Department of Psychology

PhD program in Experimental Psychology, Linguistics and Neurocognitive Sciences
Cycle XXXII
Curriculum Mind and Brain

# BILINGUALISM IN DIFFERENT CONTEXTS:

## ACADEMIC OUTCOMES IN PRIMARY SCHOOL CHILDREN

Costa Francesca

Registration number: 823646

Tutor: Professor Maria Teresa Guasti

Coordinator: Professor Marco Perugini

**ACADEMIC YEAR 2019/2020**

**Table of contents**

## Introduction

This dissertation deals with the effects of biliteracy and its relation with bilingualism. It does so by examining two contexts in which biliteracy may occur. One is the context of immersion schools, where bilingualism and biliteracy are pursued, and the second is the context of immigrant children, who learn Italian as a second language and receive literacy only in the second language or as well as in the mother language. These are very different contexts under various points of view – not only linguistically – and raise different concerns that this dissertation tries to elucidate. It investigates reading development, the relation between oral language and literacy, and narrative abilities as a bridge between oral language and literacy.

As apparent from this brief introduction, we will use the term "bilingualism" in a broad sense to refer to different situations and different degree of bilingualism (Grosjean, 2008).

To be bilingual nowadays may sound like something to aspire to, as to be connected with the entire world, in this modern frenetic international life in which we are all citizens of the same world. A lot of parents aspire for their children to know at least one additional language at a mother-tongue-like level beyond the native one, in part because they missed the same opportunity, and in part because they want to assure the best academic achievement and more social and working opportunities for their children.

The most coveted second language to acquire – for non-English-speaking individuals – is English. This is understandable. English is the language of internationality and guarantees social relationships almost all over the world.

In this spirit, parents who desire to raise their children as bilinguals choose a bilingual education program. In Italy, this is an opportunity mostly present in private schools. Essentially, bilingual education refers to any school program in which more than one language is used to teach non-language academic subject matters. This kind of education is also referred to as *immersion Program*, as children are immersed in the school environment's languages. The program's structure

and the specific second language chosen may vary even though, as said, English is the most represented.

The benefit of learning and knowing a second language is clear, given the range of social possibilities that could derive from it. However the educational outcomes of bilingual education still worry some parents and stakeholders, especially for what concerns the native language development. In Italy, the general belief is that learning two languages and the subject contents in two languages may harm literacy acquisition in Italian, which is the majority language of the community. The idea is that providing education in a language other than Italian weakens Italian. This may be one reason why immersion programs have not yet spread in public schools.

To oppose this idea, one of the aims of the present thesis is to answer the question of whether learning to read in Italian (the first language: L1) is delayed when simultaneously learning to read in English (the second language: L2). We will explore the Italian-English education outcomes in a population of primary school children attending bilingual schools in Italy, looking at their linguistic, cognitive, narrative, and reading performances.

In spite of preconceptions about early acquisition, English is still considered a language that is worth knowing. Knowing a minority language (such as a dialect or a language spoken by an immigrant community) could be regarded as useless. Bilinguals who have a minority first language and live in a foreign country – with a majority language different than their mother tongue – need to learn a second language. Typically, this population consists of children of immigrant families. They are usually referred to as *Heritage bilinguals*, and their first language (*i.e.*, their heritage language) is mostly learned at home, but never fully developed because of insufficient input from the social environment.

With respect to the previous considerations, we can distinguish between two different concepts: *bilingual education* and the *education of bilingual children*. Genesee (2004) defined bilingual education as the 'education that aims to promote bilingual competence by using both languages as media of instruction for significant portions of the academic curriculum. In contrast,

the education of bilingual children in Italy typically consists of teaching non-majority-speaking students to read and write in the majority-second language so that all the contents are learned only in Italian. They are totally immersed in their second language once they enter the Italian education system, without any graduality or transition period. Thus, as their oral Italian competence might not be age-appropriate according to monolingual norms, it may be more difficult for them to acquire Italian literacy. As a result, they often lag behind their Italian monolingual peers.

Moreover, the use of their native language, the minority language, is often relegated to their family or community context (often a low socioeconomic context), and it is acquired only in its oral form, without any other support. Despite the evidence that active support to the first language, also in its written form, could be beneficial for the second language literacy acquisition (*e.g.*, Papastefanou, Powell, & Marinis 2019; van der Velde Kremin, Arredondo, Hsu, Satterfield, & Kovelman, 2019), in Italy, only a few initiatives to support the enhancement of cultural heritage have started to spread. Such initiatives are mostly initiated by the immigrant communities themselves or by consulates of the countries of origin.

In the present thesis, we explored the effect of biliteracy – that is, learning to read and write both in Italian majority-second language and the minority-first language – on the linguistic, cognitive, and reading abilities of immigrant children in Italy. The aim is to verify whether supporting the first language influences the acquisition of Italian.

It is evident that bilingualism – and bilingual education – is an umbrella term that encompasses very heterogeneous circumstances and concerns children with different extractions and backgrounds, who may have different needs and goals. As mentioned by Bialystok (2018), bilingual education of Heritage-minority students typically aims to create an educational program for children at risk of academic failure because of low proficiency in Italian, the language of schooling. In the case of Italy, one option would be to implement biliteracy programs in which children learn to read both in Italian and in their home language. Note, however, that the success of biliteracy for immigrants is judged primarily by Italian proficiency (the majority language) literacy.

In contrast, for bilingual education in Italian-English immersion programs, the motivation is to offer an educational alternative, *i.e.*, to make Italian-majority language children bilingual. Thus, the success of these programs is judged by the extent to which children master English while maintaining proficiency in the majority language (Italian).

These two types of bilingual programs differ significantly from each other, and they compose different pieces of the complex puzzle called bilingualism. The present thesis concurs with the broadening of knowledge about bilingualism and bilingual education, and aspires to reach those responsible for the education of future generations.

**References**

Bialystok E. (2018). Bilingual education for young children: review of the effects and consequences. *Int. J. Biling. Educ. Biling*., 21(6), 666-679.

Genesee, F. (2004). What do we know about bilingual education for majority language students? In Bathia, T. K. & Richie, W. (eds.) *Handbook of bilingualism and multiculturalism,* 547-576. Malden: Blackwell.

Grosjean, F. (2008). *Studying Bilinguals*. Oxford: Oxford University Press.

Papastefanou, T., Powell, D., & Marinis, T. (2019). Language and Decoding Skills in Greek-English Primary School Bilingual Children: Effects of Language Dominance, Contextual Factors and Cross-Language Relationships Between the Heritage and the Majority Language. *Front. Commun*., 4(65).

van der Velde Kremin, L., Arredondo, M. M., Hsu, L., Satterfield, T., & Kovelman, I. (2019). The effects of Spanish heritage language literacy on English reading for Spanish-English bilingual children in the US. *International journal of bilingual education and bilingualism*, *22*(2), 192–206.

# Chapter 1

# ITALIAN-ENGLISH IMMERSION PROGRAMS IN ITALY:

# ACADEMIC OUTCOMES ON PRIMARY SCHOOL CHILDREN

## Introduction

In Italy, many families wish their children would become fluent in English to navigate the global economy in their working future. In 2000, English as a foreign language has become compulsory from primary school, but it is rather evident that learning English as a school subject is not enough to gain proper bilingual competence. Learning to use a language as an effective medium requires more language experience and *immersion* in the language itself. Thus, over time, the number of schools that offer bilingual education programs has increased, intending to promote a real bilingual competence and cultural enrichment. Rather than merely teaching English as a foreign language, bilingual academic *curricula* use English as a medium of instruction for the school subjects (Baker, 2001). Among *dual-language* approaches to bilingual education, in *immersion* programs, topics are conveyed in both the majority language of the country (L1, *i.e.*, Italian in Italy) and an additional language (L2, mainly English in Italy). The age at which children start the immersion experience and the amount of time spent in L2 immersion may vary depending on when the child enters the bilingual education system and the type of immersion program the school pursues.

A child could enter an immersion program very early, since nursery, for example, or since kindergarten. Children are effectively *immersed* in two languages, thus acquiring a second language implicitly through input during routines, daily activities, songs, and the learning of subject content. Penfield and Roberts (1959) stated that this approach is similar to the *mother's method*, and it seems more comfortable and effective in acquiring a second language than formal and explicit language instruction. When facing literacy acquisition at primary school, children learn to read and write in

two languages. Immersion teachers use only the target language in their interactions to encourage students to use the second language. Thus, children desire to learn the second language to engage in meaningful and exciting communication (Macnamara, 1973); learning is not as much grammar-driven but rather proficiency-driven, according to the learners' communication needs in the classroom (Genesee, 1985a). Teachers teach the second language both with verbal messages and (initially mostly) nonverbal significant of events, so students come to comprehend and ultimately use the second language effectively (Genesee, 1985b). Moreover, separation of languages exists, and teachers typically present content only in L1 or only in L2: the two languages are rarely used within the same lesson and are not integrated during instructional time.

The following sections will provide an overview of the immersion approach, its features, and its potential output. First, we will present different forms of immersion programs. We will discuss some advantages, which could derive from bilingualism and bilingual education, followed by a discussion about the age of first exposure to a second language, an essential variable to be considered in bilingual acquisition research. Then, we will describe different languages in terms of writing systems and other properties to keep into account when studying biliteracy. Finally, we will conclude this introduction with an overview of the Italian context.

**Immersion approaches**

In this section, we will describe the various types of immersion programs that have been developed and implemented throughout time. A review of the literature will illustrate the different forms of immersion education that developed in response to specific needs, and the debate on their effectiveness.

Immersion programs were originally introduced in 1965 in Canada, a country where bilingualism is official, carried out in two majority languages (*i.e.*, English and French). Canadian immersion programs have thus been developed in response to the practical reality of bilingualism in the country and policies of official bilingualism (Lambert & Tucker, 1972). Then, some versions of

the immersion approach have been implemented in the United States, where immigration has a secular history (*e.g.*, South California and Texas). In this case, bilingual education originated from a particular need to maintain and support the heritage language and culture. Thus, initially the rationale behind immersion education was primarily to preserve cultural and linguistic diversity of citizens with a bilingual background, because of their country policy or because of their history of immigration. Later, bilingual (or even multilingual) immersion programs have been instituted worldwide, spreading from Europe to Asia, solicited by the increasing internationalization of the world. As the world became gradually more interconnected, ruled by a global economy and mobility, implementing a second language immersion program for majority language students had become so much necessary, as an enrichment to prepare humans for a life in the modern world (Bialystok, 2009).

Generally, a distinction can be made between *partial* immersion and *total* immersion. The latter, which we are not going to discuss, provides full teaching in a second language and is typically implemented in International schools in a host country (*e.g.*, International Baccalaureate (IB) Programs). Among the partial immersion alternatives, the two main types are the 50:50 or *simultaneous* and the 90:10 or *sequential* immersion program. 50:50 or *simultaneous* is the most widespread immersion program worldwide. It ensures that children are equally exposed to two languages within the same developmental and educational time and growth window (*e.g.*, Kovelman, Baker & Petitto, 2008; Slavin & Cheung, 2005; Lara-Alecio, Galloway, Irby, Rodriguez, & Gomez, 2004), such that children consistently receive approximately equal amounts of instruction in both languages across all subjects, every day. More specifically, the instructional day is equally divided between two languages, and subjects are taught 50% in the majority language and 50% in a second language.

On the other hand, in 90:10 or *sequential* immersion programs, literacy starts to a greater extent in a minority language, with a minor quantity of instruction in the majority language. However, instruction in the majority language slowly increases throughout the school years. More

precisely, students placed in a sequential immersion program start spending approximately 90% of the instructional day learning the minority language (which it is typically their heritage language) and the remaining 10% being instructed in the majority language (*i.e.*, the language spoken in the country where they live). Then this ratio of language instruction rebalances throughout primary school, with minority language immersion gradually decreasing and the use of majority language increasing, usually reaching 50:50 by Grade 5. Following this, learning in the majority language is slowly introduced over a protracted period, until finally, the amount of learning in both languages becomes equal around ages 10-11 years old (Lara-Alecio et al., 2004). In this type of immersion program, students typically start learning how to read and write in the L1 (their dominant/home language) and, subsequently, in the majority language, which is the L2 for them.

90:10 immersion programs mostly make sense in those countries where students from immigrant families constitute a substantial part of the population in the demographic rate (like people of Hispanic or Mexican origin in California and Texas, for example). Those students have a home language (*e.g.*, in the USA, Spanish) other than the majority language of the country (in this case, English), and they enter school experiencing an initial emphasis on their minority language (L1) in the early grades and a maintenance model in which their L1 literacy is developed. Thus, the L1 could represent a solid base for literacy development in English (English as a Second Language program - ESL), and knowledge and skills acquired in the L1 can transfer to the L2 (Cummins, 1981, 1991). In any case, 90:10 programs are also thought for native English speakers too (considering the reported example of some US Countries), who are thus allowed to develop proficiency in an L2 (Spanish in this case) without sacrificing the development of academic language proficiency in their L1 (Genesee, 1987; Swain & Lapkin, 1982).

Some authors have wondered which form of immersion program – if there is a better one – is the most effective way for young bilinguals to learn reading. Bialystok, Majunder, and Martin (2003) hypothesized that early exposure to two linguistic systems *simultaneously* (*i.e.*, 50:50 immersion) might afford computational strength and processing agility in speed and accuracy,

representing a cognitive advantage. According to others, 50:50 exposure to two reading systems during the same developmental time may cause learning confusion and interference between the two reading and linguistic systems, hindering the achievement of a good proficiency level in either language (Berens Kovelman, and Petitto, 2013). In 90:10 immersion programs, where literacy is established first in one language while the other language is gradually introduced over time (Cummins, 2005), reading and language learning in the minority language may ostensibly disadvantage the child's full mastery of reading especially in the majority language (Petitto, Berens, Kovelman, Dubins, Jasinska & Shalinsky, 2012). In fact, a concern with the 90:10 immersion model is that learning to read in the majority language would inevitably occur at an older age, when plasticity in the brain for learning processes may be less effective (Dubins, Berens, Kovelman, Shalinsky & Petitto, 2009). Berens *et al.*, (2013) compared the effectiveness of 50:50 and 90:10 immersion programs on the language and reading performance of two groups of Spanish-English second and third graders (ages 7–9), who had been exposed to both Spanish and English since birth or since kindergarten (*i.e.,* from three years of age). Their results suggested that a 90:10 immersion program promotes phonological awareness in Spanish and English, but 90:10 learners lag behind 50:50 learners with respect to deeper grammatical and structural competence, which likely may be gained in upper grades. However, the choice of a 90:10 program may be appropriate for children with an L1 with a transparent orthography as Spanish has. At the same time, Berens et al. suggest that a 50:50 immersion program – paired with phonological training in the early school years – would probably be the optimal choice, especially for children who were exposed early to their two languages. According to the authors, a simultaneous approach to literacy in two languages promotes the understanding of deeper grammatical and structural aspects of language. Studies have shown that grammatical comprehension is a significant predictor of reading proficiency (*e.g.*, Baker, Stoolmiller, Good, & Baker, 2011). Their findings are consistent with the results reported by Grimalt Moraga (2010), who found that, compared to primary school children placed in 90:10 bilingual schools, a higher percentage of students in 50:50 programs reached grade-level literacy in

English. In addition, Berens et al. (2013) proposed that the specific combination of the orthographic systems of the L1 and the language to be learned should be taken into consideration when determining which immersion program would be more effective. In particular, the 50:50 immersion program – with training in phonological decoding, manipulation, and phoneme-grapheme patterns during the early school years in both languages – may ultimately prove the optimal choice for learning, so that children have a solid foundation in both surface and in-depth grammatical features from the beginning of their education. However, Berens and colleagues hypothesized that this phonological reinforcement would be particularly beneficial for children who are native speakers of languages with a deep orthography (*e.g.*, English).

**Bilingual education: advantage or disadvantage for literacy outcomes?**

In this section, we will discuss the goals of educating bilingual children, and we will review evidence on the effects and effectiveness of bilingual over monolingual education on language and literacy outcomes.

In non-English speaking countries (like Italy), English is the most common second language to be implemented in bilingual schools, because of its internationalization potential. English is, in fact, the modern shared medium of economy, travel, culture worldwide. Thus, mastering fluently at least English as a second language nowadays is seen as a springboard for the job market, a need to increase opportunities in life, mostly in industrialized countries. As claimed by Genesee (1985), immersion programs with English as a second language offer the possibility of providing an enriched educational experience without traveling abroad, and a challenging academic experience that some parents may feel is missing in regular monolingual school programs (mostly monolingual).

The overarching goal of bilingual education is to learn reading and gain language skills in both languages with high levels of mastery (Berens *et al.*, 2013), reaching grade-level academic achievements and multicultural competence (Christian, Howard, & Loeb, 2000). From a purely

social-cultural perspective, dual-language learning should represent an advantage over standard single-language education due to the globalization of communication and economy. However, the question about whether bilingual *reading* exposure may afford *reading* advantages or disadvantages, compared to traditional monolingual literacy, prevailed in numerous studies.

So far, research on biliteracy has been conducted in countries – such as Canada or South California mentioned above – where bilingualism is traditionally established. Indeed, numerous authors have investigated the effects of bilingual education on students' literacy development and achievement, often focusing on reading attainment. A conspicuous number of studies found bilingual students gained the same literacy skills as those of their monolingual peers (*e.g.*, Genesee & Jared, 2008; Lindholm-Leary & Hernández, 2011; Turnbull, Lapkin, & Hart, 2001, Abu-Rabia & Siegel, 2002). Others found children benefitted from learning in bilingual programs compared to monolingual programs (*e.g.*, Collier, 1992; Cummins, 1992; Kovelman et al. 2008; López & Tashakkori, 2004; Oller & Eilers, 2002; Rolstad, Mahoney, & Glass, 2005a; Thomas & Collier, 2003). Again, Berens *et al.* (2013), besides comparing the effectiveness of 50:50 and 90:10 English-Spanish immersion programs, also compared the reading and language performance of bilinguals – whose L1 was English – with a group of monolingual peers placed in monolingual English schools. The two groups were assessed on several tasks in English: phonological awareness, reading decoding, irregular word reading, passage comprehension, and expressive language. The authors found that students placed in immersion learning context performed significantly better in English than children from monolingual schools on two tasks, *i.e.*, irregular word reading and passage comprehension, and equally well on the other tasks. Interestingly, bilingual students managed to outperform – in some literacy skills – children placed in English monolingual schools, although they were instructed in English only partially (depending on the type of immersion program attended, namely simultaneous or progressive). Berens *et al.* (2013) concluded that a bilingual immersion approach to education does not hinder and, in some cases, favors – the development of language and reading skills in English. Thus, they showed no evidence of delay in English L1 due

to learning in two languages at once.

Slavin and Cheung (2005) conducted a systematic review of experimental studies focusing on reading attainment in bilingual primary students with different L1s or heritage languages (mostly Spanish). The studies compared biliteracy programs and English-only reading programs for *English language learners (ELLs)*. The majority of the studies found that bilingual approaches lead to better English reading measures than monolingual approaches. The remaining studies found no differences. The authors concluded that existing evidence favors children's exposure to bilingual literacy instead of fully English programs, especially bilingual strategies that teach reading in the native language and English at different times each day. Rolstad, Mahoney, and Glass (2005) carried out another meta-analysis of immersion effectiveness research on English language learners (*i.e.,* bilingual students with different heritage languages and English as a second – or majority – language) on a corpus of 17 studies, which compared bilingual education to English-only instructional programs. The meta-analysis indicated a positive effect on bilingual education programs, which resulted in promoting academic achievement in both the native and second languages (in these cases, a majority language). Finally, in a recent review, Bialystok (2018), examined research evaluating the outcomes of bilingual education (where L1s were different heritage languages, and English was the L2 or majority language) for literacy levels and academic achievement in American contexts, with a focus on early education. The author concluded that there was no credible evidence that bilingual education creates a burden for children, yet it seemed incontrovertible that it provides the advantage of learning another language and possibly some cognitive benefits. Moreover, the available evidence has proved a net benefit in many academic domains for children in bilingual programs in the early school years.

In the following section, we will discuss a fundamental variable to consider when evaluating the outcomes of a bilingual exposure: the age of first exposure.

**Age of first exposure to the second language**

When it comes to the positive effects of bilingual education on reading achievement, it is pivotal to consider the Age of first oral Exposure (AoE) to the L2. For instance, Baker (2001) found that students in an *early* immersion program, starting at least at kindergarten, could reach the same monolingual peers' competence in L1 after about five years of schooling. Immersion program students could initially lag behind their monolingual peers – and this can be expected because, in the first few years of the immersion program, instruction is focused on second-language learning – but they usually caught up with their monolingual peers in L1 competence by the end of primary school. Moreover, L2 oral language proficiency early on has been found to relate to academic success up to Grade 8 (Halle, Hair, Wandener, McNamara & Chien, 2012). In addition, moderate to strong associations have been found between proficiency in the language of education and reading and spelling in bilingual children (Prevoo, Malda, Mesman & van IJzendoorn, 2016).

Kovelman, Baker, and Petitto (2008) established that the age of first exposure to the L2 has an impact on the reading development of second- and third-graders (aged 7-9 years). In their study, students from Spanish-speaking homes attending 50:50 Spanish-English bilingual schools were recruited. The age at which they were exposed to English for the first time varied: one group had been exposed to English within the first three years of age (early bilinguals), a second group between the ages of 3 and 4 and a third group later (middle bilinguals), between 5 and 6 years of age (late bilinguals). Besides those bilinguals, they also included a group of students from English-speaking families whose first exposure to Spanish was 5-6 years. Finally, they recruited a group of English-only-speaking children who attended mainstream English education and had never been exposed to Spanish, which served as a control group. Participants were administered tasks assessing phonological awareness, reading decoding and comprehension, and expressive proficiency. Results showed that early Spanish-English bilingual children performed better than Spanish-English children exposed to the L2 after age 3, in both languages. In particular, early bilinguals reached a high level of English reading proficiency obtaining similar reading scores to their English

monolingual peers; they also exhibited a reading performance in Spanish equal to that of the Spanish-speaking children who had been monolingual until the age of 5-6. By contrast, bilinguals who had been exposed to the second language (English) after age three lagged behind their English-speaking monolingual peers in reading ability. However, an interesting exception was found: students from English-speaking homes who had been exposed to the second language, Spanish, between the ages of 5 and 6 outperformed the English monolingual control group on the tasks measuring phonological awareness in English. A possible explanation for this exception is that they have benefitted from learning Spanish (a language with a shallow orthography) and successfully transferred to English the phonological awareness skills gained in Spanish.

These findings are consistent with several other studies that have reached similar results suggesting that an early oral exposure to L2 – starting within three years of age – allows the best academic results (e.g., Flege, Yeni-Komshian & Liu, 1999; Perani, Abutalebi, Paulesu, Brambati, Scifo, Cappa & Fazio 2003). Numerous authors have identified the first three years of life as a *sensitive period* for acquiring a second language, which should yield better language competence outcomes and literacy proficiency. During their first three years, children generally experience a heightened sensitivity towards the linguistic input, and their linguistic learning reaches its full potential (Kovelman et al., 2008).

It can be concluded that the age of first oral bilingual exposure plays a crucial role in language development and has a substantial impact on literacy development.

**Differences: languages, orthographies and writing systems**

As hinted above, language systems should be taken into account when studying the possible effects of bilingual education. In the following section, we will describe some language differences, and we will review the impact these differences have on reading outcomes.

Languages have different writing systems: alphabetic, syllabic or logographic and among writing alphabetic systems, we find languages with different orthographic depth: shallow or deep.

For instance, Italian and English are languages with the same alphabetic system, but they dissociate in terms of orthographic depth. Italian has a shallow or transparent orthography while English has a deep or opaque orthography. Languages with a shallow orthography – such as Italian – rely on a high correspondence between graphemes and phonemes, while languages with a deep orthography – such as English – display a low level of grapheme-phoneme correspondence; thus, a grapheme can have various pronunciations and the same phoneme can be written by a different grapheme. In addition, several words can be read only by assessing their mining (e.g., we will *wind* up in Milan, the *wind* is blowing). Several studies found that the specific language pairs in a bilingual individual, together with their writing systems, could give rise to different reading outcomes. For instance, Bialystok, Luk, and Kwan (2005) compared early literacy tasks (decoding and phonological awareness tasks) of four groups of first-grade children: three groups of bilinguals, representing each a different combination of languages and writing systems; and a group of monolingual English controls. All the children lived in the same metropolitan area in which English was the majority language. Thus, bilinguals shared English as their L2 and differentiated for their L1, Cantonese, Hebrew, and Spanish, respectively. They spoke both languages daily, their first language at home, English at school, and were learning to read in both languages. Results showed that the progress in learning to read in English was more advanced for bilingual children who were learning two alphabetic systems: the Hebrew-English and the Spanish-English groups exhibited better early literacy skills, in English, than both the Cantonese-English group and the monolingual English children. The two-alphabetic-languages bilinguals displayed a strong correlation between their pseudoword decoding skills across their languages, suggesting an effective transfer of phonological awareness from one language to another. Therefore, bilinguals seemed to transfer literacy skills across languages only when both languages use the same writing system, and this transfer facilitated English reading. The authors concluded that the extent of the bilingual facilitation for early reading depends on the relation between the two languages and writing systems involved. D'Angiulli and Serra (2001) provided further evidence of the transferability of reading skills from a

shallow orthography language to a deep orthography language. Their study involved a group of Canadian middle-class English-Italian bilingual children attending a Heritage Italian Program at school, a group of English monolinguals, and a group of monolingual Italian children. Bilinguals were taught mainly in English, with 35-minutes of Italian instructions every day; at home, they spoke both English and Italian from birth. Results indicated that the bilingual children exhibited better performance than their English monolingual peers on English tasks involving reading and spelling, yet suggesting that the oral exposure and the minimal instruction in a language with a transparent orthography – such as Italian – can enhance English skills. However, the bilingual children lagged behind the monolingual Italian controls on all Italian tasks, but it should be noted that bilinguals' exposure to Italian was limited and certainly not as extensive as their exposure to English. Another small-scale study by Montanari (2013) obtained similar results. In this study, the biliteracy development of 60 children was investigated. Children were from first through third grade and attended an Italian-English program in Southern California. All children were systematically exposed to Italian since entering kindergarten, which means they had been exposed to Italian at least since three years of age. The Italian-English program they were attending provided 90% of Italian Instruction in kindergarten and first grade. Second and third-grade students had shifted to an 80:20 and 70:30 immersion model, respectively, with Italian being the language most represented. The author showed a long-term benefit on English reading acquisition, which should arise from exposure to Italian, a language with a transparent orthography, despite an initial delay in English reading skills in first graders. This advantage might be due to the transfer of phonological awareness - obtained in the language with a more regular grapheme-phoneme correspondence - to the language with a more complex correspondence system. In addition, these children developed strong literacy skills in English and Italian, but this time instructions were mainly in Italian.

Thus, there is evidence of the beneficial effect of Italian on English literacy acquisition. Furthermore, it seems that when both Italian and English are well supported in immersion programs, strong literacy skills could be obtained in Italian.

Nevertheless, only a few studies investigated Italian-English bilinguals' language and reading attainment in Italian (L1), when Italian is the native language of the students, and the majority language of the country. Costa, Sharley, and Guasti (2018) evaluated Italian reading decoding attainments (through reading lists of words and pseudowords) in a group of Italian-English bilingual first- and third-graders attending a private bilingual school in Italy, where *some* school subjects are taught in English, according to the CLIL (*Content and Language Integrated Learning*) methodology. Bilinguals had Italian as their majority language (L1) and had been exposed to English within the first three years of their lives in kindergarten. They were compared with a control group of Italian monolingual peers attending a regular monolingual school. Authors found that bilinguals did not lag behind the monolinguals in the reading performance in Italian, suggesting that the specific dual-language exposure did not harm the reading attainment in Italian, at least by Grade three. In this study, only reading decoding was measured, as lists of words and pseudowords were used in the assessment. Unfortunately, no measure of English reading performance was reported in this study. In another study, Costa, Guasti, and Sharley (2020) compared language and reading proficiency (measured through reading a text) in Italian of a group of first- and third graders early bilinguals attending bilingual schools in Italy (with a 50:50 immersion program) with those of an Italian monolingual control group. All the children lived in Italy and were from Italian mother tongue families. Results showed no differences in language measures of expressive vocabulary and reading speed between bilinguals and monolinguals but a small monolingual advantage in reading accuracy, both in Grades 1 and 3. However, bilinguals' reading attainment was found to be within monolingual norms, and even though the difference in reading accuracy turned out to be statistically significant, the gap between means number of reading errors was lower between third graders than between first graders. The study lacked fifth graders, but it is likely that, by the end of primary school, the gap between bilinguals and monolinguals should be filled out (accordingly to Baker, 2001). In fact, the direction of the distance between groups' performances went in that direction. Overall, it can be concluded that learning to read in

Italian and in English simultaneously does not affect reading performance in Italian – when Italian is the first language and English the second language and bilingual exposure occurs at an early age.

In Italy, attitudes to languages and bilingualism, in general, are very different from those in North America. Even if there is no evidence that bilingual education impedes the development of language and literacy skills in the majority language or creates a burden for children, the context in which education occurs is crucial. There is no universal prescription for bilingual education and no universal outcomes (Bialystok, 2018). As Baker (2011) pointed out, studies conducted in one context may have little relevance for bilingual education in another context.

In the next section, we will describe the Italian context for immersion education, where our study took place.

**The Italian context and the study in immersion programs**

As mentioned above, in Italy, the majority of bilingual schools offering immersion programs provide English as L2 (50:50 Italian-English immersion program is the most common), in virtue of being the most international language nowadays. However, most of these are private institutions with relatively high costs, and bilingual education is reserved for pupils of healthy families. Thus, it is essential to consider cultural and family variables, besides the affordability.

Nevertheless, as noted earlier, there are some real advantages to being bilingual as a result of the globalization of communication and economy that are incontrovertible. Thus, policies concerning the inclusion of students in immersion programs should be extended to children of various social backgrounds (Genesee, 2007). However, considering the peculiar Italian context, we should acknowledge that there are some limitations to this kind of program. Unlike the Canadian or American contexts described earlier, for many Italian children, using English might become a school-only experience. Moreover, immersion programs can be extremely beneficial for language development, but not very effective at promoting cultural knowledge and values linked to the L2. In any case, the right level of proficiency in two languages, if guaranteed by bilingual education,

should not be seen as a disadvantage: to say it with Bialystok (2007), “knowing more has never been a disadvantage when compared to knowing less."

Despite the encouraging results discussed above, in Italy, significant concerns still surround literacy education in a foreign language during primary school. There is a belief that is widespread among parents or stakeholders in education: learning to read in two languages simultaneously may render the young bilingual’s mastery of reading incomplete in either language or, in other words, may produce “a child who is linguistically ‘jack of all trades’ but ‘ace’ of none” (Berens *et al.*, 2013). This is probably the reason why private schools have mainly implemented bilingual education programs in Italy. Thus, only children from families who can afford a private fee can join a bilingual education. Second language immersion programs remain a matter of choice reserved for wealthy families who firmly believe in the social opportunity it represents, and in the challenging academic experience that they may feel missing in regular school programs.

There is little evidence of the double literacy effects on reading skills in Italian-English primary students from Italian speaking families, which may have contributed to the spread of mistrust towards biliteracy. Thus, a more extensive study on literacy outcomes is needed.

**Aim and hypothesis**

Exploring bilingualism is essential today to better understand its effects when it overlooks the educational environment in the form of an immersion program in Italy. It enables researches to detect literacy strengths or weaknesses – if any – in bilingual students. It informs teachers about literacy development that is to be expected when teaching bilingually. It offers knowledge for educational policy to consider the implementation of a bilingual *curriculum* in Italian public schools. Lastly, it answers to parents’ doubts and concerns about biliteracy education.

The present study is a cross-sectional research, which tried to verify whether learning to read in Italian (L1) is delayed when simultaneously learning to read in English (L2). Available evidence considering the specific combination of bilingual orthographies being acquired suggests that there

should not be adverse effects on the Italian literacy outcomes of Italian-English immersion students (at least by the end of primary school). In order to verify this hypothesis, Italian reading performance of three groups of bilinguals educated in 50:50 Italian-English immersion programs (with Italian as their L1 and majority language) in Grades 1, 3 and 5 were compared to that of three control groups of Italian monolingual peers attending mainstream monolingual Italian schools.

The outcomes on language and literacy skills in English (L2) could not be compared to those of an English monolingual group, as it was not possible to find an available monolingual group in an English-speaking country. Thus, we examined the impact of an Italian-English *immersion program* on the development of language and literacy skills in English (L2). We expect to find improvements across the three grades in English language and literacy measures. To pursue this goal, we examined the English performance of the bilingual group across Grade 1, 3, and 5.

