



# Characterization of reading errors in languages with different orthographic regularity: an Italian–English comparison

Chiara Valeria Marinelli · Cristina Romani · Victoria A. McGowan ·  
Simona Giustizieri · Pierluigi Zoccolotti

Received: 14 February 2023 / Revised: 14 February 2023 / Accepted: 20 February 2023  
© The Author(s) 2023

**Abstract** The study examined whether a classification of errors based on Hendriks and Kolk's (1997) proposal would effectively characterize the reading profile of children learning two orthographies varying for regularity, such as Italian and English. The study considered both an age-match and a grade-match comparison. Offline analysis of error production was carried out for two lists of stimuli: List 1 including regular words varying for frequency and matched non-words and List 2 including low-frequency words varying for regularity. In List 1, Italian-reading children made more multiple attempts characterized by a slow and progressive approach to the target (sounding-out behavior) than English-reading children, while the latter made relatively more word substitutions and non-word

lexicalizations. As for List 2, Italian-reading children made relatively more multiple attempts and progressive approaches to the target compared to the English-reading children (with more sounding-out behaviors and syllabifications), while the opposite occurred for phonological-visual errors, word substitutions, morphological, and semantic errors. Both groups showed a high proportion of phonological-visual and regularization errors (stress assignment in the case of Italian-reading children). Overall, the use of an error coding system specifically tuned to the characteristics of the orthographies investigated allowed a more comprehensive identification of reading difficulties which allowed the different strategies used by children of different languages to emerge more clearly (more reliance on sub-lexical routines in Italian readers and on lexical routines in English readers). These results call for more attention to error patterns in the identification of reading difficulties in children of different languages including those learning a transparent orthography where error analyses have largely been ignored.

---

C. V. Marinelli (✉)  
Learning Sciences Hub, Department of Humanities,  
University of Foggia, Foggia, Italy  
e-mail: chiaravaleria.marinelli@unifg.it

C. Romani  
Aston University, Birmingham, UK

V. A. McGowan  
University of Leicester, Leicester, UK

S. Giustizieri  
University of Salento, Lecce, Italy

P. Zoccolotti  
Department of Psychology, Sapienza University of Rome,  
Rome, Italy

**Keywords** Reading · Classification of errors · Orthography

## Introduction

It is well established that orthographic regularity has a profound effect on reading performance (Seymour

et al., 2003). Up until the '90s, most research was focused on highly irregular languages, such as English (Share, 2008). More recently, there has been a growing interest in examining reading development and the characteristics of reading deficits in languages with regular orthographies, such as Austrian/German, Dutch, Finnish, Greek, or Italian.

Wimmer (1993) made the point that children with problems in learning German, a highly regular orthography, were slow but did not make a sizeable number of errors. This led several authors to compare reading in different orthographies using time measures such as reaction times (RTs) to examine reading acquisition (Welsh–English comparison; Ellis et al., 2001; Albanian–Greek–Hiragana–Kanji–English comparison; Ellis et al., 2004; Dutch–English comparison; Patel et al., 2004; Italian–English comparison; Marinelli et al., 2014, 2016, Mauti et al., 2023) and/or to compare children with and without a reading deficit (German–English comparison; Ziegler et al., 2003). However, even though reading accuracy may be generally high in languages with regular orthographies, type of errors may also differ and require ad hoc classifications.

An original formalization of reading errors in a regular orthography (Dutch) was proposed by Bakker (1992), who distinguished between “time consuming” and “word substitution” errors. The former are characterized by slow and labored approaches to the target stimulus, while the latter represent a general category that includes most of the types of errors present in classical cognitive classifications. It should be noted, however, that Bakker’s analysis was mainly oriented toward studying dyslexia in relation to neural mechanisms (such as hemispheric lateralization) and developing a classification into P-types (perceptual: displaying accurate but slow and laborious reading) and L-types (linguistic: displaying fast reading but with many errors), and so was not focused on a characterization of the linguistic impairments.

Working on Dutch, Hendriks and Kolk (1997) developed Bakker’s (1992) original insight and proposed a more analytical and explicit categorization of reading errors. They identified two main difficulties: sounding-out behavior and word-substitution errors. In sounding-out behavior, the child gradually approaches the correct pronunciation of the word. This may occur after several attempts: for example, “complexso” (“complex”) is read

“co...com...complexso“. Sounding-out is referred to as a “behavior” rather than an error since the progressive approach to the target may lead to the correct reading of the word. In word-substitutions, a word phonologically or visually related to the target is produced in error; for example, “soffitto” (“ceiling”) is read “soffio” (“breath”). Importantly, Hendriks and Kolk (1997) posit that the presence of a sounding-out behavior is indicative of the underlying process, i.e., that, by sounding out parts of the word before uttering the whole word, the child demonstrates its reliance on the phonological, sub-lexical routine. Specifically, Hendriks and Kolk state that “*children who predominantly sound out the words during reading would have a lexical route that is relatively less efficient than the nonlexical route, whereas the reverse would hold true for children who read at a normal rate but produce substitutional errors*” (p. 323, Hendriks & Kolk, 1997).

In subsequent work, Trenta et al. (2013) showed that the classification of errors proposed by Hendriks and Kolk (1997) was useful to characterize single word and text reading in children learning Italian, a highly regular orthography. In reading lists of single words, sounding-out behaviors, errors in stress assignment, and the production of form-related nonwords were predictive of the presence of a reading disorder. In reading passages, sounding-out behaviors and form-related errors were the best predictors.

A more systematic and comprehensive coding scheme of errors may allow us to compare types of reading difficulties across different orthographies. Presently, regularization errors are generally not considered in Italian, whereas time consuming errors or sounding-out behavior are not typically examined in English-reading children. In general, the cognitive classification of reading errors in English does not consider multiple attempts, such as circumlocutions, self-corrections, and multiple responses (Coltheart, 1980). This choice was originally motivated by cases of deep dyslexia who typically have a defect in monitoring their own performance and, henceforth, produce very few multiple attempts, but it has been kept in further research. Thus, the classification described by Bakker (1992) and Hendriks and Kolk (1997) has not been applied to English-reading children although there are indications that it may be useful.

Fairbanks (1937), in a study on American college students with and without reading difficulties, carefully examined the students' reading errors together with patterns of eye movements, a procedure quite advanced at the time. Young adults with reading difficulties showed more "hesitations" (i.e., multiple attempts to word pronunciations) than typical readers and these correlated with other types of errors. Thus, Fairbanks concluded that "*inclusion of hesitations as errors raised the correlation between errors and [eye movements] regressions, suggesting that even though the determination of hesitations is a qualitative step, more central errors were included*" (for a detailed analysis of the contribution of Fairbanks see De Luca et al., 2013). In other words, Fairbanks underscored the idea that a progressive approach to pronunciation in young adults with reading problems was a selective difficulty in the decoding of text and not simply to an articulatory output problem.

### Aims of the study

In the present study, we analyzed the reading performance of Italian and English readers at different times during primary school focusing on accuracy. The work was part of a bigger study in which both reading reaction times (RT) and accuracy were examined. An analysis of reading RTs has already been published together with an in-depth description of the overall study (Marinelli et al., 2016). In the study by Marinelli et al. (2016) accuracy was examined only holistically to justify the use of time measures (as only RTs to correct responses were considered). Here, instead, we carried out an in-depth analysis of reading errors using an error classification system broadly derived from that of Hendriks and Kolk (1997). This classification was expanded by adding the categories of 'fragment' (i.e., only the first part of the stimulus is read) and 'syllabication' (in both cases independent of whether they generate a correct or incorrect response), as additional indications of sublexical reading. Moreover, the category 'regularization' was included to allow comparison with inconsistent orthographies where these errors are common. Finally, due to the presence of pseudowords in the stimulus materials, lexicalization errors have been considered.

Error profiles were examined in both languages considering both reading regular stimuli (as in

Marinelli et al., 2016), and low frequency irregular words. We carried out both an age-match and a grade-match comparison. As to the former, we examined Italian and English readers who were either about 7 years-old or 10 years old at the time of testing. Due to the different entry times in school in the two countries, younger children were in third grade in England and in second grade in Italy; older children were in fifth grade in England and in fourth grade in Italy. As to the latter, we compared children who were in fifth grade both in Italy and in England (i.e., with a similar reading experience but different ages).

Overall, the aim of the study was to examine whether a classification of errors based on Hendriks and Kolk's (1997) proposal was able to effectively characterize the accuracy profiles of children learning two orthographies with very different degrees of regularity, such as Italian and English. RT examination highlighted a greater reliance on the lexical procedure among English readers and on the sublexical procedure among Italian-reading children, as indicated by the reliance on small units of analysis (Marinelli et al., 2016). Then, in the present study, a prevalence of sounding out behaviors, syllabifications and stress errors is expected for Italian-reading children (indicative of sub-lexical processing), while a prevalence of lexical errors (i.e., word substitutions, lexicalizations, and morpho-semantic errors) is expected for English-reading children.

### Method

#### Participants

Research participants came from a social and educational environment that afforded suitable literacy opportunities. All children attended public primary schools in the provinces of Naples and Rome in Italy, and in the Birmingham area in the United Kingdom. Parents gave informed consent for their children's research participation. Non-native speakers and children who scored  $\leq 1.5$  standard deviations below the control mean in the Raven's CPM (Progressive Colored Matrices; Raven et al., 1998) were excluded from the analyses.

Age-matched subgroups included a total of 177 Italian-reading children (87 females, 90 males; Mean Raven Z score = 0.24, SE = 0.06) and 81 English-

reading children (43 females, 28 males; Mean Raven Z score = 0.35, SE = 0.08). There were 90 younger children in Italy (43 females, 47 males; mean age = 7.3 years) and 40 in England (17 females, 23 males; mean age = 7.8 years). There were 87 older children in Italy (44 females, 43 males; mean age = 9.6 years) and 41 in England (26 females, 15 males; mean age = 9.9 years). The youngest children were in third grade in England and second grade in Italy; the oldest children were in fifth grade in England and fourth grade in Italy. Matched groups did not differ by gender (all  $\chi^2$  approximately 1) and CPM Raven z score (F about 1). A quantitatively small but significant difference was present for age (for younger children:  $t(128) = 11.52$ ,  $p < 0.0001$ ; for older children:  $t(126) = 5.91$ ,  $p < 0.0001$ ); English-reading children were about four months older than Italian-reading children. Note that in England children are admitted to first grade if by August–September they have reached the expected age, unlike in Italy where this limit is set to April of the same scholastic year. Presumably this temporal difference was the cause of the slightly different age composition of the Italian- and English-reading samples. However, for the sake of presentation in the following we will refer to this as age-matched comparison.

An additional group of 30 Italian-reading fifth graders (17 females, 13 males; Mean Raven Z score = 0.15, SE = 0.15; mean age = 10.6 years) was tested. The performance of these children was tested against that of the group of English-reading fifth graders to provide for a grade-matched comparison. The two fifth-grade groups did not differ by gender ( $\chi^2 < 1$ ) or by Raven performance ( $t < 1$ ), but they did differ by age ( $t(68) = 7.86$ ,  $p < 0.0001$ ); the Italian-reading children were about seven months older.

