 (96.76)	-64.60 - 37.19	0.13	ns
Main idea units (0-5)	4.60 (0.65)	4.79 (0.41)	0.04 - 0.34	0.29	< .05
<i>Monolingual pupils</i>	4.61 (0.64)	4.82 (0.46)	0.034 - 0.38	0.33	< .05
<i>Multilingual pupils</i>	4.56 (0.70)	4.73 (0.51)	-0.49 - 0.14	0.25	ns
Secondary idea units (0-43)	17.81 (6.12)	22.56 (6.13)	3.06 - 6.45	0.78	< .0001
<i>Monolingual pupils</i>	17.41 (5.81)	23.30 (6.34)	3.91 - 7.86	1.01	< .0001
<i>Multilingual pupils</i>	19.07(7.00)	21.24 (6.68)	-5.65 - 1.32	0.31	ns

4.3. Text organization

Students in experimental classes started new paragraphs more frequently than those in control classes, with an average of 5.32 paragraphs per text versus 2.39. As a result, the mean number of words per paragraph was lower: 74.82 as opposed to 116.89. These differences are statistically significant with medium-to-high effect sizes (see Table 3).

Pupils in experimental classes used the comma more frequently (3.63/100 words vs. 2.34/100 words), while their use of periods is virtually identical to that of control classes (3.50 vs. 3.63/100 words), meaning that their sentences were about the same length.

Inappropriate commas, e.g. those separating a verb from its necessary arguments, were also scored. In spite of their more frequent use of the comma, which implies taking more risks (and probably committing fewer errors of omission, which however were not scored because of the limited reliability of this analytic category), pupils in experimental classes made slightly fewer errors than those of control classes (13.38% vs. 15.73% of all commas), and significantly fewer errors regarding the period (1.44% vs. 5.14%).

Multilingual pupils, once again, have a substantially similar profile to that of the monolingual students in each group. Those in experimental classes start about twice as many new paragraphs as monolingual and multilingual pupils in control classes. As a result, their paragraphs are on average shorter (70.58 words per paragraph). Like their monolingual peers in the experimental group, they use commas more frequently than multilingual and monolingual students in control classes: 2.63 commas/100 words vs. 2.03 and 2.44, respectively. These commas are on average more correct than those produced by the monolingual control pupils (15.18% vs. 17.35% of incorrect commas), but slightly less so than those used by the multilingual control pupils (9.67% of incorrect commas), perhaps because the latter subgroup used this punctuation mark the least.

Table 4 below reports an overview of results for each feature. Also, texts in Extract (1) exemplify two different phenomena related to paragraphing and use of punctuation in a third-grade pupil belonging to the experimental group.³

³ All texts are translated from Italian. Originals are available on the author's website: www.gabrielepallotti.it. Concerning translations, when the morpho-syntactic and orthographic mistakes made by the pupils did not have any straightforward equivalent in English, similar types of errors were introduced in order to give a better sense of the overall writing quality.

Table 4
Text organization

	Control Group Mean (SD)	Experimental Group Mean (SD)	95% CI of the difference		
			between means	Cohen's <i>d</i>	<i>p</i>
Paragraphs	2.39 (2.00)	5.32 (3.28)	-3.68 - -2.19	1.47	< .0001
<i>Monolingual pupils</i>	2.36 (1.99)	5.70 (3.41)	-2.41 - -4.27	1.68	< .0001
<i>Multilingual pupils</i>	2.48 (2.06)	4.65 (3.10)	0.87 - 3.46	1.05	< 0.005
Words/paragraph	116.89 (81.57)	74.82 (80.21)	-20.36 - -63.78	-0.52	< .0001
<i>Monolingual pupils</i>	115.96 (83.24)	77.20 (89.64)	-10.68 - -66.85	-0.47	< .001
<i>Multilingual pupils</i>	119.90 (77.37)	70.58 (61.60)	13.12 - 85.52	-0.64	< 0.01
Commas/100 words	2.34 (2.22)	3.63 (2.31)	-1.93 - -0.65	0.58	< .005
<i>Monolingual pupils</i>	2.44 (2.25)	4.19 (2.55)	0.97 - 2.53	0.78	< .0001
<i>Multilingual pupils</i>	2.03 (2.15)	2.63 (2.12)	-1.68 - 0.48	0.28	ns
% inappropriate commas	15.73 (22.77)	13.38 (16.03)	-3.54 - 8.13	-0.10	ns
<i>Monolingual pupils</i>	17.35 (24.39)	12.55 (16.14)	-0.02 - 0.12	-0.20	ns
<i>Multilingual pupils</i>	9.67 (14.22)	15.18 (16.51)	-14.66 - 3.59	0.38	ns
% inappropriate periods	5.14 (15.64)	1.44 (4.87)	0.60 - 6.78	-0.24	< .05
<i>Monolingual pupils</i>	5.88 (17.04)	0.96 (2.83)	-1.17 - -9.69	-0.29	< .05
<i>Multilingual pupils</i>	2.60 (9.26)	2.28 (6.79)	-4.03 - 4.69	-3.48	ns