Another aim was to verify whether the language and reading attainment exhibited by the bilingual children in Italian were correlated to their English performance. There is evidence of the beneficial effect of languages with a transparent orthography as Italian on English literacy acquisition, suggesting a possible transfer from one language to another. Thus, we expect that the performance in English is correlated with that in Italian. To pursue this goal, we looked for positive correlations between the same skills evaluated in the two languages in the bilingual group.

Moreover, inspired by Kovelmann *et al.* (2008), who showed that the age of first oral exposure to a language is critical to literacy achievement at the age of 8 years, we explored the relationship between oral language abilities and reading measures, considering both the bilingual and the monolingual group. We expect language performance to be correlated with reading proficiency measures.

Working memory skills have been found to predict academic attainment. It has been shown that young monolingual children's working memory scores can be a more powerful predictor of their later academic success than IQ scores (Alloway and Alloway 2010). In line with this observation, we explored the relationship between working memory and reading proficiency

measures in our bilingual pupils. We expect to find a correlation between memory and reading tasks.

Some studies have found a bilingual advantage in working memory. Marini, Eliseeva & Fabbro (2019) reported that a group of 4- to 5-years-old sequential bilinguals attending an International School with Italian as the L1 outperformed a group of age-matched monolinguals on a span task assessing short-term and working memory in Italian. This advantage did not emerge in older bilingual children (aged between 6 and 12 years) nor in young adults in a study by Bonifacci, Giombini, Bellocchi & Contento (2011). In this study, bilinguals had Italian as their L2, with different L1s, practiced at school and home. They were compared to a monolingual control group matched for age, sex, and non-verbal IQ. Bilingual and monolingual children did not differ from each other. Consistent with this study, Ratiu and Azuma (2015) did not find working memory differences between monolinguals and bilinguals adults; a recent systematic review by Lehtonen, Soveri, Laine, Järvenpää, de Bruin, and Antfolk (2018) went in the same direction. In line with these findings, we do not expect to find a bilingual advantage on working memory.

The study was approved by the Ethic Committee of the University of Milano-Bicocca (n. 280).

In sum, our expectations can be summarized as follows:

1) We do not expect to find major disadvantages in Italian reading attainments in the bilingual group with respect to the monolingual one;
2) We expect to find improvements across the three grades in English language and literacy measures;
3) We expect to find correlations between the language and reading skills in Italian and English, suggesting that there is a positive transfer from Italian to English;
4) Language abilities in one language are expected to correlate with reading abilities in the same language, supporting the idea that oral language supports reading development;

5) We do not expect any effect of bilingualism on memory tasks, but we expect to find a correlation between memory skills and reading proficiency measures.

We conclude by underlining that, in order to have a full understanding of the bilinguals' literacy attainment and to verify the positive transfer of Italian on English literacy competence, it would be useful to compare the English performances of the bilingual students with those of an English monolingual control group, but unfortunately, this was not possible.

Making parents, teachers, and institutions aware of the outcomes of an Italian-English education approach could result, in the future, in the implementation of bilingual educational paths in Italian public schools, making bilingualism an accessible educational opportunity for all children in Italy.

## Material and methods

### Participants

Participants were recruited from two private Italian-English bilingual primary schools and three Italian monolingual primary schools. Bilingual schools were in the Province of Milan and Monza-Brianza, in Northern Italy, while Italian monolingual schools were enrolled in Milan, in the Province of Monza-Brianza and Rome. In one of the bilingual schools, for first graders 75% of the schooling time consists of English immersion, while the remaining 25% is taught in Italian; subsequently, from Grade 2 to Grade 5, students attend a simultaneous immersion program, with the same amount of exposure to English and Italian, that is 50:50. In the other bilingual school, children attend a 50:50 immersion Italian-English program from Grade 1. Thus, they are exposed to English for half of the time and to Italian for the other half. In both schools, subjects are taught in both languages. 50:50 immersion programs were privileged in the study as they seem to deliver the most effective literacy outcomes.

Participants, both bilingual and monolingual, came from Italian families with an estimated medium-high SES. Due to the failure in getting all of the questionnaires filled out by the parents,

this limitation was overcome through an accurate approximation. SES was thus evaluated by considering and combining three variables: (1) family income, which was linked to the possession of financial means to enroll a child in a private school (bilingual group); (2) geographical living area: enlisted primary schools, both bilingual and monolingual, are located in areas where the standards of living are known to be high, such as central districts of Milan and Rome, and the well-to-do area of Monza-Brianza Province; (3) level of education, measured with a short questionnaire to be filled out by parents (on a scale from 1 to 5, where five corresponds to a university degree or a higher level of education). Only children whose parents had both ticked at least a level of education corresponding to 4 (*i.e.,* high school diploma) in the questionnaire – when available – were included in the final sample. Due to these circumstances, to include only families with an estimated medium-high SES was the best attempt to ensure an adequate level of homogeneity between groups. Thus, it should have been found differences between the two groups, these could not be linked to their belonging to different socio-economic-status. SES homogeneity is an essential point to be considered because some studies investigating differences in performance between monolingual and bilingual groups have failed to adequately control for ethnicity and socioeconomic background (for instance studies comparing Canadian English monolinguals and Canadian immigrant bilinguals as Bialystok, 1986, Bialystok and Senman, 2004, and Bialystok and Shapero, 2005). In their study, Morton and Harper (2007) investigated performances in a cognitive control task of bilingual and monolingual children of identical ethnic and socioeconomic backgrounds and found the two groups performing identically.

Inclusion criteria for bilingual participants were to have Italian as their mother tongue and to have been exposed to English within the first three years of their life, which means at least since formal kindergarten in Italy (in this case, they had to have attended an Italian-English bilingual kindergarten). Considering the limit of three as the age of first exposure to English, bilingual students could be all considered as early bilinguals, according to Genesee, Paradis, and Cargo (2004). As has been previously argued, early bilinguals seem to reach better literacy outcomes in L2

than late bilinguals, which are usually exposed after three to four years of age (Bonifacci & Tobia, 2016; Kovelman et al., 2008). To summarize: only early bilingual pupils were included in the study; children – both bilingual and monolingual – were selected from families with a medium-high SES to ensure the right level of homogeneity within and between the experimental and control groups.

Both bilinguals and monolinguals diagnosed with a Developmental Language Disorder or a Learning Disability, or with any special needs were excluded from the sample. At the moment of data collection, children who were suspected of having a Learning Disability or were undergoing a diagnostic assessment were excluded. The final sample comprised a total of 192 children.

Before starting testing, directors of school were deeply informed about the study goals and procedure. With their collaboration, teachers and families who wished to participate had been made aware of the study project. An informative form was also provided. Parents were asked to sign an informed consent form in order to authorize their children's participation. Thus, a parent's written consent was obtained for every child to be tested.

The bilingual group consisted of 121 children, attending Grade 1 ($N$ = 48), Grade 3 ($N$ = 43) and Grade 5 ($N$ = 30), whereas the Italian monolingual group included 71 children attending Grade 1 ($N$ = 21), Grade 3 ($N$ = 34) and Grade 5 (N = 16). Considering Grades, the final sample comprised in total 69 first graders, 77 third graders, and 46 fifth graders. As could be noted, the sample is slightly unbalanced in the three classes. Unfortunately, it was impossible to get the designed control subgroup of monolingual students attending Grade 5, partially because of the difficulties encountered in arranging data collection with schools, partially because of the sanitary emergency due to the Covid-19 epidemic.

First graders mean age was 83.8 months ($SD$ = 4.7), third graders mean age was 106.4 months ($SD$ = 4.8), and fifth-graders mean age was 128.6 ($SD$ = 4.9), calculated on the total number of subjects. The mean age of the monolingual group was 101.7 ($SD$ = 19), and the mean age of the bilingual children was 106.8 ($SD$ = 15.5), calculated in months. The mean ages of the two groups did not differ ($t(190) = 1.9$, $p = 0.07$. However, looking at individual classes, the bilingual and

monolingual subgroups in Grades 1 and 3 differed significantly according to their age. Monolingual first-graders ($M$ = 87.3, $SD$ = 4.1) were observed to be significantly older (t(67) = 4.74, $p < 0.001$) than bilingual first-graders ($M$ = 82.3, $SD$ = 4) and monolingual third-graders ($M$ = 108.8, $SD$ = 4.7) were found to be significantly older (t(75) = 4.13, $p < 0.001$) than bilingual third-graders ($M$ = 104.6, $SD$ = 4.1). As we are going to explain in detail in the procedure, they were tested in different periods, due to organizational reasons.

The sample was balanced for sex, with 105 participants (54.4% of the sample) being female and 88 being male (45.6%). More specific details regarding the subgroups can be found in *Table 1*.

| **Grades** | **Bilinguals** ($N$ = 121) | **Monolinguals** ($N$ = 71) |
|---|---|---|
| **Grade 1** | $N$ = 48 | $N$ = 21 |
| Mean age in months | $M$ = 82.3 ($SD$ = 4) | $M$ = 87.3 ($SD$ = 4.1) |
| Females | $N$ = 23 | $N$ = 12 |
| Males | $N$ = 25 | $N$ = 9 |
| **Grade 3** | $N$ = 43 | $N$ = 34 |
| Mean age in months | $M$ = 104.6 ($SD$ = 4.1) | $M$ = 108.8 ($SD$ = 4.7) |
| Females | $N$ = 25 | $N$ = 19 |
| Males | $N$ = 18 | $N$ = 15 |
| **Grade 5** | $N$ = 30 | $N$ = 16 |
| Mean age in months | $M$ = 128.8 ($SD$ = 5.3) | $M$ = 128.2 ($SD$ = 4.1) |
| Females | $N$ = 18 | $N$ = 8 |
| Males | $N$ = 13 | $N$ = 8 |

**Table 1** shows the number of subjects, the average ages in months, and the number of females and males of the following subgroups: bilingual first-graders, bilingual third-graders, bilingual fifth-graders, monolingual first-graders, monolingual third-graders, and monolingual fifth-graders.

## Materials

Children were tested on a battery of tests, which included indexes of reading proficiency (accuracy, speed, and passage comprehension) and some control measures assessing language abilities and working memory. We measure reading by reading a text. Different competencies are required to read a text: decoding skills, vocabulary knowledge, morpho-syntactic competence, as expressed in sentences, short term memory (to keep in mind the content of the text), and inferencing skills (to connect different sentences of the text). As mentioned above, many authors have investigated the

association between oral language and literacy skills. More specifically, Duff, Reen, Plunket & Nation (2015) found significant relationships between vocabulary knowledge and reading skills, supporting the theory that vocabulary is a cognitive foundation of both reading accuracy and reading comprehension. Siegel (2008) examined the relation of morphological awareness to reading skills of children with dyslexia, typical readers, and English language learner children (ELLs). She found morphological competence related to reading ability, as typically achieving readers were significantly better than children with dyslexia on some morphological awareness tasks. No differences were reported between the ELL and the English typical development students. Recently, Robertson & Deacon (2019) found that morphological awareness appears to have a significant role in word-level reading across the early to middle elementary school grades. Additionally, Chik, Ho, Yeung, Wong, Chan, Chung & Lo (2012) found that senior primary school children's reading ability was largely accounted for by the text-level reading-related skill of morpho-syntax, demonstrating that older children pay more attention to morpho-syntactic features, especially in reading comprehension. Finally, as mentioned above, Kovelman et al. (2008) proposed that differences in children's literacy outcomes depends on early language exposure in bilingual children. In line with this, the expressive vocabulary was identified as an index of lexical proficiency and sentence repetition as a measure of morpho-syntactic competence.

As discussed above, working memory ability has been found to predict reading skills (White, 2019). The battery included both a pseudoword repetition task and a digit span task to test working memory. A pseudoword repetition task was included because it is considered a process measure.[1] That means a less language-dependent and more memory-reliant measure, making it a less biased method of evaluating phonological working memory (*e.g.*, Henry & Botting, 2016; Kohnert, 2010). Moreover, memory recall tasks are frequently used as a control to ensure that different language groups are cognitively similar (Schmidt *et al.*, 2019).

[1] It should be pointed out that the pseudoword task is not a pure measure of memory, as it is dependent on the phonology of a given language and respects the phonotactic constraints of the language in questions (Chiat, 2015).

Italian tasks were presented to both groups of participants, while English tasks were presented to bilinguals only.

***Italian tasks***

*Prove MT-3-Clinica: La valutazione delle abilità di Lettura e Comprensione per la scuola primaria e secondaria di I grado* (Cornoldi & Carretti, 2016) evaluates the reading skills of children aged between 6 and 14 years. It is a standardized test that assesses reading rate (*i.e.*, velocity, measured in syllables/seconds), accuracy (*i.e.*, number of errors), and comprehension with two separate passages. Children are asked to read a passage aloud to assess velocity and accuracy. Reading comprehension instead consists of the silent reading of another passage, followed by a multiple-choice questionnaire about the story. Children are allowed to go back to the passage whenever they need to individuate the correct answer. Both passages differ according to the grade the child is attending.

*Test Neuropsicologico Lessicale per l'età evolutiva* (TNL; Cossu, 2013) is designed to measure expressive vocabulary in children between 3 and 9 years (although it should be noted that it was administered to older subjects as well). The naming task, which is part of the expressive vocabulary subtest, was administered to participants. Children are presented with 50 panels (A4 size) representing pictures they have to name aloud. The test examines the nouns category only. Nouns range from low to high frequency and are randomly ordered. Scores range between 0 and 50.

*LITMUS SRep task* – Italian version (Marinis & Armon-Lotem, 2015 – Italian adaptation by Levorato – not published) is a non-standardized school-age sentence repetition task that evaluates the grammatical competence in Italian. The task has been developed within the European project Cost Action IS0804 *'Language Impairment in a Multilingual Society*,' whose aim was to create a range of parallel tasks across a large number of languages to provide comparable results when assessing two or more languages. The task consists of 48 sentences pre-recorded by native Italian speakers and incorporated into a PowerPoint presentation for administration. The sentences include

simple and complex syntactic structures, both language independent (complex across languages), and language-specific (complex in the specific language), and are presented with increasing difficulty. The child listens to the sentences via headphones and then has to recall and repeat them aloud. Children are explicitly invited to say what they heard precisely. It was used the 0-1 score scheme: a score of 1 was allocated if the sentence was repeated entirely verbatim while a score of 0 if the sentence was repeated with one or more changes. Scores range between 0 and 48.

*Pseudoword Repetition Task* (Ciccarelli, 1998) measures phonological working memory. Children are asked to repeat a list of 40 pseudowords structured of two, three, four, and five syllables, which are presented randomly in the list. The experimenter pronounces the pseudowords to be repeated by the child. Scores range from 0 to 40. The test is non-standardized and was created as part of a doctoral research project.

*Digit Span Tasks* – Italian version (non-standardized test created by Tsimpli et al., University of Cambridge, PowerPoint implemented by Costa) measures working memory skills using digits recall paradigm. Participants were presented with the Forward and Backward Digit Span tests. Tasks consist of the recall of increasing sequences of numeric strings the child hears. Sequences of numbers were audio-recorded by a native speaker and presented to the child, one at a time, throughout headphones. Each Digit Span task is divided into groups of four sequences of numbers. Participants have to correctly repeat at least three of the sequences in each set of numeric sequences to move on to the next series; the task is interrupted when the participant makes more than one mistake within the same series. Possible scores on the Forward and Backward Digit Span tasks range from 0 to 32.

***English tasks***

*York Assessment of Reading Comprehension Passage Reading* (YARC; Snowling et al., 2009) assesses reading rate (measured in seconds), accuracy, and comprehension of children aged between 5 and 11 years. Unlike the Italian reading task, this standardized test assesses reading

decoding and comprehension in the same passage. Firstly, children are asked to read the passage aloud to assess the reading rate and accuracy; secondly, they have to answer orally to the examiner's open-ended comprehension questions. Participants are invited to read the passage another time silently before starting comprehension questions, but they can also go back to the passage if needed. The test provides passages with seven levels of reading complexity (range from Beginners to Level 6); the right passage for each child is determined based on the performance on a word reading task, which is administered before the passage: the higher the number of words reads correctly, the higher the reading level.

*Word Finding Vocabulary Test: The Renfrew Language Scales* (Renfrew & Mitchell, 2010) assesses expressive vocabulary in children aged between 3 and 8 years (although it should be noted that it was administered to older subjects as well), with possible scores between 0 and 50. Children are shown 50 panels on a screen, representing pictures they have to name aloud, one at a time. Pictures represent both low and high-frequency nouns and are randomly ordered.

*LITMUS SRep task* – English version (Marinis, Chiat, Armon-Lotem, Gibbons & Gipps, 2010) evaluates the grammatical competence in English. The task is structured similarly to the Italian version of the test, with pre-recorded sentences presented into a PowerPoint presentation to be repeated aloud by the child. The English sentence repetition task has 60 sentences targeting 15 syntactic structures of increasing complexity. The same 0-1 score scheme was used. Scores range between 0 and 60.

*Children's Test of Nonword Repetition* (CN REP; Gathercole & Baddeley, 1996) measures the phonological short-term memory of children aged between 4 and 8 years (although it should be noted that we used it to assess older children as well). Participants are asked to repeat a list of 40 pseudowords of two, three, four, and five syllables. Pseudowords of different phonological complexity follow a random order. Recorded items were played and presented to the children through headphones; the child had to repeat the pseudoword immediately after hearing it. Scores range from 0 to 40.

*Digit Span Tasks* – English version (created by Tsimpli et al., University of Cambridge, PowerPoint implemented by Costa) is structured exactly like the Italian version of the test.

**Procedure**

Data collection has been started in April 2017 and finished in February 2020. Due to the Covid-19 epidemic alert in Italy and the subsequent closure of all schools of all grades, monolingual group data collection could not be completed as planned.

Participants were tested individually on all tasks in a quiet room made available at school. One exception was passage reading comprehension, which was administered during school in small groups of two-to-three children, again in a quiet room during school hours. Children were outdistanced enough not to interfere with each other during the task. When it was not possible to administer the task collectively, it was administered individually. As mentioned above, bilinguals and monolinguals were tested at different moments. Bilinguals were tested firstly. They were tested in Italian and English on separate days during the final part of the school year (springtime, between April and June). Monolinguals were tested only in Italian. Children from Italian monolingual schools in Rome and Milan (5 children attending Grade 5) were approximately tested during the same period as bilingual; Monolinguals from school in Monza-Brianza, which substantially covered Grade 1 and Grade 3 Italian monolingual control sample, were tested at the beginning of the following school year (September). Thus, at the time of testing, these monolingual first graders had just begun their second year of elementary school, and monolingual third graders had just started their fourth year. This explains the significant age differences between the monolingual and bilingual groups in Grade 1 and Grade 3, as shown above. This different testing period will be taken into account in the statistical analysis.

The original design of the study project was planned to include an English monolingual control group also, in order to obtain a complete comparison between performances of bilinguals and monolingual counterparts. Unluckily, because of the difficulty encountered in finding a

monolingual English school in an English-speaking country that was willing to participate in the study, it was not possible.

Although a proficient English person administered English tests, English instructions and language items (*e.g.*, pseudowords, sentences, or digits) were recorded by a native English speaker and presented to the participants as audio recordings through headphones.

Instructions were repeated as many times as necessary to ensure that children understood the task.

**Statistical analyses**

To compare the performance on Italian tasks of the bilingual and the Italian monolingual groups in Grade 1, 3, and 5, an analysis of variance was performed, with *Group* (bilingual and monolingual) and *Grade* (first, third and fifth) as independent variables. Scores at a particular test were fixed as a dependent variable, and testing time (calculated in terms of the number of days from the beginning of the year) as a covariate. By using the testing time as a covariate, the different ages between bilingual and monolingual first- and third-graders were controlled. These bilingual and monolingual subgroups were, in fact, tested at different times in the school years. Significant main effects and interactions will be reported. Bonferroni *post hocs* were used to make post hoc comparisons. An analysis of variance was also used to compare the three Grades' English tasks in the bilingual group to explore English development across the three grades.

Correlations were used to verify whether the bilingual group's performance on Italian tasks was related to that on English tasks.

**Results**

The reading comprehension scores of 9 bilingual first-graders were missing because of the limited time available at the end of the school year; thus, it was not possible to administer the task individually or collectively. Missing data were replaced with the average score of the bilingual first-grade target group.

***Comparison between bilinguals and monolinguals on the Italian tasks***

To answer the first research question, whether an Italian-English *simultaneous immersive* education has an impact on the acquisition of Italian (L1) and the literacy development in Italian, we examined the Italian performance of the bilingual group, compared to that of the monolinguals'.

First, we analyzed measures assessing oral language abilities. As evident in *Figure 1,* expressive vocabulary was found to improve across the three Grades, both in monolingual and bilingual children.

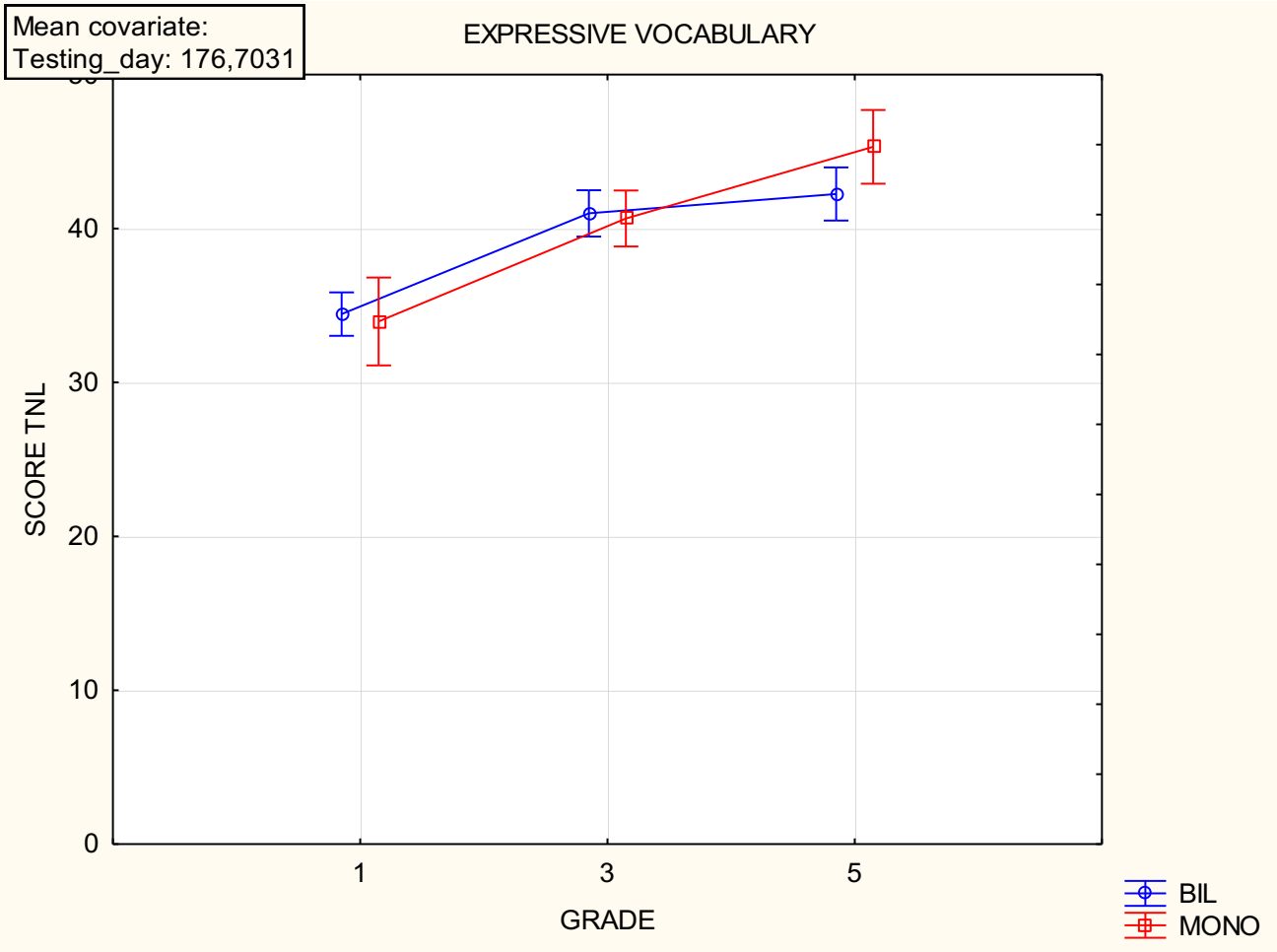



**Figure 1** shows raw scores (Max=50) at Italian expressive vocabulary (TNL) in Grade 1 and 5 of monolingual and bilingual children. Bars indicate a confidence interval at 0.95.

An ANOVA with Group (Monolingual and Bilingual) and Grades (1,3,5) as the independent variables, raw scores at the *Test Neuropsicologico Lessicale per l'età evolutiva* (TNL) as the dependent variable, and testing time as a covariate confirmed this observation: only an effect of

Grade was found F(2,185) = 40.5, $\eta^2 = 0.30$, $p < 0.001$.[2] Bonferroni *post hoc* confirmed that all three Grades differed from each other ($p < 0.05$). No significant Group difference was found.

As shown in *Figure 2*, performances at sentence repetition task improves from first to third Grade, in both Groups of children. In third Grade, we observe a ceiling effect (Total number of responses being = 48).

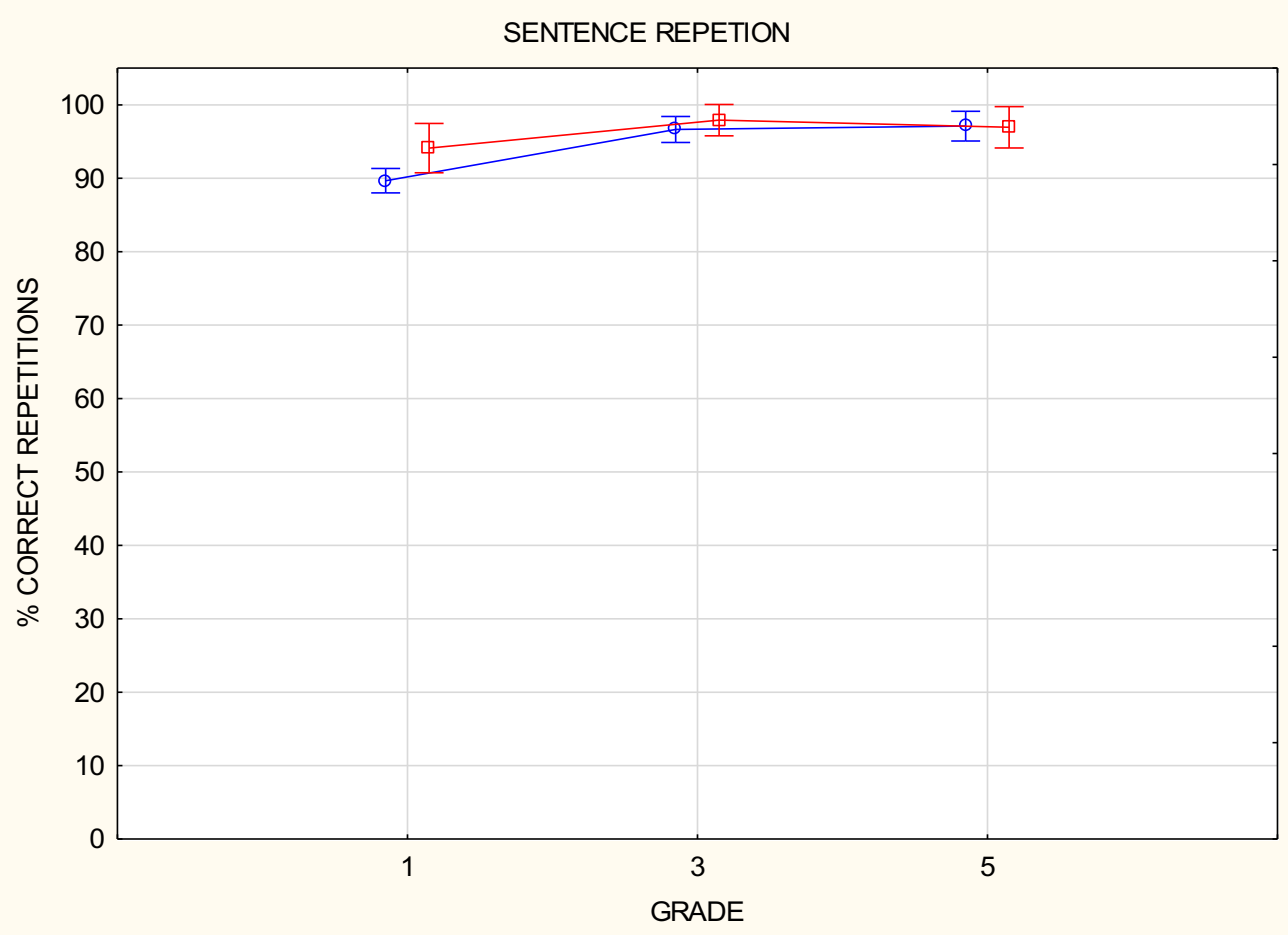



**Figure 2** shows scores (max=48) at Italian sentence repetition task in Grade 1, 3, and 5 of monolingual and bilingual children. Bars indicate a confidence interval at 0.95.

As expected, an ANOVA with Group and Grades as the independent variables, scores at the *LITMUS SRep task* – Italian version as the dependent variable, and testing time as a covariate, found only an effect of Grade F(2,185) = 14.6, $\eta^2 = 0.14$, $p < 0.005$. Besides, Bonferroni *post hoc*

[2] In our analyses, we report the Eta-square ($\eta^2$), which is a measure of effect size for use in ANOVA. It ranges between 0 and 1. Effect sizes follow this interpretation: .01 ~ small; .06 ~ medium; >.14 ~ large.

showed that first-graders differed from third- and fifth-graders ($p < 0.001$).

To summarize, mixed-model ANOVAs revealed an effect of Grade on language proficiency, indicating positive gains in both Groups at lexical and morpho-syntactic levels.

Then, we moved to analyze passage reading proficiency measures. As shown in *Figure 3*, reading speed, which in *Prove MT-3-Clinica* is measured in syllables/seconds, similarly increased across Grades in the two Groups.

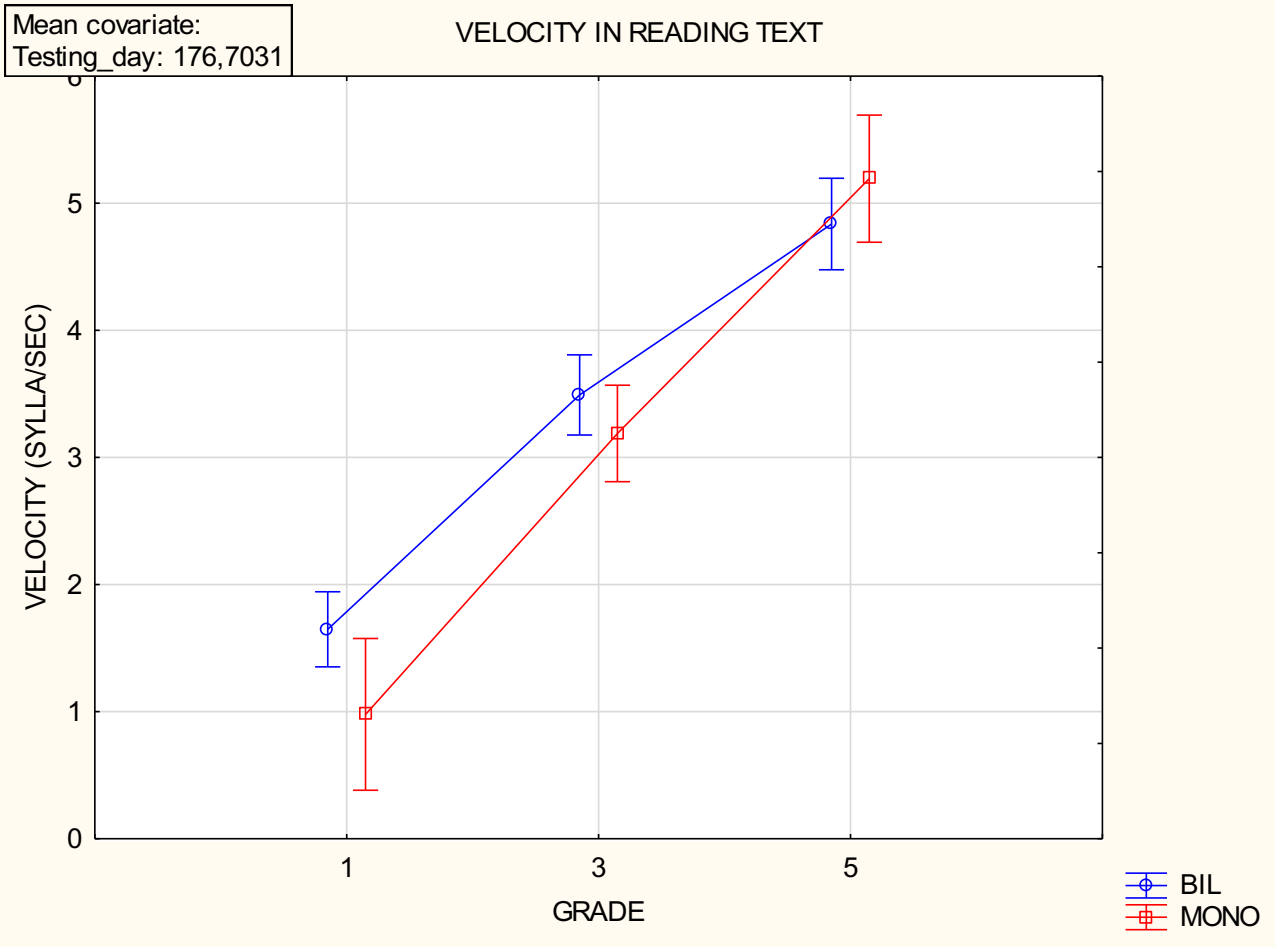



**Figure 3** shows reading velocity (syllable/seconds) of a text (MT) in Italian in Grade 1, 3, and 5 monolingual and bilingual children. Bars indicate a confidence interval at 0.95.