## Materials

### *List 1: Regular words varying for frequency and matched non-words*

For each language, 120 stimuli evaluated the effects of stimulus type (N = 40 high-frequency words, N = 40 low-frequency words, and N = 40 non-words) and length (4, 5, 6, and 7–9 letters) with 10 stimuli for each length. Only Italian words with regular stress (i.e., on

the penultimate syllable) and English words with regular correspondences (no unusual letter-sound correspondences) were included in this list. High-frequency words had a mean frequency of 106.2 (SD = 94.2) in English (CELEX lexical database, Baayen et al., 1993) and 63.7 (SD = 55.6) in Italian (CoLFIS database, Bertinetto et al., 2005); low-frequency words had a mean frequency of 2.9 (SD = 1.4) in English and 3.2 (SD = 3.2) in Italian. Both Italian and English word frequencies were calculated over 1,000,000 occurrences. The Italian and English lists were matched for ortho-syllabic difficulty (i.e., presence of double consonants, consonant groups, and contextual rules; Barca et al., 2007), point of articulation of the first phoneme and word frequency, but not for the number of syllables, which was systematically higher in the Italian sets. Non-words were derived from the corresponding high-frequency words by changing a few letters.

### *List 2: Words varying for regularity*

For each language, 60 low-frequency words assessed the effect of regularity by contrasting 30 regular and 30 irregular nouns (overall mean frequency = 14.3; SD = 18, for English; and 9.5, SD = 10.6, for Italian with an average length of seven letters, SD = 1). Regular and irregular words were matched for ortho-syllabic difficulty (presence of double consonants, clusters of consonants and contextual rules), articulation point of the first phoneme, number of letters and word frequency (frequency norms for Italian: from the Colfis database, Bertinetto et al., 2005; for English: CELEX database, Baayen et al., 1993).

In the Italian list, regularity was established in terms of frequency and consistency of stress pattern, where consistency considered the proportion of words with the same final orthographic–phonemic sequence sharing stress pattern (e.g., the final sequence -ola is predominantly associated with the antepenultimate stress; Burani et al., 2004, 2014). Regular words were words with a with stress on the penultimate syllable (a pattern which occur about 80% of times in the Italian language) and a consistent neighbourhood. Irregular words were words with stress on the anti-penultimate (a pattern which occurs about 18%—in the Italian language) and an inconsistent stress neighbourhood. For irregular words accurate lexical retrieval is necessary to produce the right stress.

In the English list, regularity was established by considering the frequency of letter-sound correspondences. Words were considered irregular if they included one or more letter-sound correspondences with a frequency of less than 5% in the database of English words by Hanna et al. (1966). Also in this case, lexical retrieval is needed to accurately read irregular words (while regular words might be read correctly according to both lexical and sub-lexical procedure). In this vein, a regularity effect (better performance on regular words) is an index of the reliance on the sub-lexical procedure, while the lexical procedure will ensure similar accuracy between regular and irregular words. Note that data on this list were not included in our original report (Marinelli et al., 2016).

### Procedure

Children were tested individually in a quiet room at their school using a laptop computer. They were seated approximately 60 cm from the computer screen. Stimuli were presented using the E-prime2 software. Each trial began with a fixation point that remained on the screen for 500 ms. Subsequently, a word appeared at the same location. The stimulus remained on the screen until the child responded.

Words and non-words were presented in separate blocks. To make the task less tiring, the words were divided into three separate blocks with a short break between them. Since non-words were derived from words in the same list, they were presented blocked before the presentation of the words to avoid an increase in word substitutions (due to priming a wrong response). Six practice stimuli preceded both the word and nonword reading trials and the list of regular and irregular words. In both the word and nonword blocks, the order of presentation of the trials was randomized for each child.

Children were asked to read the stimulus as quickly and accurately as possible. Vocal RTs were recorded using a voice key (S-R Box). The experimenter manually recorded any error. Responses were recorded to allow for offline rechecking of children's production.

### Error analysis

Errors were classified according to the following mutually exclusive categories:

1. Sounding-out behavior with correct response: the child progressively approaches the correct response, with multiple attempts to decode the orthographic string and final success. This type of "error" might include both hesitations (mestiere "JOB" → me...mestiere; limousine → limo...limousine) or self-corrections (mestiere "JOB" → mo...mestiere; limousine → isc...limousine).
2. Sounding-out behavior with error: the child progressively approaches the word with multiple attempts to decode the orthographic string, but the final production is incorrect due to the presence of one or more errors (mestiere "JOB" → mes...mestiare; limousine → limou...limousife). The errors might be all errors reported below at categories 7–13.
3. Syllabication with no error: the child read the stimulus correctly, but the decoding is syllabized (mestiere "JOB" → me-stie-re; limousine → li-mou-si-ne), but no part of the stimulus is re-read.
4. Syllabication with error: the stimulus is both syllabized and it is incorrect for the presence of one or more errors reported below at categories 7–13 (mestiere "JOB" → me-stia-re; limousine → li-mou-si-fe).
5. Fragment: only the first part of the stimulus is read. The fragment may be correct (the first part of the stimulus, e.g., mestiere "JOB" → mest...; limousine → lim...) or incorrect (i.e., with phonological-visual errors, e.g., mestiere "JOB" → mesp...; limousine → lima...).
6. Omission-I do not know response: the child omits the stimulus or says he/she doesn't know what it is.
7. Phonological-visual error. Errors are in this category if they cannot be classified according to categories 8–13. The response produced shares more than half of the letters with the target stimulus, but includes one or more of the following: insertion of letter/letters (mestiere "JOB" → mestierte; limousine → limoustine), deletion of letter/letters (mestiere



“JOB” → mesiere; limousine → limoune), substitution of letter/letters (mestiere “JOB” → mestiele; limousine → limoudine), transposition of letter/letters (mestiere “JOB” → restieme; limousine → limuosine) and/or contextual errors (bicchiere “glass” → bicchiere; juple → jumple; naver → narver).

8. Regularization error: these are phonologically plausible errors arising from the application of sub-lexical routines, and stress errors. Note that phonological plausible errors are only possible in English, due to the absence of irregular words in reading for Italian, while stress errors are only possible in Italian at least with our lists (cadavere “CORPSE” → cadavère; apricot → apricot pronounced with the A as in bat).
9. Regularization and phonological-visual error: the child makes both a regularization and a phonological-visual errors (cadavere “CORPSE” → cadavèra; apricot → pricof pronounced with the A as in bat).
10. Word substitution: the target word is substituted with another word that is not semantically or morphologically related to the target words (mestiere “JOB” → mestolo “LADLE”; limousine → lemon).
11. Morphological error: the target is substituted with a morphologically-related word (mestiere “JOB” → mestierino “SMALL JOB”; mestieri “JOBS”; limousine → limousines).
12. Semantic error: the target is substituted with a semantically-related word (mestiere “JOB” → lavoro “WORK”; limousine → car).
13. Non-word lexicalization: the target is read as a real word sharing several letters with the target stimulus (fibestre → finestre; tinection → infection).

Note that errors in categories 7–12 were only possible in reading words, while errors in category 13 were only possible for non-words.

Finally, we counted the number of errors for each child resulting in a real word or a nonword (from now errors resulting in words: wordER and errors resulting in nonword: nonwordER). Note that here we only considered complete attempts and not sounding-out behaviors (with or without correct response), fragment productions, syllabifications (with or without errors) and omission-I do not know responses.

## Results

### *List 1: Regular words varying for frequency and matched non-words: age match comparisons*

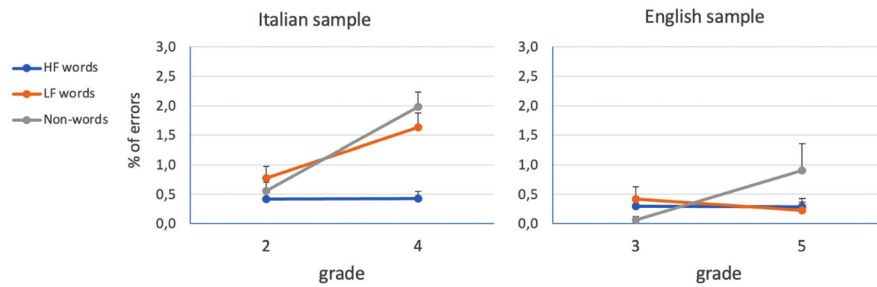
Descriptive statistics (average error rates) for Italian and English readers matched for age (2nd and 4th Italian grades with 3rd and 5th English grades) are presented in “[Appendix A](#)”. Separate ANOVAs were carried out for different types of reading errors. In these analyses, the type of stimulus was the repeated measure, and age (younger, older) and language (Italian, English) the unrepeated measures. Depending on the type of error, the type of stimulus factor might involve comparing high-frequency (HF) words, low-frequency (LF) words, and non-words, or only words (i.e., HF words and LF words) as in the case of word substitution, morphological, and semantical error categories. In one case (i.e., non-word lexicalizations), the ANOVA was confined to non-word stimuli and the type of stimulus factor was not included. These ANOVAs are presented in detail in “[Appendix B](#)”.

A synthesis of these analyses is presented in [Table 1](#). Inspection of the table indicates several major findings:

- Italian-reading children made more sounding-out behaviors (with correct responses; see [Fig. 1](#)), syllabifications (with no error), syllabifications (with errors), and fragments;
- English-reading children made more word substitutions and non-word lexicalizations; the pattern of word substitutions varied as a function of age and language (see [Fig. 2](#));
- There was no language difference in omission-I do not know responses, phonological-visual errors, regularization or stress errors, regularization and phonological-visual errors, morphological and semantic errors;
- In all these types of errors (except for syllabifications with no error), there was a significant type of stimulus main effect, with more errors for non-words than LF frequency words and more errors for LF than HF words;
- Word substitutions and morphological errors were more frequent for LF than HF words for younger children, especially in the English-reading sample;

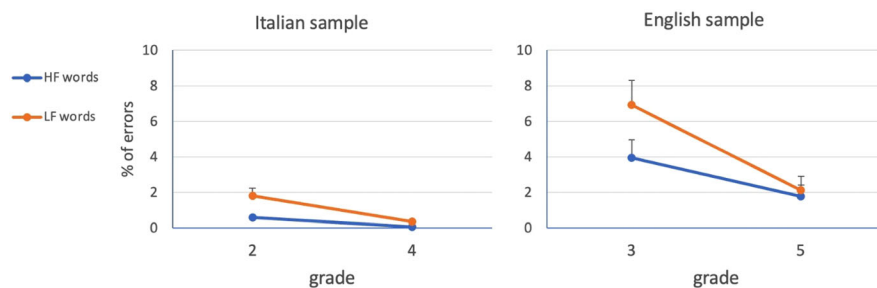
**Table 1** Summary of the ANOVAs results for different types of errors (age comparison). Legend: NW = Pseudowords; LFW = Low frequency words; HFW = High frequency words