Extract (1) Text organization at the beginning and at the end of the school year, pupil 3OL60

3OL60, October

First Charlot shows a ticket to a man after that the man welcomes him Charlot then Charlot walks around with a man after that Charlot takes a lady by the hand and takes her into the warehouse after that they put on roller skates then they start skating after that Charlot blindfolds himself and starts to rollerskate then it looks like Charlot is falling down from the upper floor after that Charlot shows a bed to the lady then Charlot moves around a bit on his rollerskates then Charlot closes the girl in the warehouse after that 3 gentlemen look for Charlo who in the meantime is in a totally different place

3OL60, May

At first Harry gets on a steam train then a girl climbs in with a hat and a Pomeranian dog.

The girl dropped the Pomeranian dog and they left without him but Harry took it with a cane and gave it to its owner named Giulia, she was very happy to get back her Pomeranian dog, and told him thank you very much.

At some point they entered a tunnel and everyone started looking for the dog.

Harry found it and gave it to Giulia once again, s/he was really really happy to have it back, and then Giulia covered it with a blanket of hers/his.

While both paraphrasing and punctuation marks are absent in the first text, at the end of the school year the child organizes contents in thematic blocks and experiments with the use of punctuation, even though her command of commas is not fully developed (see the last two instances).

4.4. Reference to entities

At the end of the school year, the texts written in the experimental classes were also clearer than the texts written in the control classes in terms of reference to entities. In fact, the introduction, maintenance, and reintroduction of referents were less ambiguous, that is, they allowed the reader to correctly identify the entity to which they refer more often. This difference was not very substantial as regards introductions and maintenances, but reached statistical significance for reintroductions ($p < 0.05$).

As for multilingual pupils, the picture is slightly more complex. Their reintroductions were generally less ambiguous than those of both monolingual and multilingual students in control classes, with a fairly noticeable, though not statistically significant, difference (6.37% vs. 8.42% and 9.36%, respectively). Their introductions were less ambiguous than those of monolingual pupils in the control group (1.79% vs. 2.39%), but significantly more so than those made by multilingual peers in the control classes, who did not produce a single ambiguous introduction. A similar situation can be seen for maintenances: here, too, multilingual pupils in experimental classes produced fewer ambiguities than monolingual students in the control group (5.66% vs. 5.84%), but slightly more than multilingual peers (5.50%). Especially in this last case, however, differences are quite small and never reach statistical significance (Table 5). Some examples of ambiguous introductions, maintenances, and reintroductions taken from both experimental and control classes are shown in Extract (2), (3), and (4).