Statistical analysis confirmed this observation: an effect of Grade was found $F(2,185) = 121.3$, $\eta^2 = 0.56$, $p < 0.001$. Bonferroni *post hoc* revealed that all three grades differed from each other ($p < 0.001$). Furthermore, the testing time was significant $F(1,185) = 16.43$, $\eta^2 = 0.08$, $p < 0.001$, although this effect is low, as indicated by the $\eta^2$. No significant difference was found in the

performance of monolingual and bilingual Groups.

In *Figure 4*, we can observe scores at reading accuracy expressed in terms of the number of errors. As can be seen, bilingual children made slightly more errors than the monolinguals, especially in Grade 1. Also, there is a definite improvement in performance from Grade 1 to Grade 3 in both Groups.

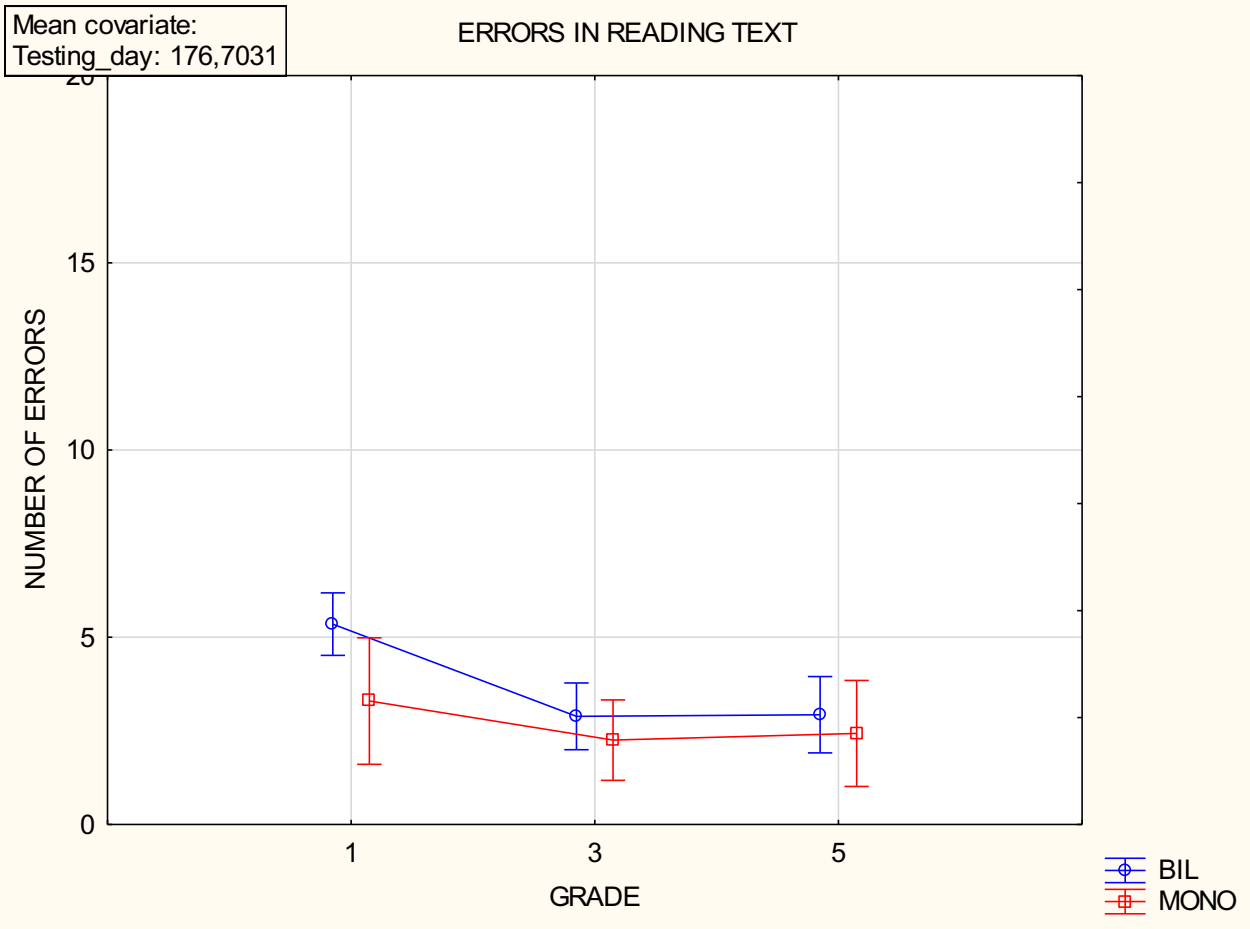



**Figure 4** shows errors at reading a text (MT) in Italian in Grade 1, 3, and 5 monolingual and bilingual children. Bars indicate a confidence interval at 0.95.

The statistical analysis revealed only an effect of Grade $F(2,185) = 6.06$, $\eta^2 = 0.06$, $p < 0.005$. Bonferroni *post hoc* showed that first-graders differed from third- and fifth-graders ($p < 0.01$). Differences in the performance of monolingual and bilingual Groups did not turn out to be significant.

Finally, we analyzed passage reading comprehension. *Figure 5* illustrates the performance of the participants.

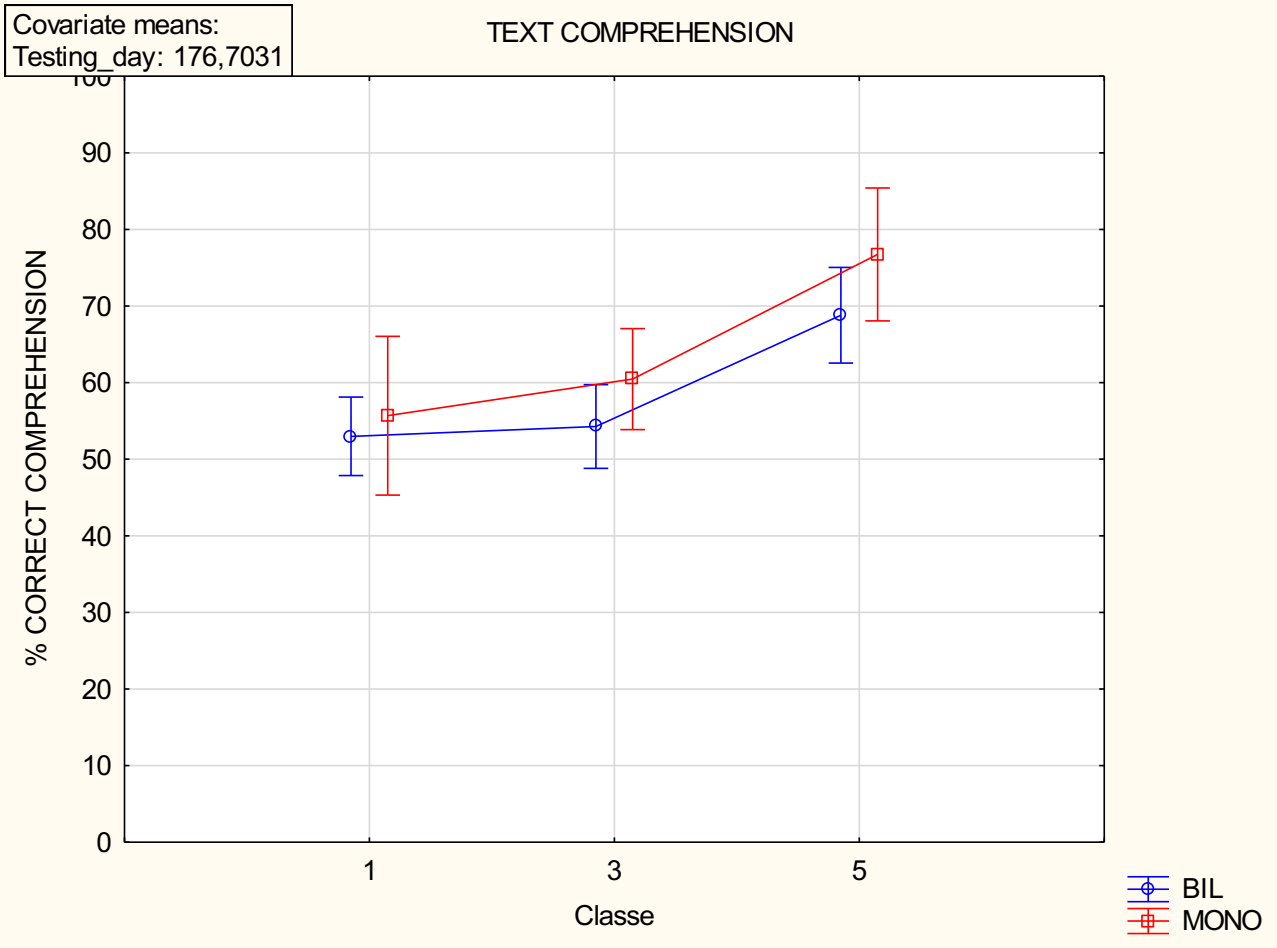



**Figure 5** shows raw scores at the reading comprehension (MT) in Italian in Grade 1, 3, and 5, of monolingual and bilingual children. Bars indicate a confidence interval at 0.95.

The ANOVA revealed an effect of Grade $F(2,185) = 11.36$, $\eta^2 = 0.10$, $p < 0.001$. Bonferroni *post hoc* showed that first- and third-graders differed from fifth-graders ($p < 0.001$).

Lastly, we analyzed working memory tasks. *Figure 6* reports the results in the pseudoword repetition task.

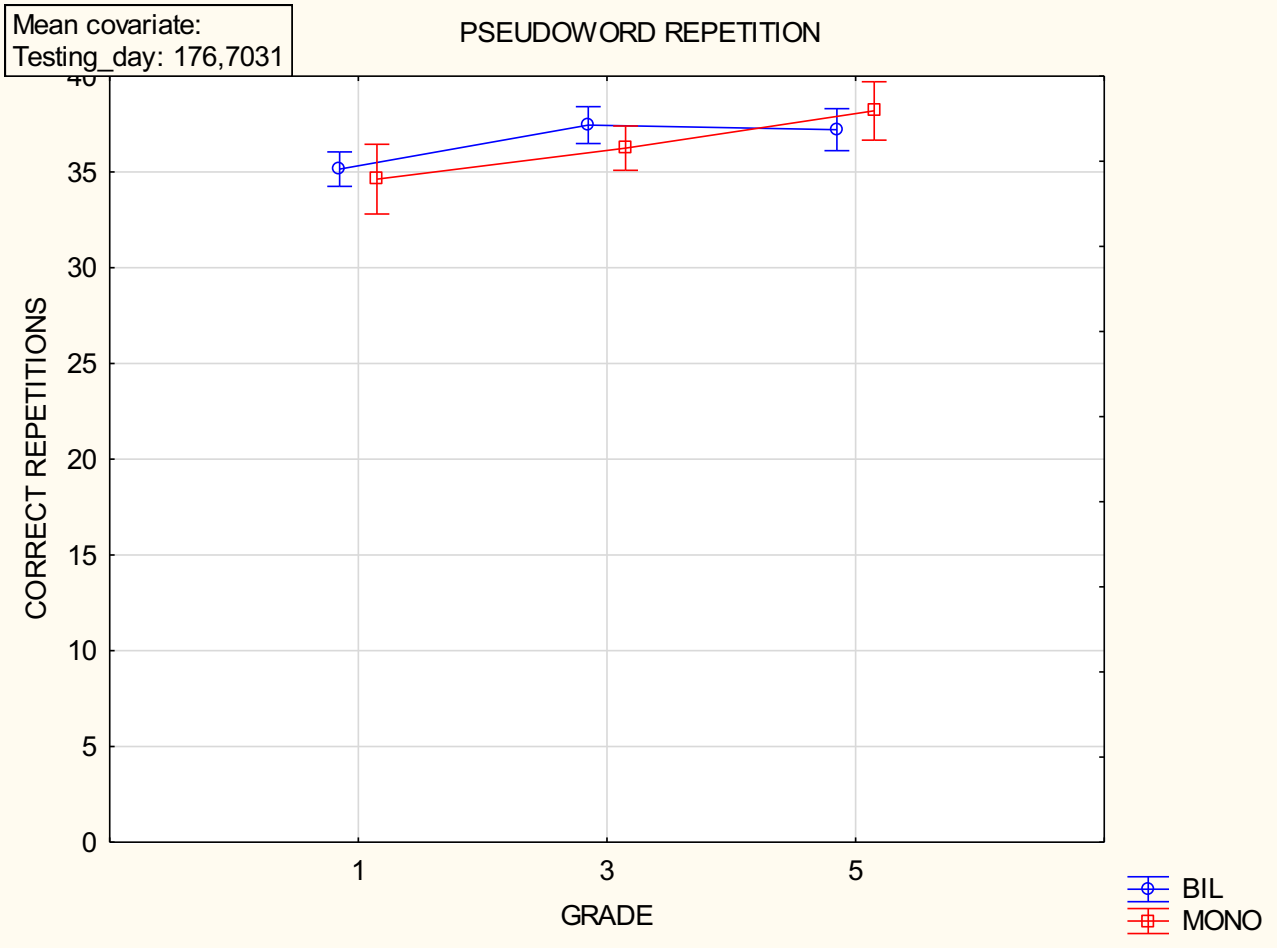



**Figure 6** shows raw scores (Max=40) at the pseudoword repetition task in Italian in Grade 1, 3, and 5, of monolingual and bilingual children. Bars indicate a confidence interval at 0.95.

The ANOVA showed an effect of Grade on *Pseudoword Repetition Task* $F(2,185) = 8.67$, $\eta^2 = 0.09$, $p < 0.001$ and Bonferroni *post hoc* showed that first-graders differed from third- and fifth-graders ($p < 0.01$). The day of testing resulted to be significant in pseudoword repetition task $F(1,185) = 4.62$, $\eta^2 = 0.03$, $p < 0.05$, although this effect was small. Group difference was not found to be significant.

On *Forward Digit Span Task* we only found an effect of Grade $F(2,185) = 12.9$, $\eta^2 = 0.12$, $p < 0.001$ and Bonferroni *post hoc* showed that first- and third-graders differed from fifth-graders ($p < 0.001$). *Figure 7* illustrates the performance of the participants.

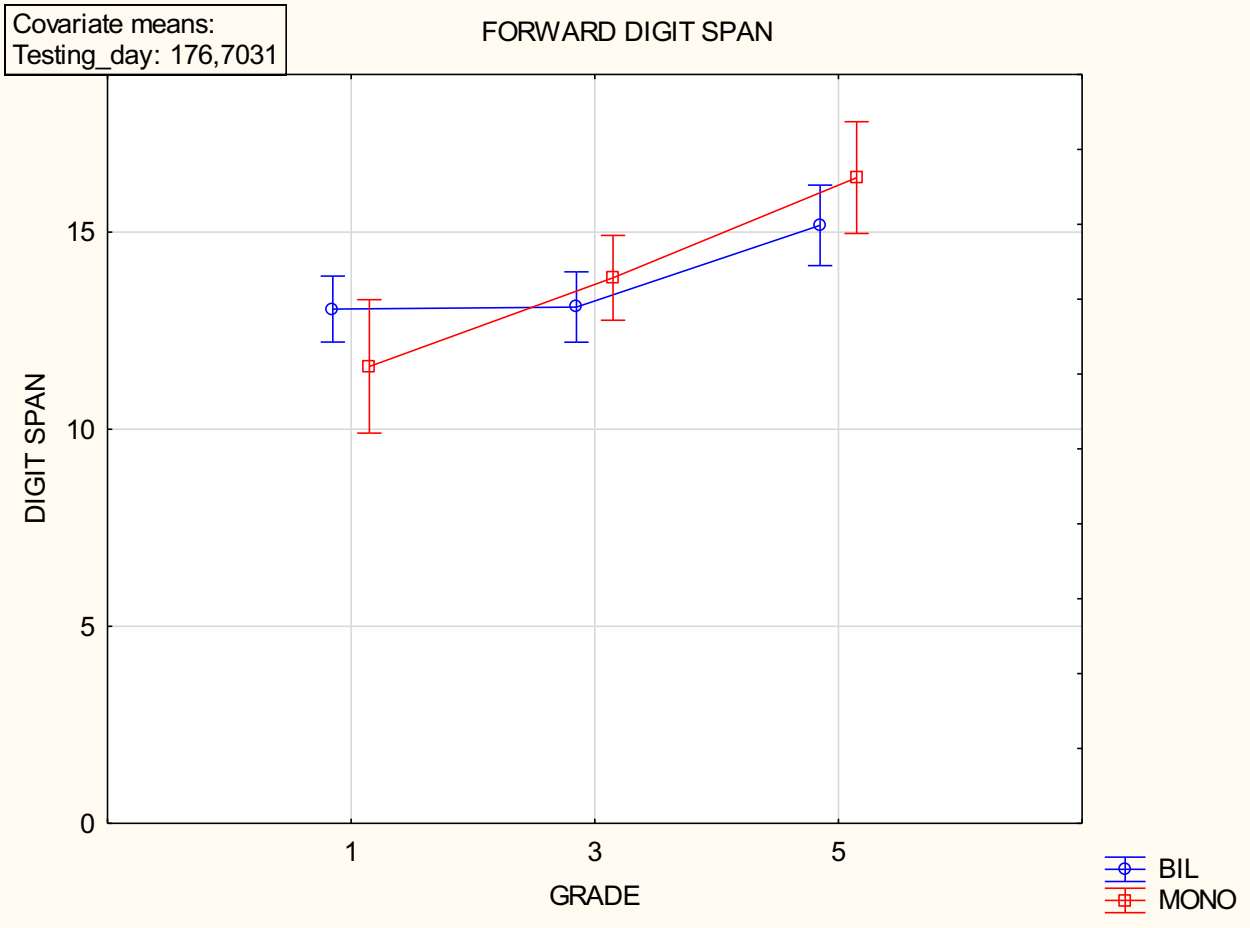



**Figure 7** shows scores at the forward digit span task in Italian in Grade 1, 3, and 5, of monolingual and bilingual children. Bars indicate a confidence interval at 0.95.

As shown in *Figure 8*, performances at *Backward Digit Span Task* seem to improve from first to third Grade, in both Groups of children. Moreover, in fifth Grade, bilinguals seem to improve their performance further, outperforming monolingual peers. Conversely, monolinguals reverse their trend in Grade 5.

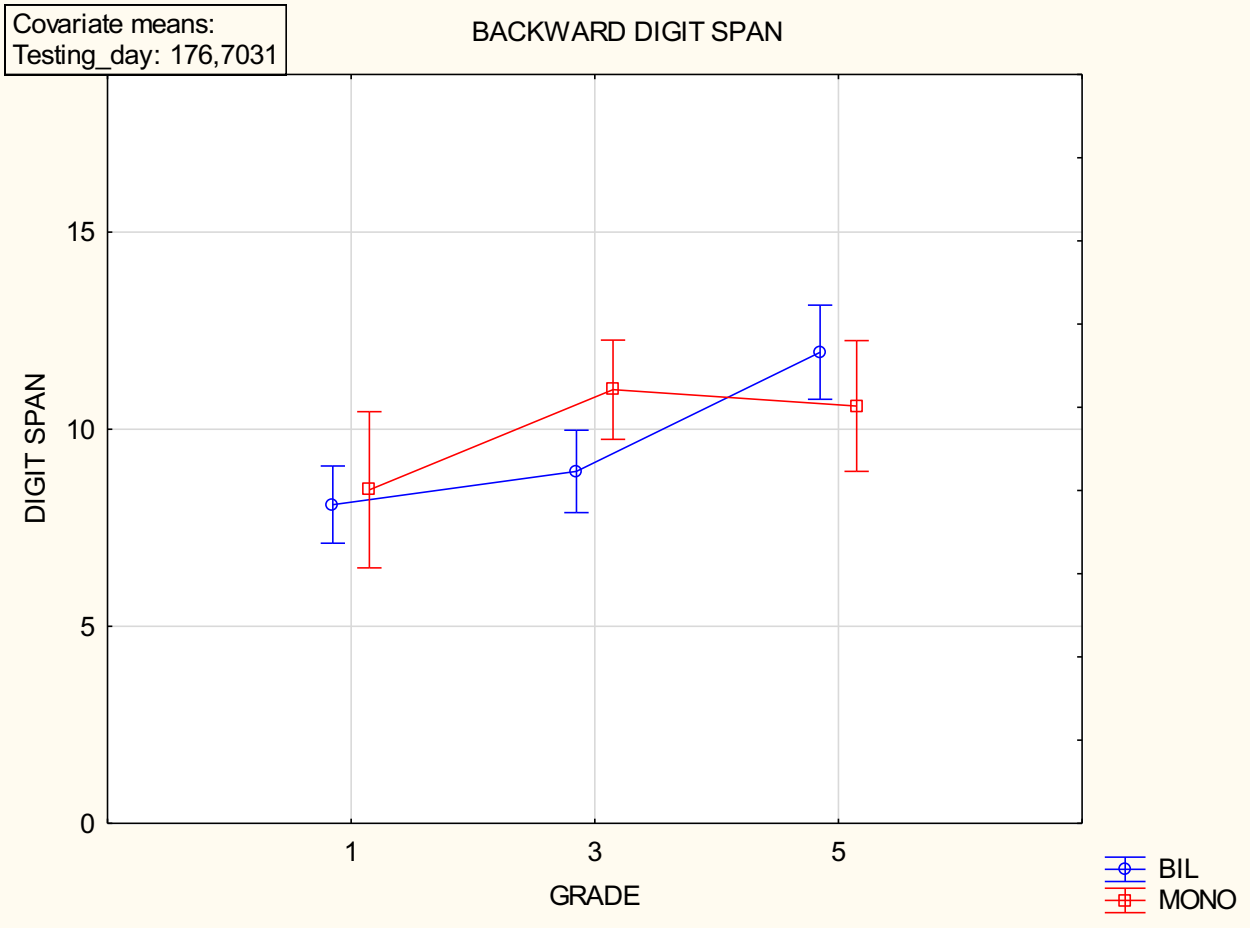



**Figure 8** shows scores at the backward digit span task in Italian in Grade 1, 3, and 5, of monolingual and bilingual children. Bars indicate a confidence interval at 0.95.

In the statistical analysis we found an effect of Grade $F(2,185) = 7.28$, $\eta^2 = 0.07$, $p < 0.01$. Bonferroni *post hoc* confirmed that all three Grades differed from each other ($p < 0.005$). No significant Group difference was found. An effect of Interaction was also found, $F(2,185) = 3.56$, $\eta^2 = 0.04$, $p < 0.05$, arguably because the performance of the bilinguals in Grade 1 is slightly lower than that of monolingual children, while the role between the two Groups inverted in Grade 5. Nevertheless, the difference between the two groups was not significant.

In conclusion, robust effects of Grade were found in all of the Italian tasks evaluated. When controlling for the time of testing, which was different for monolingual children in grade 1 and 3, we did not find Group differences on any measure.

***English language and reading development within the bilingual group***

To observe the impact of an Italian-English *simultaneous immersion program* on the development of language and literacy skills in English (L2), we examined the English performance of the bilingual group across Grade 1, 3, and 5.

First, we analyzed English expressive vocabulary and the sentence repetition task in English. Only 23% of bilingual children had English vocabulary scores within the monolingual normal limits, but this is a generally common bilingual feature (*e.g.*, Bialystok, Luk, Peets & Yang, 2010; Hoff, Rumiche, Burridge, Ribot & Welsh, 2014). We must point out that in Grade 5, standard scores were just indicative, as the test is standardized for children up to 8.5 years old.

As shown in *Figures 9 and 10*, both skills were found to improve across Grades, although to a greater extent from Grade 1 to 3.

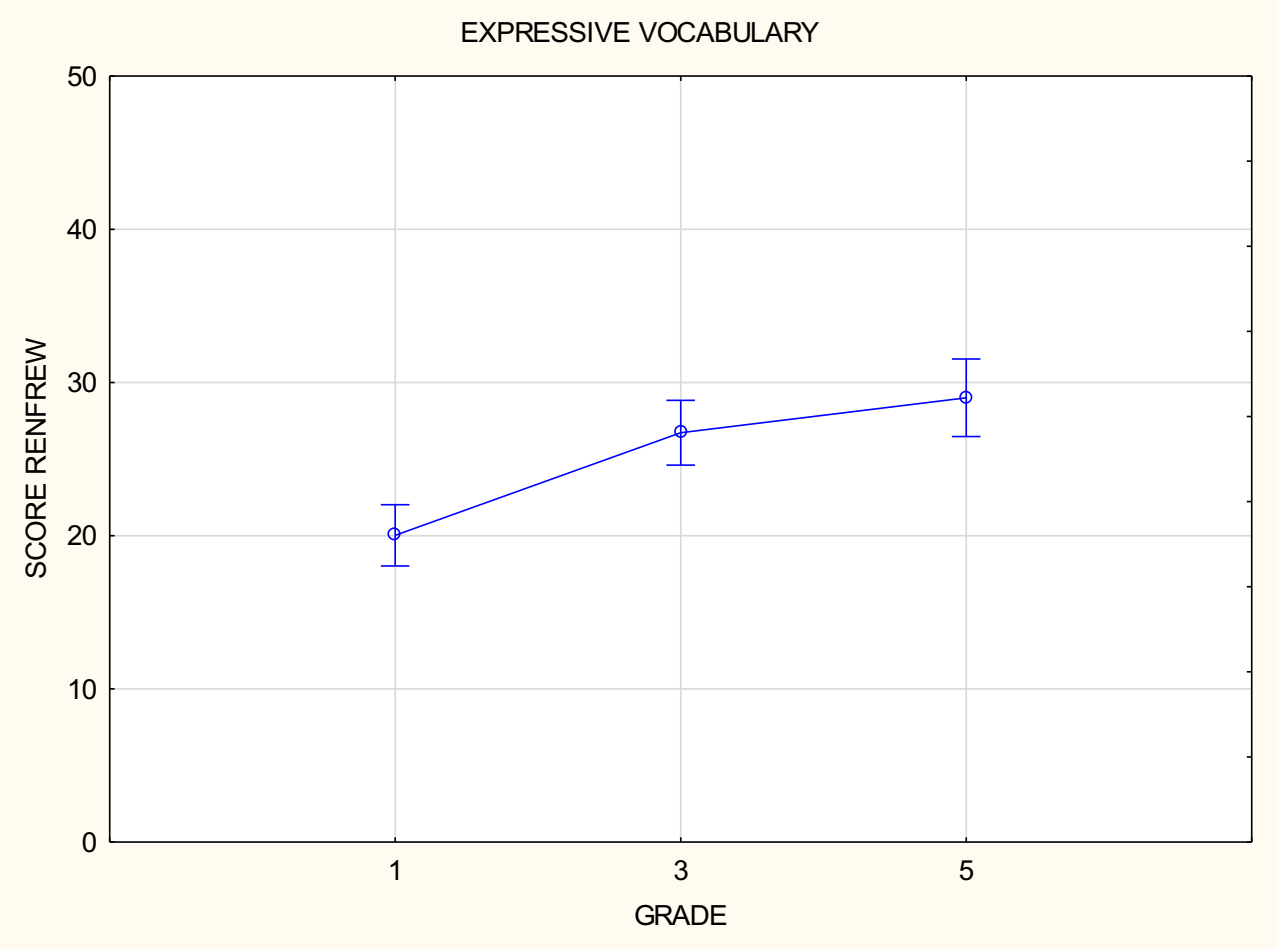



**Figure 9** shows raw scores (Max=50) at English expressive vocabulary (Renfrew) in Grade 1, 3, and 5 of bilingual children. Bars indicate a confidence interval at 0.95.

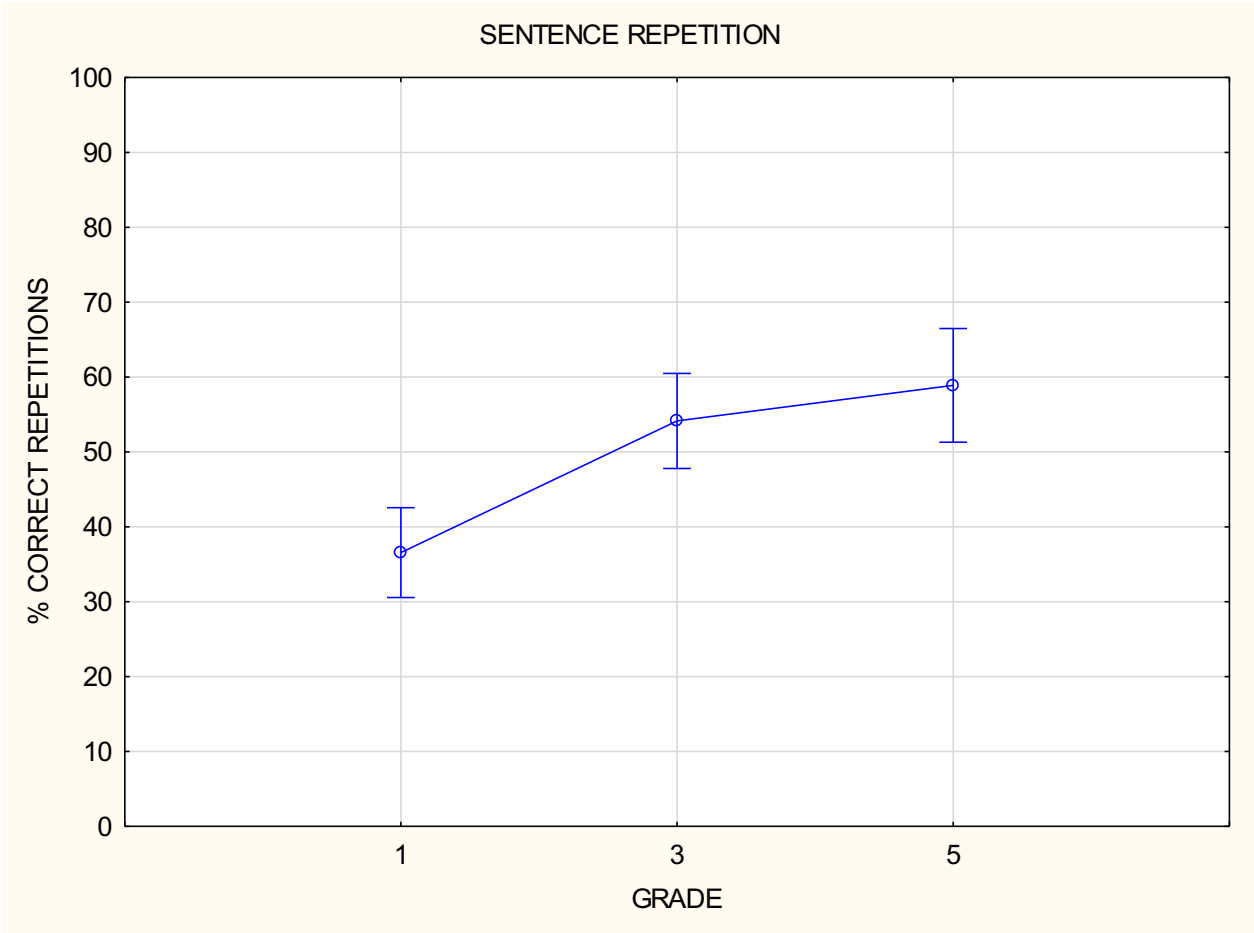



**Figure 10** shows scores (Max=60) at English sentence repetition task in Grade 1, 3, and 5 of bilingual children. Bars indicate a confidence interval at 0.95.

An ANOVA with Grades (1,3,5) as the independent variables and raw scores on the *Word Finding Vocabulary Test* as the dependent variable, showed an effect of Grade F (2,118) = 18.2, $\eta^2 = 0.24$, $p < 0.001$. The Bonferroni *post hoc* test confirmed that bilingual first-graders differed from third- and fifth-graders ($p < 0.001$). An effect of Grade was also found at the *LITMUS SRep task* – English version F(2,118) = 13, $\eta^2 = 0.18$, $p < 0.001$. Bonferroni *post hoc* showed that first-graders differed from third- and fifth-graders ($p < 0.001$).