Type of error	Language effect	Age	Language × age	Type of stimulus	Type of stimulus × age	Type of stimulus × language	Type of stimulus × age × language
Sounding-out behavior with correct responses	< 0.001 Italian > English	p < 0.001 increase with age	< 0.001	< 0.001 NW > LFW > HFW	< 0.001	< 0.001	n.s.
Sounding-out behavior with errors	n.s.	n.s.	< 0.05	n.s.	n.s.	n.s.	n.s.
Syllabifications (with no error)	< 0.01 Italian > English	< 0.05 decrease with age	< 0.01	n.s.	n.s.	n.s.	n.s.
Syllabifications (with errors)	< 0.05 Italian > English	< 0.001 decrease with age	< 0.05	< 0.001 NW > LFW > HFW	< 0.001	n.s.	n.s.
Fragments	< 0.001 Italian > English	n.s.	n.s.	< 0.001 NW > LFW > HFW	n.s.	< 0.001	n.s.
Omission—I do not know responses	n.s.	n.s.	n.s.	< 0.01 NW > LFW > HFW	n.s.	n.s.	< 0.05
Phonological-visual errors	n.s.	< 0.001 decrease with age	n.s.	< 0.001 NW > LFW > HFW	< 0.001	< 0.01	n.s.
Regularization errors (including stress errors)	n.s.	< 0.001 decrease with age	n.s.	< 0.001 LFW > HFW	< 0.001	n.s.	n.s.
Regularization and phonological-visual errors	n.s.	< 0.01 decrease with age	n.s.	< 0.05 LFW > HFW	< 0.05	n.s.	n.s.
Word substitutions	< 0.001 English > Italian	< 0.001 decrease with age	< 0.05	< 0.001 LFW > HFW	< 0.001	< 0.05	< 0.05
Morphological errors	n.s.	< 0.05 decrease with age	n.s.	n.s.	< 0.05	< 0.01	n.s.
Semantic errors	n.s.	n.s.	n.s.	n.s.	< 0.05	n.s.	n.s.
Non-word lexicalizations	< 0.001 English > Italian	< 0.001 decrease with age	n.s.	n.s.	< 0.05	n.s.	n.s.
Error resulting in non-word	n.s.	< 0.001 decrease with age	n.s.	< 0.001 NW > LFW > HFW	< 0.001	< 0.05	n.s.
Error resulting in word	< 0.001 English > Italian	< 0.001 decrease with age	n.s.	< 0.001 NW = LFW > HFW	< 0.01	< 0.001	n.s.



**Fig. 1** List 1—sounding-out behavior with correct responses as a function of the type of stimulus (high-frequency words, low-frequency words, and non-words) and grade. Error bars indicate

standard errors. 2nd and 4th Italian grades are approximately matched to 3rd and 5th English grades for chronological age. *HF* high frequency, *LF* low frequency words



**Fig. 2** List 1—word substitutions as a function of the type of stimulus (high-frequency words, and low-frequency words) and grade. Error bars indicate standard errors. 2nd and 4th Italian

grades are approximately matched to 3rd and 5th English grades for chronological age. *HF* high frequency, *LF* low frequency words

- Fragments and sounding-out behaviors with correct responses were more frequent for non-words and LF words than for HF words, especially in the Italian-reading sample;
- There were substantial differences in the proportion of error types; thus, phonological-visual errors were most frequent as expected while morphological and semantic errors were rare (please refer to “Appendix A”). Also, regularization errors were quite infrequent (but note that List 1 only contained regular words);
- Wherever significant, the effect of age indicated a reduction of errors with older age/greater reading experience except for sounding-out behavior (with correct responses) which tended to increase with age (particularly in the Italian readers; see Fig. 1).

A second set of analyses compared English- and Italian-reading children matched for age on types of wordER or nonwordER. These ANOVAs are

presented in detail in “Appendix B”; a synthesis is shown in the lower part of Table 1.

NonwordER were more frequent among younger (10.83%) than older (3.31%) children. The effect of the type of stimulus indicated more nonwordERs for non-words (13.36%) than LF words (5.61%) than HF words (2.28%). The type of stimulus  $\times$  age interaction indicated a decrease with age in nonwordERs particularly for non-words and LF words. The type of stimulus  $\times$  language interaction indicated a greater number of nonwordERs for LF words in English- (7.54%) than in Italian-reading (4.66%) children; the number of nonwordERs for HF words and non-words was very similar in the two groups.

WordERs were more frequent in English- (4.87%) than in Italian-reading (1.31%) children and in younger (3.67%) than older (1.26%) children. The main effect of the type of stimulus indicated more wordERs for non-words (2.88%) and LF words (2.81%) than for HF words (1.72%). The type of



stimulus  $\times$  age interaction ( $F_{(1,259)} = 6.69$   $p < 0.01$ ) indicated a decrease with age in wordERs, particularly for LF words. The type of stimulus  $\times$  language interaction was significant ( $F_{(1,259)} = 28.97$   $p < 0.001$ ) indicating that Italian-reading children made considerably fewer wordERs than English-reading children in general and particularly in the case of non-words. By contrast, English-reading children made many more wordERs, and these were sensitive to the type of stimulus; thus, they made wordER particularly in the case of non-words and LF words (see Fig. 3).

*List 1: Regular words varying for frequency and matched non-words: grade-matched comparisons*

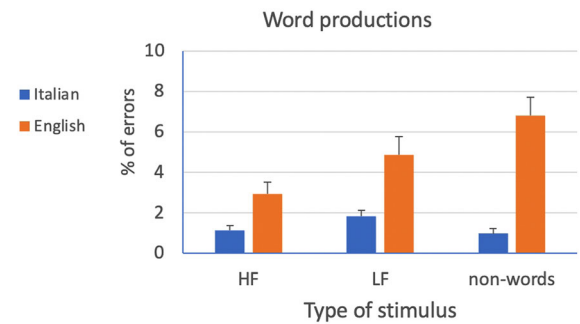
Here, we compared children matched for number of years of schooling (i.e., fifth grade Italian and English readers). A synthesis of these analyses is presented in Table 2.

Inspection of the table indicates a few major findings:

- In general, the absolute number of errors was lower than in the previous analyses owing to the older age of children;
- Italian-reading children made more sounding-out behavior (with correct responses), especially in reading LF words and non-words;
- Fragments were more numerous in the Italian than in the English readers, but only in the case of non-words;
- English-reading children made more non-word lexicalizations; the pattern for word substitutions was in the same direction but only a statistical trend was present;
- For sounding-out behavior (with correct responses) and phonological-visual errors, there was a significant type of stimulus effect, indicating more errors for non-words than LF and for LF than HF words.

Another set of analyses compared English- and Italian-reading children matched for schooling (i.e., fifth grade Italian- and English-reading children) in the production of wordERs and nonwordERs. These ANOVAs are presented in detail in “Appendix B”; a synthesis is presented in the lower part of Table 2.

NonwordERs were more frequent for non-words (7.57%) than LF words (2.09%) and more frequent



**Fig. 3** ERwords as a function of the type of stimulus (high-frequency words, low-frequency words, and non-words) across grades. Error bars indicate standard errors. *HF* high frequency, *LF* low frequency words

for LF than HF words (0.61%). The main effect of language was not significant and did not interact with the type of stimulus. WordERs were more frequent in English- (3.01%) than in Italian-reading (0.43%) children. There were more wordERs for non-words (2.0%) than LF words (1.57%) and for LF than HF words (0.43%). The type of stimulus  $\times$  language interaction indicated that Italian-reading children made few wordERs with all types of stimuli. The English-reading children made mostly wordERs in reading all stimuli, but especially in reading non-words (see Fig. 4).

*List 1: Regular words varying for frequency and matched non-words: comments*

Results for List 1 (including only regular stimuli) indicate that the error coding scheme adopted effectively captured reading problems in both Italian and English-reading children. The Italian readers showed a prevalence of multiple attempts characterized by a slow and progressive approach to the target (which may eventually be correctly read or not). The English-reading children showed a prevalence of errors involving word substitutions or non-word lexicalizations. Again, there was a clear tendency for English reading children to produce more erroneous words than Italian-reading children.

With increasing age, errors, and multiple utterances (including syllabifications) generally decreased except for sounding-out behaviors with correct responses that increased with age, due to increasing self-corrections with older age/more experience. The pattern of

**Table 2** Summary of the ANOVAs results for different types of errors (grade-matched comparison). Legend: NW = Pseudowords; LFW = Low frequency words; HFW = High frequency words

Type of error	Language effect	Type of stimulus	Type of stimulus × language
Sounding-out behavior with correct responses	< 0.05 Italian > English	< 0.001 NW > LFW > HFW	< 0.05
Sounding-out behavior with errors	n.s.	n.s.	n.s.
Syllabifications (with no error)	n.s.	n.s.	n.s.
Syllabifications (with errors)	n.s.	n.s.	n.s.
Fragments	n.s.	n.s.	< 0.05
Omission—I do not know responses	n.s.	n.s.	n.s.
Phonological-visual errors	n.s.	NW > LFW > HFW	n.s.
Regularization errors (including stress errors)	n.s.	n.s.	n.s.
Regularization and phonological-visual errors	n.s.	n.s.	n.s.
Word substitutions	0.07 English > Italian	0.09	n.s.
Non-word lexicalizations	< 0.001 English > Italian		
Type of production	Language effect	Type of stimulus	Type of stimulus × language
Error resulting in non-word	n.s.	< 0.001 NW > LFW > HFW	n.s.
Error resulting in word	< 0.01	< 0.01	< 0.001

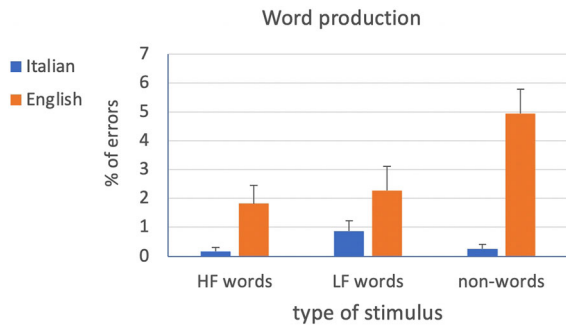
findings was similar in the age-matched and in the grade-matched comparisons although the former set of analyses was more sensitive. Therefore, it appears that the cross-linguistic differences cannot be easily explained in terms of age or reading experience. Further comments will be presented in the Discussion section.

#### *List 2: Words varying for regularity: age-match comparisons*

Descriptive statistics for Italian- and English-reading children matched for age are presented in “[Appendix C](#)”. Here we compared 2nd and 4th Italian readers with 3rd and 5th English readers which are approximately matched for chronological age. Separate ANOVAs were carried out for the different types of reading errors as described for List 1, with the exception that the type of stimulus factor contrasted irregular vs regular words. The results of these ANOVAs are presented in detail in “[Appendix D](#)”.