Table 5
Reference to entities

	Control Group Mean (SD)	Experimental Group Mean (SD)	95% CI		
			of the difference between mean	Cohen's <i>d</i>	<i>p</i>
% Ambiguous introductions	1.83 (6.49)	1.27 (3.47)	-0.86 - 1.99	-0.09	ns
Monolingual pupils	2.39 (7.35)	0.98 (3.47)	-0.36 - 3.19	-0.19	ns
Multilingual pupils	0.00	1.79 (4.69)	0.23 - 3.33	n/a	<0.05
% Ambiguous maintenances	5.76 (9.22)	5.34 (5.70)	-1.70 - 2.47	-0.05	ns
Monolingual pupils	5.84 (9.33)	5.16 (5.64)	-1.80 - 3.03	-0.07	ns
Multilingual pupils	5.50 (.04)	5.66 (6.80)	-4.37 - 4.10	0.02	ns
% Ambiguous reintroductions	8.64 (9.70)	5.68 (6.85)	-0.71 - -5.28	-0.31	< .05
Monolingual pupils	8.42 (9.31)	5.29 (6.89)	-0.56 - -5.75	-0.34	< .05
Multilingual pupils	9.36 (11.05)	6.37 (8.04)	-2.01 - -8.05	-0.27	ns

Extract (2). *Example of ambiguous references to entities*

Ambiguous introduction **4VM19_October**
 [...] when he came down with the elevator the men [new entity] hide from Charlot.

Extract (3). *Example of ambiguous references to entities*

Ambiguous maintenance **3EP05_May**
 [...] then the ledy looked at the clock and tooked the bag where the dog was then \emptyset [= she/he/it?] begun to bark and the ledy opened the bag and faund the dog but not her handkerchiefs [...]

Extract (4). *Example of ambiguous references to entities*

Ambiguous reintroduction **5ARR03_May**
 [...] Harry saw the ticket controller arrive so he took his suitcase, which was the woman's he began to take out all the clothes, then he put the dog in and put it back when the ticket controller passed \emptyset [= he/it?] started barking so the ticket controller became suspicious, Harry, began to bark to create confusion. [...]

4.5. Verb tense cohesion

The one aspect in which experimental classes performed more poorly than the controls is verb tense cohesion, where unjustified tense changes, here labelled “tense shifts,” were more common: 2.43/100 words vs. 1.58/100 words. Extract (5) reports an example from the experimental group.

Extract (5). Tense shifts in a fourth grader, experimental group

4GC43_May

Harry a man has to take a train and he saw a lady and lets her pass first. The lady start talking to the ticket controller and Harry came in quickly. The lady's little dog has jumped off the train and she wanted to get it back. Harry picked up a can and catches the little dog and hid it in his jacket. [...]

It is worth noting that the text this extract comes from is of average length (248 words), is quite rich in terms of primary and secondary information units (5 and 29, respectively), and presents a good quantity of cohesive devices (28 connectives). This means that verb tense cohesion may still pose problems, even when overall text quality is satisfactory.

The difference seen between experimental and control classes as a whole remains basically the same when multilingual and monolingual pupils are analyzed separately: multilingual students in experimental classes produce a higher proportion of tense shifts (3.91/100 words) than multilingual (2.63) and monolingual (1.26) peers in the control group; monolingual pupils in experimental classes, with their 1.59 shifts per 100 words, perform better than multilingual and slightly worse than monolingual controls. It should be noted that, in both experimental and control classes, this phenomenon is more apparent in multilingual than in monolingual pupils. Tense shifts are, in fact, more closely related to linguistic competence in a narrow sense, that is, to grammatical control rather than to the more communicative aspects of writing ability. This finding resonates with a number of studies which have reported that LX writers have less control over verb forms compared to L1 peers, both during compulsory education (Cameron & Besser, 2004) and at the university (Hinkel, 2004).

5. Discussion and conclusion

As discussed above, this program was proposed to the whole class, without dividing monolingual and multilingual pupils. The underlying principle is that everybody benefits from good quality language education and that promoting peer-to-peer interaction through group work gives better results than having pupils follow separate tracks.

Although there was no separation of learners in terms of teaching, it is legitimate to discuss the program's effects on the two sub-groups of monolingual and multilingual students. Not many results at this level are statistically significant, both because the sub-samples were not very large and because in several cases differences were not very pronounced. Thus, this brief discussion will not aim at reaching broad generalizations, rather it will attempt to stimulate reflection on the relationships between multilingualism and language education in today's Italian classrooms, which are possibly applicable to other contexts as well. The study's main findings, based on the research questions formulated in section 3.1, can be summarized as follows.