Secondly, we analyzed reading speed and accuracy on the *York Assessment of Reading Comprehension Passage Reading* (YARC). In *Figures 11 and 12*, we can observe a similar trend across Grades for both measures. Differently from Italian MT-3 test for reading (which measures velocity in syllables/seconds), in YARC reading speed is the total time in seconds that the child needed to read the passage.

Bilingual children in Grade 3 took more time to read the passage than bilingual first-graders, but the performance of bilingual children resulted to definitely improve in Grade 5.

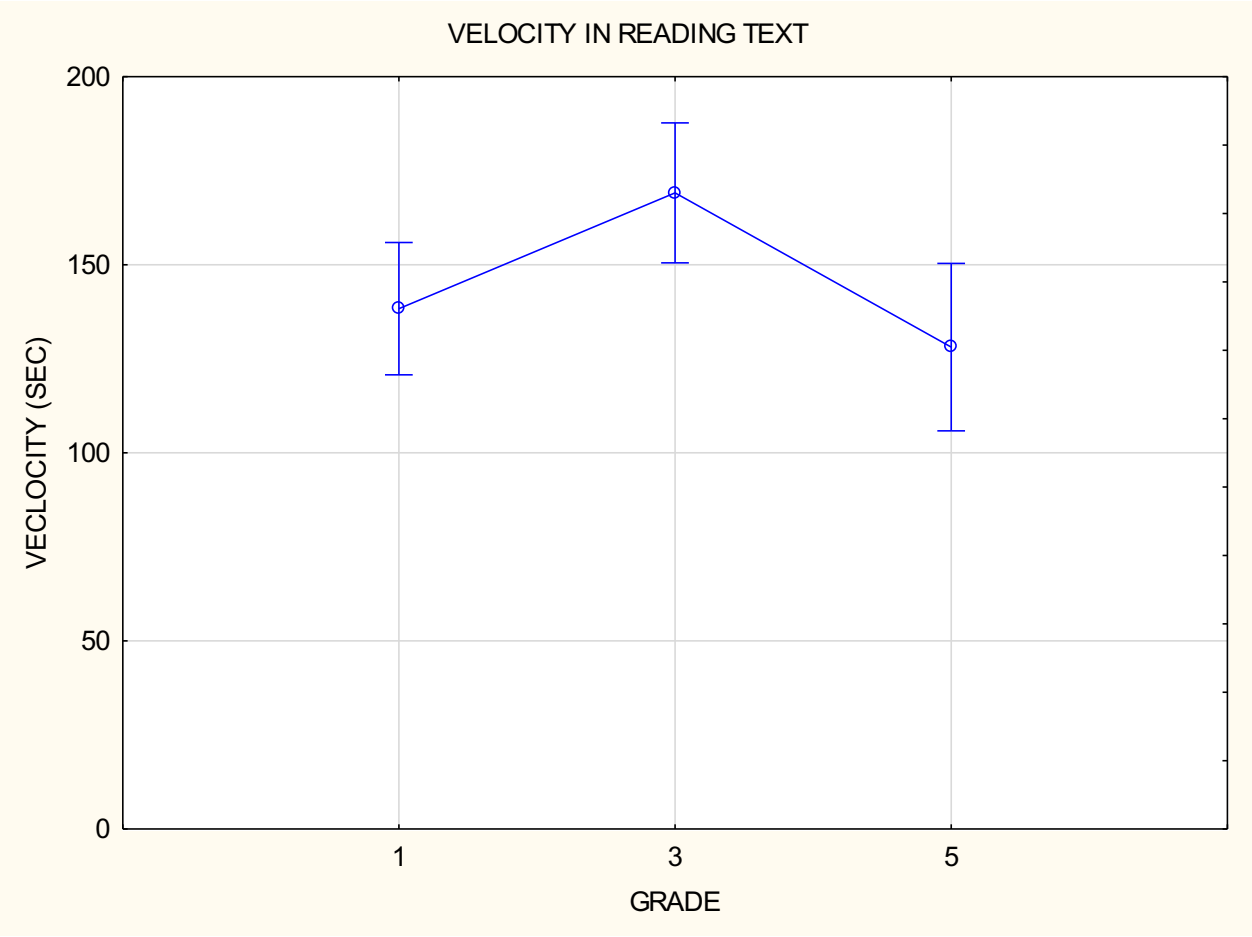



**Figure 11** shows reading velocity of a text (Yark) in English in Grade 1, 3, and 5 bilingual children. Bars indicate a confidence interval at 0.95.

As far as accuracy is concerned, third graders made more errors while reading a text (accuracy score is an error score), with respect to both first and fifth graders.

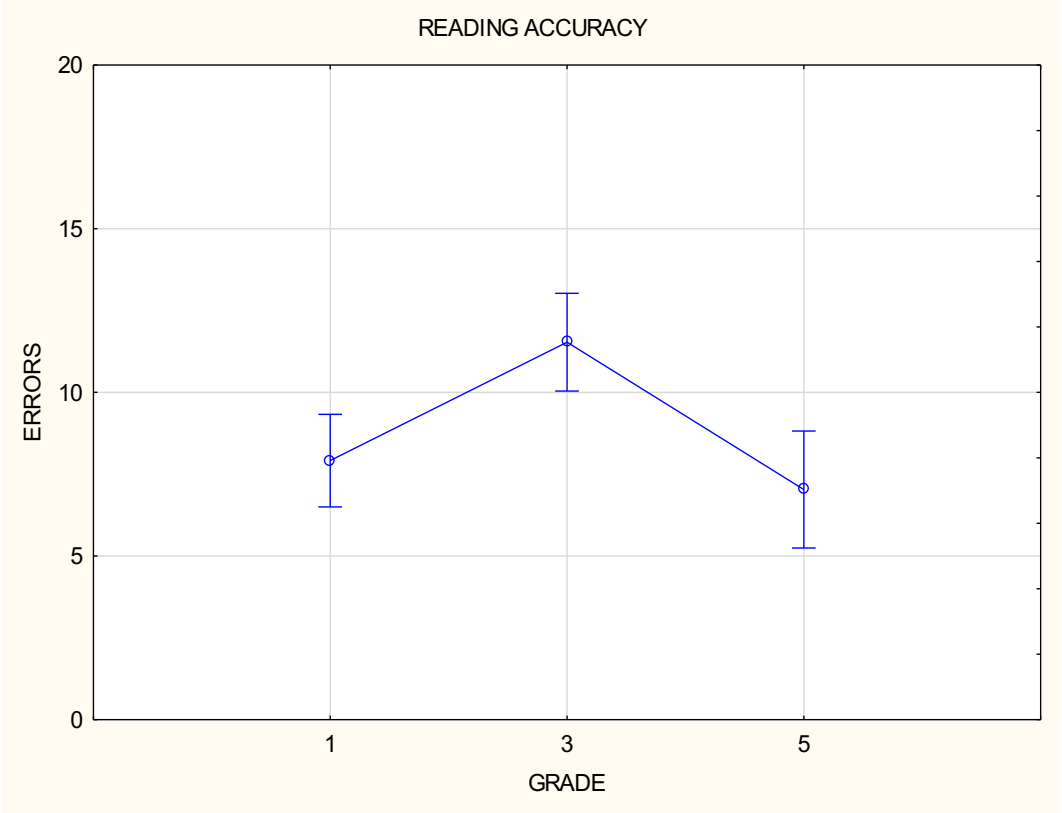



**Figure 12** shows error in reading an English text (Yark) in Grade 1, 3, and 5 bilingual children. Bars indicate a confidence interval at 0.95.

Statistical analyses found an effect of Grade both in reading rate, $F(2,118) = 4.66$, $\eta^2 = 0.07$, $p < 0.05$ and in reading accuracy, $F(2,118) = .918$, $\eta^2 = 0.13$, $p < 0.001$. Bonferroni *post hoc* revealed that third-graders differed from fifth-graders ($p < 0.05$) in reading rate, with fifth-graders outperforming third-graders; in reading accuracy, Bonferroni *post hoc* showed that first-graders differed from third-graders ($p < 0.005$), being more accurate than them (less errors), and third-graders differed from fifth-graders ($p < 0.001$), with third graders being less accurate (more errors) than fifth graders. 82% of the children were more than 1.50 standard deviations below the mean on reading velocity, and 97% were more than 1.50 standard deviations below the mean on reading accuracy, calculated on the basis of English monolingual norms.

Reading comprehension performance is displayed in *Figure 13*. As can be seen, there is an improvement from Grade 3 to 5, although the difference was not statistically significant. We could not estimate standard scores for reading comprehension as the test provides norms on the basis of two reading passages, while we administered only one.

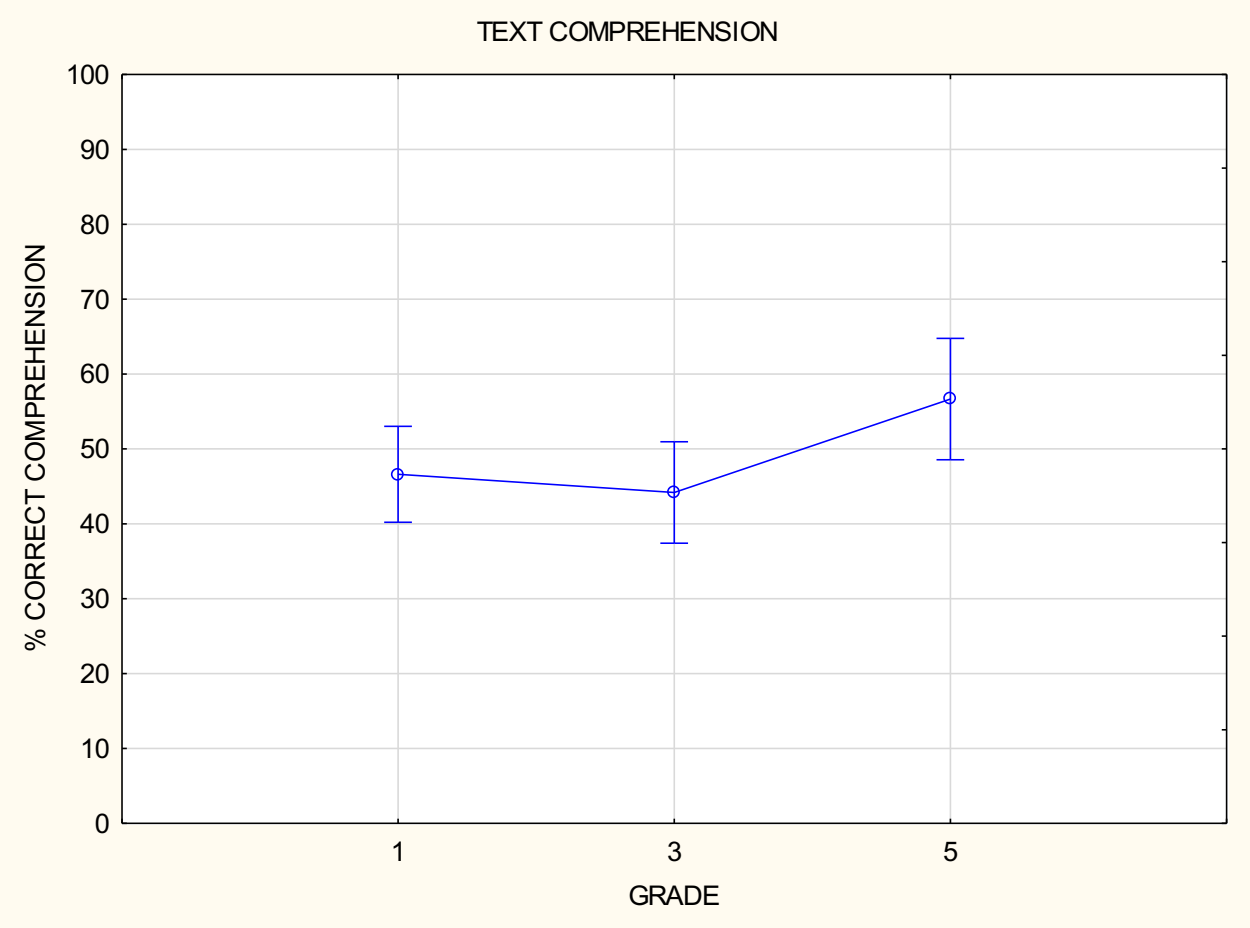



**Figure 13** shows English reading comprehension (Yark) scores in Grade 1, 3, and 5 of bilingual children. Bars indicate a confidence interval at 0.95.

Finally, we analyzed English working memory tasks. ANOVA showed an effect of Grade on *Children's Test of Nonword Repetition* (CN REP) $F(2,118) = 9.77$, $\eta^2 = 0.14$, $p < 0.001$, and Bonferroni *post hoc* showed that first-graders differed from third- and fifth-graders ($p < 0.05$). Bilingual third- and fifth-graders were better than first-graders, as evident in *Figure 14*.

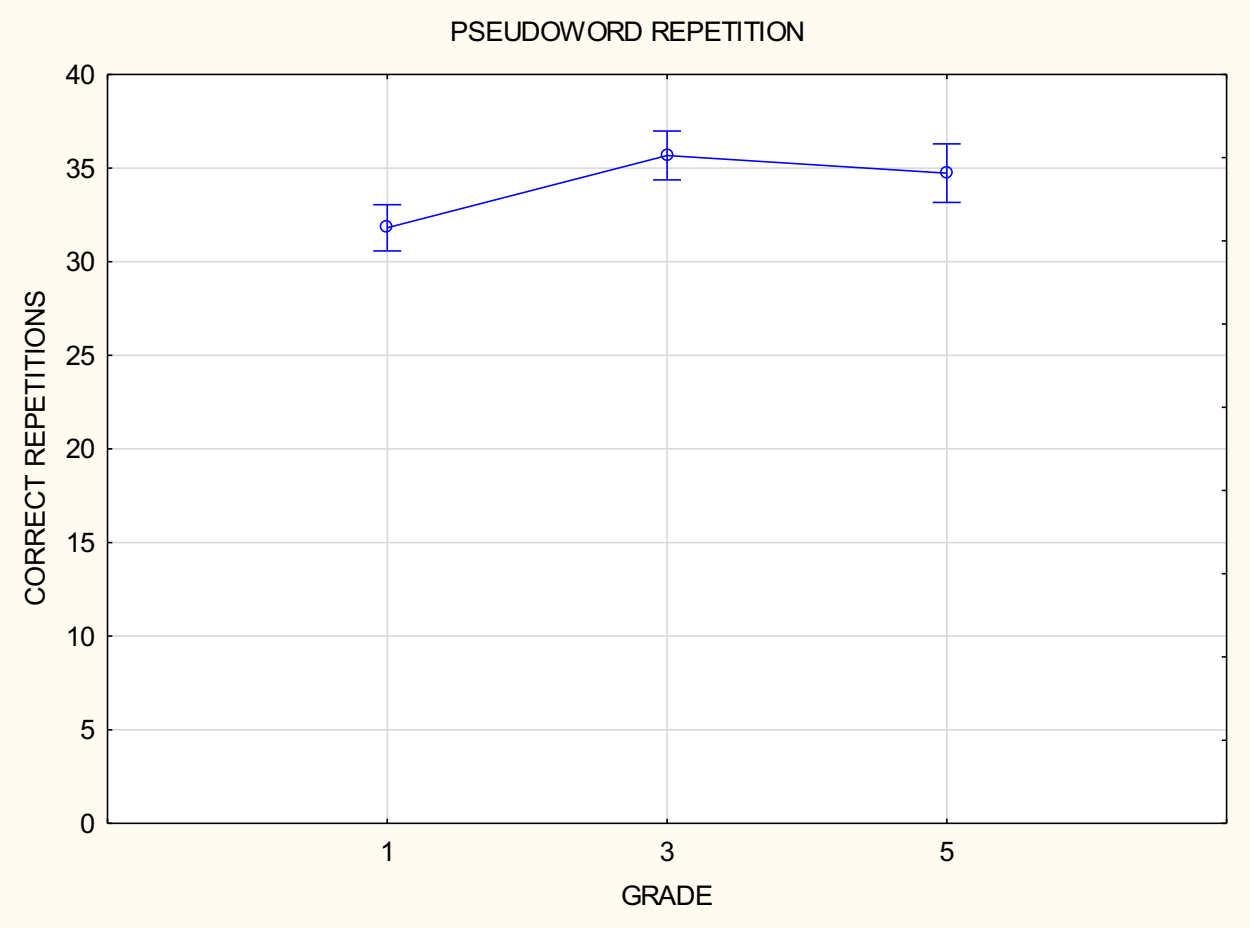



**Figure 14** shows scores (Max=40) at English pseudoword repetition task in Grade 1, 3, and 5 of bilingual children. Bars indicate a confidence interval at 0.95.

*Figure 15* shows the performance on the forward digit span task, and *Figure 16* illustrates children's performance on the backward digit span task. As can be seen, there is an improvement across grades in both cases. An effect of Grade was found both on English *Forward Digit Span Task,* $F(2,118) = 7.83$, $\eta^2 = 0.12$, $p < 0.001$, and on *Backward Digit Span Task,* $F(2,118) = 12.73$, $\eta^2 = 0.18$, $p < 0.001$.

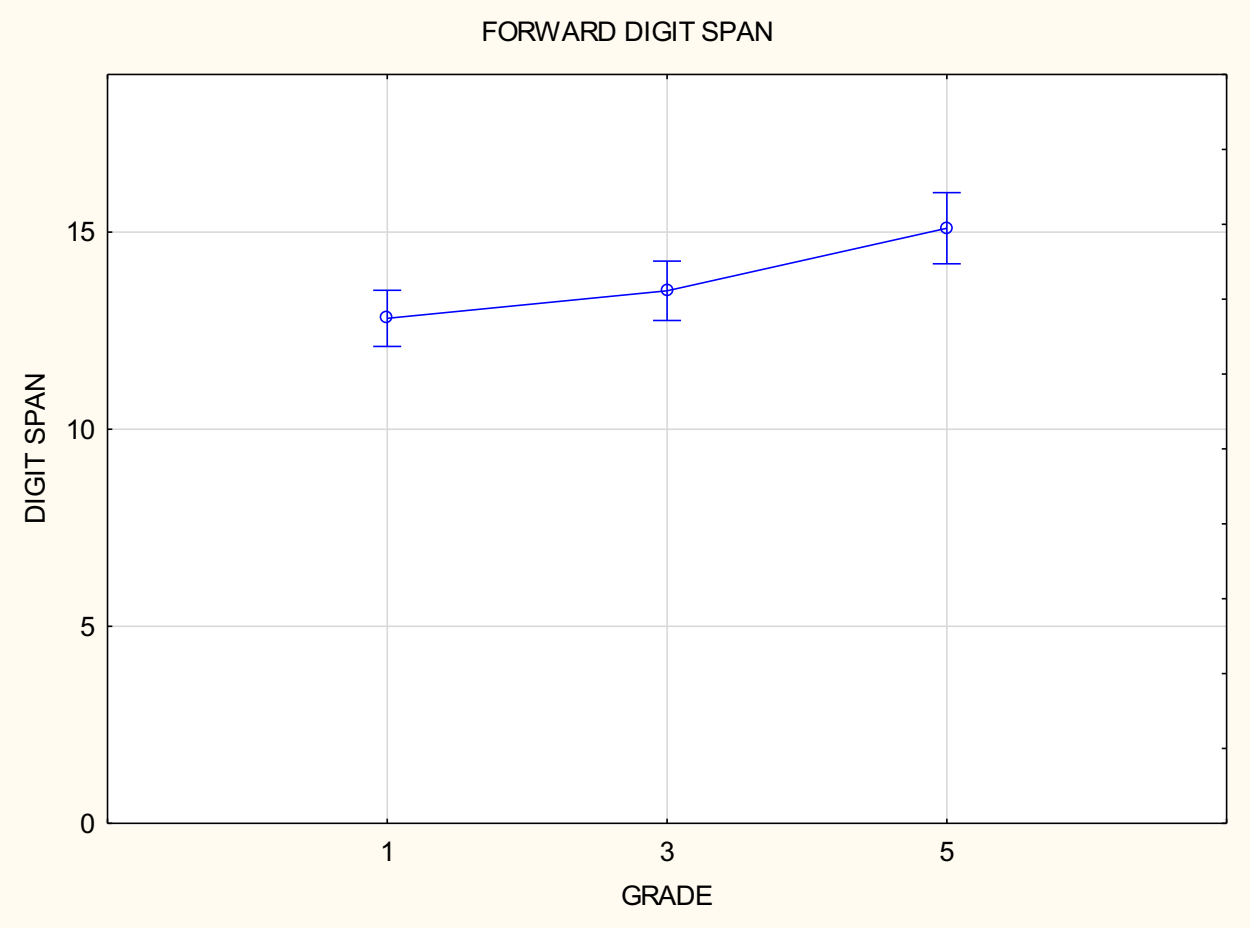



**Figure 15** shows scores at the forward digit span task in English in Grade 1, 3, and 5 of bilingual children. Bars indicate a confidence interval at 0.95.

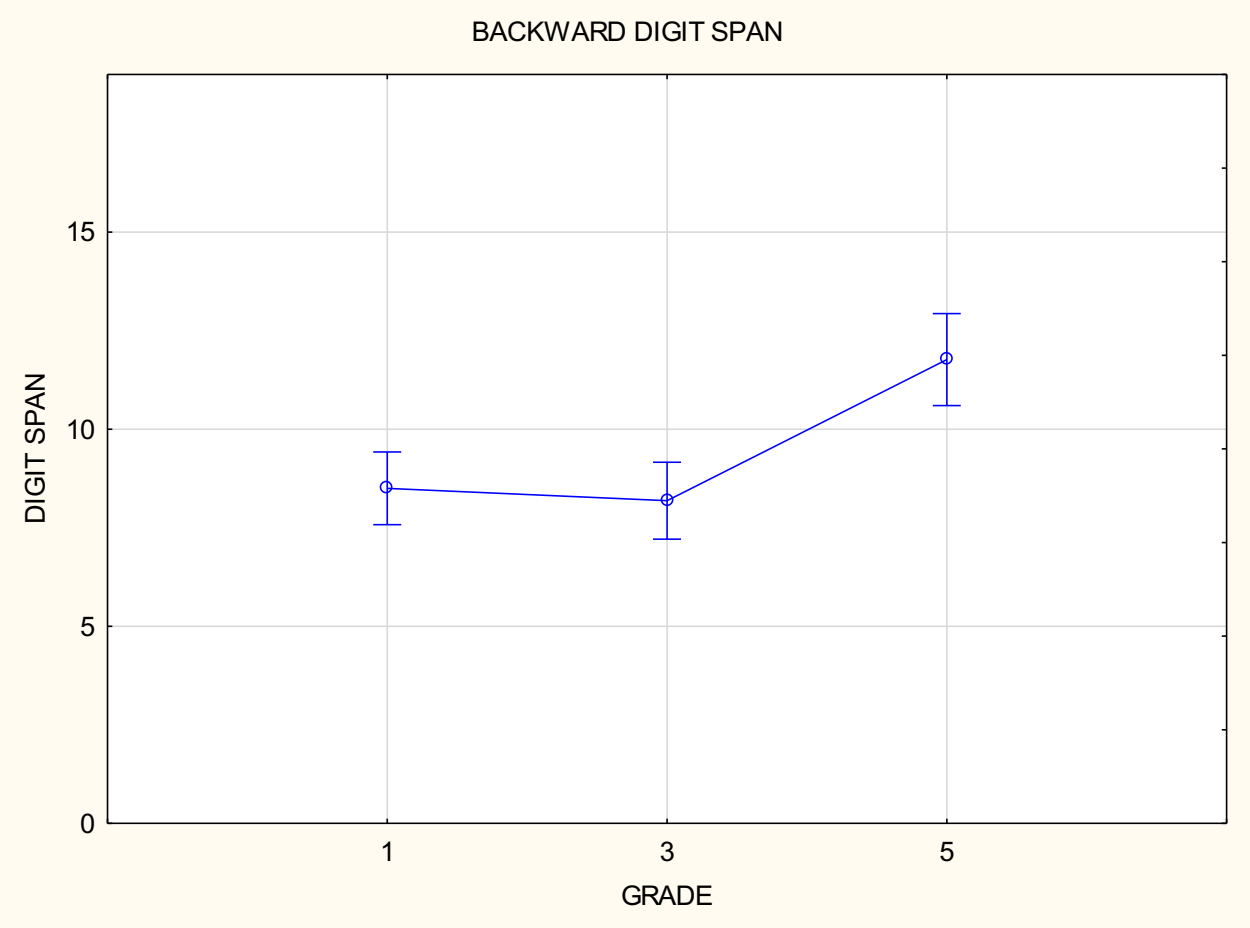



**Figure 16** shows scores at the backward digit span task in English in Grade 1, 3, and 5 of bilingual children. Bars indicate a confidence interval at 0.95.

Bonferroni *post hoc* showed that first-graders differed from fifth-graders (with fifth-graders being better than first-graders), and third-graders differed from fifth-graders (with fifth-graders being better than third-graders) both on Forward ($p < 0.05$) and Backward ($p < 0.001$) digit span tasks. Thus, no improvement was evident between first- and third-graders.

In conclusion, the effect of Grade was found on all of the English measures in the Italian-English bilingual group, with the only exception of reading comprehension. However, regarding English reading proficiency measures (*i.e.*, reading rate and accuracy), a weaker performance was revealed in Grade 3 children. However, there was a final improvement in Grade 5, by the end of primary school.

***Correlations between Italian and English skills in the bilingual group***

To verify whether the bilinguals' performance in Italian was related to English performance, we carried out correlation analyses. In these analyses, performances on each ability were correlated in the two languages. Since, in some cases, the two tasks in the two languages were not directly comparable because of raw scores from different tests, we used standard scores (as in the case of reading accuracy and velocity measures), calculated on the basis of monolingual norms provided by the test, or the percentage scores of accuracy (as in the case of vocabulary or reading comprehension). As the Yark test for reading assessment does not provide norms for velocity in the case of Beginner level of reading, we excluded from the correlational analysis of reading velocity first-grader bilinguals, as a major part of them were tested with the Beginner level text.

As far as language measures are concerned, a significant positive correlation was found between expressive vocabulary measures in Italian and English $r = 0.47$, $p < 0.001$. A significant positive correlation was also found between the scores on the *LITMUS SRep tasks* (Italian and English version) $r = 0.40$, $p < 0.001$.

Reading speed in Italian and English – involving only third- and fifth-graders – correlated significantly, $r = 0.55$, $p < 0.001$: those who read faster in Italian also read faster in English; reading errors also correlated $r = 0.37$, $p < 0.05$: those who made more errors in Italian also did so in English. Reading comprehension marginally correlated, $r = 0.21$, $p = 0.06$, but we must remark the difference between the Italian and English reading comprehension tasks. While in Italian, the responses to the questions were written and different texts were used to test technical reading and reading comprehension, in English they were administered orally and the same passage was used to test both technical reading and reading comprehension.

Lastly, we compared Italian working memory tasks to the English ones. We found a significant positive correlation between the *Pseudoword Repetition* scores in Italian and those in the *Children's Test of Nonword Repetition* (CN REP) in English $r = 0.53$, $p < 0.001$, between scores

obtained in Italian and in English on the *Forward Digit Span Task* $r = 0.65$, $p < 0.001$, and on the *Backward Digit Span Task* $r = 0.70$, $p < 0.001$.

***Correlations among the various abilities in the monolingual and bilingual groups***

To gain further information about the development of language, memory skills, and literacy in Italian, we performed several correlation analyses among the various abilities in the two groups separately.

We first analyzed the correlations within the monolingual group. As evident in the correlation matrix in *Table 2*, vocabulary and reading speed correlated with most of the other variables. Specifically, vocabulary correlated with reading, sentence repetition, and memory measures. Reading speed correlated with other reading measures, sentence repetition, and memory. Memory measures correlated with each other.

| | Vocab. | Read speed | Read acc. | Read compr. | Sent. rep. | NW rep. | DS forw. | DS back. |
|---|---|---|---|---|---|---|---|---|
| **Vocab.** | 1,0000 | 0,5625 | -0,1535 | 0,4315 | 0,5584 | 0,3349 | 0,5339 | 0,3522 |
| **Read speed** | 0,5625 | 1,0000 | -0,2967 | 0,3457 | 0,2990 | 0,2574 | 0,5708 | 0,3257 |
| **Read acc.** | -0,1535 | -0,2967 | 1,0000 | -0,1704 | 0,0027 | -0,2567 | -0,0653 | -0,0123 |
| **Read compr.** | 0,4315 | 0,3457 | -0,1704 | 1,0000 | 0,2207 | 0,1549 | 0,3109 | 0,1430 |
| **Sent. rep.** | 0,5584 | 0,2990 | 0,0027 | 0,2207 | 1,0000 | 0,3907 | 0,3136 | 0,3393 |
| **NW rep.** | 0,3349 | 0,2574 | -0,2567 | 0,1549 | 0,3907 | 1,0000 | 0,0832 | 0,1888 |
| **DS forw.** | 0,5339 | 0,5708 | -0,0653 | 0,3109 | 0,3136 | 0,0832 | 1,0000 | 0,4516 |
| **DS back.** | 0,3522 | 0,3257 | -0,0123 | 0,1430 | 0,3393 | 0,1888 | 0,4516 | 1,0000 |

**Table 2** shows the correlation matrix of the Italian measures within the monolingual group. Significant correlations are marked in red ($p < 0.05$).

The same picture emerges in the case of the bilingual children, as evident in *Table 3*: vocabulary and reading speed correlated with most of the other variables, the reading measures correlated with each other, and so did the memory measures.

| | Vocab. | Read speed | Read acc. | Read compr. | Sent. rep. | NW rep. | DS forw. | DS back. |
|---|---|---|---|---|---|---|---|---|
| **Vocab.** | 1,0000 | 0,5493 | -0,2700 | 0,2980 | 0,4587 | 0,2032 | 0,1969 | 0,2124 |
| **Read speed** | 0,5493 | 1,0000 | -0,5362 | 0,3514 | 0,4151 | 0,3113 | 0,2434 | 0,3378 |
| **Read acc.** | -0,2700 | -0,5362 | 1,0000 | -0,2449 | -0,1983 | -0,2391 | -0,0786 | -0,1474 |
| **Read compr.** | 0,2980 | 0,3514 | -0,2449 | 1,0000 | 0,2169 | 0,1012 | 0,2518 | 0,2926 |
| **Sent. rep.** | 0,4587 | 0,4151 | -0,1983 | 0,2169 | 1,0000 | 0,3914 | 0,3628 | 0,3253 |
| **NW rep.** | 0,2032 | 0,3113 | -0,2391 | 0,1012 | 0,3914 | 1,0000 | 0,3153 | 0,2437 |
| **DS forw.** | 0,1969 | 0,2434 | -0,0786 | 0,2518 | 0,3628 | 0,3153 | 1,0000 | 0,5665 |
| **DS back.** | 0,2124 | 0,3378 | -0,1474 | 0,2926 | 0,3253 | 0,2437 | 0,5665 | 1,0000 |

**Table 3** shows the correlation matrix of the Italian measures within the bilingual group. Significant correlations are marked in red ($p < 0.05$).

In sum, these correlations, which are similar for the mono- and bilingual children, show that the various reading, language (vocabulary and sentence repetition), and memory measures develop similarly in both groups.

Finally, *Table 4* reports the correlations among the various abilities in English in the bilingual group. We found various correlations between vocabulary, reading speed and comprehension, sentence repetition, and memory measures (pseudoword repetition and digit span forward). Reading speed correlated with language measures (vocabulary and sentence repetition), with verbal short-term memory measure but not with digit span measures. Memory measures correlated with each other, and some reading measures correlated with other reading measures.

| | Vocab. | Read speed | Read acc. | Read compr. | Sent. rep. | NW rep. | DS forw. | DS back. |
|---|---|---|---|---|---|---|---|---|
| **Vocab.** | 1,0000 | 0,2103 | 0,0578 | 0,2458 | 0,6146 | 0,4865 | 0,3147 | 0,1553 |
| **Read speed** | 0,2103 | 1,0000 | 0,4394 | -0,1353 | 0,2421 | 0,3697 | -0,0701 | -0,0424 |
| **Read acc.** | 0,0578 | 0,4394 | 1,0000 | -0,1903 | 0,0883 | 0,2346 | -0,1314 | -0,0312 |
| **Read compr.** | 0,2458 | -0,1353 | -0,1903 | 1,0000 | 0,0948 | -0,0015 | 0,3363 | 0,2349 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Sent. rep.** | 0,6146 | 0,2421 | 0,0883 | 0,0948 | 1,0000 | 0,6669 | 0,3836 | 0,3495 |
| **NW rep.** | 0,4865 | 0,3697 | 0,2346 | -0,0015 | 0,6669 | 1,0000 | 0,1899 | 0,0954 |
| **DS forw.** | 0,3147 | -0,0701 | -0,1314 | 0,3363 | 0,3836 | 0,1899 | 1,0000 | 0,4706 |
| **DS back.** | 0,1553 | -0,0424 | -0,0312 | 0,2349 | 0,3495 | 0,0955 | 0,4706 | 1,0000 |

**Table 4** shows the correlation matrix of the English measures within the bilingual group. Significant correlations are marked in red ($p < 0.05$).