A synthesis of these analyses is presented in Table 3. Inspection of the table indicates several major findings:

- Italian-reading children showed more sounding-out behaviors with and without correct responses, and made more syllabifications with and without additional errors;
- English-reading children made more omissions-I do not know responses, phonological-visual errors, word substitutions, morphological and semantic errors;
- There was no language difference for fragments;
- As expected, regularization errors were more frequent in List 2 than in List 1 particularly in the case of stress assignment errors (see “[Appendix C](#)”);
- Italian-reading children made more regularization errors than English-reading children, but note that these were stress assignment errors for the Italian readers and phonologically plausible errors for the English readers; crosslinguistic differences were larger for younger than older children;
- Wherever significant, the effect of age indicated a reduction of errors with older age/greater reading experience except for sounding-out behavior (with correct responses) which tended to increase with age.



**Fig. 4** WordERs in the Italian and in the English readers as a function of the type of stimulus (high-frequency words, low-frequency words, and non-words). Data refer to fifth grade children in both English- and Italian-reading samples. Error bars indicate standard errors. *HF* high frequency, *LF* low frequency words

- Word substitutions, morphological and semantic errors were very rare in Italian readers, while they were frequent in English readers and modulated by age (with higher rates in younger children) and by type of stimulus (with higher rates for irregular words).
- In English-reading children, phonological-visual errors decreased with age and were more frequent for irregular than regular words, especially for younger children; for Italian-reading children, these errors decreased with age but without a significant difference between the two types of words.

A second set of analyses compared English- and Italian-reading children matched for age on different types of wordERs or nonwordERs. These ANOVAs are presented in detail in “Appendix D”; a synthesis of these analyses is presented in the lower part of Table 3. NonwordERs were more frequent in younger (10.83%) than older (3.31%) children. Children produced more nonwordERs with irregular (20.28%) than regular words (8.31%). There was no main effect of language. However, nonwordERs reduced with age/experience more in the case of irregular words, particularly among English-reading children (as indicated by the significant type of stimulus  $\times$  language  $\times$  age interaction). WordERs were also more frequent in younger (4.65%) than older (2.13%) children. There were more wordERs in the case of irregular (4.84%) than regular (1.94%) words. English-reading children

made many more wordERs (5.53%) than Italian-reading (1.25%) children. The type of stimulus  $\times$  age  $\times$  language interaction showed that the effect of regularity was present in English-reading children for whom it reduced with age/experience while it was small and not significant in Italian readers (see Fig. 5).

#### *List 2: Words varying for regularity: grade-match comparisons*

Here, we compared children matched for number of years of schooling (i.e., 5th grade Italian- and English-reading children). A synthesis of these analyses is presented in Table 4.

Inspection of the table indicates a few major findings:

- Italian-reading children made more sounding-out behavior (with correct responses);
- English-reading children made more phonological-visual errors, regularization with phonological-visual errors, and word substitutions.
- Semantic errors were absent in Italian-reading children with both regular and irregular words; in English-reading children they were more frequent with irregular than regular words.

Another set of analyses compared English and Italian readers matched for schooling (i.e., 5th grade Italian- and English-reading children) for wordERs and nonwordERs. These ANOVAs are presented in detail in “Appendix D”; a synthesis is shown in the lower part of Table 4. NonwordERs were more frequent for irregular (9.00%) than regular words (2.98%). There was no difference between the two language groups. WordERs were more frequent in English- (3.51%) than in Italian-reading (0.69%) children. There were more wordERs for irregular (3.20%) than regular (1.00%) words. The type of stimulus  $\times$  language interaction was not significant.

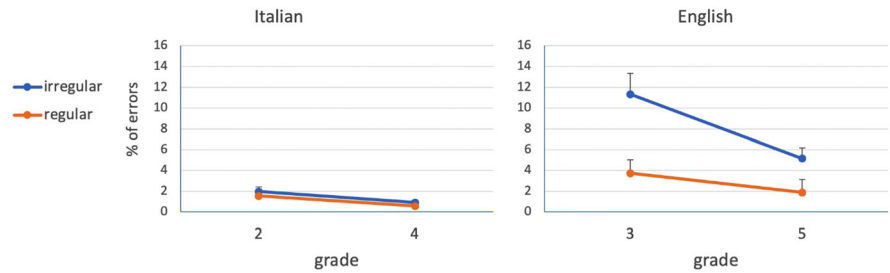
#### *List 2: Words varying for regularity: comments*

List 2 included only low frequency words and examined the effect of regularity. Error rates were generally higher in English- than Italian-reading children, in particular, phonological-visual errors, as well as all errors indexing a reliance on a lexical procedure, such as word substitutions, morphological

**Table 3** Summary of the ANOVAs results for different types of errors (age comparison). Legend: IR = irregular words; R = regular words

Type of error	Language effect	Age	Language × age	Type of stimulus	Type of stimulus × age	Type of stimulus × language	Type of stimulus × age × language
Sounding-out behavior with correct responses	< 0.001 Italian > English	< 0.01 increase with age	< 0.01	n.s.	n.s.	n.s.	n.s.
Sounding-out behavior with errors	< 0.01 Italian > English	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Syllabifications (with no error)	< 0.01 Italian > English	< 0.01 decrease with age	< 0.01	< 0.05 R > IR	n.s.	n.s.	n.s.
Syllabifications (with errors)	< 0.001 Italian > English	< 0.01 decrease with age	< 0.01	n.s.	n.s.	n.s.	n.s.
Fragments	n.s.	n.s.	n.s.	< 0.01 IR > R	n.s.	n.s.	n.s.
Omission—I do not know responses	< 0.05 English > Italian	n.s.	n.s.	< 0.01 IR > R	n.s.	n.s.	n.s.
Phonological-visual errors	< 0.001 English > Italian	< 0.001 decrease with age	< 0.01	< 0.001 IR > R	n.s.	< 0.001	< 0.001
Regularization errors (including stress errors)	< 0.001 Italian > English	< 0.001 decrease with age	n.s.	< 0.001 IR > R	< 0.001	< 0.001	n.s.
Regularization and phonological-visual errors	< 0.01 Italian > English	< 0.001 decrease with age	< 0.001	n.s.	n.s.	n.s.	< 0.01
Word substitutions	< 0.001 English > Italian	< 0.001 decrease with age	< 0.01	< 0.001 IR > R	< 0.001	< 0.05	< 0.05
Morphological errors	< 0.05 English > Italian	< 0.01 decrease with age	n.s.	< 0.001 IR > R	< 0.01	< 0.05	< 0.05
Semantic errors	< 0.05 English > Italian	n.s.	n.s.	< 0.001 IR > R	n.s.	< 0.001	< 0.05
Error resulting in non-word	n.s.	< 0.001 decrease with age	n.s.	< 0.001 IR > R	< 0.001	< 0.01	< 0.01
Error resulting in word	< 0.001 English > Italian	< 0.001 decrease with age	< 0.05	< 0.001 IR > R	< 0.0	< 0.001	< 0.01

**Fig. 5** List 2—WordERs as a function of the type of stimulus (irregular and regular words). 2nd and 4th Italian grades are approximately matched to 3rd and 5th English grades for chronological age. Error bars indicate standard errors



**Table 4** Summary of the ANOVAs results for different types of errors (grade-match comparison). Legend: IR = irregular words; R = regular words

Type of error	Language effect	Type of stimulus	Type of stimulus × language
Sounding-out behavior with correct responses	< 0.001 Italian > English	n.s.	n.s.
Sounding-out behavior with errors	n.s.	n.s.	n.s.
Syllabifications (with no error)	n.s.	n.s.	n.s.
Syllabifications (with errors)	n.s.	n.s.	n.s.
Fragments	n.s.	n.s.	n.s.
Omission – I do not know responses	n.s.	n.s.	n.s.
Phonological-visual errors	< 0.01 English > Italian	< 0.05 IR > R	n.s.
Regularization errors (including stress errors)	n.s.	< 0.001 IR > R	n.s.
Regularization and phonological-visual errors	< 0.001 English > Italian	n.s.	n.s.
Word substitutions	< 0.05 English > Italian	< 0.01 IR > R	n.s.
Semantic errors	n.s.	< 0.05 IR > R	< 0.05
Type of production	Language effect	Type of stimulus	Type of stimulus × language
Error resulting in non-word	n.s.	< 0.001 IR > R	n.s.
Error resulting in word	< 0.05	< 0.01 IR > R	n.s.

and semantic errors. These errors decreased with age and were more frequent with irregular than regular words, especially in younger children. As with List 1, Italian-reading children made more multiple attempts to the target, with a higher number of sounding-out behaviors (with and without correct responses) and syllabifications (with and without error).

NonwordERs reduced with age/experience more for irregular words, particularly for English readers.

WordERs were more frequent in English-reading children (particularly in the case of irregular words) than Italian readers. The regularity effect decreased with age/experience in English-reading children and was minimal in the Italian readers.

Overall, the difficulty of the orthography made English-reading children more error prone. Results also indicated different reading strategies across orthographies: i.e., a reliance on a more holistic

procedure in English readers (possibly based on lexical-semantic retrieval) and an effortful and progressive approach to the target in Italian readers. Thus, the present findings generally confirm the results of List 1, with the presence of only low frequency words in this list making the cross-linguistic differences generally more evident.

The specific interest for this list was to examine regularization errors. As expected, they were more frequent with irregular than regular words. This pattern was evident even in a very regular orthography such as Italian where irregularity only affects stress pattern in multisyllabic words. The number of stress assignment errors made by Italian-reading children was actually higher than the number of regularization errors made by English-reading children, especially when younger samples were compared. List 2 irregular words were selected with the aim to generate as many regularization errors as possible and results are consistent with this aim. However, the intrinsic different nature of the stimuli in the two languages suggests caution in any direct quantitative comparison. Further comments will be presented in the “Discussion” section.

As in the case of List 1, errors and multiple attempts generally decreased with age except for sounding out behaviors with correct responses that increased with age.

In the grade-match comparisons with 5th grade children, crosslinguistic differences were still evident showing a higher rate of sounding-out behaviors (with correct responses) in Italian-reading children, and more word substitutions, semantic errors, phonological-visual errors, and regularizations with phonological-visual errors in English-reading children. As in the age-matched comparisons, cross-linguistic differences in nonwordER were not evident, while wordERs were more frequent in English- than Italian-reading children.

## Discussion

The experimental results indicated that the error coding scheme we adopted effectively captured reading difficulties in both Italian- and English-reading children. The Italian-reading children showed a prevalence of multiple attempts characterized by a slow and progressive approach to the target (which may eventually be correctly read or not). The English

children, instead, showed a prevalence of word substitutions and non-word lexicalizations. This tendency was confirmed when we examined the type of errors produced in terms of resulting in either a real word or a non-word. Again, there was a clear tendency to produce more wordERs (but not nonwordERs) in English- than Italian-reading children.

This general pattern of findings was similar in both lists examined, and in the age- as well as in the grade-matched comparisons. The former set of analyses was more sensitive, possibly because of the range of ages considered. Therefore, it appears that the cross-linguistic differences cannot be easily explained in terms of age or reading experience instead pointing to more fundamental differences in the progression towards reading proficiency in different languages. However, a limitation of this study is the absence of grade comparison for children at the early stage of literacy acquisition. Furthermore, as stated above, the age-matching was imprecise presumably because of the different procedure for school admittance in first grade in England and Italy.