RQ1. In a primary school context, what are the effects on whole classes participating in a long-term pedagogical intervention based on the process approach to writing, formative assessment, and cooperative learning? The effects on whole classes are largely positive. Their pupils outperform those of control classes on all dimensions assessed with both analytic and holistic measures, the sole exception being verb tense shifts. This result is thus in line with previous research (reviewed in section 2.2) showing the effectiveness of instructional practices like teaching students strategies for planning, drafting, and revising their texts, promoting peer collaboration, and providing constructive and specific feedback.

RQ2. If there are any effects, do they differ for monolingual and multilingual pupils? In experimental classes, monolingual children achieve better results than their multilingual peers with regard to all aspects of writing with one exception, the word per paragraph ratio. It should be noted, though, that experimental classes hosted a higher number of multilingual students who were not born in Italy and who had resided in the country for a shorter period of time; hence, their relatively poorer performance might be due to language learning per se, as many of these children were true additional language learners, whose Italian was still developing.

Despite this initial disadvantage, the program's effects were equally beneficial for both monolingual and multilingual students. In particular, multilingual children in the experimental group outperformed both monolingual and multilingual students in control classes on the vast majority of dimensions, with just a few exceptions. One of these is tense shifts, which were consistently more frequent in experimental classes (with the sole exception of monolingual pupils in experimental classes performing better than multilingual controls); this shows that tense shifts may be more directly related to language proficiency than to other dimensions. Another aspect that may be related to language proficiency is text comprehensibility, which is lower in experimental multilingual pupils when compared to both monolingual and multilingual control students. Finally, multilingual control children outperform experimental (both monolingual and multilingual) peers as regards appropriate use of the comma and ambiguous introductions.

RQ3. What are the differences, as regards writing skills, between monolingual and multilingual pupils in control classes, that is, with no experimental treatment? In control classes, multilingual and monolingual children's performances seem to be rather balanced. Monolingual pupils achieve consistently better scores in holistic assessment and, as regards analytic measures, tense shifts, their use of commas, and they have a lower rate of ambiguous reintroductions. Multilingual students, on the other hand, write longer texts with more secondary information units, with fewer problems in the use of periods and commas, and with fewer ambiguous introductions and maintenances.

The tentative conclusions that can be drawn from the study are that, at least in this sample, participation in the experimental program seems to have had a larger impact on writing performance than monolingual/multilingual status. Multilingual children in experimental classes contribute as much as their monolingual peers to the good results obtained by the class as a whole, and outperform both monolingual and multilingual pupils in control classes in most dimensions, especially those related to broader "cognitive-academic" (Cummins, 1986) aspects. Linguistic proficiency, however, does play a role, especially for those who have had a shorter immersion in the new language, as was the case for several multilingual children in experimental classes.

In the context investigated here, our general conclusion is that students' writing skills are impacted by both intrinsic factors, such as the length of exposure to the target language, and extrinsic factors, such as the type of pedagogical intervention. This study can be taken as evidence that a good language curriculum has more weight in a student's academic success than his or her status as an L1 or LX learner.

Acknowledgments

We wish to acknowledge the contribution of Fabiana Rosi, who was involved in the design of the educational intervention and in the data collection, as well as the contributions of Giulia Borrelli, Elisa Caretta, and Elena Russo, who carried out some of the data coding and analysis.

References

- Andrade, Maureen Snow, & Evans, Norman W. (2013). *Principles and practices for response in second language writing*. New York, New York: Routledge.
- Boscolo, Pietro, Arfé, Barbara, & Quarisa, Mara (2007). Improving the quality of students' academic writing: an intervention study. *Studies in Higher Education*, 32(4), 419-438. <http://dx.doi.org/10.1080/03075070701476092>.
- Brakel Olson, Vicki L. (1990). The revising processes of sixth-grade writers with and without peer feedback. *The Journal of Educational Research*, 84(1), 22-29.
- Cameron, Lynne, & Besser, Sharon (2004). Writing in English as an additional language at key stage 2. *Research report RR586*. Leeds, United Kingdom: Department for Education and Skills, University of Leeds.
- Cook, Vivian J. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly*, 33(2), 185-209. <http://dx.doi.org/10.2307/3587717>.

- Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge, United Kingdom: Cambridge University Press.
- Cumming, Alister (2001). Learning to write in a second language: Two decades of research. *International Journal of English Studies*, 1(2), 1–23. <http://dx.doi.org/10.6018/ijes.1.2.48331>.
- Cumming, Alister, Rebuffot, Jacques, & Ledwell, Monica (1989). Reading and summarizing challenging texts in first and second languages. *Reading and Writing*, 2, 201–219. <http://dx.doi.org/10.1007/BF00377643>
- Cummins, Jim (1986). Empowering minority students: A framework for intervention. *Harvard Educational Review*, 56(1), 18–37.
- Dewaele, Jean-Marc (2017). Why the dichotomy ‘L1 versus L2 user’ is better than ‘native versus non-native speaker’. *Applied Linguistics*, 39(2), 236–240. <http://dx.doi.org/10.1093/applin/amw055>.
- De Winter, Joost C. F., & Dodou, Dimitra (2010). Five-point Likert items: t test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research & Evaluation*, 15(11), 1–12.
- Ede, Lisa S., & Andrea C. Lunsford (1990). *Singular texts/plural authors: perspectives on collaborative writing*. Carbondale and Edwardsville, Illinois: Southern Illinois University Press.
- Englert, Carol Sue, Mariage, Troy V., & Dunsmore, Kailonnie (2006). Tenets of sociocultural theory in writing instruction research. In Charles A. MacArthur, Steve Graham, and Jill Fitzgerald (Eds.), *Handbook of writing research* (pp. 208–221). New York, New York: Guilford Press.
- Fagerland, Morten W., & Sandvik, Leiv (2009). Performance of five two-sample location tests for skewed distributions with unequal variances. *Contemporary Clinical Trials*, 30(5), 490–496. <http://dx.doi.org/10.1016/j.cct.2009.06.007>.
- Ferretti, Ralph P., Lewis, William E., & Andrews-Weckerly, Scott (2009). Do goals affect the structure of students’ argumentative writing strategies? *Journal of Educational Psychology*, 101, 577–589. <http://dx.doi.org/10.1037/a0014702>
- Fitzgerald, Jill, & Teasley, Alan B. (1986). Effects of instruction in narrative structure on children's writing. *Journal of Educational Psychology*, 78(6), 424–432. <http://dx.doi.org/10.1037/0022-0663.78.6.424>.
- Giuliano, Patrizia (2007). *Abilità narrativa ed emarginazione sociale*. Napoli, Italia: Liguori.
- Graham, Steve, Harris, Karen R., & Chambers, Amber B. (2016). Evidence-based practice and writing instruction: A review of reviews. In Charles A. MacArthur, Steve Graham, and Jill Fitzgerald (Eds.), *Handbook of writing research* (pp. 211–226). New York, New York: Guilford Press.
- Graham, Steve, Harris, Karen R., & Santangelo, Tanya (2015). Research-based writing practices and the common core: Meta-analysis and meta-synthesis. *The Elementary School Journal*, 115(4), 498–522. <http://dx.doi.org/10.1086/681964>.
- Graham, Steve, McKeown, Debra, Kihara, Sharlene, & Harris, Karen R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, 104(4), 879–896. <http://dx.doi.org/10.1037/a0029185>.
- Hall, Chris (1990). Managing the complexity of revising across languages. *TESOL Quarterly*, 24(1), 43–60. <http://dx.doi.org/10.2307/3586851>.
- Hendriks, Henriëtte (1998). Reference to person and space in narrative discourse: A comparison of adult second language and child first language acquisition. *Studi Italiani di Linguistica Teorica e Applicata*, XXVII(1), 67–86.
- Hickmann, Maya (1995). Discourse organisation and the development of reference to person, space, and time. In Paul Fletcher & Brian MacWhinney (Eds.), *The handbook of child language* (pp. 194–218). Oxford, United Kingdom: Wiley-Blackwell.
- Hinkel, Eli (2001). Matters of cohesion in L2 academic texts. *Applied Language Learning*, 12(2), 111–132.
- Hinkel, Eli (2004). Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research*, 8(1), 5–29. <http://dx.doi.org/10.1191/1362168804lr132oa>.
- Hyland, Ken (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16, 148–164. <http://dx.doi.org/10.1016/j.jslw.2007.07.005>.
- Jones, Stan, & Tetroe, Jacqueline (1987). Composing in a second language. In Ann Matsuhashi (Ed.), *Writing in real time: Modelling production processes* (pp. 34–57). Norwood, New Jersey: Ablex.