## Discussions

The study aimed to investigate the effects of bilingual education on literacy attainment of primary school Italian-English bilinguals. In Italy, some concerns and queries surround eventual benefits of bilingual education on literacy success and raise the question of whether learning to read in Italian (L1) is delayed when simultaneously learning to read in English (L2). Thus, we were primarily interested in verifying whether the Italian reading performance of early bilingual primary school children differed significantly from that exhibited by their Italian monolingual peers. To do so, language and reading performance in Italian of three groups of bilinguals educated in 50:50 Italian-English immersion programs (with Italian as their L1 and majority language) in Grades 1, 3, and 5 were compared to that of Italian monolingual peers living in Italy. Secondly, we expected to find improvements in language and reading English across the three grades in the bilingual group. Therefore, we compared the performance of the three grades. Third, we were interested in verifying whether the bilinguals' language and reading attainment in Italian were related to their English performance. To do so, we performed correlations between the same skills evaluated in the two languages in the bilingual group. Fourth, we aimed at verifying the hypothesis that competence in oral language supports literacy development, and we tested this hypothesis by correlating language and reading measures both in Italian and English. Lastly, we investigated the relation between memory, language, and literacy in the two languages. In the rest of the discussion, we are going to take up each hypothesis.

***Learning to read two languages simultaneously***

Regarding whether Italian literacy is influenced by learning to read in two languages simultaneously, we predicted that there should not have been major differences between the monolinguals' and bilinguals' competence in Italian. In line with this prediction, we found that there was a development across grades in language and literacy skills in both groups, and this mirrors a consolidation in language and reading skills. On all language and reading measures, the two groups did not differ. This finding is in agreement with Costa *et al.* (2018), who measured decoding skills through reading lists of words and pseudowords. These authors found that the monolingual and bilingual groups did not differ in any measures of reading. In our study, we measured reading skills through a text. This choice has the effect that not only decoding is assessed, but also other skills, such as vocabulary knowledge, morpho-syntactic competence, and the ability to make inferences. Regarding reading accuracy and reading comprehension, our results differ with the findings of Costa *et al.* (2020) that showed a small monolingual advantage in reading accuracy measure and comprehension, in early years of primary school. The authors compared reading proficiency in Italian of a group of bilingual first- and third-graders with the same educational background (they were exposed to English within three years of age, and they were attending an Italian-English immersion program in Italy), with a group of monolingual Italian children. According to Genesee (1983), it is not unusual for students placed in immersion programs to lag behind their monolingual peers in the first three or four years of literacy development. In general, the gap in reading proficiency should be filled out by the end of schooling, and bilingual children usually catch up with their monolingual peers (Baker, 2001). Nevertheless, in our sample, bilinguals did not differ from monolinguals in Italian reading, even in first and third grade. Despite the study of Costa *et al.* (2020), we should notice that we controlled for the different ages of monolinguals and bilinguals in the analyses – with bilinguals being significantly younger than monolinguals. Thus, this could be one reason for different results. Moreover, both studies have a cross-sectional design, which could give rise to sampling effects.

We found an improvement across grades in all the tasks considered. However, we should notice that in some cases the improvement was significant only from Grade 1 to Grade 3 (that was the case of sentence repetition task and reading accuracy), or from Grade 3 to Grade 5 (reading comprehension). This could be attributed to the cross-sectional nature of the study.

In sum, bilingual and monolingual children performed similarly in Italian, in line with the literature. Even if we did not always find a significant development across all grades, this was present, even if minimal. However, this is a weakness of a cross-sectional study. A longitudinal study could overcome this limitation. Thus, we can conclude that learning to read in Italian and in English simultaneously does not affect reading performance in Italian – when bilingual exposure occurs by three years of age.

***Development in English language and literacy***

As regards the impact of an Italian-English *immersion program* on the development of language and literacy skills in English (L2), we predicted that there should be an improvement across the three grades.

Considering English language measures, that is expressive vocabulary and morpho-syntactic abilities, we found an improvement from Grade 1 to Grade 3, but the difference between Grade 3 and Grade 5 was not found to be statistically significant. But, this may be due to sampling effects given the cross-sectional character of our study.

Looking at English reading proficiency, bilinguals were shown to improve their reading velocity by Grade 5, even if performance in Grade 3 were slower. For reading accuracy we observed the same trend, as third-graders were less accurate than first-graders, but the performance improved by Grade 5. Moreover, almost all of children achieved performance within English monolingual norms. As regard English reading comprehension, the performance across Grades did not show a significant improvement.

Bilingual third- and fifth-graders were better than first-graders at repeating pseudowords. Moreover, bilingual fifth-graders were found to better perform for the digit span tasks in English, compared to their younger peers. This is not unsurprising, as memory span has been shown to increase with age (refer to Huttenlocher & Burke 1976, for a review of this matter). Further research should compare the performance at short-term memory task – as digit span – in English of the bilingual group with that of a monolingual English control group to verify the effect of bilingualism on short-term memory in L2 too.

In conclusion, the downfalls we observed in the bilingual group could plausibly be attributed to the cross-sectional design of the present study. We will discuss in detail the limits of the cross-sectional study in the *Limitations* section below.

***Language and literacy development in Italian and English***

The third aim of the study was to verify whether the language and reading attainment exhibited by the bilingual children in Italian were related to their English performance. Significant correlations were found between the Italian and the English performances on all of the measures. This indicates a similar performance in the two languages and a similar development across grades. If we overview these findings, we can conclude that the significant correlations between Italian and English scores on all measures indicate that proficiency in one language is mirrored by competence in another language. These correlations also suggest that certain language and reading skills might be transferred from one language to the other. Plausibly, this cross-linguistic transfer is more likely to occur from Italian to English because, as we know, Italian has a shallow orthography and a less complex phonological structure compared to English, and thus may facilitate learning to read and write. An English monolingual control group is needed to verify a possible beneficial effect of Italian on English literacy.

### *The relation between language and reading*

As discussed above, it is well established that oral language competence is a predictor of reading outcomes. Thus, another aim of the study was to verify whether oral language abilities were related to reading performance. Considering Italian, we found that vocabulary was related with most of the reading measures in the monolingual group, and with all of the reading measures in the bilinguals, as expected (Duff *et al.,* 2015). Morpho-syntax, as measured by sentence repetition, was also found to relate with reading measures in the bilingual group. This result was in line with Kovelmann *et al.* (2008) and Prevoo *et al.* (2016), who found strong associations between language proficiency and reading in bilingual children. In the monolingual group, morpho-syntactic measure was related only with reading speed, among the reading measures.

Looking at the bilinguals' English performance, vocabulary showed to be related to most of the reading measures, and the morpho-syntactic measure was related to some of the reading measures.

Thus, it stands to reason that proficiency in the language of school instruction is not only a necessity for understanding the teacher but also a forerunner to literacy attainment (Prevoo *et al.* 2016).

### *Memory, language, and reading*

Measures investigating working memory skills showed not to differ between groups. All the tasks showed a general improvement, even if in the pseudowords repetition and Backward Digit Span Task the improvement was significant from Grade 1 to Grade 3, while in Forward Digit Span it was significant from Grade 3 to Grade 5. Once again, this may be due by chance to sampling.

These results are consistent with Ratiu *et al.* (2015) and Lehtonen *et al.* (2018), who showed that the bilingual condition *per se* does not provide an advantage on working memory skills. Moreover, our groups were matched for medium-high socio-economic background, thus no differences were expected.

Briefly considering the pseudoword repetition task, the performance between the two groups approximately overlapped. This outcome was expected since a typically developing population was tested: pseudoword repetition task is a *process measure* only slightly affected by bilingualism and is often used to identify dyslexia in bilingual children (refer to Vender, Delfitto & Melloni 2019, for a review of this matter).

To conclude, we found working memory skills to be correlated to reading attainment in most cases, as expected. As discussed above, working memory is implicated when reading and supports reading performance (Swanson *et al.*, 2004). We found span tasks correlated with reading comprehension in all cases (monolingual and bilingual groups, in Italian and in English), and pseudoword repetition task to correlate more often with reading proficiency measures (velocity and accuracy). Even if there is not yet an overall consensus about which particular memory skill is correlated explicitly with reading performance (White, 2019), it is plausible to think that short-verbal memory implicated in a pseudoword repetition task better correlates with the words-reading level in a text, for its phonological component. In contrast, memory processes implicated in span tasks better correlate with the comprehension of a text, as it involves the storage of verbal and contextual information.

***Limitations***

The current study has some limitations. Firstly, the study has a cross-sectional design instead of a longitudinal design, which would have been preferable for more valuable analysis. Research using longitudinal, randomized designs is needed to better understand the development of reading proficiency for Italian-English bilingual learners in an *immersion program*. Moreover, the two samples (bilingual and monolingual control group) were slightly unbalanced. This was partially due to some difficulties we encountered in recruiting Italian monolingual schools, probably for lack of interest in the study's goals, partially because of the Covid-19 sanitary emergency, which determined the schools' closure in Italy since February 2020. Also, the study did not include an English monolingual control group. It would be interesting to verify a potential beneficial effect of

Italian – a shallow orthography language – on English – a deep orthography language, to have a complete insight of the reading proficiency in bilingual Italian-English population.

Children's family SES could only be estimated based on the two observable economic variables outlined above, and parents' title of study, when available. SES approximation was assumed to be a reliable estimation of a medium-high SES of our entire sample. Nevertheless, it would have been better to have all of the questionnaires filled out. The compilation of questionnaires should probably be done by the experimenter, online with parents, which could directly ask for the information needed and not by parents themselves. It could be the occasion to eventually further clarify the importance of gathering SES data, if necessary. Of course, collecting SES information in person takes a long time, and it is not sure that all parents are available to provide information.

As already pointed out above, another limitation was that some monolingual groups were tested in different moments during the school year with respect to bilinguals. In particular monolingual first- and third-graders were tested at the beginning of Grade 2 and Grade 4, respectively. Thus, bilingual and monolingual groups were not perfectly matched for age. To control for different ages, we used the testing day as a covariate. We considered monolingual students as an appropriate control group because their performance was assessed right after the summer break (whereas bilingual participants were tested before summer vacation). On the other hand, some researches have shown that the summer holiday could have a positive impact on academic achievement on high SES children (Downey, Von Hippel & Broh, 2004). In a meta-analysis conducted on 13 studies, Cooper, Nye, Charlton, Lindsay & Greathouse (1996) found that students from middle-class families displayed gains in reading attainment over the summer, probably because they spent more time practicing reading in that period. Our monolingual students' reading performance may have benefited from the summer holidays, which could justify the slight monolingual reading advantage found. Of course, it would have been better to test both

monolingual and bilingual groups during the same period, but this was not possible. Future studies should try to test all children at the same time in the school year.

**Conclusions**

Our predictions were all confirmed by the results of the present study. To summarize, we found no evidence that a bilingual learning environment hinders the L1 language and reading proficiency in Italian mother-tongue students in a *simultaneous Italian-English immersion program* when they are exposed to English L2 from age 3. We found bilingual students' improvements in language, reading, and memory skills, both in Italian and English. Nevertheless, in some English reading skills (velocity and accuracy), we found a downfall in Grade 3 performance, but this could be attributed to the study's sampling. We found positive correlations between the bilinguals' Italian and English performances, conjecturing a positive transfer of skills from one language to another – presumably from Italian, the dominant language, which could support the same ability in English, the second language. Language abilities in one language were found to correlate with reading abilities in the same language, supporting the idea that oral language supports reading development. Finally, we found working memory skills to develop similarly in the two groups and to predict reading performance in both groups and both languages.

Further research using longitudinal, randomized designs is needed to confirm our results. An English monolingual control group is also needed to determine English literacy outcomes more accurately in the bilingual group and, eventually, how best to ensure reading success in both languages for bilingual learners. We could expect some reading beneficial effects of Italian over English or immersion students to catch up to English monolingual students within few years (Montanari, 2013), confirming that skills acquired in Italian can be transferred to English, and possibly *vice versa*. These *lacunae* will motivate our research in understanding bilingualism and biliteracy.

We could conclude that, when it comes to an immersive and early L2 exposure, Italian-English biliteracy does not seem to impede a good proficiency in both languages (Costa *et al.*, 2020), and consistent with Kovelmann *et al*. (2008), the age of first bilingual exposure has an impact on literacy development. In fact, our children have been exposed to English from age 3 in a school setting for half of the time.

Today bilingualism should be the norm to navigate the global world. The research, education, and public community need to be well informed about the bilingual education outcomes by documenting its overall academic success. This is the final aim of our research. We believe bilingual education should be fostered in modern societies and should be an available educational opportunity for all children in Italy. In the future, it could be implemented in public schools also. Surely, knowing more than one language is always a good idea.

**References**

Alloway, T. P., & Alloway, R. G. (2010). Investigating the Predictive Roles of Working Memory and IQ in Academic Attainment. *Journal of Experimental Child Psychology,* 106, 20-29.

Alloway, T., Gathercole, S. E., Adams, A. M., Willis, C., Eaglen, R., & Lamont, E. (2005). Working Memory and Phonological Awareness as Predictors of Progress Towards Early Learning Goals at School Entry. *British Journal of Developmental Psychology*, 23, 417-426.

Abu-Rabia, S., & Siegel, L. S. (2002). Reading, syntactic, orthographic, and working memory skills of bilingual Arabic-English speaking Canadian children. *Journal of Psycholinguistic Research*, 31(6), 661-678.

Baddeley, A. D., Gathercole, S., Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychological Review*, 105, 158-173.

Baker, C. (2001). The Effectiveness of Bilingual Education. *Foundations of Bilingual Education and Bilingualism* (3rd ed, pp. 228-243). Clevedon, England: Multilingual Matters.

Baker, D. L, Stoolmiller, M., Good, R. H., Baker, S. K. (2011). Effect of reading comprehension on passage fluency in Spanish and English for second-grade English learners. *School Psychology Review*, 40(3), 331–351.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development*, *57*(2), 498-510.

Bialystok, E., Majumder, S., Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics*, 24(1), 27-44.

Bialystok, E., Senman, L. (2004). Executive processes in appearance-reality tasks: the role of inhibition of attention and symbolic representation. *Child Dev*., 75(2), 562‑579.

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, *9*(1), 43-61.

Bialystok, E., Shapero, D. (2005). Ambiguous benefits: the effect of bilingualism on reversing ambiguous figures. *Dev Sci*., 8(6), 595‑604.

Bialystok, E. (2007). Acquisition of literacy in bilingual children: A framework for research. *Language Learning*, *57*(s1), 45-77.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525-531.

Bialystok E. (2018). Bilingual education for young children: review of the effects and consequences. *Int. J. Biling. Educ. Biling*., 21(6), 666‑679.

Blom E., Boerma T., Bosma E., Cornips L., & Everaert E. (2017). Cognitive Advantages of Bilingual Children in Different Sociolinguistic Contexts. *Frontiers in Psychology* 8, 552.

Bonifacci, P., Giombini, L., Bellocchi, S., Contento, S. (2011). Speed of processing, anticipation, inhibition and working memory in bilinguals. *Developmental Science*, 14(2), 256-269.

Bonifacci, P., & Tobia, V. (2016). Crossing barriers: Profiles of reading and comprehension skills in early and late bilinguals, poor comprehenders, reading impaired, and typically developing children. *Learning and Individual Differences*, *47*, 17-26.

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278-288.

Chiat, S. (2015). *Non-Word Repetition*. In Assessing Multilingual Children: Disentangling Bilingualism form Language Impairment, edited by Armon-Lotem, S., de Jong, J., & Meir, N., 125-147. Bristol: Multilingual Matters.

Chik, P. P, Ho, C. S., Yeung, P., Wong, Y., Chan, D. W., Chung, K. K. & Lo L. (2010). Contribution of discourse and morphosyntax skills to reading comprehension in Chinese dyslexic and typically developing children. *Ann. of Dyslexia*, 62, 1-18.

Ciccarelli, L. (1998). Comprensione del linguaggio, dei processi di elaborazione e memoria di lavoro: uno studio in età prescolare. *Unpublished doctoral dissertation, University of Padua*.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*(3), 227-268.

Cornoldi, C., & Carretti, B. (2016). *Prove MT-3-Clinica: La valutazione delle abilità di Lettura e Comprensione per la scuola primaria e secondaria di I grado.* Florence, Italy: Giunti Edu.

Cossu, G. (2013). *TNL - Test Neuropsicologico Lessicale per l'età evolutiva*. Florence, Italy: Hogrefe.

Costa, A., Hernández, M., Costa-Faidella, J., and Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: now you see it, now you don't. *Cognition* 113, 135-149.

Costa F., Sharley, S. & Guasti M. T. (2019). Does English delay learning to read Italian in primary school children? *Rivista di grammatica generativa*.

Costa, F., Guasti, M. T., & Sharley, S. (2020). Double Literacy Effects on Language and Reading Skills in Italian–English Primary School Children. *Journal of Monolingual and Bilingual Speech* 2(2), 185-218.

Cummins, J. (1981). *The role of primary language development in promoting educational success for language minority students: A theoretical framework.* Los Angeles: California State University, Evaluation, Dissemination and Assessment Center.

Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70-89). Cambridge, England: Cambridge University Press.

Cummins, J. (2005). A proposal for action: Strategies for recognizing heritage language competence as a learning resource within the mainstream classroom. *Modern Language Journal*, 89(4), 585-592.

D'Angiulli, A., & Serra, E. (2001). The development of reading in English and Italian bilingual children. *Applied Psycholinguistics*, 22, 479-507.

Downey, D. B., Von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, *69*(5), 613-635.

Dubins, M. H., Berens, M. S., Kovelman, I., Shalinsky, M. H., Petitto, L. A. (2009). *An oddball investigation of monolingual and bilingual infants' Phonetic discrimination using functional Near-Infrared Spectroscopy (fNIRS)*. Presented at annual Cognitive Neuroscience Society, San Francisco, CA, 21-24.

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes?. *Journal of child psychology and psychiatry, and allied disciplines*, 56(8), 848–856.

Engel de Abreu, P. M. J., Gathercole, S. E., Martin, R. (2011). Disentangling the Relationship Between Working Memory and Language: The Roles of Short-Term Storage and Cognitive Control. *Learning and Individual Differences,* 21, 569-574.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*(1), 78-104.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., Emslie, H. (1994). The Children's Test of Nonword Repetition: a test of phonological working memory. *Memory*, 2(2), 103‑127.

Gathercole, S. E., Brown, L., & Pickering, S. J. (2003). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment levels. *Educational and Child Psychology, 20*(3), 109-122.

Gathercole, S. E., Pickering, S. J., Knight, C., Stegmann, Z. (2004). Working Memory Skills and Educational Attainment: Evidence From National Curriculum Assessments at 7 and 14 Years of age. *Applied Cognitive Psychology,* 18, 1-16.

Genesee, F. (1985). Second Language Learning Through Immersion: A Review of US Programs. *Review of Educational Research*, 55(4), 541-561.

Genesee, F. (1985). *The baby and the bathwater: What immersion has to say about bilingual education.* Unpublished manuscript, Department of English as a Second Language, University of Hawaii at Manoa.

Genesee, F. (1987). *Learning through two languages: Students of immersion and bilingual education.* Cambridge, MA: Newbury House.

Genesee, F. (2007). French immersion and at-risk students: a review of research evidence, *Canadian Modern Language Review*, 63(5), 655–687.

Genesee, F., Jared, D. (2008). Literacy development in early French immersion programs. *Canadian Psychology/Psychologie Canadienne, 49*(2), 140-147.

Halle, T., Hair, E., Wandener, L., McNamara, M., & Chien, N. (2012). Predictors and Outcomes of Early Versus Later English Language Proficiency among English Language Learners. *Early Childhood Research Quarterly,* 27, 1-20.

Henry, L. A., Botting, N. (2016). Working Memory and Developmental Language Impairments. *Child Language Teaching and Therapy*, 33, 19-32.

Hilchey, M. D., &Klein, R. M. (2011). Are there bilingual advantages on non-linguistic interference tasks? Implications for the plasticity of executive control processes. *Psychon. Bull. Rev.,* 18, 625-658.

Hitch, G. J., Towse, J. N. and Hutton, U. (2001). What Limits Children's Working Memory Span? Theoretical Accounts and Applications for Scholastic Development. *Journal of Experimental Psychology: General*, 130, 184-198.

Hoff, E., Rumiche, R., Burridge, A., Ribot, K. M., & Welsh, S. N. (2014). Expressive vocabulary development in children from bilingual and monolingual homes: A longitudinal study from two to four years. *Early Childhood Research Quarterly*, *29*(4), 433-444.

Huttenlocher, J., & Burke, D. (1976). Why does memory span increase with age? *Cognitive Psychology,* 8(1), 1-31.

Kane, M. J., Engle, R. W. (2003). Working-memory Capacity and the Control of Attention: The Contributions of Goal Neglect, Response Competition, and Task set to Stroop Interference. *Journal of Experimental Psychology: General,* 132, 47–70.

Kovelman, I., Baker, S. A., & Petitto, L. A. (2008). Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism: Language and Cognition*, *11*(2), 203-223.

Lambert, W. E., & Tucker, G. R. (1972). *The bilingual education of children.* Rowley, MA: Newbury House.

Lindholm-Leary, K.J., Hernández, A. (2011). Achievement and language proficiency of Latino students in dual language programmes: native English speakers, fluent English/previous ELLs, and current ELLs. *Journal of Multilingual and Multicultural Development*, 32(6), 531-545.

Macnamara, J. (1973). Nurseries, streets and classrooms. *Modern Language Journal, 57,* 250- 254.

Marini, A., Eliseeva, N. & Fabbro, F. (2019). Impact of early second-language acquisition on the development of first language and verbal short-term and working memory. *International Journal of Bilingual Education and Bilingualism*, 22(2), 165-176.

Marinis, T., Chiat, S., Armon-Lotem, S., Gibbons, D., & Gipps, E. (2010). *School-Age Sentence Imitation Test (SASIT)* University of Reading. Reading.

Marinis, T., & Armon-Lotem, S. (2015). Sentence Repetition. In: Armon-Lotem, S., de Jong, J. & Meir, N. (Eds.). *Methods for assessing multilingual children: disentangling bilingualism from Language Impairment*. Multilingual Matters.

Montanari S. (2013). A Case Study of Bi-literacy Development among Children Enrolled in an Italian–English Dual Language Program in Southern California. *International Journal of Bilingual Education and Bilingualism*, 17, 509-525.

Morton, J. B., and Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Dev. Sci.,* 10, 719-726.

Paap, K. R., Johnson, H. A., and Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex,* 69, 265-278.

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., & Fazio, F. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping*, *19*(3), 170-182.

Petitto, L. A., Berens, M. S., Kovelman, I., Dubins, M. H., Jasinska, K., Shalinsky, M. (2012). The "Perceptual Wedge" hypothesis as the basis for bilingual babies' phonetic processing advantage: New insights from fNIRS brain imaging. *Brain and Language*, 121(2), 142-155.

Prevoo, M. J. L., Malda, M., Mesman, J., & van IJzendoorn, M. H. (2016). Within- and Cross-Language Relations Between Oral Language Proficiency and School Outcomes in Bilingual

Children with an Immigrant Background: A Meta- Analytical Study. *Review of Educational Research*, 86, 237-276.

Renfrew, C. E., & Mitchell, P. (2010). *Word Finding Vocabulary Test: The Renfrew Language Scales* (1st ed). London, England: Speechmark (Routledge).

Richardson, J. T.E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6(2), 135-14.

Robertson, E. K., & Deacon, S. H. (2019). Morphological awareness and word-level reading in early and middle elementary school years. *Applied Psycholinguistics,* 40, 1051-1071

Rolstad, K., Mahoney, K. S., & Glass, G. V. (2005a). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy*, *19*(4), 572-594.

Rolstad, K., Mahoney, K. S., & Glass, G. V. (2005b). Weighing the evidence: A meta-analysis of bilingual education in Arizona. *Bilingual Research Journal*, *29*(1), 43-67.

Schmidt, E., Pérez, A., Cilibrasi, L., & Tsimpli, I. (2020). Prosody facilitates memory recall in L1 but not in L2 in highly proficient listeners. *Studies in Second Language Acquisition, 42* (1), 223-238.

Siegel, L. S. (2008). Morphological Awareness Skills of English Language Learners and Children With Dyslexia. *Topics in Language Disorders*, 28(1), 15-27.

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research*, *75*(2), 247-284.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Hulme, C. (2009). *YARC York Assessment of Reading for Comprehension Passage Reading*. GL Publishers.

Swain, M., & Lapkin, S. (1982). *Evaluating bilingual education.* Clevedon, England: Multilingual Matters.

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The Relationship Between Working Memory and Mathematical Problem Solving in Children at Risk and not at Risk for Math Disabilities. *Journal of Education Psychology,* 96, 471-491.

Vender, M., Delfitto, D., Melloni, C. (2019). How do bilingual dyslexic and typically developing children perform in pseudoword repetition? Evidence from a study on Italian L2 children. *Bilingualism: Language and Cognition*, 1–13.

White, M. J. (2019). The development of English proficiency and working memory in 5–6 year old ELLs in their first year of formal education. *International Journal of Bilingual Education and Bilingualism*, DOI: 10.1080/13670050.2019.1571009.

## Chapter 2

# BILITERACY EFFECTS ON ACADEMIC SKILLS IN HERITAGE BILINGUALS AT PRIMARY SCHOOL

### Introduction

Like other European countries, Italy has become the host country for a significant number of immigrants during the last years.

As reported in the annual document of the Italian Ministry of Education (MIUR, 2020), in the 2018/2019 school year, Italian schools welcomed a total of 8.580.000 students, of which approximately 860.000 with non-Italian citizenship. As evident, 10% of the school population is of immigrant origin. The majority of these international students are second-generation students, *i.e.*, children born in Italy from immigrant parents. Thus, in the last years, the number of bilingual heritage children attending mainstream schools in Italy has increased dramatically.

Primary schools – where children approach literacy and learn to read and write – absorb the largest number of these international students, representing 11.4% of the total number of primary school pupils, the higher percentage among the different degrees of education. The national distribution of foreign students is not homogeneous. Lombardy is the Italian region with the highest number of immigrant students, about a quarter of Italy's total present. Among Lombardy provinces, Milan is the first in the list. This frame reflects immigrant families' privileged settlement in those contexts where the job market is more favourable and offers more possibilities.

Immigrant families in Italy come from different countries – more than 200 – even if some communities are far more represented than others: European countries are the most represented (46.3%), followed by students from Africa (25.7%) and Asia (20.1%). Among the Asian communities, the Chinese one is undoubtedly the most numerous and has more than doubled in the last decade.

There is a limit of international students *per* class, fixed at 30% on the total, even with some exceptions[3]. Nevertheless, in the last years, the 30% threshold has been exceeded in some schools – mostly primary schools – reaching even 50%. This is the case of the Lombardy region.

As evident, mainstream classes' framework has changed in the Italian school, with bilingual children of immigrant origin – a very heterogeneous population – representing a critical relatively new entry, with new didactic dynamics, inclusion, and integration needs. It has become of too high importance to be aware of the multicultural presence implications. Children in classes have to face other peers with different backgrounds, history, culture, habits, and language. The teacher is the conductor of this cultural melting pot.

Multiculturalism is not yet a spread or official policy in Italy, arguably because immigration has not ancient history. Teachers often need to find their strategies to drive the class to the academic outcomes – to allow effective participation of second-language students, to adapt subjects and material, to provide access to the content (Sears, 1998) – in the not effortless endeavour to keep the class at the same literacy level and to avoid slowdowns in the fixed educational program. In fact, research has shown that L2 students' are at risk of low literacy, which might exert a social cost (Vernice, Matta, Tironi, Caccia, Lombardi, Guasti, Sarti, & Lang, 2019).

Immigrants are bilinguals who need to be bilingual. Children of immigrants speak and hear their first language in their homes (and neighborhoods in some cases), but are educated in the majority language (Abu-Rabia & Siegel, 2002). They may speak two languages from birth or begin learning the second language later. Thus, when they approach the school, they are often disadvantaged, for their Italian level, and need more support, which unfortunately is not always guaranteed.

The early section will introduce relevant arguments about bilinguals, with a focus on bilingualism dictated by immigrant status. A general background concerning relevant second

[3]. This limit could be lowered if there are some particular and documented complexities or if the Italian proficiency of the immigrant students is not yet adequate for the class's participation. Oppositely, it can be increased if adequate Italian linguistic competence is yet possessed.

language learning issues and possible disadvantages due to the immigrant condition in acquiring literacy will be provided. The subsequent section will discuss the possible advantages of this bilingual population. We will then illustrate the research on bilingualism's academic outcomes supported by biliteracy, and we will introduce the bilingual education programs for minority speaking students. We will conclude the review reconsidering some crucial variables to consider when facing reading acquisition in an L2, as the Age of First Exposure to a language and the kind of languages involved in the bilingualism/biliteracy. We already discussed in detail these topics in the first chapter of the present thesis. Thus, some concepts will be repeated, but, being each chapter independent, we assumed it was useful to recall some content. Lastly, we will give an insight into the Italian context in this regard.

**Immigrant bilinguals in Italy**

Frequently, bilingual children with an immigration background are exposed to their mother tongue (L1) just orally at home. Regular exposure to the majority language (Italian-L2) could start at the age of 3, if they enter kindergarten, or even not before the age of 6 when they start primary schooling. Although L2 learners start their literacy acquisition at the same time as their monolingual peers, they learn to read in L2 at a time when their oral L2 language skills (i.e., vocabulary, phonological awareness) might not be fully mastered. Thus, they often lag in literacy achievements, especially in the first years of primary school.

As we will discuss in detail in the dedicated section, one possible reason is that reading acquisition relies on the age of oral language first exposure and oral language proficiency (Kovelmann, Baker, & Petitto, 2008). In this context, it is important to clarify the concept of proficiency in relation to reading. Cummins (1979) proposed that there are two L2-competences to consider, one called Basic Interpersonal Communication Skills (BICS) and the other Cognitive Academic Language Proficiency (CALP). The BICS refers to oral conversational competence in everyday life, whereas CALP is more complex and abstract and include learning technical

vocabularies and the language used to describe academic content. While an L2-learning child could get the conversational competence after two years of regular exposure to the L2, for academic proficiency, up to 5-7 years of formal education are needed to reach monolingual literacy skills (Cummins, 2008). It follows that L2 students regularly exposed to L2 starting from primary school will need more time to catch up with their monolingual peers in academic achievements. If not adequately sustained and properly understood in their needs, literacy and learning difficulties of these L2 students could misinterpreted as learning diseases. In fact, in Italy, Developmental Dyslexia is commonly diagnosed at the age of 8, when dual-language typically developing children may still lag behind their monolingual peers. This situation may result in an inaccurate diagnosis, also motivated by the fact that the educational delay is more widespread among L2 pupils than L1 Italian pupils. In this respect, it is essential to verify when the strategies for reading words and pseudowords start to diverge in this population, as it is a standard reading test used for diagnosing Developmental Dyslexia in Italian children (Guasti, Costa, Chailleux, Pagliarini, Sarti, Granocchio, & Stucchi, *submitted*).

Another reason that could influence foreign students' literacy achievement delay is the socio-cultural and socioeconomic (SES) factor. Immigrant children, in fact, often come from low SES families. Lots of evidence has been provided on the link between a low level of children's family SES and literacy outcomes. Van Steensel (2006), for example, demonstrated the relation between children's home literacy environments (HLE) and their literacy development in the first phase of primary school (first and second grade). They found that majority (Dutch) children and children from high SES families had, in general, the most stimulating HLEs (rich, child-directed). On the other hand, they observed considerable variability in HLEs within ethnic minority and low SES groups (often with a poor HLE). After controlling for relevant background characteristics, they found that the HLE affected children's vocabulary scores and general reading comprehension during first and second grades. A famous study of Hart and Risley (1995), conducted to look for the cause of the disparity in linguistic/academic progress among children from different socioeconomic

backgrounds, analyzed daily life parent-child interactions of 42 families for 2.5 years. The authors found that parent talkativeness accounted for a correlation between SES and the children's later academic development. In particular, they highlighted the positive relationship between a middle-high SES and a richer, conversational, and more varicoloured mother's linguistic input. Conversely, lower SES families demonstrated to be less talkative (they spoke about a 30-million-word gap). They found a strong connection between vocabulary size and reading performance, and children's SES-talkativeness. This study was one of the first to explicitly link vocabulary size to socioeconomic status and demonstrated that children's early experiences greatly impact their future ability to learn. Children's experiences in the early years of life are pivotal for language development and literacy readiness. If the child does not benefit from a significant linguistic experience in L1, within an active dialogic relationship with the parents, even the L2 risks to develop poorly for a deficiency of cognitive comparative-events (Murineddu, Duca & Cornoldi, 2006), with all the consequences in the literacy development.