The literature on cross-linguistic comparisons has already indicated that the nature of the orthography plays an important role in shaping the reading profile of children with and without a reading disorder (Seymour et al., 2003), as well as the ability to learn new words (lexical learning; see Marinelli et al., 2020). The present study extends previous finding by showing the importance of error analyses. We have extended previous investigations on Dutch (Bakker, 1992; Hendriks & Kolk, 1997) and Italian readers (Trenta et al., 2013) by comparing the errors made by English and Italian readers using a comparable coding system. Results have shown that error profiles are quite different in different orthographies. However, even in a regular orthography, types of errors are informative of reading difficulties.

English-reading children, like Italian-reading ones, showed behaviors such as sounding-out or syllabifications; however, they did so much more rarely. This is consistent with the proposal that children from regular orthographies (such as German) start to read sublexically and develop the ability to access the lexicon only at a later stage. In contrast, children learning to read irregular orthographies (such as English) would rely on a direct, lexical access route right away (Landerl et al., 1997; Wimmer & Goswami, 1994), thus becoming skilled in lexical reading earlier on



(e.g., Marinelli et al., 2021). The present findings are in keeping with this general proposal and confirm the findings we previously reported on reading RTs and spelling (see Marinelli et al., 2016 and Marinelli et al., 2015, respectively). In agreement with Hendriks and Kolk (1997), our results underscore that the presence of sounding-out behaviors does not merely identify a generic difficulty, but rather specifically points to a reliance on the phonological (sub-lexical) routine. Results of the present study are in keeping with the “central” nature of the sounding-out behavior (as originally noted by Fairbanks, 1937). Thus, sounding-out responses were closely associated to the nature of stimulus, i.e., they were produced most often in reading non-words and rarely when reading high frequency words. Children approach learning to read in regular and irregular orthographies with a differential reliance on lexical and sub-lexical routines, and the use of a coding scheme, such as the one used here, is effective in capturing this difference.

Interestingly, we were able to capture regularization errors in both languages. List 2 was specifically devised to maximize the presence of such errors. In English, this typically occurs when a child reads an irregular word based on grapheme-to-phoneme conversion yielding a so-called “valid” or phonologically plausible error (Temple, 1985). In Italian, the only form of irregularity in reading is given by the assignment of stress in words with three-syllables or more which is not governed by rule. Both the Italian and the English versions of List 2 were effective in producing regularization errors and, in fact, the Italian-reading children made more regularization errors in terms of stress-assignment than the English-reading children made in terms of conversion rules. Given the different nature of the stimuli, however, any direct comparison in rate of regularization across the two languages is not possible. More generally, one should keep in mind that, despite best efforts, matching stimuli in two languages will always have limitations. For example, the syllabic structure of English and Italian words is quite different (Burani et al., 2017; Perfetti & Harris, 2017), and this may influence the likelihood to detect syllabifications. Furthermore, neighborhood size is larger in English than in Italian orthography. This creates a greater opportunity to produce word substitutions.

In summary, the pattern of reading errors observed might be influenced by at least two general orders of

factors. On the one hand, it would depend upon the structural characteristics of a language, so that some errors (such as syllabifications and word substitutions) will be more likely in Italian or English, respectively. On the other hand, the characteristics of a language and its orthography will shape the modality of reading acquisition. Thus, a language, such as English, where a large proportion of words are irregular, and there is a large availability of orthographic neighbors, will foster a strong reliance on a lexical-access procedure even at early stages of reading acquisition. An orthography, such as Italian, where all words are regular (i.e., obey orthographic rules, apart from stress assignment), there is large proportion of long and multi-syllabic words, and orthographic neighbors are relatively few (largely confined to short words), will favor a greater reliance on sub-lexical procedures, at least in the early phase of acquisition. Notably, the relative weight of these two factors cannot be easily disentangled based on empirical data such as the ones described here, and further ad hoc research would be needed to this aim.

It is well known that an effective measure of reading proficiency in regular orthographies is provided by reading speed (Wimmer & Schurz, 2010). However, accuracy measures may also be important provided that a comprehensive error classification scheme is used as shown here (see also Bakker, 1992; Hendriks & Kolk, 1997). With this approach, clear cross-linguistic differences emerge, such that Italian-reading children show a predominance of errors characterized by a slow and progressive approach to the target while English-reading children present a predominance of word substitutions (as well as non-word lexicalizations). From a clinical perspective, error analyses may add important information to the diagnosis of reading difficulties especially in the early phases of reading acquisition.

The present findings may have important clinical implications. While diagnosis in regular orthographies is predominantly based on speed measures, error analyses may add relevant information to the diagnosis of reading difficulties especially in the early phases of reading acquisition. Focusing on accuracy will allow characterizing the reading deficit in children speaking a consistent orthography, provided that the coding scheme considers the production of multiple attempts. Future research should examine whether error analyses can increase our ability to capture

improvements after the administration of rehabilitation programs in children with reading difficulties. Moreover, the analysis of errors profile might also be informative of crosslinguistic differences in the reading profiles shown by dyslexic children in consistent and inconsistent orthographies.

**Acknowledgements** We would like to thank Prof. Daniela Traficante for her help in setting up the statistical analyses, Dr. Katarina Rigova and Dr. Kate Ashley for their contribution in data collection, Dr. Martina Marra for her help in error analysis, and Dr. Cristina Burani for her precious comments to the manuscript and her contribution in creating List 2.

**Funding** Open access funding provided by Università di Foggia within the CRUI-CARE Agreement. The was partially supported by the grant PRIN (Research Project of National Interest) n. 20128YAFKB\_004 (prof. Pierluigi Zoccolotti). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Data availability** Data archiving is not mandated but data will be made available on reasonable request.

**Code availability** Not applicable.

#### Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Ethics approval** This research complies with the dictates of the Declaration of Helsinki and has been approved by the local committee of Departments and school authorities. It has also received approval from the Research Committee of Aston University and the Ethics Committee of IRCCS Fondazione Santa Lucia Rome (Prot. CE-PROG.480).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

#### Appendix A

Means and SDs for all error types (List 1: age comparison). Data indicate average percentage error rates and are separately presented as a function of grade and language. Legend: LFW = Low frequency words; HFW = High frequency words

Types of error	Type of stimulus	Italian						English			
		2nd grade		4th grade		5th grade		3rd grade		5th grade	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Sounding-out behavior with correct responses	HF	0.42	1.14	0.43	1.09	0.43	1.17	0.30	0.82	0.28	0.97
	LF	0.78	1.90	1.64	2.25	1.55	2.26	0.42	1.34	0.23	0.91
	Non-words	0.56	1.39	1.98	2.39	2.16	2.65	0.06	0.39	0.91	2.96
Sounding-out behavior with errors	HF	0.08	0.59	0.00	0.00	0.00	0.00	0.12	0.54	0.06	0.38
	LF	0.28	0.95	0.14	0.59	0.09	0.46	0.06	0.39	0.17	1.13
	Non-words	0.36	0.96	0.11	0.53	0.26	1.02	0.00	0.00	0.17	0.64
Syllabifications (with no error)	HF	0.69	2.43	0.11	0.84	0.09	0.46	0.00	0.00	0.06	0.38
	LF	1.14	2.68	0.06	0.38	0.26	1.02	0.00	0.00	0.06	0.38
	Non-words	1.22	2.32	0.20	0.78	0.34	0.88	0.06	0.39	0.11	0.53
Syllabifications (with errors)	HF	0.86	2.81	0.11	0.84	0.09	0.46	0.00	0.00	0.06	0.38
	LF	1.56	3.41	0.06	0.38	0.34	1.45	0.48	0.99	0.23	0.73
	Non-words	1.94	3.24	0.26	1.01	0.34	0.88	0.95	1.83	0.11	0.53
Fragments	HF	0.14	0.69	0.09	0.46	0.09	0.46	0.18	0.65	0.06	0.38
	LF	0.33	1.01	0.52	1.38	0.17	0.64	0.36	0.89	0.23	0.91
	Non-words	0.97	2.01	0.95	1.88	0.52	1.03	0.30	0.82	0.06	0.38

Types of error	Type of stimulus	Italian						English			
		2nd grade		4th grade		5th grade		3rd grade		5th grade	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Omission—I do not know responses	HF	0.11	0.52	0.11	0.53	0.09	0.46	0.12	0.54	0.28	1.11
	LF	0.39	1.18	0.29	1.23	0.17	0.64	0.65	1.47	0.63	1.95
	Non-words	0.69	1.84	0.17	0.64	0.17	0.64	0.18	0.85	0.57	2.40
Phonological-visual errors	HF	2.00	0.00	4.00	0.00	0.52	1.40	3.00	0.00	5.00	0.00
	LF	6.14	9.35	0.92	2.50	0.95	1.55	3.75	5.90	0.97	2.54
	Non-words	7.58	7.86	1.26	2.55	7.24	8.19	10.42	10.09	3.18	4.89
Regularization errors (including stress errors)	HF	2.00	0.00	4.00	0.00	0.00	0.000	3.00	0.00	5.00	0.00
	LF	0.08	0.00	0.00	0.00	0.00	0.000	0.00	0.00	0.06	0.37
Regularization and phonological-visual errors	HF	2.00	0.00	4.00	0.00	0.00	0.000	3.00	0.00	5.00	0.00
	LF	0.25	0.00	0.00	0.00	0.09	0.464	0.17	0.84	0.11	0.52
Word substitutions	HF	2.00	0.00	4.00	0.00	0.09	0.46	3.00	0.00	5.00	0.00
	LF	0.58	0.00	0.06	0.38	0.60	1.28	3.95	6.60	1.78	4.15
Morphological errors	HF	2.00	0.00	4.00	0.00	0.00	0.00	3.00	0.00	5.00	0.00
	LF	0.08	0.45	0.09	0.46	0.00	0.00	0.00	0.00	0.00	0.00
Semantic errors	HF	2.00	0.00	4.00	0.00	0.00	0.00	3.00	0.00	5.00	0.00
	LF	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Non-word lexicalizations	Non-words	1.67	4.24	0.32	1.07	0.26	0.77	8.55	10.06	4.83	5.63
Error resulting in non-word	HF	3.94	5.79	0.66	1.50	0.43	1.17	3.63	5.55	0.80	2.30
	LF	7.92	9.18	1.41	2.63	0.78	1.35	11.67	10.82	3.41	4.66
	Non-words	18.94	14.34	6.75	6.27	7.24	8.19	20.77	14.09	7.90	12.85
Error resulting in word	HF	1.97	4.41	0.29	0.80	0.17	0.64	4.05	6.65	1.82	1.13
	LF	3.14	5.37	0.52	1.15	0.86	1.92	7.44	9.92	2.27	1.83
	Non-words	1.64	4.24	0.32	1.07	0.26	0.77	8.69	10.14	4.94	0.98

**Appendix B**

*B.1. Age comparison (List 1): ANOVAS on different types of reading errors*

Significant effects are marked in bold.