- Kersten, Kristin (2009). *Verbal inflections in L2 child narratives*. Trier, Germany: Wissenschaftlicher Verlag Trier.
- Koster, Monica, Tribushinina, Elena, de Jong, Peter F., & van den Bergh, Huub (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research*, 7(2), 299-324.
- Kuiken, Folkert, & Vedder, Ineke (2017). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing*, 34(3), 321-336. <http://dx.doi.org/10.1177/0265532216663991>
- Leki, Ilona, Cumming, Alister, & Silva, Tony (2006). Second-language composition teaching and learning. In Peter Smagorinsky (Ed.), *Research on composition: Multiple perspectives on two decades of change* (pp. 141-169). New York, New York and London: Teachers College Press.
- Montrul, Silvina (2016). *The acquisition of heritage languages*. Cambridge, United Kingdom: Cambridge University Press.
- Oi, Kyoko Mizuno (1984). Cross-cultural differences in rhetorical patterning: A study of Japanese and English. *Dissertation Abstracts International*, 45(8), 2511A.
- Olson, Carol Booth, Scarcella, Robin, & Matuchniak, Tina (2013). Best practices in teaching writing to English learners. In Steve Graham, Charles A. MacArthur, and Jill Fitzgerald, (Eds.), *Best practices in writing instruction* (pp. 381-402). New York, New York: Guilford Press.
- Pallotti, Gabriele (2010). Doing interlanguage analysis in school contexts. In Inge Bartning, Maisa Martin & Ineke Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 159-190). Amsterdam: Eurosla Monographs.
- Pallotti, Gabriele (2017a). Osservare l'interlingua: percorsi di educazione linguistica efficace per ridurre le diseguglianze. In Massimo Vedovelli (Ed.), *L'italiano dei nuovi italiani. Atti del XIX Convegno Nazionale GISCEL* (pp. 505-520). Roma, Italia: Aracne.
- Pallotti, Gabriele (2017b). Applying the interlanguage approach to language teaching. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 55(4), 393-412. <http://dx.doi.org/10.1515/iral-2017-0145>.
- Pallotti, Gabriele, & Rosi, Fabiana (2017). Educazione linguistica inclusiva nella scuola primaria: percorsi di sperimentazione e analisi dell'efficacia. In Loredana Corrà (Ed.), *Educazione linguistica in classi multietniche* (pp. 117-142). Roma, Italia: Aracne.
- Pallotti, Gabriele, Borghetti, Claudia, & Rosi, Fabiana (in preparation). *Insegnare a scrivere nella scuola primaria. Percorsi educativi e indagini sperimentali*.
- Reid, Joy M. (1988). Quantitative differences in English prose written by Arabic, Chinese, Spanish, and English students. *Dissertation Abstracts International*, 50(3), 672A.
- Rietdijk, Saskia, Janssen, Tanja, van Weijen, Daphne, van den Bergh, Huub, & Rijlaarsdam, Gert (2017). Improving writing in primary schools through a comprehensive writing program. *The Journal of Writing Research*, 9(2), 173-225. <http://dx.doi.org/10.17239/jowr-2017.09.02.04>
- Schunk, Dale H., & Swartz, Carl W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, 18, 337-354. <http://dx.doi.org/10.1006/ceps.1993.1024>.
- Selinker, Larry (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(2), 209-231.
- Silva, Tony (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27(4), 657-677. <http://dx.doi.org/10.2307/3587400>.
- Solarino, Rosaria (2009). *Imparare dagli errori*. Napoli, Italia: Tecnodid.
- Stevenson, Marie, Schoonen, Rob, & de Glopper, Kees (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL, *Journal of Second Language Writing*, 15, 201-233, <http://dx.doi.org/10.1016/j.jslw.2006.06.002>.
- Storch, Neomy (2013). *Collaborative writing in L2 classrooms*. Bristol, United Kingdom: Multilingual Matters.
- Tracy, Brenda, Reid, Robert, & Graham, Steve (2009). Teaching young students strategies for planning and drafting stories: The impact of self-regulated strategy development. *Journal of Educational Research*, 102, 323-331. <http://dx.doi.org/10.3200/JOER.102.5.323-332>.