Nevertheless, SES and bilingualism conditions are not two strictly interdependent factors. Engel de Abreu, Cruz-Santos, Tourinho, Martin, and Bialystok (2012) tested 40 Portuguese-Luxembourgish bilingual children from low-income immigrant families in Luxembourg, and 40 matched monolingual children from Portugal on some cognitive tasks and found that children from low SES families obtained lower cognitive performance than high SES children, regardless of their bilingualism condition. On the contrary, when matched on SES, bilingualism constitutes a protective factor. The authors demonstrated that the bilingual advantage (often associated with bilingualism in executive functioning) is neither confounded nor limited by socioeconomic and cultural factors, and it extends to young immigrant children challenged by poverty.

We will discuss the cognitive effects of bilingualism in the following section.

**Cognitive effects of bilingualism**

Whether acquiring a second language affords any cognitive advantage has been a matter of fierce scientific debate for decades, and still is. Some studies have shown bilinguals to outperform monolinguals in some general cognitive abilities: Hilchey and Klein (2011) reviewed the literature on bilingual long-term cognitive performance, compared to monolinguals, founding that bilinguals often enjoy a widespread cognitive advantage. On the contrary, other studies did not find any bilingual cognitive advantage (*e.g.*, Duñabeitia, Hernández, Antón, Macizo, Estévez, Fuentes, *et al.,* 2014; Paap, 2014). Some researchers questioned the validity of results showing a bilingual advantage (see de Bruin, Treccani & Della Sala, 2015; Valian, 2015), while others interpreted the null results as a typical manifestation of variation, where different types of bilingual experience, of language pairs, and societal context may lead to different effects on cognitive abilities (Garaffa, Beveridge & Sorace, 2015). In a very recent study, Nichols, Wild, Stojanoski, Battista, and Owen (2020) assessed more than 11,000 bilingual participants on a broad battery of cognitive tasks and found a bilingual advantage over monolinguals just on working memory test (Digit Span). However, this effect disappeared when the groups were matched to remove potentially confounding factors. DeLuca, Rothman, Bialystok, and Pliatsikas (2019) suggested to treat bilingualism as a *continuous* factor (or a combination of factors), where individual experience, made up of Age of Acquisition, language combination, patterns of bilingual use (home use and social use), educational level, societal attitudes, *Etc.*, brings to differential effects on the brain structure. Ultimately, any bilingualism benefit is pertinent to the specific sample tested with its nuances, and findings from traditional cross-sectional comparisons are not easily generalizable to all bilinguals. Participants are always embedded in different sociolinguistic contexts, and this may influence outcomes (Surrain & Luk, 2019). With this in mind, it is essential to gather data from different types of bilinguals, with different language backgrounds and in different bilingual settings, to gain a fuller picture of the effects of bilingualism in cognitive domains (Garaffa *et al*., 2015). We should also point that while

some studies find advantages and other do not, no study report a clear and lasting disadvantage due to bilingualism.

Besides these considerations, some aspects of cognitive assessment in bilingual studies – especially with low SES immigrants – is often used as a control measure to ensure that different groups are cognitively matched. Robust effects of SES have been found across cognitive skills, including memory and intelligence (Bradley & Corwyn, 2002), showing a relation between higher SES and better outcomes. Calvo and Bialystok (2014) found no effect of SES in basic cognitive skills for all children – regardless of their bilingual or monolingual condition – on nonverbal reasoning and visual search (they used Matrices Subtest of the Kaufmann Brief Intelligence Test, Kaufman & Kaufman, 2004, similar to Raven). Thus a non-verbal cognitive task is not expected to be affected by a low SES level.

Rothou and Tsimpli (2017) introduced the biliteracy variable in the *advantage debate*, looking at a possible effect on bilingualism's cognitive functioning when supported by biliteracy. The authors compared working memory performances with a digit span task of three groups of children in grades 3 and 6 (mean age 9–11 years old): Albanian-Greek monoliterates, Albanian-Greek biliterates, and Greek monolinguals. They found that Albanian-Greek biliterates had a higher performance in working memory than their Albanian-Greek monoliterate peers. The authors argued that a biliteracy effect appeared in the enhanced working memory skills, compared to monoliteracy education. Nevertheless, they acknowledged the need for further investigation with a larger sample of participants and with a variety of verbal working tasks.

Our study will expand the biliteracy *advantage debate*, investigating the effect of biliteracy on two cognitive skills: fluid intelligence and verbal short-term memory. In addition, it will investigate the effects of biliteracy on literacy outcomes in the L2.

The following section will illustrate biliteracy in terms of academic outcomes.

**Dual language education: when bilingualism means biliteracy**

Growing up bilingually and acquiring two languages in their spoken and sometimes written form has been shown to influence literacy development positively. Several studies of children from different first-language backgrounds and educational settings have suggested that first-language phonological processing skills are related to the development of word reading skills in English. A study by Quiroga, Lemos-Britten, Mostafapour, Abbott, and Berninger (2002), for example, explored the relation between phonological awareness in L1 and decoding skills in L2 in a group of Spanish-English (Spanish = L1; English = L2) first-graders from immigrant families. Bilinguals received their school instruction in English and had not had systematic instruction in Spanish. The researchers found that phonological awareness in Spanish (their first language) was highly correlated with word and pseudoword reading in English (second-majority language), the language of their instruction. Thus, phonological awareness may transfer across first and second languages and across oral and written language.

Mumtaz and Humphreys (2002) considered L2 literacy's effects when L1 literacy is also acquired. The authors explored Urdu's effects on the acquisition of L2 English literacy skills by comparing the reading, memory, and phonological processing skills in English of biliterate Urdu-English children, with those of a group of monolingual English school-aged children (7- to 8-year old). Urdu is a language with a shallow alphabetic orthography, while English has a deeper orthography. They found that bilingual children were better at reading regular words and pseudowords, compared to monolinguals. The authors interpreted this finding as the result of enhanced phonological skills transferred from their first language to the reading development in their second language. Learning to read in a language with a shallow alphabetic orthography, such as Urdu, may depend primarily on phonological processing skills, whereas learning to read in a deeper orthography, such as English, may place more reliance on visual processing skills. Arguably, in this biliterate population, first language skills facilitated the development of both lexical and phonological routes to reading. Niolaki and Masterson (2012) went in the same direction in their

study, suggesting that in biliterates, decoding processes are transferred from one language to the other. The authors investigated word spelling performance in English L2 in a group of biliterate Greek-English bilingual children (aged from 6 to 10 years), compared with two groups of monolingual English- and monolingual Greek-speaking children. The results showed that in the biliterate children with stronger Greek literacy skills, phonological awareness significantly predicted English spelling. This finding suggests that spelling skills in English – a language with a deep orthography – could be enhanced by adequate decoding skills in a language with a transparent orthography, like Greek. Phonological awareness was predictive of spelling ability also in the monolingual Greek group but not in the English monolingual group.

A very recent study by Papastefanou, Powell, and Marinis (2019) involving forty Greek-English bilingual primary school children living in the UK confirmed the significant cross-language association: phonological awareness in the heritage language (Greek) was positively related to decoding skills in the majority language (English). Children involved in Papastefanou *et al.* (2019) attended an English mainstream primary school and a Greek supplementary school, thus they were biliterate bilinguals, and came mostly from families of average and above-average socioeconomic status.

The literature so far would suggest that literacy in the majority (L2) language might be supported by the skills acquired in the minority (L1) language, arguably further when the minority language is also acquired in its written form. This evidence is valid, at least when the majority L2 is a language with a deep orthography, such as English, and the minority L1 a language with a shallow orthography. What is not clear is whether learning to read in the heritage language supports literacy skills in the majority language, even when the majority language has a transparent orthography, such as Italian. Geva (2006) argued that the effect might vary, depending on the similarity and differences between the first- and second-language orthography. In addition, students' level of first-language oral proficiency and literacy is likely to influence the relationship.

Of course, the positive effect of heritage language literacy acquisition is vital for maintaining the heritage and cultural identity and should always be encouraged. However, its possible beneficial effect for developing the literacy skills to be acquired in the majority language could have significant practical implications on educational policy.

Given the recognized importance of the heritage language maintenance, *Heritage* and *Dual Language Education Programs* have proliferated, especially in school contexts with many language-minority students (*e.g.*, Spanish-English teaching in USA schools). We will present in detail this type of education in the following dedicated section.

**Bilingual Education Programs**

The influx of immigrant students' presence in classes during the past several decades has presented challenging pedagogical issues to educators in public schools, and the number of immigrants continues to grow in more industrialized countries. The lack of L2-majority proficiency is an obstacle for L2-learners, not only in terms of social relations and inclusion in society (Cummins, 1989) and for the academic survival of these students. As a result, a significant challenge that schools face is to prepare immigrant L2-learners to be able to participate competitively in academic life and job market (Tong, Lara-Alecio, Irby, Mathes, & Kwok, 2008).

A great number of educational programs for bilingual children have been implemented in countries with a high presence of minority language-speaking citizens, declined in various models and approaches, but with the same underlying aim to boost literacy acquisition in the majority language (L1). Baker and de Kanter (1981) classified some bilingual education programs, the most represented in the USA scenario. They identified the Transitional Bilingual Education (TBE), English as a Second Language (ESL) – which they defined as a form of submersion – and the Structured Immersion (SI). TBE is a program in which subject matter is taught in the children's home language until their English-L2 is strong enough to participate in an all-English classroom, with the native language gradually phasing out and the use of English gradually phasing in. The aim

of TBE is a shift from bilingual students' L1 (a minority language) to an L2 (the majority language). The ESL is a program in which children are placed in regular English classes for most instructional days. In the remaining days or part of the day, they are withdrawn from mainstream classes to concentrated instruction classes in order to attend compensatory lessons still in the majority language. Thus, the learning focus is only English. Lastly, SI is structured so that students may acquire the language of instruction (majority-second language) while simultaneously learning content. In the SI approach, the teacher knows the children's first language; however, she rarely speaks children's L1 in the classroom. According to Baker (2001), these bilingual education approaches are *weak* and do not promote children's L1 or strengthen their cultural identity. Instead, they aim to foster monolingualism. Baker also identified some *strong* forms of dual-language education that seek to promote bilingualism, maintenance, pluralism, and cultural enrichment: *immersion* programs educate students in both L1 and L2, and they vary according to the amount of time spent in L2 immersion; *Heritage Language Education* occurs when instruction is carried out in two languages, with one of them being a minority language; *Two-Way* or *Dual-Language Education* provides bilingual instruction and is usually applied to school contexts which display equal numbers of language-minority and language-majority students.

Dual-Language Programs, in which formal instruction occurs in L2 and L1, have started to arise in those countries that faced immigration earlier and massively, as Hispanic immigrants in certain States of the US. The purpose was to support the heritage culture, but also to enhance the learning of literacy in the second-majority language (English). As Baker (2001) pointed out, children in Heritage programs often show a positive attitude towards this kind of bilingual education. Moreover, they can maintain their home language without loss in academic achievements, as L2 performance is comparable to that displayed by their mainstream peers. The author concludes that strong bilingual education programs are to be preferred to mainstream monolingual education because they seem to result in overall higher academic achievements.

More recent research has provided convergent evidence that instruction through immigrants' L1-minority language could facilitate (or even accelerate) the reading achievement in their L2 (e.g., Francis, Lesaux, & August, 2006). A meta-analysis of seventeen studies by Rolstad, Mahoney, and Glass (2005) provided empirical evidence that strong bilingual education is more beneficial for English Language Learners students than all-English approaches such as ESL and SI. From the review, Heritage or Dual-Language approach to bilingual education resulted in better performance on various academic achievement measures in L2. Tong *et al.* (2008) found that Spanish-English-Language Learners (ELLs) acquired academic English oracy[4] at a constant rate within Dual-Language program models. They concluded that instruction in the first-minority language did not impede the learning of the second-majority language, but contrarily, it enhanced and accelerate L2 literacy acquisition, regardless of a lower initial level of English-language proficiency. Therefore, enhancements and best practices in dual-language education programs should be fostered to diminish possible disadvantages of low levels of English proficiency that ELLs should bring at school entry.

As far as second-language learning is concerned, the age of first exposure might vary from one child to another: students could face the second language in kindergarten or nursery school, or later, starting from primary school. Secondly, the child's first-minority language could be more or less typologically close to the second-majority language in terms of orthographic systems, phonology, and morpho-syntactic rules, modulating the effect of the *transfer* between the languages. We will discuss the Age of First exposure to L2 and the specific language pair as variables to consider in the next section.

**Bilingualism: Age of First exposure and languages involved**

The language proficiency depends on a range of contextual factors, besides the use of each language (Silva-Corvalán & Treffers-Daller, 2015). Among these are whether bilinguals use each

[4] The ability to express oneself fluently and grammatically in speech.

language at the same rate (Montrul, 2008, 2016) and whether children develop literacy skills in one or both languages (Papastafanou *et al.*, 2019).

In this section we will reconsider some topics we already discussed in the first chapter of the present thesis.

The Age of First Exposure may also be a factor that influences the relationship between second-language oral proficiency and various reading skills among language-minority children (Geva, 2006). Kovelmann *et al.* (2008) demonstrated that reading acquisition relies on the age of oral language first exposure and oral language proficiency. The authors compared a sample of 8-years-old bilingual children exposed to English before age 3 (*early* bilinguals) and a group exposed later (*late* bilinguals) to a group of English monolingual peers. They found that early bilinguals achieved the same reading proficiency as monolingual English children, whereas late bilinguals lagged behind them in decoding, phonological awareness, and passage comprehension tasks. Although most studies focused on English, a few studies have reported similar results in other languages.

In a study on the Italian-L2 literacy development of immigrant primary school children in Italy, Bonifacci and Tobia (2016) compared the reading speed and accuracy, and reading and language comprehension performance of early bilinguals (EB), late bilinguals (LB) – with different L1s – and a control group of monolingual primary school students in grades 1 through 5. EB and LB groups were selected using Kovelman *et al.*'s (2008) criteria: EB were exposed to Italian-L2 from birth or within the first four years of age; LB after 4 years of age. In particular, they found highly discrepant decoding skills between early and late bilinguals, regardless of their L1. Early bilingual children's performance was very similar to that displayed by the control group; on the other hand, late bilinguals were outperformed by the control group. In another study on Italian-L2 literacy acquisition, Bellocchi, Bonifacci, and Burani (2016) found that early (exposed before age 4) and late (exposed after age 4) Italian-learning children aged between 8 and 10 achieved within monolingual norms decoding skills. However, those exposed after age 4 were relatively slower,

made more errors, and were more sensitive to frequency effects than those exposed earlier. So, the age of first exposure seems to play an essential role in achieving adequate proficiency in L2 learners. It should be noted that this study focused on the acquisition of an L2 with a transparent orthography (Italian), complementing previous research conducted mainly on children acquiring an opaque orthography, such as English, as a second language.

Another critical factor, which has been seen to influence bilingualism outcomes, is the relatedness of language pairs as writing systems, or 'language proximity', as called by Grohmann (2014). That is, the closeness or distance between the two languages a child acquires.

Languages have different writing systems. Without claiming to be exhaustive in definitions herein (see Sampson, 2015 for a broader disquisition on the topic), we could summarily distinguish between logographic, syllabic, or alphabetic systems. This first tripartite scheme has remained the most popular within the study of writing (Daniels & Bright, 1996). Taking the logographic category first, the most common example to cite is the Mandarin Chinese script. Simplistically, the Chinese writing system is based on *characters* containing the whole word meaning and hints at its pronunciation. Thus, Chinese characters have a bipartite structure, combining a *phonetic* with a *signific* element (Sampson, 2015). These two individual parts forming a character are called radicals, one hinting at its pronunciation and the other to its meaning (Shakkour, 2014). Syllabic systems, such as Japanese or Korean, establish correspondences between signs and various consonant-vowel groups. Thus syllables have a specific graphical representation. It derives that children need to rely on graphic-symbol recognition and storage abilities when learning to read in a language based on a logographic or a syllabic writing system. Alphabetic languages ground on written symbols (or symbol groups), the *graphemes,* which univocally represent a single sound of that language (*phonemes*). Phonemes combined together – following language-specific phonotactic rules – can create infinite words (Perfetti & Marron, 1998). Graphemes could have different *scripts*, depending on the specific alphabet the language relies on. Different types of alphabets exist – for example, Latin, Greek, Cyrillic, Arabic – with own peculiarities. Even though different scripts can

be used, they share the essential feature that graphemes represent phonological segments. The Arabic alphabet, for example, is called *consonant*, as it has not graphemes representing vowel sounds. Moreover, letters could be depicted in different modes, depending on their position in the word (initial, middle, final, or isolated), as it is for some graphemes in the Greek alphabet. However, Arabic is considered an alphabetic writing system with a regular nature. The first crucial step when learning to read an alphabetic script is, therefore, to acquire a basic understanding of the system that maps symbols to sounds (Ziegler & Goswami, 2005), i.e., graphemes into phonemes (and then *phonos*, when articulated). This is the decoding ability.

Neuroimaging reading studies have identified unique patterns of activation for individuals reading in languages with different writing systems, suggesting the involvement of different processes and neural networks (Perfetti, Liu, Fiez, Nelson, Bolger & Tan, 2007). Moreover, the decoding ability requires less learning time to establish, compared to the memorization of all the logograms necessary for reading acquisition in languages with a logographic system (Bialystok, McBride-Chang & Luk, 2005).

Several studies found that biliteracy could give rise to different reading outcomes, depending on the languages' specific writing systems. In general, most of the studies of the last two decades have examined the relationships between first-language (L1) alphabetic or logographic orthography and English-L2 in terms of cross-linguistic transfer of orthographic knowledge, phonological awareness, morphological awareness, and reading skills (see Shakkour, 2014 for a review). Bialystok, Luk, and Kwan (2005) compared the early English literacy skills of a group of English monolinguals with three groups of biliterate bilingual peers, each differing from their first language writing systems: Cantonese, Hebrew, and Spanish. The authors found a substantial advantage for children learning two alphabetic systems, suggesting a significant *transfer* of literacy skills across their languages, or from one language to another. Another study by Bialystok *et al.* (2005) went in the same direction. The authors compared phonological awareness and word decoding tasks in English in three groups of 5- and 6-year-old children: monolingual English,

bilingual English-Mandarin Chinese, and Cantonese-speaking children beginning to learn English (2nd-language learners). The former two groups lived in Canada, where the predominant language and language of schooling is English; the third group lived in Hong Kong, where the predominant language and language of schooling is Cantonese. The latter group of children spoke Cantonese at home and in the community and learned English as a school subject. School instruction was English only during language classes. Results established that the decoding ability developed separately as a mere function of proficiency and instructional experiences in English, with no transfer from Mandarin or Cantonese (both logographic systems) to English. Therefore, there was no overall influence of bilingualism on learning to read in English (even if we must recognize that only the bilingual English-Mandarin-learners could be considered biliterate at some extent). Previous research with bilingual children has reported transfer of decoding skills across languages in the early stages of literacy (*e.g.*, Geva & Siegel, 2000) for languages with the same writing system; for languages with different systems, there is no evidence for such transfer (Gottardo, Yan, Siegel & Wade-Woolley, 2001; Huang & Hanley, 1994). An exception to the transfer of decoding skills between languages with the same writing systems is the previously mentioned study by Rothou *et al.* (2017), who did not find such a transfer. The authors aimed to investigate the effect of biliteracy on working memory, as we said earlier, but they also looked at Greek word recognition and reading comprehension. To recall, they examined the performance of three groups of children in grades 3 and 6: monoliterates Albanian-Greek, biliterates Albanian-Greek, and Greek monolinguals. Greek and Albanian have both an alphabetic system (although they use different alphabets, Greek and Latin, respectively). Biliterate children learned to read in Greek (L2) and Albanian (L1) while monoliterate in the second language only, *i.e.*, Greek. With regard to word recognition, it was found that monoliterate Albanian-Greek children did not differ from their monolingual peers, despite differences in oral language proficiency. On the opposite, biliterate Albanian-Greek children had lower performance than their monolingual peers. Rothou and Tsimpli interpreted this result as a competition between the two languages because of Albanian and Greek orthographic systems'

similar level of transparency. They suggested that monoliterate children had better performance on word reading than biliterate bilinguals, as the latter, when presented with written stimuli, experienced higher competition and interference from the minority language than their monoliterate peers, who are exposed to Albanian only orally. Concerning reading comprehension, biliterate children had lower performance on the reading comprehension task than both their monoliterate and monolingual peers. Thus, results from this study might suggest that being able to read in both Albanian and Greek may not improve Greek literacy skills, although, as we said earlier it enhances working memory performance.

To summarize, whether bilingualism benefits children in learning to read in an L2 seems to depend on the system of scripts: if they both employ the same writing system, a cross-linguistic transfer can be more effective (even though some exceptions were found). In addition to this, another factor that play a role in biliteracy acquisition are the age of exposure to the L2.

**The Italian context and the study on Italian L2 pupils**

According to the most current data from the Italian Educational Ministry (MIUR, 2020), in the 2018/2019 school year, 30.1% of non-Italian students in mainstream Italian *curriculum* were delayed in academic achievements, against the 9.1% of Italian pupils. As said earlier, children of immigration background typically get regular exposure to Italian only when they enter the public education system, between age 3 and 6, becoming literate in Italian and maintaining their mother tongue only in oral form in the family context. This gives rise to a disadvantage regarding their level of Italian and a subsequent literacy gap between them and their monolingual peer in the same class.

At the current time, this fragile situation in classes is entirely managed by the teachers, who feel the need to become skilled in dealing with immigrants' educational special needs (*i.e.*, more intensive support to literacy), but who also have to teach to a whole class at the same time. Thus, a feeling of overcharge is quite understandable. Moreover, it is possible that foreign children, not

being sufficiently supported in their literacy needs in Italian, are invited to consult clinical services by teachers, with clinicians facing an increment in assessments and consequent stress imposed on the Public Clinical Units for Infancy. New language support strategies for these children are needed. Surely, they need to boost their Italian oral competence to access literacy. Unfortunately, a dedicated learning-support extra-time is not always guaranteed in Italian mainstream schools.

Moreover, the literature, which investigated the role of the first language – also in its written acquisition – suggests that actively supporting L1 with a Heritage-Dual-Language literacy would bring benefit to literacy in the L2-majority language (*e.g.*, Papastefanou *et al.,* 2019; van der Velde Kremin, Arredondo, Hsu, Satterfield, & Kovelman, 2019). Despite a lot of evidence, only a few initiatives to support heritage languages have been explored or implemented in mainstream schools in Italy. There are still limited witnesses of the effort and the recognition of the importance of promoting multiculturalism and sustaining all the languages spoken by foreign students. This is due to the lack of resources, which afflict educational policies, the lack of understanding about the importance of supporting the heritage language, or the misconception that support of the heritage languages may be at the majority language's expense and may affect integration in the society. We should acknowledge that some studies did not find actual advantages of biliteracy and this may be due to the different writing systems and in one case to other individual and environmental factors (*e.g.*, Rothou *et al.*, 2017). While the bilingualism factor has been previously studied with Italian L2 learners (*e.g.*, Bonifacci *et al.*, 2016), the biliteracy factor has rarely been investigated in a systematic and quantitative way, particularly with students with an immigration background. The present study contributes to this *lacuna*.

**Aim and hypothesis**

Children with immigrant backgrounds facing Italian literacy at school often struggle more than their Italian monolingual peers, and learning appears to be slower or more difficult for them, compared to monolingual peers. Besides specific, innovative, and up-to-date didactic strategies to be

implemented by teachers, and extra support to Italian L2, it might be useful to sustain and boost children's minority L1 in both its oral and written form (see Papastefanou *et al.* 2019). Our study was based on related research and extended it by exploring the effect of biliteracy on reading abilities of bilingual immigrant children – of different L1 extractions – who learn to read in an orthographically transparent language as Italian.

Exploring biliteracy effects on immigrant students in the Italian context would have significant implications both for the Italian educational and health policy: it could be considered as a national standard to guarantee in schools, even involving communities and public institutions such as Consulates; it could reduce the unnecessary clinical assessment, thus improving the management of users of the Public Clinical Services.

In the present study, we first asked whether biliteracy could raise any strengths or advantage (or weaknesses) in bilingual students on Italian literacy and literacy-related achievements. Thus the first aim was to investigate how primary school bilingual children of immigrant extraction and with different first languages perform in some cognitive and language domains, and reading skills, when their first language is or is not acquired in its written form, besides the oral one. To address this aim, we compared performance on cognitive, linguistic, and reading abilities in Italian of a group of bilingual-biliterate children – attending an Italian mainstream primary school and a Heritage language supplementary school – with the performance of a group of bilingual-monoliterate children, attending only an Italian mainstream primary school. This study investigated, for the first time, the effect of biliterate bilingualism on cognitive, linguistic, and reading skills in Italian L2 immigrants.

As regard to cognitive effects, biliterate children are supposed to be more skilled bilingual because of their higher exposure to their first language, both at home, within their parents or siblings, and at the community school, where teachers speak with them and teach them the written language. Monoliterate bilinguals may result in a relatively lower degree of bilingual proficiency (Blom, Boerma, Bosma, Cornips & Everaert, 2017) since they only have oral and in-home exposure

to L1. Given previous observations that a certain level of bilingual proficiency is required for cognitive effects to develop (see Videsott, Della Rosa, Wiater, Franceschini & Abutalebi, 2012; Weber, Johnson, Riccio & Liew, 2016), the biliterate group may gain a cognitive benefit because of their higher bilingual proficiency.

Bilinguals' expressive vocabulary has been found to be commonly weak, compared to monolinguals, if it is assessed in one single language only (*e.g.*, Bialystok, Luk, Peets & Yang, 2010; Hoff, Rumiche, Burridge, Ribot & Welsh, 2014). Biliteracy should not influence the vocabulary width in Italian L2 – as to be literate also in the L1 should reinforce L1 vocabulary, if anything – so we expect similar performance between bilingual groups.

We plausibly expected a beneficial effect of biliteracy on reading measures tested following Papastefanou *et al.* (2019). Nevertheless, being aware of Rothou *et al.* (2017)'s findings, of the heterogeneous bilingual population (in terms of different first and second languages, age of the first exposure to Italian, socioeconomic background), and due to the paucity of evidence on the specific languages couple, we reckoned with the possibility that no effects would emerge. In particular, we expected to find a possible advantage in reading performance in biliterate children, compared to monoliterate, at least because of their more extensive reading exercise.

Secondly, we wanted to ascertain whether bilingual students automatized reading around age 8 (Grade 3), following a reading development as in monolingual children. This is of crucial relevance in order to identify reading disorder among this population. To address this question we analyzed when strategies for reading words and pseudowords start to diverge in biliterate and monoliterate bilinguals. We predicted that both groups reached mature decoding skills by Grade 3 – with words read faster and better than pseudowords –, replicating Bellocchi *et al.* (2016), who found that Italian-L2 children aged between 8 and 10 years old performed equally within norms, regardless of their early or later exposition to Italian L2.

We also took into account the role of cognitive skills and oral language proficiency in children with Italian L2. Research has repeatedly demonstrated the contribution of cognitive and

language skills to literacy development; thus the examination of the relationship between bilingualism, working memory, and oral language skills seems to be a key piece in the puzzle of biliteracy (Rothou *et al.,* 2017). We aimed to address whether there were relationships between the cognitive, linguistic, and reading abilities tested, and identify positive or negative associations. We explored the correlations between cognitive and oral language abilities and reading measures in both groups of bilinguals. As cognitive and linguistic factors have been shown to contribute to the development of the children's reading skills, we expected to find cognitive and language performance to correlate with reading proficiency measures (see Gathercole, Brown & Pickering, 2003; White, 2019, for link between cognitive abilities and reading skills; see Duff, Reen, Plunkett & Nation, 2015, for link between vocabulary and reading ability).

To summarize, analyzing with a cross-sectional study a biliterate group of bilinguals and a monoliterate group of bilinguals (being schooled only in Italian L2), we aimed at establishing:

1) Whether bilingual-biliterate children gain some advantage with respect to bilingual-monoliterate children in cognitive, vocabulary and reading in Italian (L2);
2) Whether the course of reading development for both bilingual-biliterate and bilingual-monoliterate proceeds as in monolingual children, with a dissociation between reading words and pseudowords around the age of 8 years (Bellocchi *et al.*, 2016);
3) Whether there are correlations between the various cognitive, memory, language, and reading measures.

Our biliterate groups was not homogeneous as it comprised biliterate children with alphabetic (Polish, Arabic, Russian) scripts and biliterate children with a logographic (Chinese Mandarin) script. While in the first set of analyses, we kept them all together, in a further analysis, we kept them separate to verify whether there was any difference between the two groups. We did expect to find the alphabetic biliterate group to perform better than the Chinese Mandarin L1 biliterate group, as for the first group only we expected effects of a cross-linguistic transfer (Bialystok *et al.*, 2005; Geva *et al.*, 2000). We did these analyses separately, as our ML

population's size was larger than that of the two BL samples, while the two BL Groups of were comparable in size.

The study was approved by the Ethic Committee of the University of Milano-Bicocca (n. 281).

## Material and methods

### Participants

One hundred and ninety-one bilingual students were recruited from four different public schools in the Province of Milan and six *community schools* in the Milan and Rome area. We recruited an Italian monolingual control group, but the testing had to be discontinued due to the pandemic Covid-19. The sample tested turned out to be too unbalanced in terms of the size of the sample, with respect to the bilingual and therefore we did not include in our analysis a control group of Italian monolinguals. Ten bilingual children were excluded because they did not complete the testing session. The final sample consisted of 181 participants. The sample was divided into two groups: 95 bilingual children attending an Italian mainstream primary school (*L2 monoliterate, ML*); 86 bilingual children attending an Italian mainstream primary school, and a Heritage language supplementary school (*biliterate, BL*). Exclusion criteria included being an immigrant living in Italy for less than three years, for first Graders. Third- and fifth Graders had to start primary school in Italy since Grade 1.

ML group consisted of 27 children from Year 1 (Mean age = 88.2 months, SD = 5.7, 10 boys and 17 girls), 49 children from Year 3 (Mean age = 110.3 months, SD = 4.5, 20 boys and 29 girls), and 19 children from Year 5 (Mean age = 128.6 months, SD = 4.4, 8 boys and 11 girls).

BL group consisted of 19 children from Year 1 (Mean age = 84.1 months, SD = 4.3, 10 boys and 9 girls), 35 children from Year 3 (Mean age = 108.9 months, SD = 5.3, 16 boys and 19 girls), and 32 children from Year 5 (Mean age = 130.3 months, SD = 6.1, 16 boys and 16 girls). As could be noted, both samples are slightly unbalanced in the three classes.

Within the BL group, there are two groups of biliterate children: one comprising 36 children with an alphabetic system (Arabic, Polish, and Russian) and one including 50 children with a logographic system (Chinese Mandarin). The two groups did not differ in terms of age or cognitive score (Raven's CPM). Details are reported in *Table 1*.

| Group | Mean Age (months) | Mean Raven score |
|---|---|---|
| **BL alphabetic (N = 36)** | 108.5 (SD = 16.7) | 30.5 (SD = 4.1) |
| **BL logographic (N = 50)** | 113.5 (SD = 19.2) | 31 (4.7) |

**Table 1** shows the number of participants, the average ages in months, and the average scores at the Raven's CPM test of the following subgroups: biliterate with an alphabetic L1, and biliterate with a logographic L1.

Heritage language schools were differently organized. Arabic and Chinese literacy were taught within one of the Italian primary schools in Milan involved in the study, implemented as part of a Heritage program to promote the Heritage culture and language. Children are usually engaged in L1 classes as an after-school activity some afternoons a week and during weekends. These mother-tongue courses are organized by immigrant communities or associations and are hosted within the school. Other Chinese biliterate students were engaged from two community schools in the Chinatown district in Milan. They usually attended Mandarin classes on weekends during winter-springtime (together with the Italian school) and almost daily in summertime (as Italian mainstream schools end in early June). Polish biliterate students were recruited from the Polish community school in Rome, as part of the Italian Polish Embassy, which promotes and supports the Heritage culture and dual-language identity. They took part in lessons two to three afternoons per week and on weekends. Finally, Russian biliterate students were recruited from the Russian community in Milan. They attended Russian classes during weekends. Thus, L1 literacy amount is unbalanced with respect to Italian L2 literacy, which occurred daily, from early September till early June.

Most children were born in Italy (second-generation immigrants), but some were born in a foreign country and moved to Italy before the primary schooling's commencement, as reported by teachers. Parents did not always fill out questionnaires about the age of first exposure to Italian L2 and language background, thus we often trusted teachers. Teachers' reports indicated that all the children were free of hearing, neurological and behavioral problems, and none of them had any history of speech or language delay or impairment. No concerns were reported about their language or literacy development.