ES: effect size is expressed as partial eta squared. Large ESs (> 0.14) for main effects and interactions are marked in bold.

Factor	df	F	p	ES
<b>Sounding-out behavior with correct responses</b>				
Type of stimulus	2,518	8.56	<b>0.000</b>	0.032
Type of stimulus × language	2,518	4.78	<b>0.009</b>	0.018
Type of stimulus × age	2,518	9.79	<b>0.000</b>	0.036
Type of stimulus × language × age	2,518	1.87	0.155	0.007
Language	1,259	14.15	<b>0.000</b>	0.052
Age	1,259	9.45	<b>0.002</b>	0.035
Language × age	1,259	2.98	0.086	0.011
<b>Sounding-out behavior with errors</b>				
Type of stimulus	2,518	1.80	0.167	0.007

Factor	df	F	p	ES
Type of stimulus × language	2,518	1.49	0.226	0.006
Type of stimulus × age	2,518	0.13	0.875	0.001
Type of stimulus × language × age	2,518	1.39	0.249	0.005
Language	1,259	1.55	0.214	0.006
Age	1,259	0.57	0.452	0.002
Language × age	1,259	4.42	<b>0.037</b>	0.017
<b>Syllabication (with no error)</b>				
Type of stimulus	2,518	1.62	0.199	0.006
Type of stimulus × language	2,518	0.83	0.437	0.003
Type of stimulus × age	2,518	0.91	0.401	0.004
Type of stimulus × language × age	2,518	0.91	0.405	0.003
Language	1,259	10.43	<b>0.001</b>	0.039
Age	1,259	6.68	<b>0.010</b>	0.025
Language × age	1,259	8.58	<b>0.004</b>	0.032
<b>Syllabication (with error)</b>				
Type of stimulus	2,518	9.60	<b>0.000</b>	0.036
Type of stimulus × language	2,518	0.13	0.881	0.000
Type of stimulus × age	2,518	6.47	<b>0.002</b>	0.024
Type of stimulus × language × age	2,518	0.46	0.631	0.002
Language	1,259	5.32	<b>0.022</b>	0.020
Age	1,259	14.87	<b>0.000</b>	0.054
Language × age	1,259	5.08	<b>0.025</b>	0.019
<b>Fragments</b>				
Type of stimulus	2,518	9.84	<b>0.000</b>	0.037
Type of stimulus × language	2,518	8.46	<b>0.000</b>	0.032
Type of stimulus × age	2,518	0.32	0.725	0.001
Type of stimulus × language × age	2,518	0.18	0.834	0.001
Language	1,259	8.77	<b>0.003</b>	0.033
Age	1,259	0.39	0.532	0.002
Language × age	1,259	0.95	0.331	0.004
<b>Omission—I do not know responses</b>				
Type of stimulus	2,518	7.51	<b>0.001</b>	0.028
Type of stimulus × language	2,518	2.09	0.124	0.008
Type of stimulus × age	2,518	0.48	0.621	0.002
Type of stimulus × language × age	2,518	3.37	<b>0.035</b>	0.013
Language	1,259	0.72	0.398	0.003
Age	1,259	0.01	0.903	0.000
Language × age	1,259	2.15	0.143	0.008
<b>Phonological-visual errors</b>				
Type of stimulus	2,518	239.79	<b>0.000</b>	<b>0.481</b>
Type of stimulus × language	2,518	6.69	<b>0.001</b>	0.025
Type of stimulus × age	2,518	37.20	<b>0.000</b>	0.126
Type of stimulus × language × age	2,518	1.77	0.171	0.007
Language	1,259	0.93	0.337	0.004
Age	1,259	69.54	<b>0.000</b>	<b>0.212</b>

Factor	df	F	p	ES
Language × age	1,259	0.02	0.897	0.000
<b>Regularization errors (including stress assignment errors)</b>				
Type of stimulus	2,518	48.39	<b>0.000</b>	<b>0.156</b>
Type of stimulus × language	2,518	0.37	0.542	0.001
Type of stimulus × age	2,518	24.75	<b>0.000</b>	0.087
Type of stimulus × language × age	2,518	1.88	0.172	0.007
Language	1,259	0.49	0.483	0.002
Age	1,259	21.98	<b>0.000</b>	0.078
Language × age	1,259	0.27	0.601	0.001
<b>Regularization and phonological-visual errors</b>				
Type of stimulus	2,518	4.17	<b>0.042</b>	0.016
Type of stimulus × language	2,518	0.30	0.587	0.001
Type of stimulus × age	2,518	6.31	<b>0.013</b>	0.024
Type of stimulus × language × age	2,518	0.01	0.941	0.000
Language	1,259	0.03	0.870	0.000
Age	1,259	12.00	<b>0.001</b>	0.044
Language × age	1,259	1.23	0.268	0.005
<b>Word substitutions</b>				
Type of stimulus	2,518	35.53	<b>0.000</b>	0.120
Type of stimulus × language	2,518	4.92	<b>0.027</b>	0.018
Type of stimulus × age	2,518	19.55	<b>0.000</b>	0.070
Type of stimulus × language × age	2,518	4.43	<b>0.036</b>	0.017
Language	1,259	34.54	<b>0.000</b>	0.117
Age	1,259	19.40	<b>0.000</b>	0.069
Language × age	1,259	6.02	<b>0.015</b>	0.023
<b>Morphological errors</b>				
Type of stimulus	2,518	0.91	0.342	0.003
Type of stimulus × language	2,518	7.74	<b>0.006</b>	0.029
Type of stimulus × age	2,518	5.46	<b>0.020</b>	0.020
Type of stimulus × language × age	2,518	1.95	0.163	0.007
Language	1,259	0.00	0.960	0.000
Age	1,259	3.77	<b>0.053</b>	0.014
Language × age	1,259	1.50	0.222	0.006
<b>Semantic errors</b>				
Type of stimulus	2,518	3.17	0.076	0.012
Type of stimulus × language	2,518	6.14	<b>0.014</b>	0.023
Type of stimulus × age	2,518	1.03	0.311	0.004
Type of stimulus × language × age	2,518	0.10	0.752	0.000
Language	1,259	3.17	0.076	0.012
Age	1,259	0.10	0.752	0.000
Language × age	1,259	1.03	0.311	0.004
<b>Non-word lexicalizations</b>				
Language	1,259	67.81	<b>0.000</b>	<b>0.206</b>
Age	1,259	13.39	<b>0.000</b>	0.049
Language × age	1,259	2.91	0.089	0.011

*B.2. Age comparison (List 1): ANOVAs on different types of reading productions (non-word, word utterances) when children made an error (Significant effects are marked in bold)*

ES: effect size is expressed as partial eta squared. Large ESs (> 0.14) for main effects and interactions are marked in bold.

Factor	df	F	p	ES
<b>Non-word production</b>				
Type of stimulus	2,518	238.78	<b>0.000</b>	<b>0.480</b>
Type of stimulus × language	2,518	3.95	<b>0.020</b>	0.015
Type of stimulus × age	2,518	40.42	<b>0.000</b>	0.135
Type of stimulus × language × age	2,518	0.54	0.582	0.002
Language	1,259	2.62	0.107	0.010
Age	1,259	75.76	<b>0.000</b>	<b>0.226</b>
Language × age	1,259	0.14	0.707	0.001
<b>Word production</b>				
Type of stimulus	2,518	24.84	<b>0.000</b>	0.088
Type of stimulus × language	2,518	28.97	<b>0.000</b>	0.101
Type of stimulus × age	2,518	6.69	<b>0.001</b>	0.025
Type of stimulus × language × age	2,518	2.13	0.120	0.008
Language	1,259	37.24	<b>0.000</b>	0.126
Age	1,259	23.01	<b>0.000</b>	0.082
Language × age	1,259	2.49	0.116	0.010

*B.3. Grade-match comparison (List 1): ANOVAs on different types of reading errors (Significant effects are marked in bold.)*

ES: effect size is expressed as partial eta squared. Large ESs (> 0.14) for main effects and interactions are marked in bold.

Factor	df	F	p	ES
<b>Sounding-out behavior with correct responses</b>				
Type of stimulus	2,146	10.66	<b>0.000</b>	0.131
Type of stimulus × language	2,146	3.34	<b>0.038</b>	0.045
Language	1,73	6.03	<b>0.017</b>	0.078
<b>Sounding-out behavior with errors</b>				
Type of stimulus	2,146	1.17	0.313	0.016
Type of stimulus × language	2,146	0.29	0.749	0.004
Language	1,73	0.03	0.866	0.000
<b>Syllabifications (with no error)</b>				

Factor	df	F	p	ES
Type of stimulus	2,146	1.27	0.285	0.018
Type of stimulus × language	2,146	0.60	0.548	0.008
Language	1,73	2.75	0.102	0.037
<b>Syllabifications (with errors)</b>				
Type of stimulus	2,146	1.51	0.225	0.021
Type of stimulus × language	2,146	0.31	0.732	0.004
Language	1,259	1.29	0.261	0.018
<b>Fragments</b>				
Type of stimulus	2,146	2.32	0.102	0.032
Type of stimulus × language	2,146	3.77	<b>0.025</b>	0.050
Language	1,73	1.76	0.189	0.024
<b>Omission—I do not know responses</b>				
Type of stimulus	2,146	1.44	0.240	0.020
Type of stimulus × language	2,146	0.48	0.621	0.007
Language	1,73	1.13	0.292	0.016
<b>Phonological-visual errors</b>				
Type of stimulus	2,146	45.49	<b>0.000</b>	<b>0.391</b>
Type of stimulus × language	2,146	0.86	0.426	0.012
Language	1,73	1.26	0.265	0.017
<b>Regularization errors (including stress errors)</b>				
Type of stimulus	2,146	0.51	0.478	0.007
Type of stimulus × language	2,146	0.51	0.478	0.007
Language	1,73	0.04	0.835	0.001
<b>Regularization and phonological-visual errors</b>				
Type of stimulus	2,146	0.21	0.648	0.003
Type of stimulus × language	2,146	0.21	0.648	0.003
Language	1,73	2.02	0.160	0.027
<b>Word substitutions</b>				
Type of stimulus	2,146	2.94	0.091	0.039
Type of stimulus × language	2,146	0.14	0.712	0.002
Language	1,73	3.39	0.070	0.045
<b>Non-word lexicalizations</b>				
Language	1,73	9.67	<b>0.000</b>	<b>0.210</b>

*B.4. Grade-match comparison (List 1): ANOVAs on different types of reading productions when children made an error (Significant effects are marked in bold.)*

ES: effect size is expressed as partial eta squared. Large ESs (> 0.14) for main effects and interactions are marked in bold.



Factor	df	F	p	ES
<b>Non-word production</b>				
Type of stimulus	2,518	45.03	<b>0.000</b>	<b>0.388</b>
Type of stimulus × language	2,518	1.28	0.282	0.018
Language	1,259	1.66	0.201	0.023
<b>Word production</b>				
Type of stimulus	2,518	7.24	<b>0.001</b>	0.093
Type of stimulus × language	2,518	9.12	<b>0.000</b>	0.114
Language	1,259	9.20	<b>0.003</b>	0.115

### Appendix C

Means and SDs for all error types (List 2). Data indicate average percentage error rates and are separately presented as a function of grade and language.