- Uzawa, Kozue (1996). Second language learners' processes of L1 writing, L2 writing, and translation from L1 into L2. *Journal of Second Language Writing*, 5(3), 271–294. [http://dx.doi.org10.1016/S1060-3743\(96\)90005-3](http://dx.doi.org10.1016/S1060-3743(96)90005-3).
- van Waes, Luuk, & Leijten, Mariëlle (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition*, 38(A), 79–95. <http://dx.doi.org10.1016/j.compcom.2015.09.012>.
- van Weijen, Daphne (2009). *Writing processes, text quality, and task effects: Empirical studies in first and second language writing*. Utrecht, The Netherlands: LOT.
- White, Ron, & Arndt, Valerie (1991). *Process Writing*. Harlow, United Kingdom: Longman.

Appendix A

RATING SCALES FOR HOLISTIC ASSESSMENT

Content: Is the number of information units provided in the text adequate and relevant?

- 1) The number of ideas is *not at all adequate* and insufficient and the ideas are unrelated to each other.
- 2) The number of ideas is *scarcely adequate* and the ideas lack consistency.
- 3) The number of ideas is *somewhat adequate*, even though they are not very consistent.
- 4) The number of ideas is *adequate* and they are sufficiently consistent.
- 5) The number of ideas is *very adequate* and they are very consistent to each other.
- 6) The number of ideas is *absolutely adequate* and they look very consistent to each other.

Comprehensibility: How much effort is required to understand text purpose and ideas?

- 1) The text is *not at all comprehensible*. Ideas and purposes are unclearly stated and the efforts of the reader to understand it are ineffective.
- 2) The text is *scarcely comprehensible*, its purposes are not clearly stated and the reader struggles to understand the ideas of the writer. The reader has to guess most of the ideas and purposes.
- 3) The text is *somewhat comprehensible*, some sentences are hard to understand on a first reading. A second reading helps clarify text purposes and the ideas conveyed, but some doubts persist.
- 4) The text is *comprehensible*. Only a few sentences are unclear but are understood, without too much effort, after a second reading.
- 5) The text is *easily comprehensible* and reads smoothly. Comprehensibility is not an issue.
- 6) The text is *very easily comprehensible* and highly readable. The ideas and the purpose are clearly stated.

Coherence and cohesion, 1: Is the text coherent and cohesive (e.g. cohesive devices, strategies)?

- 1) The text is *not at all coherent*. Unrelated progressions and coherence breaks are very common. The writer does not use any anaphoric device (pronouns, sentences with a clearly interpretable null subject). The text is not at all cohesive, connectives are hardly ever used, and ideas are unrelated.
- 2) The text is *scarcely coherent*. The writer often uses unrelated progressions; when coherence is achieved, it is often done through repetitions. Few anaphoric devices and some coherence breaks. The text is not very cohesive: ideas are not well linked by connectives, which are rarely used.
- 3) The text is *somewhat coherent*. Unrelated progressions and/or repetitions are frequent. More than two sentences in a row can have the same subject, even when the subject is understood. Some anaphoric devices are used. There can be few coherence breaks. The text is somewhat cohesive: some connectives are used, but they are mostly conjunctions.
- 4) The text is *coherent*. Unrelated progressions are somewhat rare, but the writer sometimes relies on unnecessary repetitions to achieve coherence. A sufficient number of anaphoric devices is used. There may be some coherence breaks. The text is cohesive. The writer makes good use of connectives, sometimes not limiting this to conjunctions.
- 5) The text is *very coherent*: when the writer introduces a new topic, it is usually done by using connectives or connective phrases. Repetitions are very infrequent. Anaphoric devices are numerous. There are no coherence breaks. The text is very cohesive and ideas are well linked by adverbial and/or verbal connectives.
- 6) The writer ensures *extreme coherence* by integrating new ideas in the text with connectives or connective phrases. Anaphoric devices are used regularly. There are few incidences of unrelated progressions and no coherence breaks. The structure of the text is extremely cohesive, thanks to a skilful use of connectives (especially linking chunks, verbal constructions, and adverbials), often used to describe relationships between ideas.