The children came mostly from families with a low or average socioeconomic status. We tried to get SES information uniformly with a questionnaire about the academic parental level. We created translated versions of the questionnaire for each language included, and we organized a meeting to clarify and help parents understand the request. Teachers and some parents with a useful competence of Italian helped us. Unfortunately, parents' participation in most cases was poor, and we could not obtain all the questionnaires filled out. The parents' level of education was measured on a scale from 1 to 5, where 1 corresponds to no schooling and 5 to a university degree or a higher education level. Among the questionnaires filled out, none ticked a 5 level of education. The majority had a 2 to 3 level of education, very few (mostly Polish families) a 4 level (*i.e.*, high school diploma). Nevertheless, this limitation was overcome through an accurate approximation. When the questionnaire did not define SES, it was evaluated by considering the geographical living area: primary schools involved in the study are located in some peripheral neighborhoods of Milan, where the living standards are known to be medium-low. Thus, this was the best attempt to exclude the confounding effects of SES.

To summarize, the bilingual population of the present study was heterogeneous. Besides the *monoliterate* and *biliterate* status, different cultural backgrounds, mother-tongues, family's immigration history, and exposition to Italian as the L2 characterized our sample, as it is common when it comes to interface with immigrant populations. Nevertheless, we tried to ensure the best homogeneity level within and between the groups in terms of SES and age (as shown below). The

sample was also fairly balanced by sex, with 101 female participants (55.8% of the sample) and 80 males (44.2%).

## Materials

Children were tested on language and reading proficiency measures, together with a non-verbal cognitive assessment and a measure of verbal working memory. Standardized and non-standardized tasks were used. Full details of the tasks used – for the cognitive part and the language and reading part, respectively – are given below.

### *Cognitive assessment*

*Raven's Colored Progressive Matrix test*

All children completed the Raven's Colored Progressive Matrix test of general intelligence (CPM; Italian standardization by Belacchi, Scalisi, Cannoni, & Cornoldi, 2008). Cognitive assessment in bilingual studies is often used as a control measure to ensure that different groups are cognitively matched (Garaffa *et al.*, 2015). The task is regarded as a non-verbal estimate of fluid intelligence. It is internationally recognized as a culture-fair standardized test designed for children between the ages of 5 and 11 years.

Children are presented with 36 matrix pictures, made up of 6 pieces each (one of which is missing). Perceptual and conceptual matching exercises are divided into three sets of twelve items, presented to the child in sequence throughout a laptop screen (pictures were implemented in a PowerPoint file). The items increase in difficulty within a given set. Participants have to find out the right missing piece of every matrix – as part of a kind of "logical puzzle" – choosing one of six alternatives the child sees at the bottom of the matrix picture. She/he has to write down on a table the right choice with its corresponding number. The raw score is the total number of correct items out of 36.

*Pseudowords Repetition Task*

Recent uses of the pseudowords repetition task seem to accept it as a measure of phonological working memory capacity (Gathercole, Willis, Baddeley, & Emslie, 1994). The task taps many language processes, including speech perception, phonological encoding, phonological memory, phonological assembly, and articulation (Coady & Evans, 2008). Even if an overall consensus about the specific working memory ability which better correlates with reading skills has not yet been reached, pseudowords repetition is one of the most commonly used tasks, for its underlying processes linked to reading ability (Alloway, Gathercole, Adams, Willis, Eaglen & Lamont, 2005; Engel de Abreu, Gathercole & Martin, 2011), and because it has been found to be implicated in reading skills (*e.g.*, Gathercole *et al.*, 2003; White, 2019).

Phonological memory was therefore assessed using a non-standardized pseudowords repetition test (Ciccarelli, 1998) built according to Italian-phonotactic rules. It comprises 40 pseudowords of two, three, four, and five syllables, randomly ordered. The experimenter pronounces the pseudowords to be repeated, one at a time. The child has to repeat the pseudoword aloud. Scores range from 0 to 40. The test was created as part of a doctoral research project.

***Italian language and reading tasks***

*Expressive vocabulary*

We measured expressive vocabulary because of its relationship with the reading ability (Duff *et al.*, 2015). T*he Test Neuropsicologico Lessicale per l'età evolutiva* (TNL; Cossu, 2013) was used. This is a standardized test for assessing expressive vocabulary in children aged from 3 to 9 years. The test examines the nouns category only. We only administered the first step of the naming task, which is part of the *expressive vocabulary* subtest. Children have to name aloud 50 pictures that are presented with A4 size panels, one at a time. They have 5 seconds to name each picture, and in case of *anomia*, they step forward to the following picture. Some pictures have two compatible naming alternatives[5]. Nouns range from low to high frequency and are randomly ordered. Scores range

[5]. For example, *alligatore* ("alligator"*)* and *coccodrillo* ("crocodile") are both acceptable.

between 0 and 50. As we compared groups' performance, we used raw scores, and the task was administered to older subjects as well.

*Reading tasks*

Considering reading, we examined reading decoding skills and passage reading and comprehension. Decoding skills are the ability to assign the right sound of the language to the right letter of the alphabet, or sound-to-letter correspondence (Berens, Kovelman, & Petitto, 2013), and it is commonly assessed through a task of reading words and pseudowords (Compton, 2000). Reading a text is a more complex task, and it implies more competencies besides decoding: vocabulary and morpho-syntactic competence, working memory, and the ability to make inferences, comprehend, keep in mind and connect the contents of the text.

To measure decoding skills, we used the *DDE-2: Batteria per la valutazione della dislessia e della disortografia evolutiva-2* (Sartori & Job, 2007) – *subtests 2 and 3*. It is a standardized test of reading decoding skills of children aged between 7 and 14 years (although it should be noted that it was administered to younger subjects as well). *Subtest 2* is a regular word reading task made up 112 of different words. Words are subdivided into four columns to be read aloud by the child, one at a time. Each word set has an increasing level of difficulty; the first series is characterized by more concrete, high-frequency words, and the last series by more abstract, less frequent, and longer words. It assesses the *lexical way* of reading. *Subtest 3* is a pseudowords reading task organized as same as the words-reading subtest. Pseudowords to be read aloud are 48 subdivided into three series of increasing difficulty: from shorter and simpler to longer and more phonotactically difficult. Pseudowords conform to the phonotactic rules of Italian. This subtest assesses the *phonological* (or *sub-lexical) way* of reading. In both tasks, children are asked to read aloud at their best, the most accurate and rapid. The two subtests measure the reading rate (*i.e.*, velocity, calculated in syllables/seconds) and accuracy (*i.e.*, number of errors[6]).

[6]). The errors' score corresponds to the number of words and pseudowords read incorrectly. Self-correction is not counted as a mistake.

We used the *Prove MT-3-Clinica: La valutazione delle abilità di Lettura e Comprensione per la scuola primaria e secondaria di I grado* (Cornoldi & Carretti, 2016) to evaluate the text reading skills of children. This test is a standardized test that assesses reading rate (syllables/seconds), accuracy (number of errors), and passage comprehension in children aged between 6 and 14 years. Children are asked to read a passage aloud to assess velocity and accuracy. Reading comprehension instead consists of the silent reading of another passage, followed by a multiple-choice questionnaire about the story. Children are allowed to go back to the passage whenever they need to individuate the correct answer. Passages differ according to the grade the child is attending.

**Procedure**

The data collection has lasted three years, from April 2017 until June 2019. All children were tested during springtime to ensure an almost full year of formal instruction, especially for first graders.

Before starting testing, school directors were informed about the goals and procedures, and they subscribed their consent to the study. Teachers involved and informed families of the three grades to test (with a translator's help in some cases) and provided them with an informative form to subscribe. Parents who agree with their child to participate signed their informed consent.

ML and BL children were assessed via one session during school hours, in a schoolroom. ML and some BL children were tested morning time at their mainstream primary schools during regular school hours, while most BL children were tested weekend time (or weekly in the afternoon) at their community school. Children were tested individually excepted for Raven's Coloured Matrices (for intelligence index) and passage reading comprehension, administered in small groups of two-to-three children. When it was not possible to administer the task collectively, it was administered individually.

A well-prepared Italian native speaker administered all the tasks, and instructions were repeated as often as necessary, to ensure that children understood what they had to do.

## Results

### Comparison between Monoliterate and Biliterate bilinguals

In the first set of analyses, we compared the ML and BL groups' performance in Grade 1, 3, and 5. Although L1s were different in terms of the writing system (alphabetic *vs*. logographic), we firstly kept biliterate children together in a single Group (BL), matched for age and intelligence measure (see Table 1 above) to the ML children.

ML and BL did not differ in terms of age. Age details are reported in *Table 2*.

| Grade | 1 | | 3 | | 5 | |
|---|---|---|---|---|---|---|
| | Mean Age | SD | Mean Age | SD | Mean Age | SD |
| **ML** | 88.2 | 5.7 | 110.3 | 4.5 | 128.6 | 4.4 |
| **BL** | 84.1 | 4.3 | 108.9 | 5.3 | 130.3 | 6.1 |

**Table 2** shows the average ages in months and the Standard Deviations of the ML and BL Groups in Grade 1, 3, and 5.

An analysis of variance was performed, with *Group* (ML and BL) and *Grade* (first, third, and fifth) as independent variables. Scores at a particular test were the dependent variable. We will report significant main effects and interactions. Bonferroni *post hocs* were used to make post hoc comparisons.

One BL bilingual first-grader's reading comprehension score was missed and was replaced with the average score of the BL first-grade group.

More specific details regarding the groups can be found in the *Appendix* section.

### *Cognitive measures*

First, we analyzed the nonverbal intelligence measure (Raven's CPM). Details are reported in *Table 3*.

| Grade | 1 | | 3 | | 5 | |
|---|---|---|---|---|---|---|
| | Mean Raven score | SD | Mean Raven score | SD | Mean Raven score | SD |
| **ML** | 22.4 | 6.4 | 27.9 | 4.5 | 29.5 | 5.4 |
| **BL** | 28.2 | 4.6 | 31.1 | 4.3 | 32 | 3.9 |

**Table 3** shows the average Raven scores and the Standard Deviations of the ML and BL Groups in Grade 1, 3, and 5 (max 36).

Raven score improved across the three Grades, both in ML and BL bilinguals. Moreover, BLs were better than MLs at the IQ test.

An ANOVA with Raven score as dependent variable, and Group and Grades as the independent variables showed an effect of Group $F(1,175) = 25.6$, $\eta^2 = 0.13$, $p < 0.001$, with BL being better than ML, and an effect of Grade $F(2,175) = 17.1$, $\eta^2 = 0.16$, $p < 0.001$.[7] Bonferroni *post hoc* showed that first-graders differed from third- and fifth-graders ($p < 0.01$).

Second, we analyzed the verbal short-term memory measure. *Figure 1* reports the results in the pseudowords repetition task. As evident, ML bilinguals seemed to be slightly better than BLs. Besides, the skill seems to improve across grades.

[7] In our analyses, we report the Eta-squared ($\eta^2$), which is a measure of effect size for use in ANOVA. It ranges between 0 and 1. Effect sizes follow this interpretation: .01 ~ small; .06 ~ medium; >.14 ~ large.

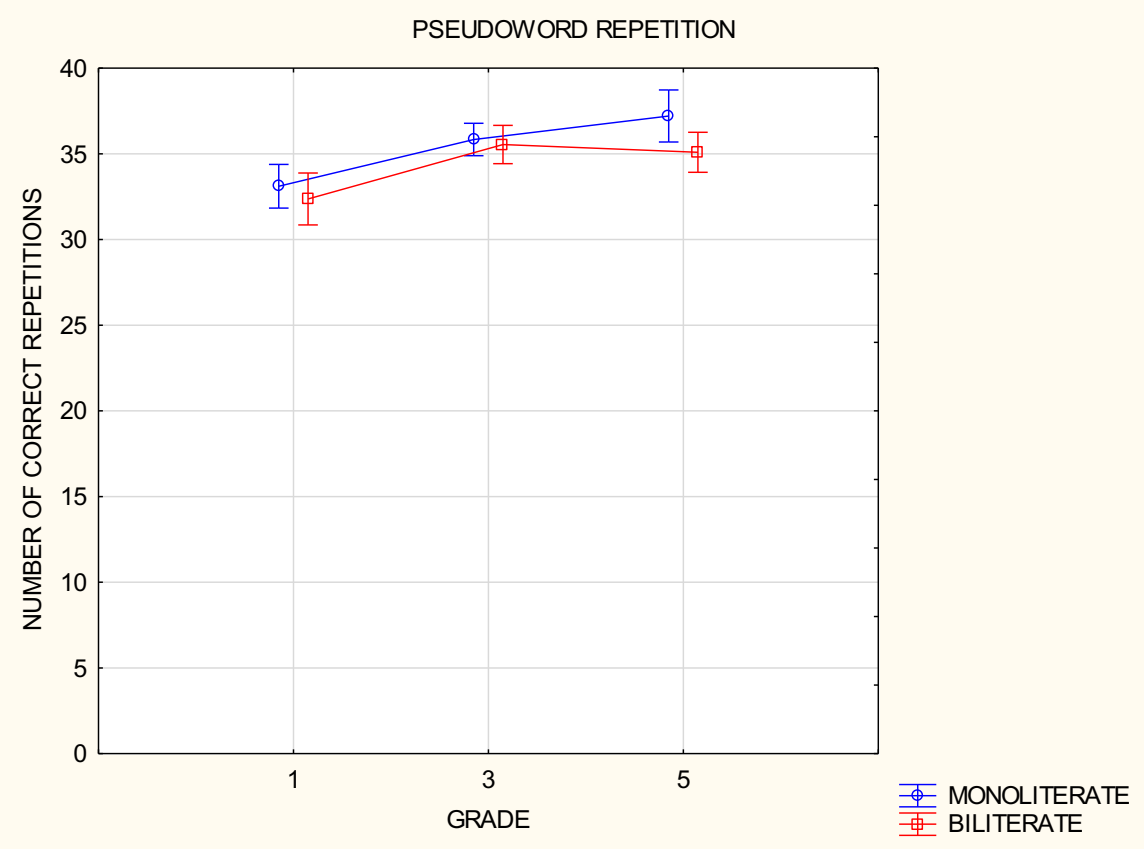



**Figure 1** shows raw scores (Max=40) at the pseudowords repetition task in Italian in Grade 1, 3, and 5, of ML and BL bilingual children. Bars indicate a confidence interval at 0.95.

The ANOVA confirmed an effect of Group $F(1,175) = 3.98$, $\eta^2 = 0.02$, $p < 0.05$, and of Grade $F(2,175) = 14.69$, $\eta^2 = 0.14$, $p < 0.001$ on *Pseudowords Repetition Task.* Bonferroni *post hoc* showed that first-graders differed from third- and fifth-graders and that monoliterate were better than biliterature children, although the effect was small ($p < 0.05$).

### *Vocabulary measure*

Secondly, we analyzed expressive vocabulary. In *Figure 2*, we can observe scores at expressive vocabulary test. Vocabulary improved from first to third Grade, in both Groups of children.

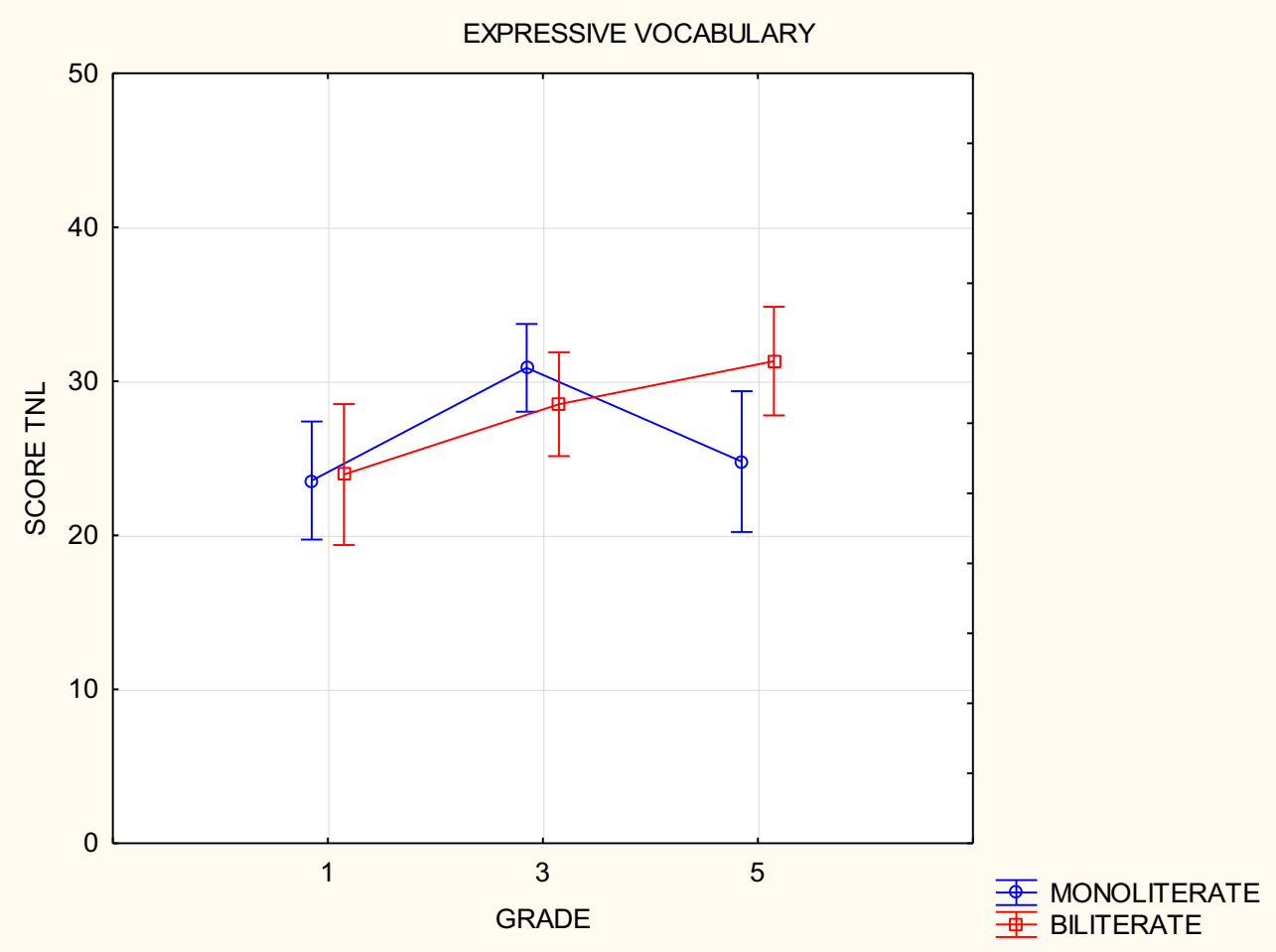



**Figure 2** shows raw scores (Max=50) at expressive vocabulary in Grade 1, 3, and 5, of ML and BL bilingual children. Bars indicate a confidence interval at 0.95.

The ANOVA with Group and Grades as the independent variables, and scores at the *Test Neuropsicologico Lessicale per l'età evolutiva* (TNL) as the dependent variable, found only an effect of Grade $F(2,175) = 5.03$, $\eta^2 = 0.05$, $p < 0.05$. Bonferroni *post hoc* showed that first-graders differed from third- and fifth-graders ($p < 0.05$). No Group effect was found.

***Reading measures***

Third, we analyzed reading proficiency measures. We will first present analyses on decoding tasks, and then we will introduce passage reading measures.

*Figure 3* illustrates the two Groups' reading speed performance at word and pseudoword reading, measured in syllables/seconds. As evident, reading decoding skills similarly increased across Grades in the two Groups (with more syllables read per second). In first Grade, as children

have started to approach the process of reading, we can notice a similar level of reading velocity in both items (words and pseudowords). From third Grade, word reading increased more than pseudoword reading, due to automation of the reading strategy, as it would be expected, until reaching a full reading competence in fifth grade.

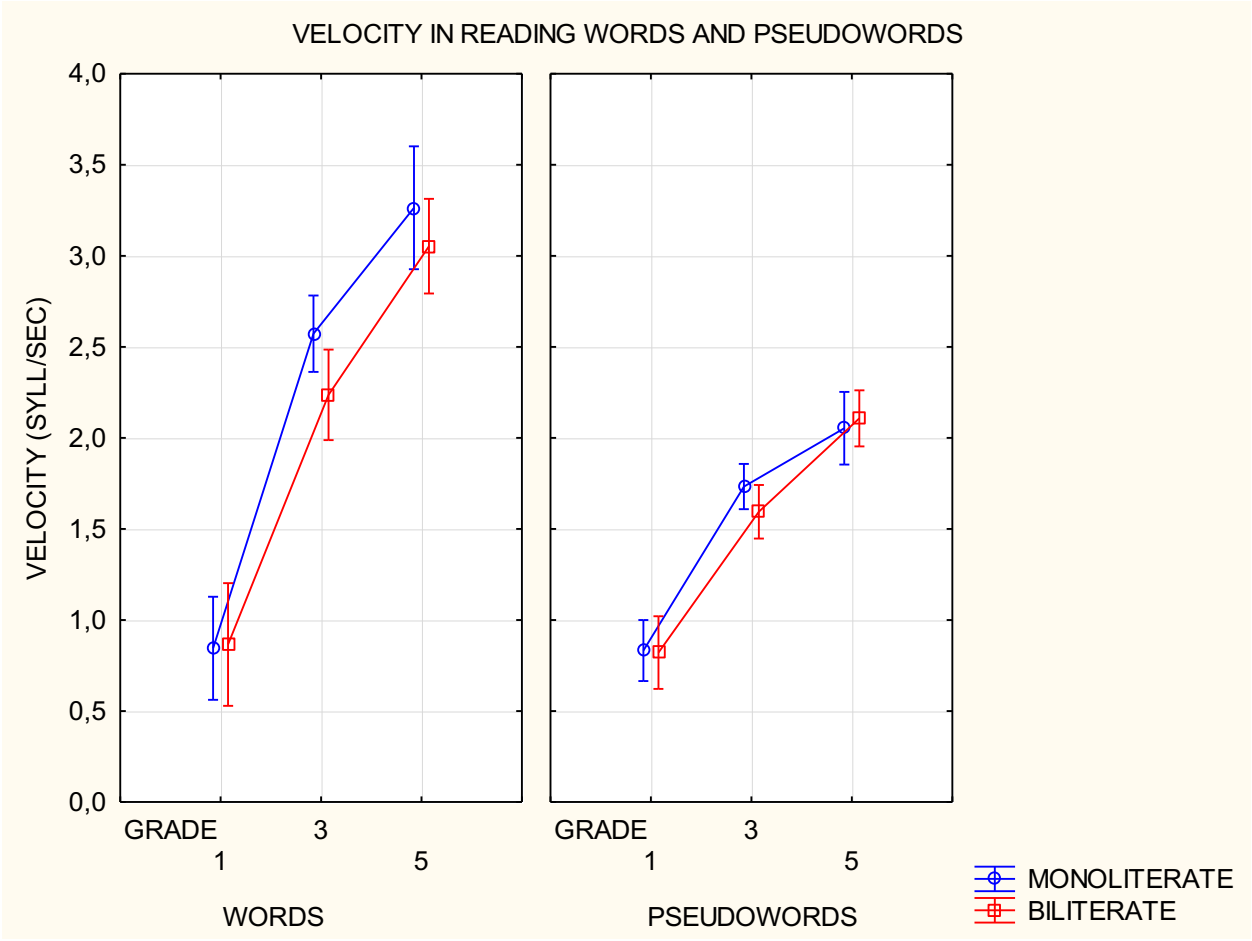



**Figure 3** shows raw scores at the word and pseudoword reading rate measure in Grade 1, 3, and 5, of ML and BL bilingual children. Bars indicate a confidence interval at 0.95.

The repeated measures ANOVA with Group (ML, BL) and Grade as between-subject factors and Item (Words, Pseudowords) as within-subject factors showed an effect of Grade $F(2,175) = 117.61$, $\eta^2 = 0.57$, $p < 0.001$. Bonferroni *post hoc* showed that classes differed from each other ($p < 0.01$). In addition, we found an effect of Item $F(1,175) = 302.18$, $\eta^2 = 0.63$, $p < 0.001$, with word read faster than pseudowords. The interaction between Item and Group was significant $F(1,175) = 4.11$, $\eta^2 = 0.02$, $p < 0.05$, due to a deviation of the parallelism. The ML Group was faster that BL in reading words (even though the difference was not significant), whereas the two Groups'

performance overlapped in reading pseudowords. An effect of Interaction between Item and Grade was also found $F(2,175) = 65.64$, $\eta^2 = 0.43$, $p < 0.001$ because words and pseudowords in first Grade were read at the same pace, whereas from Grade 3, words were read faster than pseudowords (Bonferroni *post hoc* tests $p < 0.05$).

Accuracy at word and pseudoword reading (*i.e.*, the number of errors that the children committed while reading) was analyzed in the two Groups. *Figure 4* shows the performance of ML and BL Groups. We can observe that both Groups committed more errors in Grade 1, especially with words, whereas the performance decreased from Grade 3.

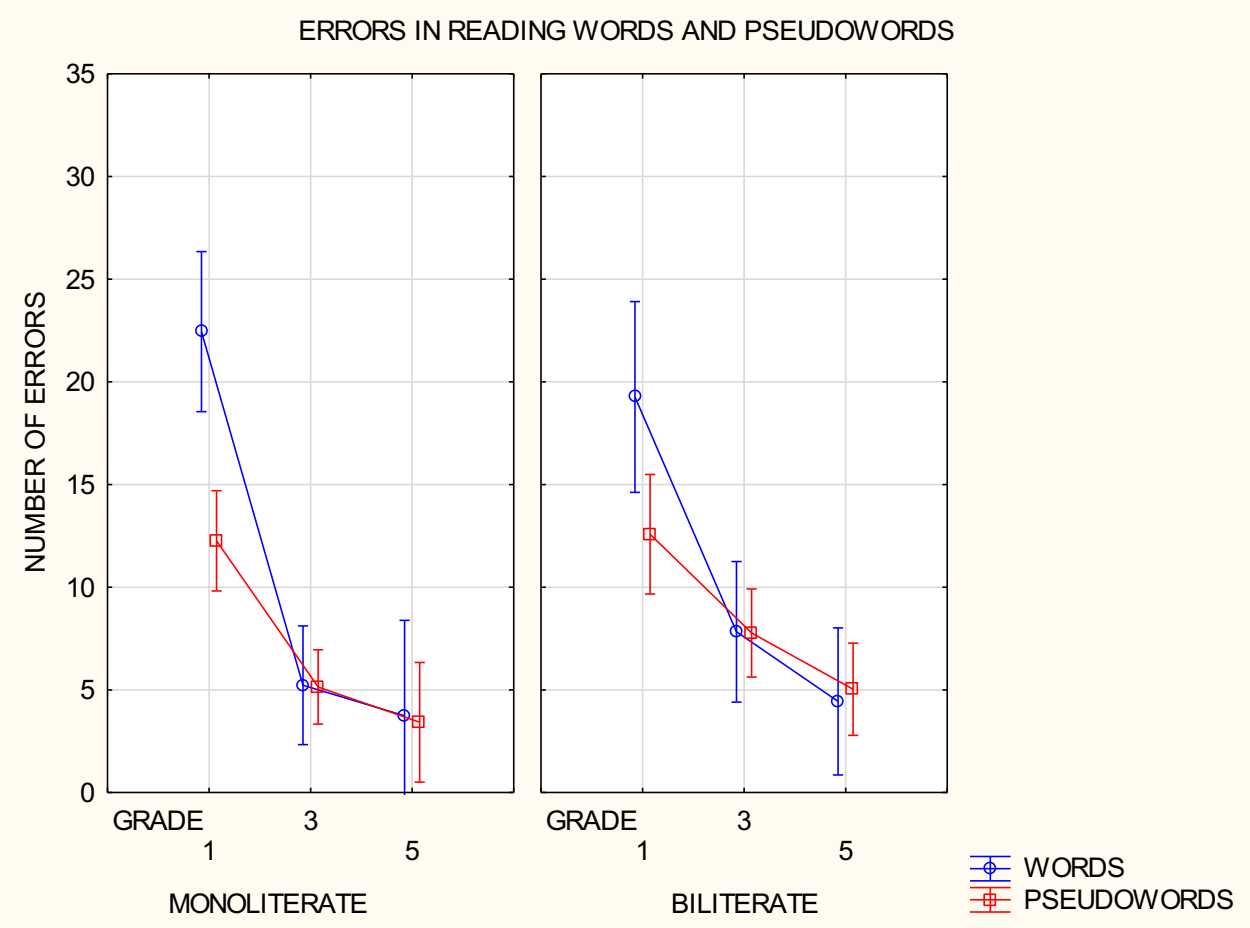



**Figure 4** shows raw scores at the word and pseudoword reading accuracy measure in Grade 1, 3, and 5, of ML and BL bilingual children, separately. Bars indicate a confidence interval at 0.95.

The analyses revealed an effect of Grade $F(1,175) = 32.49$, $\eta^2 = 0.27$, $p < 0.001$. Bonferroni *post hoc* showed that first-graders differed from third and fifth-graders ($p < 0.01$). An effect of Item

$F(1,175) = 35.14$, $\eta^2 = 0.17$, $p < 0.001$ and of the Interaction between Item and Grade $F(2,175) = 33.04$, $\eta^2 = 0.27$, $p < 0.001$ were also found. Bonferroni *post hoc* confirmed that in Grade 1 words reading accuracy differed from pseudowords reading accuracy ($p < 0.01$), but not in Grade 3 and 5.

Then, we analyzed passage reading proficiency measures. As shown in *Figure 5*, reading speed similarly increased across Grades in the two Groups.

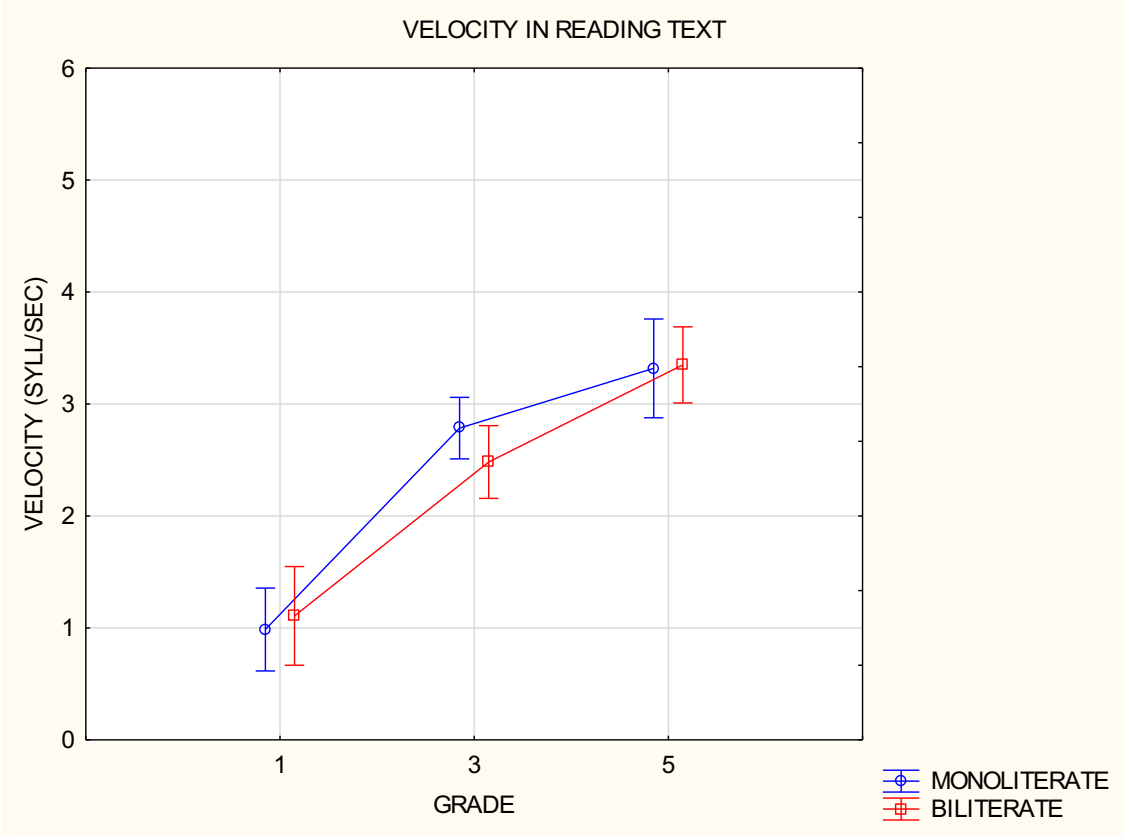



**Figure 5** shows raw scores at the reading rate measure in Grade 1, 3, and 5, of ML and BL bilingual children. Bars indicate a confidence interval at 0.95.