Types of error	Type of stimulus	Italian						English			
		2nd grade		4th grade		5th grade		3rd grade		5th grade	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Sounding-out behavior with correct responses	Irregular	1.33	2.68	3.37	3.79	2.76	3.79	0.33	1.47	0.43	1.33
	Regular	1.19	2.56	2.72	3.50	3.22	3.39	0.42	1.35	0.43	1.81
Sounding-out behavior with errors	Irregular	0.67	1.43	0.57	1.62	0.11	0.62	0.00	0.00	0.29	1.18
	Regular	0.63	2.39	0.42	1.22	0.23	0.86	0.00	0.00	0.00	0.00
Syllabifications (with no error)	Irregular	2.04	3.86	0.11	0.61	0.11	0.62	0.00	0.00	0.07	0.49
	Regular	2.26	4.68	0.23	0.99	0.23	0.86	0.17	0.74	0.07	0.49
Syllabifications (with errors)	Irregular	3.41	6.32	0.23	0.85	0.11	0.62	1.17	1.93	0.58	1.28
	Regular	3.04	5.80	0.31	1.21	0.11	0.62	0.25	0.89	0.07	0.49
Fragments	Irregular	1.11	2.45	1.57	3.33	0.23	0.86	1.42	2.81	0.80	2.01
	Regular	0.96	1.95	0.92	2.31	0.34	1.36	0.92	2.13	0.14	0.69
Omission—I do not know responses	Irregular	0.41	1.64	0.54	1.96	0.11	0.62	1.00	2.64	1.30	3.55
	Regular	0.26	1.25	0.27	1.04	0.23	0.86	0.42	1.72	0.58	1.76
Phonological-visual errors	Irregular	5.93	8.39	1.88	3.51	0.00	0.00	17.67	12.90	5.43	7.68
	Regular	7.44	8.35	1.07	2.41	0.23	0.86	11.92	12.49	3.55	4.84
Regularization errors (including stress errors)	Irregular	17.52	8.73	8.77	5.43	5.98	5.59	8.58	5.11	2.90	3.82
	Regular	4.11	4.12	2.72	3.39	0.00	0.00	0.75	1.92	0.00	0.00
Regularization and phonological-visual errors	Irregular	2.41	3.81	0.00	0.00	0.00	0.00	0.25	1.17	0.14	0.69
	Regular	1.15	2.13	0.00	0.00	0.07	0.49	0.50	1.42	0.07	0.49
Word substitutions	Irregular	1.52	4.19	0.65	1.59	1.03	2.37	9.08	10.94	4.28	6.39
	Regular	1.37	2.54	0.54	1.60	0.11	0.62	3.67	7.91	1.74	7.43
Morphological errors	Irregular	0.22	0.84	0.11	0.61	0.11	0.62	0.92	1.85	0.29	0.95
	Regular	0.19	0.77	0.04	0.36	0.00	0.00	0.08	0.53	0.07	0.49
Semantic errors	Irregular	0.00	0.00	0.00	0.00	0.00	0.00	1.00	2.03	0.51	1.21
	Regular	0.07	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.49

Types of production	Type of stimulus	2nd grade		4th grade		5th grade		3rd grade		5th grade	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
		Error resulting in non-word	Irregular	25.67	14.13	10.50	7.31	7.13	7.22	34.08	18.22
Regular	12.81		10.09	3.79	4.52	2.41	3.55	13.08	13.19	3.55	4.84
Error resulting in word	Irregular	1.96	4.32	0.92	1.88	1.26	2.73	11.33	12.69	5.14	6.98
	Regular	1.56	3.01	0.57	1.70	0.11	0.62	3.75	8.07	1.88	8.39

**Appendix D**

Significant effects are marked in bold.

ES: effect size is expressed as partial eta squared. Large ESs (&gt; 0.14) for main effects and interactions are marked in bold.

*D.1. Age comparison (List 2): ANOVAs on different types of reading errors.*

Factor	df	F	p	ES
<b>Sounding-out behavior with correct responses</b>				
Type of stimulus	1,259	0.84	0.359	0.003
Type of stimulus × language	1,259	1.28	0.259	0.005
Type of stimulus × age	1,259	0.57	0.453	0.002
Type of stimulus × language × age	1,259	0.29	0.591	0.001
Language	1,259	33.06	<b>0.000</b>	0.113
Age	1,259	9.23	<b>0.003</b>	0.034
Language × age	1,259	8.07	<b>0.005</b>	0.030
<b>Sounding-out behavior with errors</b>				
Type of stimulus	1,259	1.01	0.315	0.004
Type of stimulus × language	1,259	0.04	0.835	0.000
Type of stimulus × age	1,259	0.72	0.395	0.003
Type of stimulus × language × age	1,259	0.13	0.716	0.001
Language	1,259	10.94	<b>0.001</b>	0.041
Age	1,259	0.00	0.986	0.000
Language × age	1,259	0.95	0.331	0.004
<b>Syllabication (with no error)</b>				
Type of stimulus	1,259	5.61	<b>0.019</b>	0.021
Type of stimulus × language	1,259	2.43	0.120	0.009
Type of stimulus × age	1,259	1.39	0.239	0.005
Type of stimulus × language × age	1,259	0.00	0.959	0.000
Language	1,259	7.49	<b>0.007</b>	0.028
Age	1,259	13.83	<b>0.000</b>	0.051
Language × age	1,259	8.22	<b>0.004</b>	0.031
<b>Syllabication (with error)</b>				
Type of stimulus	1,259	0.71	0.401	0.003
Type of stimulus × language	1,259	0.08	0.776	0.000
Type of stimulus × age	1,259	0.21	0.648	0.001
Type of stimulus × language × age	1,259	0.01	0.921	0.000
Language	1,259	12.64	<b>0.000</b>	0.047
Age	1,259	10.65	<b>0.001</b>	0.039
Language × age	1,259	10.42	<b>0.001</b>	0.039
<b>Fragments</b>				
Type of stimulus	1,259	7.72	<b>0.006</b>	0.029
Type of stimulus × language	1,259	0.25	0.616	0.001
Type of stimulus × age	1,259	0.87	0.352	0.003
Type of stimulus × language × age	1,259	0.25	0.618	0.001
Language	1,259	1.51	0.220	0.006
Age	1,259	0.87	0.353	0.003

Factor	df	F	p	ES
Language × age	1,259	2.97	0.086	0.011
<b>Omission—I do not know responses</b>				
Type of stimulus	1,259	9.81	<b>0.002</b>	0.037
Type of stimulus × language	1,259	2.62	0.107	0.010
Type of stimulus × age	1,259	0.23	0.635	0.001
Type of stimulus × language × age	1,259	0.00	0.969	0.000
Language	1,259	4.70	<b>0.031</b>	0.018
Age	1,259	0.51	0.474	0.002
Language × age	1,259	0.15	0.697	0.001
<b>Phonological-visual errors</b>				
Type of stimulus	1,259	16.99	<b>0.000</b>	0.062
Type of stimulus × language	1,259	24.72	<b>0.000</b>	0.087
Type of stimulus × age	1,259	0.84	0.359	0.003
Type of stimulus × language × age	1,259	13.59	<b>0.000</b>	0.050
Language	1,259	36.91	<b>0.000</b>	0.125
Age	1,259	71.74	<b>0.000</b>	<b>0.217</b>
Language × age	1,259	7.72	<b>0.006</b>	0.029
<b>Regularization errors (including stress assignment errors)</b>				
Type of stimulus	1,259	290.37	<b>0.000</b>	<b>0.529</b>
Type of stimulus × language	1,259	24.27	<b>0.000</b>	0.086
Type of stimulus × age	1,259	48.10	<b>0.000</b>	<b>0.157</b>
Type of stimulus × language × age	1,259	1.86	0.173	0.007
Language	1,259	104.68	<b>0.000</b>	<b>0.288</b>
Age	1,259	65.85	<b>0.000</b>	<b>0.203</b>
Language × age	1,259	3.28	0.071	0.013
<b>Regularization and phonological-visual errors</b>				
Type of stimulus	1,259	3.20	0.075	0.012
Type of stimulus × language	1,259	5.65	<b>0.018</b>	0.021
Type of stimulus × age	1,259	2.40	0.123	0.009
Type of stimulus × language × age	1,259	6.84	<b>0.009</b>	0.026
Language	1,259	10.65	<b>0.001</b>	0.040
Age	1,259	26.58	<b>0.000</b>	0.093
Language × age	1,259	14.53	<b>0.000</b>	0.053
<b>Word substitutions</b>				
Type of stimulus	1,259	50.60	<b>0.000</b>	<b>0.163</b>
Type of stimulus × language	1,259	44.32	<b>0.000</b>	<b>0.146</b>
Type of stimulus × age	1,259	6.36	<b>0.012</b>	0.024
Type of stimulus × language × age	1,259	6.08	<b>0.014</b>	0.023
Language	1,259	84.67	<b>0.000</b>	<b>0.246</b>
Age	1,259	35.01	<b>0.000</b>	0.119
Language × age	1,259	11.55	<b>0.001</b>	0.043
<b>Morphological errors</b>				
Type of stimulus	1,259	18.05	<b>0.000</b>	0.065
Type of stimulus × language	1,259	11.69	<b>0.001</b>	0.043
Type of stimulus × age	1,259	4.42	<b>0.036</b>	0.017
Type of stimulus × language × age	1,259	5.72	<b>0.017</b>	0.022

Factor	df	F	p	ES
Language	1,259	5.64	<b>0.018</b>	0.021
Age	1,259	6.98	<b>0.009</b>	0.026
Language × age	1,259	1.29	0.257	0.005
<b>Semantic errors</b>				
Type of stimulus	1,259	24.34	<b>0.000</b>	0.086
Type of stimulus × language	1,259	29.93	<b>0.000</b>	0.104
Type of stimulus × age	1,259	3.17	0.076	0.012
Type of stimulus × language × age	1,259	5.37	<b>0.021</b>	0.020
Language	1,259	30.52	<b>0.000</b>	0.105
Age	1,259	3.29	0.071	0.013
Language × age	1,259	1.61	0.205	0.006

D.2. Age comparison (List 2): ANOVAs on different types of reading productions (non-word, word utterances) when children made an error

Significant effects are marked in bold.

ES: effect size is expressed as partial eta squared. Large ESs (> 0.14) for main effects and interactions are marked in bold.