Adapted from Kuiken & Vedder 2017

Coherence and cohesion, 2

- 1) Can link words or groups of words with very basic linear connectors like 'and' or 'then'.
- 2) Can link groups of words with simple connectors like 'and', 'but', and 'because'.
- 3) Can use the most frequently occurring connectors to link simple sentences in order to tell a story or describe something as a simple list of points.
- 4) Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
- 5) Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some 'jumpiness' in a long contribution.
- 6) Can use a variety of linking words efficiently to clearly mark the relationships between ideas.
- 7) Can produce clear, smoothly flowing, well-structured speech, showing controlled use of organizational patterns, connectors, and cohesive devices.
- 8) Can create coherent and cohesive text making full and appropriate use of a variety of organizational patterns and a wide range of cohesive devices.

(Council of Europe, 2001, par. 5.2.3.1)

Gabriele Pallotti, Università di Modena e Reggio Emilia

gabriele.pallotti@unimore.it

- EN** | **Gabriele Pallotti** is a professor of language-teaching methodology at the University of Modena and Reggio Emilia. He is a member of the Executive Committee of the European Second Language Association (EuroSLA) and the editor of the EuroSLA Studies Series. His research focuses on L2 interaction and socialization, interlanguage analysis, linguistic complexity, methodology, and epistemology in applied linguistics
- ES** | **Gabriele Pallotti** es profesor de metodología de enseñanza de idiomas en la Universidad de Módena y Reggio Emilia. Es miembro del Comité Ejecutivo de la Asociación Europea de Segundo Idioma (EuroSLA, por sus siglas en inglés) y editor de la serie *EuroSLA Studies*. Su investigación se centra en la interacción y la socialización en L2, el análisis entre idiomas, la complejidad lingüística, la metodología y la epistemología en lingüística aplicada.
- IT** | **Gabriele Pallotti** è professore di didattica delle lingue moderne presso l'Università di Modena e Reggio Emilia. È membro del Comitato Esecutivo della European Second Language Association (EuroSLA) e direttore della serie EuroSLA Studies. La sua ricerca si concentra sull'interazione e socializzazione in L2, analisi dell'interlingua, complessità linguistica, metodologia ed epistemologia nella linguistica applicata.

Claudia Borghetti, Università di Bologna

claudia.borghetti@unibo.it

- EN** | **Claudia Borghetti** is a research fellow in language learning and teaching at the il Dipartimento di Lingue, Letterature e Culture Moderne, Università di Bologna, Italy. Her primary research interests are intercultural language education, student mobility and multilingualism, teaching Italian as a second language, and (academic) writing.
- ES** | **Claudia Borghetti** es investigadora de aprendizaje y enseñanza de idiomas en el il Dipartimento di Lingue, Letterature e Culture Moderne de la Università di Bologna, Italia. Sus principales intereses de investigación son la educación intercultural de idiomas, la movilidad estudiantil y el multilingüismo, la enseñanza de italiano como segunda lengua y la escritura académica.
- IT** | **Claudia Borghetti** è ricercatrice di apprendimento ed insegnamento delle lingue moderne presso il Dipartimento di Lingue, Letterature e Culture Moderne dell'Università di Bologna, in Italia. I suoi principali interessi di ricerca sono l'insegnamento interculturale delle lingue, la mobilità degli studenti e il multilinguismo, l'insegnamento dell'italiano come lingua seconda e la scrittura (accademica).