Factor	df	F	p	ES
<b>Non-word production</b>				
Type of stimulus	1,259	356.41	<b>0.000</b>	<b>0.579</b>
Type of stimulus × language	1,259	11.94	<b>0.001</b>	0.044
Type of stimulus × age	1,259	61.13	<b>0.000</b>	<b>0.191</b>
Type of stimulus × language × age	1,259	8.83	<b>0.003</b>	0.033
Language	1,259	3.08	0.081	0.012
Age	1,259	128.42	<b>0.000</b>	<b>0.331</b>
Language × age	1,259	2.90	0.090	0.011
<b>Word production</b>				
Type of stimulus	1,259	76.79	<b>0.000</b>	<b>0.229</b>
Type of stimulus × language	1,259	58.16	<b>0.000</b>	<b>0.183</b>
Type of stimulus × age	1,259	10.98	<b>0.001</b>	0.041
Type of stimulus × language × age	1,259	10.36	<b>0.001</b>	0.038
Language	1,259	39.18	<b>0.000</b>	0.131
Age	1,259	13.61	<b>0.000</b>	0.050
Language × age	1,259	4.87	<b>0.028</b>	0.018

D.3. Grade-match comparison (List 2) ANOVAs on different types of reading errors.

Significant effects are marked in bold.

ES: effect size is expressed as partial eta squared. Large ESs (> 0.14) for main effects and interactions are marked in bold.

Factor	df	F	p	ES
<b>Sounding-out behavior with correct responses</b>				
Type of stimulus	1,73	0.52	0.473	0.007
Type of stimulus × language	1,73	0.52	0.473	0.007
Language	1,73	24.66	<b>0.000</b>	<b>0.252</b>
<b>Sounding-out behavior with errors</b>				
Type of stimulus	1,73	0.54	0.465	0.007
Type of stimulus × language	1,73	2.89	0.093	0.038
Language	1,73	0.03	0.854	0.000
<b>Syllabifications (with no error)</b>				
Type of stimulus	1,73	2.02	0.159	0.027
Type of stimulus × language	1,73	5.08	<b>0.027</b>	0.065
Language	1,73	0.93	0.339	0.013
<b>Syllabifications (with errors)</b>				
Type of stimulus	1,73	0.00	1.000	0.000
Type of stimulus × language	1,73	0.00	1.000	0.000
Language	1,73	0.15	0.703	0.002
<b>Fragments</b>				
Type of stimulus	1,73	1.39	0.242	0.019
Type of stimulus × language	1,73	2.84	0.096	0.037

Factor	df	F	p	ES
Language	1,73	0.62	0.434	0.008
<b>Missing—I do not know responses</b>				
Type of stimulus	1,73	1.25	0.267	0.017
Type of stimulus × language	1,73	2.37	0.128	0.031
Language	1,73	2.83	0.097	0.037
<b>Phonological-visual errors</b>				
Type of stimulus	1,73	4.76	<b>0.032</b>	0.061
Type of stimulus × language	1,73	1.76	0.189	0.024
Language	1,73	9.76	<b>0.003</b>	0.118
<b>Regularization errors (including stress errors)</b>				
Type of stimulus	1,73	0.21	0.651	0.003
Type of stimulus × language	1,73	0.21	0.651	0.003
Language	1,73	1.97	0.165	0.026
<b>Regularization and phonological-visual errors</b>				
Type of stimulus	1,73	48.42	<b>0.000</b>	0.399
Type of stimulus × language	1,73	1.98	0.164	0.026
Language	1,73	13.73	<b>0.000</b>	0.158
<b>Word substitutions</b>				
Type of stimulus	1,73	8.71	<b>0.004</b>	0.107
Type of stimulus × language	1,73	1.91	0.172	0.025
Language	1,73	4.27	<b>0.042</b>	0.055
<b>Semantic errors</b>				
Type of stimulus	1,73	1.31	0.577	0.004
Type of stimulus × language	1,73	1.61	0.206	0.042
Language	1,73	3.13	<b>0.049</b>	0.078

D.4 Grade-match comparison (List 2): ANOVAs on different types of reading productions (non-word, word utterances) when children made an error.

ES: effect size is expressed as partial eta squared. Large ESs (> 0.14) for main effects and interactions are marked in bold

Factor	df	F	p	ES
<b>Non-word production</b>				
Type of stimulus	1,73	45.44	<b>0.000</b>	<b>0.384</b>
Type of stimulus × language	1,73	2.13	0.149	0.028
Language	1,73	2.58	0.112	0.034

Factor	df	F	p	ES
<b>Word production</b>				
Type of stimulus	1,73	11.39	<b>0.001</b>	0.135
Type of stimulus × language	1,73	2.61	0.110	0.035
Language	1,73	4.63	<b>0.035</b>	0.060

## References

- Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. University of Pennsylvania, Linguistic Data Consortium.
- Bakker, D. J. (1992). Neuropsychological classification and treatment of dyslexia. *Journal of Learning Disabilities*, 25(2), 102–109. <https://doi.org/10.1177/002221949202500203>
- Bakker, D. J. (2006). Treatment of developmental dyslexia: A review. *Pediatric Rehabilitation*, 9(1), 3–13. <https://doi.org/10.1080/13638490500065392>
- Barca, L., Ellis, A. W., & Burani, C. (2007). Context-sensitive rules and word naming in Italian children. *Reading and Writing: An Interdisciplinary Journal*, 20, 495–509.
- Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, C., Ratti, D., Rolando, C., & Thornton, A.M. (2005). *CoLFIS (Corpus e Lessico di Frequenza dell’Italiano Scritto) [Corpus and Frequency Lexicon of Written Italian]*. <http://www.istic.cnr.it/groupage/databases>.
- Burani, C., & Arduino, L. S. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain and Language*, 90(1–3), 318–325. [https://doi.org/10.1016/S0093-934X\(03\)00444-9](https://doi.org/10.1016/S0093-934X(03)00444-9)
- Burani, C., Paizi, D., & Sulpizio, S. (2014). Stress assignment in reading Italian: Friendship outweighs dominance. *Memory and Cognition*, 42, 662–675. <https://doi.org/10.3758/s13421-013-0379-5>
- Burani, C., Thornton, A., & Zoccolotti, P. (2017). Learning to read Italian. In L. Verhoeven & C. Perfetti (Eds.), *Learning to read across languages and writing systems* (pp. 211–242). Cambridge University Press.
- Coltheart, M. (1980). The semantic error: Types and theories. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 146–159). Kegan Paul.
- De Luca, M., Pontillo, M., Primativo, S., Spinelli, D., & Zoccolotti, P. (2013). The eye-voice lead during oral reading in developmental dyslexia. *Frontiers in Human Neuroscience*, 7, 696. <https://doi.org/10.3389/fnhum.2013.00696>
- Ellis, N. C., & Hooper, A. M. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics*, 22(4), 571–599. <https://doi.org/10.1017/S0142716401004052>
- Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Van Daal, V. H., Polyzoe, N., & Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, 39(4), 438–468. <https://doi.org/10.1598/RRQ.39.4.5>

- Fairbanks, G. (1937). The relation between eye-movements and voice in the oral reading of good and poor silent readers. *Psychological Monographs*, 48(3), 78–107. <https://doi.org/10.1037/h0093394>
- Hanna, P. R., Hanna, J. S., Hodges, R. E., & Rudorf, E. H. (1966). *Phoneme-grapheme correspondences as cues to spelling improvement*. U.S. Department of Health, Education, and Welfare.
- Hendriks, A. W., & Kolk, H. H. J. (1997). Strategic control in developmental dyslexia. *Cognitive Neuropsychology*, 14(3), 321–366. <https://doi.org/10.1080/026432997381510>
- Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German–English comparison. *Cognition*, 63(3), 315–334. [https://doi.org/10.1016/S0010-0277\(97\)00005-X](https://doi.org/10.1016/S0010-0277(97)00005-X)
- Marinelli, C. V., Home, J. K., McGeown, S., Zoccolotti, P., & Martelli, M. (2014). Does the mean adequately represent reading performance? Evidence from a cross-linguistic study. *Frontiers in Psychology*, 5(903), 1–16. <https://doi.org/10.3389/fpsyg.2014.00903>
- Marinelli, C. V., Romani, C., Burani, C., & Zoccolotti, P. (2015). Spelling acquisition in English and Italian: A cross-linguistic study. *Frontiers in Psychology*, 6, 1843. <https://doi.org/10.3389/fpsyg.2015.01843>
- Marinelli, C. V., Romani, C., Burani, C., Mc Gowan, V. A., & Zoccolotti, P. (2016). Costs and benefits of orthographic inconsistency in reading: Evidence from a cross-linguistic comparison. *PLoS ONE*, 11(6), e0157457. <https://doi.org/10.1371/journal.pone.0157457>
- Marinelli, C. V., Zoccolotti, P., & Romani, C. (2020). The ability to learn new written words is modulated by language orthographic consistency. *PLoS ONE*, 15(2), e0228129. <https://doi.org/10.1371/journal.pone.0228129>
- Mauti, M., & Marinelli, C.V., O'Connor, R.J., Zoccolotti, P., & Martelli M. (2023). Decision times in orthographic processing: A cross-linguistic study. *Experimental Brain Research*. <https://doi.org/10.1007/s00221-022-06542-0>
- Patel, T. K., Snowling, M. J., & de Jong, P. F. (2004). A cross-linguistic comparison of children learning to read in English and Dutch. *Journal of Educational Psychology*, 96(4), 785–797. <https://doi.org/10.1037/0022-0663.96.4.785>
- Perfetti, C., & Harris, L. (2017). Learning to read English. In L. Verhoeven & C. Perfetti (Eds.), *Learning to read across languages and writing systems* (pp. 347–370). Cambridge University Press.
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Section 4, Advanced progressive matrices*. Oxford Psychologists Press Ltd.
- Seymour, P. H., Aro, M., Erskine, J. M., Collaboration with COST Action A8 Network. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2), 143–174. <https://doi.org/10.1348/000712603321661859>
- Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an “outlier” orthography. *Psychological Bulletin*, 134(4), 584–615. <https://doi.org/10.1037/0033-2909.134.4.584>
- Temple, C.M. (1985). Surface dyslexia: variations within a syndrome. In K.E. Patterson, J.C. Marshall, M. Coltheart (Eds.) *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading*, 1st ed. pp. 269–288. Mahwah: Lawrence Erlbaum Associates.
- Trenta, M., Benassi, M., Di Filippo, G., Pontillo, M., & Zoccolotti, P. (2013). Analysis of error profile in a regular orthography: Role of reading deficit and strategic control. *Cognitive Neuropsychology*, 30, 147–171. <https://doi.org/10.1080/02643294.2013.814569>
- Wimmer, H. (1993). Characteristics of developmental dyslexia in regular writing system. *Applied Psycholinguistics*, 14, 1–33. <https://doi.org/10.1017/S0142716400010122>
- Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition*, 51(1), 91–103. [https://doi.org/10.1016/0010-0277\(94\)90010-8](https://doi.org/10.1016/0010-0277(94)90010-8)
- Wimmer, H., & Schurz, M. (2010). Dyslexia in regular orthographies: Manifestation and causation. *Dyslexia*, 16(4), 283–299. <https://doi.org/10.1002/dys.411>
- Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology*, 86(3), 169–193. [https://doi.org/10.1016/S0022-0965\(03\)00139-5](https://doi.org/10.1016/S0022-0965(03)00139-5)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.