Statistical analysis confirmed this observation: an effect of Grade was found $F(2,175) = 67.3$, $\eta^2 = 0.43$, $p < 0.001$. Bonferroni *post hoc* revealed that all three grades differed from each other ($p < 0.01$). No significant difference was found in the performance of ML and BL Groups.

In *Figure 6*, we can observe scores at reading accuracy (*i.e.*, number of errors). As can be seen, BL bilingual children made more errors than the ML bilinguals, especially in Grade 3, where ML improved their performance while BL Group had a slight downfall. Nevertheless, both Groups seem not to improve from Grade 1 to Grade 5.

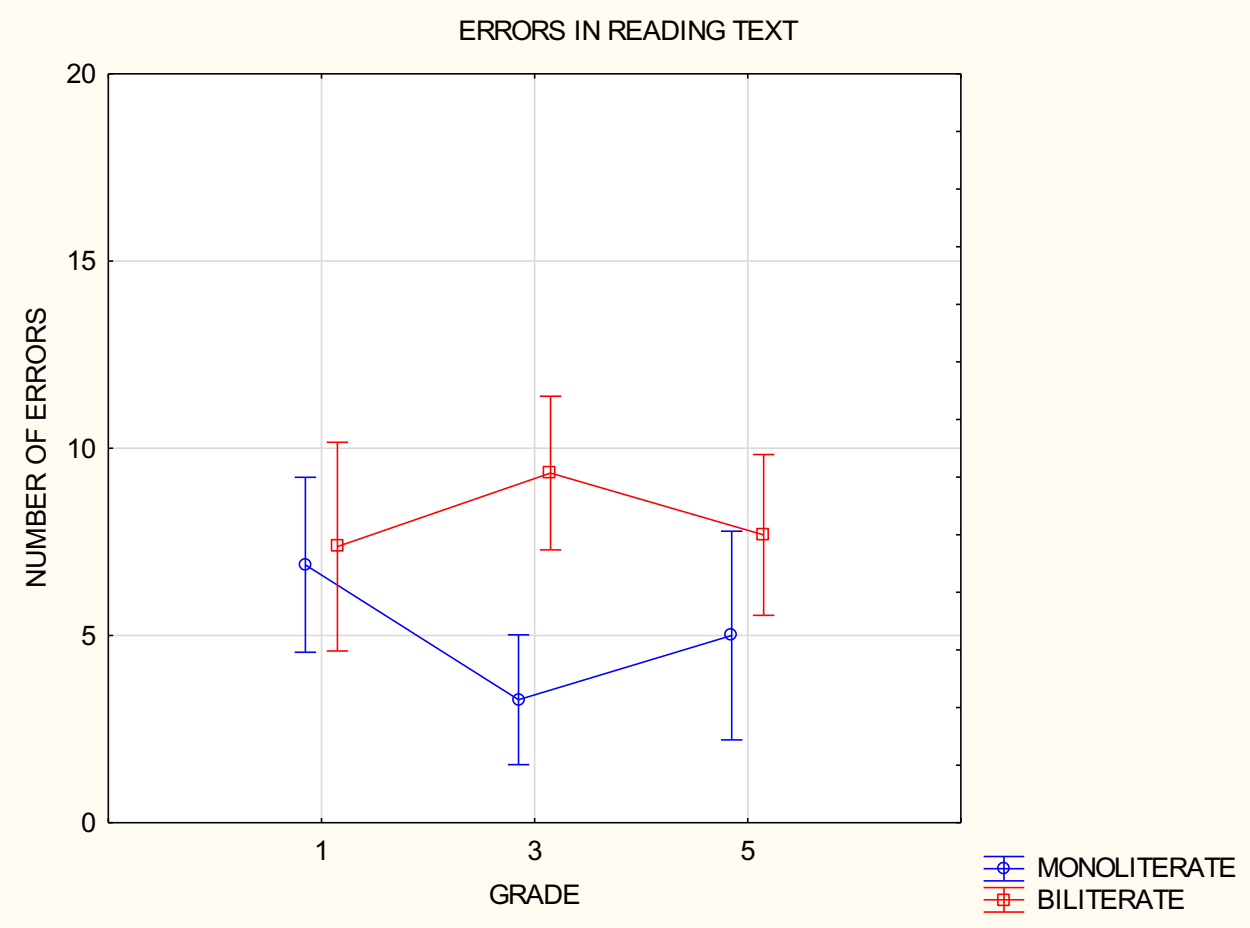



**Figure 6** shows raw scores at the reading accuracy measure in Grade 1, 3, and 5, of ML and BL bilingual children. Bars indicate a confidence interval at 0.95.

The statistical analysis confirmed these observations: the difference in the performance of ML and BL bilingual children turned out to be significant $F(1,175) = 10.1$, $\eta^2 = 0.05$, $p < 0.05$, and no effect of Grade was found. The Interaction between Group and Grade was significant $F(2,175) = 3.18$, $\eta^2$

= 0.04, $p < 0.05$, because MLs and BLs differed in Grade 3, with BL Group making more errors than ML Group ($p < 0.01$). Thus, the effect of Group is due to third-grade children.

Finally, we analyzed reading comprehension. *Figure 7* illustrates the performance of the participants.

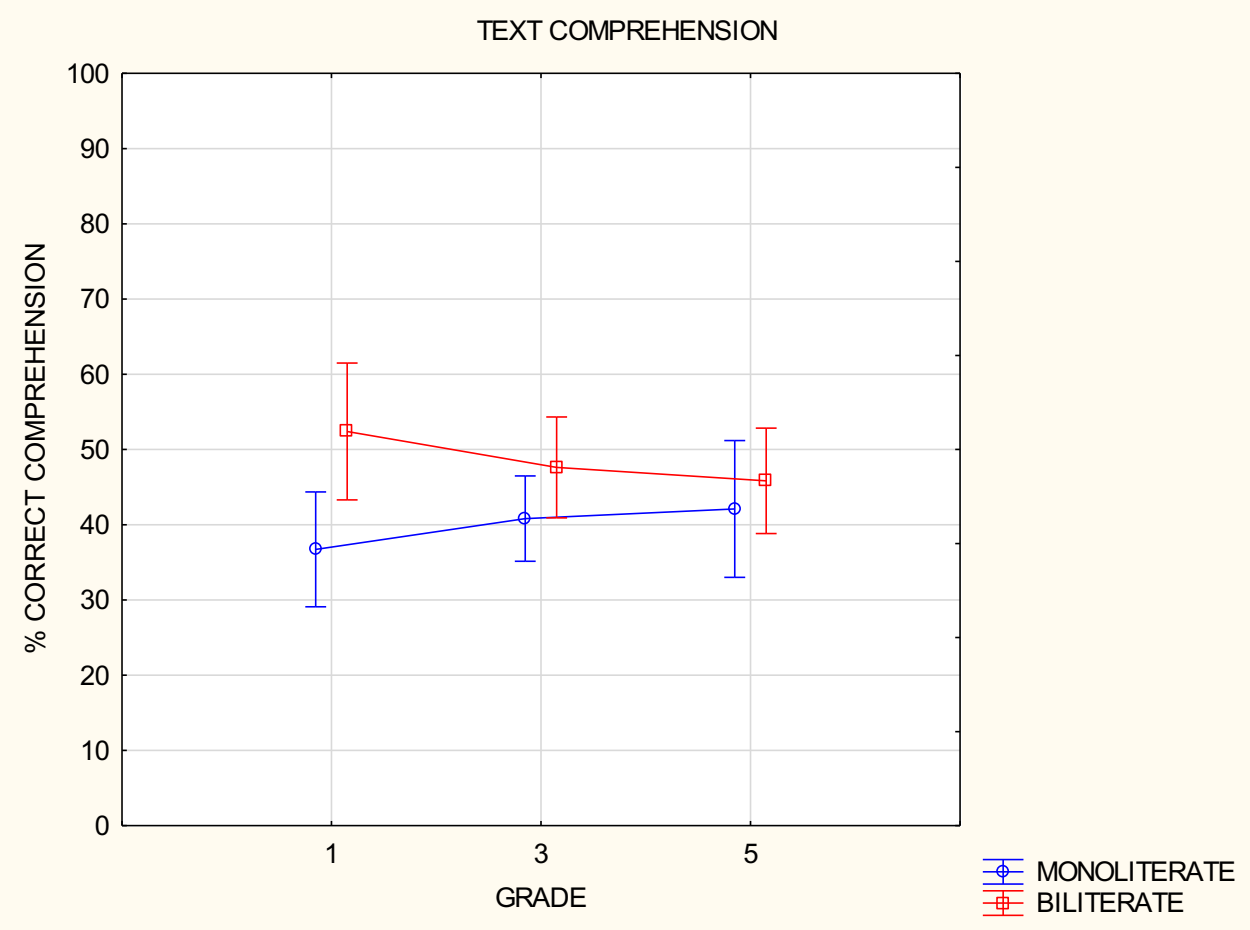



**Figure 7** shows the scores (percentages) at the reading comprehension in Grade 1, 3, and 5 of ML and BL bilingual children. Bars indicate a confidence interval at 0.95.

As can be seen, a small advantage in favor of BL bilinguals is evident, especially in Grade 1. In fact, Group turned out to be significant $F(1,175) = 7.65$, $\eta^2 = 0.04$, $p < 0.05$ in favor of BL bilinguals, although the effect size is small.

To summarize, we did not find huge differences between the two Groups. BL bilinguals were better at the IQ measure (Raven score), while ML bilinguals were better at the pseudowords repetition task. We found an effect of Group in passage reading accuracy, essentially due to BL

Grade 3 children who made more errors than their ML peer, and reading comprehension, with BL children being better than their ML peers. However, no improvement in reading comprehension was observed across Grades in both the Groups. Considering reading velocity, the two groups did not differ, although the interaction Group*Item was significant, due to a tendency for ML children to read words faster than BL (but ML and BL children did not differ on Bonferroni post hoc), with no similar discrepancy for pseudowords.

To conclude, starting from Grade 3, ML and BL bilingual children, read words faster than pseudowords, thus proving they have learned to use the lexical route for reading, and the reading strategy is age-appropriate.

**Correlation among the different measures**

To verify relations among the performance at the different tasks, we carried out correlation analyses. As can be seen in *Table 4*, almost all the measures correlated with each other. We report the results of the two Groups (ML and BL) together, as the same correlations were found considering the Groups separated.

| | Raven | Vocab | W_vel | W_acc | NW_vel | NW_acc | Text_vel | Text_acc | Text_compr | NW_rep |
|---|---|---|---|---|---|---|---|---|---|---|
| Raven | 1,0000 | 0,2396 | 0,4307 | -0,3698 | 0,3888 | -0,2771 | 0,4298 | -0,0284 | 0,3395 | 0,2363 |
| Vocab | 0,2396 | 1,0000 | 0,2999 | -0,3849 | 0,2908 | -0,3177 | 0,3660 | -0,2379 | 0,3720 | 0,3056 |
| W_vel | 0,4307 | 0,2997 | 1,0000 | -0,6353 | 0,9227 | -0,5674 | 0,9020 | -0,3402 | 0,2529 | 0,4559 |
| W_acc | -0,3698 | -0,3849 | -0,6353 | 1,0000 | -0,6307 | 0,8653 | -0,6220 | 0,5591 | -0,2949 | -0,5515 |
| NW_vel | 0,3888 | 0,2908 | 0,9227 | -0,6307 | 1,0000 | -0,5364 | 0,8702 | -0,3245 | 0,2253 | 0,4225 |
| NW_acc | -0,2771 | -0,3177 | -0,5674 | 0,8653 | -0,5364 | 1,0000 | -0,5451 | 0,6464 | -0,2255 | -0,5538 |
| Text_vel | 0,4298 | 0,3660 | 0,9020 | -0,6220 | 0,8702 | -0,5451 | 1,0000 | -0,3591 | 0,3074 | 0,4823 |
| Text_acc | -0,0284 | -0,2379 | -0,3402 | 0,5591 | -0,3245 | 0,6464 | -0,3591 | 1,0000 | -0,2447 | -0,4769 |
| Text_compr | 0,3395 | 0,3720 | 0,2529 | -0,2949 | 0,2253 | -0,2255 | 0,3074 | -0,2447 | 1,0000 | 0,2672 |
| NW_rep | 0,2363 | 0,3056 | 0,4559 | -0,5515 | 0,4225 | -0,5538 | 0,4823 | -0,4769 | 0,2672 | 1,0000 |

**Table 4** shows the correlation matrix of the different measures tested. ML and BL are grouped. Significant correlations are marked in red ($p < 0.05$).

Raven score correlates with almost all the other measures (with the only exception of passage reading accuracy).

To better understand the role of Raven IQ measure and its possible interactions with our reading and language measures, we grouped the measures empirically, as follows:

- *Velocity*: the three velocity measures (reading rates calculated in syllables/seconds) in reading words, pseudowords and text, and passage comprehension measure (under the assumption that reading fluently a text helps its comprehension).
- *Vocabulary*: expressive vocabulary measure, pseudowords repetition measure (under the assumption that some of the skills are related to learning new words, see Gathercole, 2006), and the three accuracy measures in reading words, pseudowords and text (under the assumption that reading accurately means that the child has a correct orthographic representation).
- *Raven*.

Then, we carried out three partial correlations. In the first partial correlation analyses (see *Table 5*), we found that *Raven* correlated with the *Velocity* (including the three reading speeds and text comprehension), partialling out *Vocabulary*.

| | Raven | W_vel | NW_vel | Text_vel | Text_compr |
|---|---|---|---|---|---|
| Raven | 1,0000 | 0,2582 | 0,1962 | 0,2592 | 0,2657 |
| W_vel | 0,2582 | 1,0000 | 0,8704 | 0,8366 | 0,0654 |
| NW_vel | 0,1962 | 0,8704 | 1,0000 | 0,7874 | 0,0284 |
| Text_vel | 0,2592 | 0,8366 | 0,7874 | 1,0000 | 0,1080 |
| Text_compr | 0,2657 | 0,0654 | 0,0284 | 0,1080 | 1,0000 |

**Table 5** shows the correlation matrix between *Raven* and *Velocity* partialling out *Vocabulary*. Significant correlations are marked in red ($p < 0.05$).

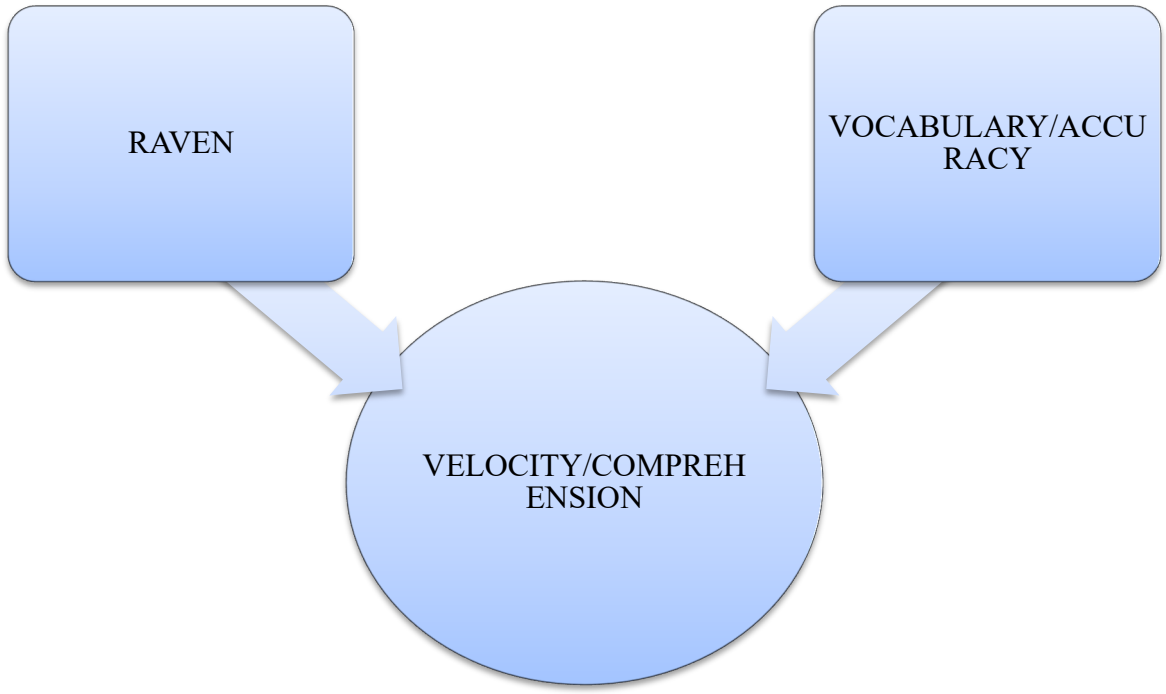
In the second partial correlation analysis, we found that *Raven* did not correlate with the *Vocabulary* (including the three measures of reading accuracy, pseudowords repetition, and expressive vocabulary), partialling out *Velocity*.

In the third and last partial correlation analyses (see *Table 6*), we found that *Vocabulary* correlated with *Velocity,* partialling out *Raven*.

| | Vocab | W_vel | NW_vel | Text_vel | Text_compr |
|---|---|---|---|---|---|
| Vocab | 1,0000 | 0,2245 | 0,2209 | 0,3001 | 0,3182 |
| W_vel | 0,2245 | 1,0000 | 0,9082 | 0,8797 | 0,1256 |
| NW_vel | 0,2209 | 0,908248 | 1,000000 | 0,845154 | 0,107659 |
| Text_vel | 0,3001 | 0,8797 | 0,8452 | 1,0000 | 0,1901 |
| Text_compr | 0,3182 | 0,1256 | 0,1077 | 0,1901 | 1,0000 |

**Table 6** shows the correlation matrix between *Vocabulary* and *Velocity*. Significant correlations are marked in red ($p < 0.05$).

Our results of the partial correlations can be schematically summarized as follows:



Hence, we found that Raven (IQ) and vocabulary contribute to reading speed and comprehension, but we did not find a direct link between Raven (IQ) and vocabulary.

Following the suggestion of the first reviewer, we conducted a Principal Component Analysis (PCA) on all our variables that could contribute to variability in our data set. In this way, we wanted to see which variables accounted the most for data variability and, in case, to eliminate parameters that could create confounding effects. Since the two groups of children, monoliterate and biliterate had different Raven scores, we conducted two PCAs, as it is possible that this fact has different effects on the variability. As already observed in the partial correlations, Raven score contributed to reading velocity and comprehension, as did vocabulary, but no direct link was observed between Raven and vocabulary. Both PCAs included the following variables: Raven, pseudowords repetition, expressive vocabulary, reading words accuracy and velocity, reading pseudowords accuracy and velocity, reading text accuracy, velocity and text comprehension. As for monoliterate children, all the variables, but text comprehension and number of correctly repeated pseudowords, contributed to 74% of variability of the dataset. Of this 74%, the first component alone accounted for 51% and the second component for 13% of the variability. In *Table 7* we see the variables' contribution to the PCs.

| Variable | Power | Importance |
|---|---|---|
| Reading words velocity | 0,929867 | 1 |
| Reading words accuracy | 0,882085 | 2 |
| Reading pseudoword velocity | 0,873164 | 3 |
| Reading pseudowords accuracy | 0,854138 | 4 |
| Reading text velocity | 0,844464 | 5 |
| Expressive vocabulary | 0,796303 | 6 |
| Reading text accuracy | 0,776718 | 7 |
| Raven | 0,547448 | 8 |
| Text comprehension | 0,499729 | 9 |
| Pseudowords repetition | 0,409307 | 10 |

**Table 7** shows the variable importance in the monoliterate group. Text comprehension and pseudowords repetition did not contribute to the variability of the dataset.

As for the biliterate group, all the variables, but expressive vocabulary, contributed to 76% of variability of the dataset. Of this 76%, the first component alone accounted for 53% of the variability and the second component for 14%. In *Table 8* we see the variables' contribution to the PCs.

| Variable | Power | Importance |
|---|---|---|
| Reading words velocity | 0,938048 | 1 |
| Reading text velocity | 0,931054 | 2 |
| Reading pseudoword velocity | 0,921673 | 3 |
| Raven | 0,885551 | 4 |
| Reading pseudowords accuracy | 0,784348 | 5 |
| Reading words accuracy | 0,762880 | 6 |
| Reading text accuracy | 0,687400 | 7 |
| Text comprehension | 0,647539 | 8 |
| Pseudowords repetition | 0,615304 | 9 |
| Expressive vocabulary | 0,498781 | 10 |

**Table 8** shows the variable importance in the biliterate group. Expressive vocabulary did not contribute to the variability of the dataset.

Comparing the two tables, we can observe that the different variables contributed differently. Beyond reading measures, which contributed in both groups, Raven turned out to be much more important for biliterates than for monoliterates.

**Comparison within the Biliterate bilingual Group**

Lastly, we wanted to explore whether there was a difference between the two Groups of BL children (alphabetic or A-BL *vs* logographic BL or L-BL). Although alphabetic languages considered in the study (Arabic, Polish, and Russian) are not identical in terms of the alphabet, we grouped them together as the grapheme-phoneme correspondence is present in all of them.

We did this analysis separately, as our ML population's size was larger than that of the two BL samples, while the two BL Groups were comparable in size. Biliterate groups were also cognitively matched, as shown above. We will present the tasks separately, as made above, for the

first set of analyses. Complete descriptive statistics of the two BL Groups can be found in the *Appendix* section.

***Vocabulary measure***

*Figure 8* shows scores of alphabetic (A-BL) and logographic (L-BL) biliterates at expressive vocabulary test. Vocabulary improved sharply across Grades in the L-BL Group of children, yet worse than the A-BL Group, except for grade 5. On their side, A-BL Group improved slightly from first to third Grade, but without further improvements in fifth Grade.

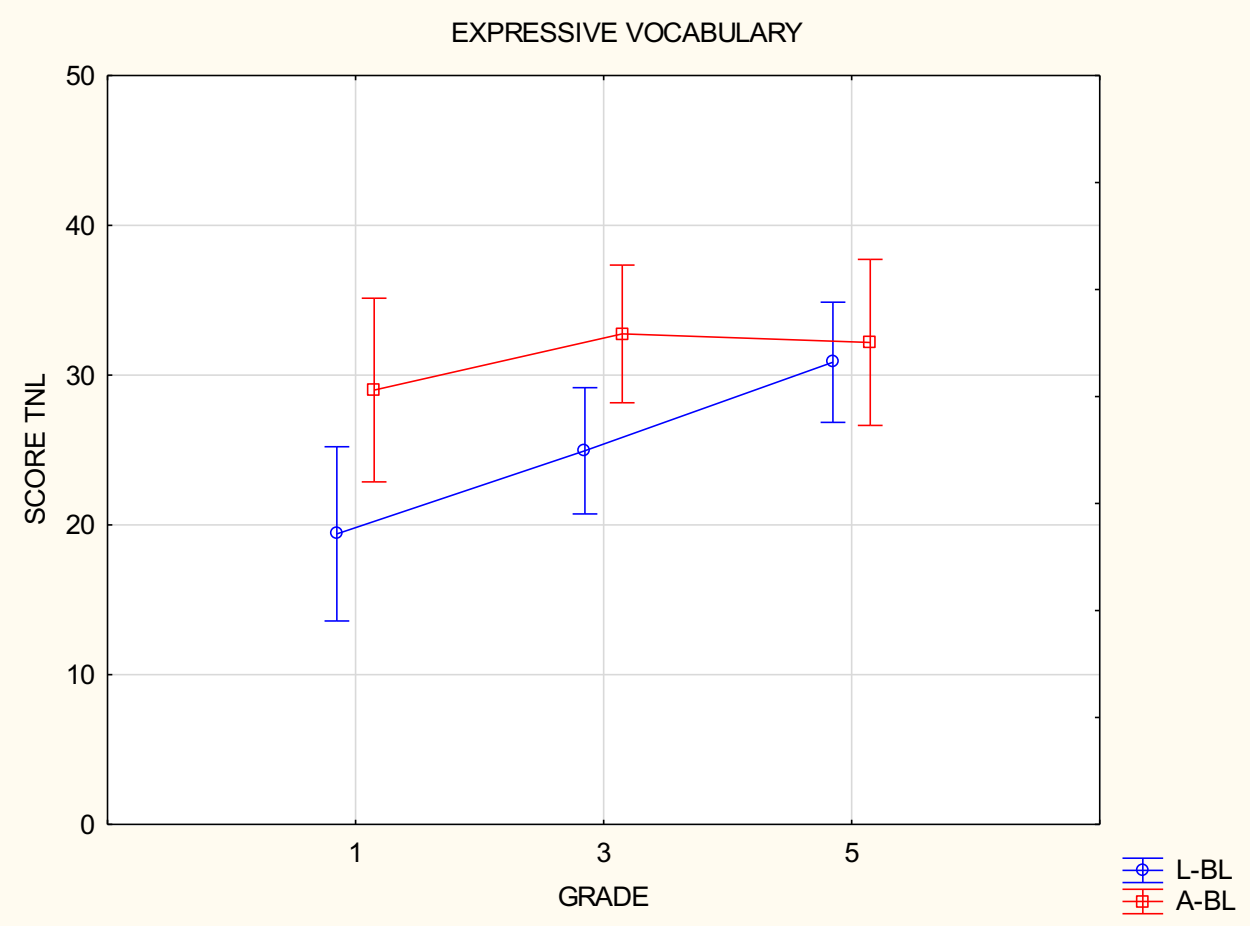



**Figure 8** shows raw score (Max=50) at expressive vocabulary in Grade 1, 3, and 5, of L-BL and A-BL bilingual children. Bars indicate a confidence interval at 0.95.

The ANOVA with Group (L-BL and A-BL) and Grades (1, 3, and 5) as the independent variables, scores at the *Test Neuropsicologico Lessicale per l'età evolutiva* (TNL) as the dependent variable, found an effect of Group $F(1,80) = 8.84$, $\eta^2 = 0.10$, $p < 0.05$, with A-BL being better than L-BL,

and Grade F(2,80) = 3.60, $\eta^2 = 0.08$, $p < 0.05$. Bonferroni *post hoc* showed that Grade 1 differed from Grade 5 ($p < 0.05$).

***Reading measures***

*Figure 9* illustrates the word and pseudoword reading speed performance (measured in syllables/seconds) of the two Groups. Reading speed increased across Grades in both Groups, with words being read faster from Grade 3, meaning that reading has automatized. Improvement in speed is less sharp in the A-BLs, especially for words, than in the L-BL Group.


**Figure 9** shows raw scores at the word and pseudoword reading accuracy measure in Grade 1, 3, and 5, of L-BL and A-BL bilingual children, separately. Bars indicate a confidence interval at 0.95.

The analyses revealed an effect of Grade F(2,80) = 50.01, $\eta^2 = 0.6$, $p < 0.001$, and an effect of Item F(1,80) = 117.59, $\eta^2 = 0.6$, $p < 0.001$. Some Interactions were significant: (1) between Group and Grade F(2,80) = 5.15, $\eta^2 = 0.1$, $p < 0.05$, because in the L-BL Group all three grades differed from

each others (Bonferroni $p < 0.01$), while in the A-BL only first-graders differed from third- and fifth-graders (Bonferroni $p < 0.01$); (2) between Item and Grade $F(2,80) = 22.32$, $\eta^2 = 0.4$, $p < 0.001$, because both Groups in Grade 1 read words and pseudowords at the same rate, whereas in Grade 3 and 5 words rate differed from pseudowords rate ($p < 0.01$); (3) between Item, Group and Grade $F(2,80) = 4.91$, $\eta^2 = 0.1$, $p < 0.05$ due to a deviation from the parallelism. As could be seen in *Figure 9* L-BL fifth-graders seemed faster in reading than the A-BL peers, but the difference was not significant in the *post hoc*.

Then, we analyzed word and pseudoword reading accuracy. *Figure 10* shows the performance of L-BLs and A-BLs. Both Groups committed more errors in Grade 1 with L-BL making many errors especially with words. L-BL children gradually improved across Grades, while A-BL's improvement from Grade 3 to 5 is less evident.

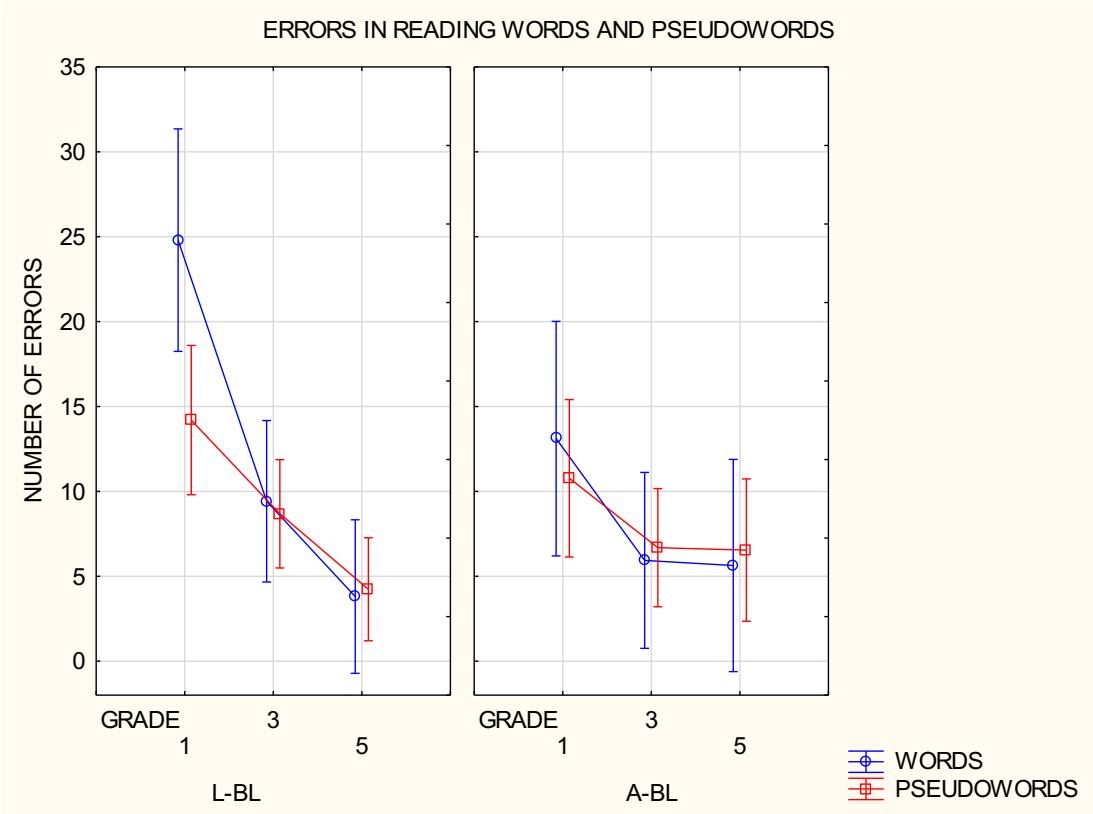



**Figure 10** shows raw scores at the word and pseudoword reading accuracy measure in Grade 1, 3, and 5, of L-BL and A-BL bilingual children, separately. Bars indicate a confidence interval at 0.95.

We found an effect of Grade $F(2,80) = 9.54$, $\eta^2 = 0.2$, $p < 0.001$, and an effect of Item $F(1,80) = 9.31$, $\eta^2 = 0.1$, $p < 0.05$, with more words errors. Bonferroni *post hoc* revealed that Grade 1 differed from Grade 3 and 5 ($p < 0.01$). The interaction between Item and Group was significant $F(1,80) = 7.27$, $\eta^2 = 0.1$, $p < 0.05$, because L-BL Group committed more errors reading words than pseudowords in Grade 1 (Bonferroni $p<0.05$). Moreover the interaction between Item and Grade was significant $F(2,80) = 10.91$, $\eta^2 = 0.2$, $p < 0.001$, because there is significant difference between word and pseudoword reading accuracy in Grade 1 (due especially to the L-BL Group, Bonferroni $p < 0.05$), but not in Grade 3 and 5. Lastly, the interaction between Item, Group and Grade was significant $F(2,80) = 3.15$, $\eta^2 = 0.1$, $p < 0.05$ again because in Grade 1 L-BL committed more errors with words than pseudowords (Bonferroni $p < 0.05$).

Analyzing speed while reading a text, we found an effect of Grade $F(2,80) = 33.73$, $\eta^2 = 0.5$, $p < 0.001$ and of an interaction between Group and Grade $F(2,80) = 3.2$, $\eta^2 = 0.1$, $p < 0.05$. As shown in *Figure 11*, the performance of A-BL Group was slightly better than that of L-BL in Grade 1 and 3, whereas in Grade 5, the graph lines cross, indicating that L-BL were better than A-BL. This justifies the significant Interaction. Bonferroni *post hoc* showed that all Grades differed from each other in the L-BL Group, while in A-BL Group only Grade 1 differed from Grade 3 and 5, no difference was found from Grade 3 to Grade 5 (as could be guessed by *Figure 11*). Thus, the effect of grade is due in large part to the L-BL Group.

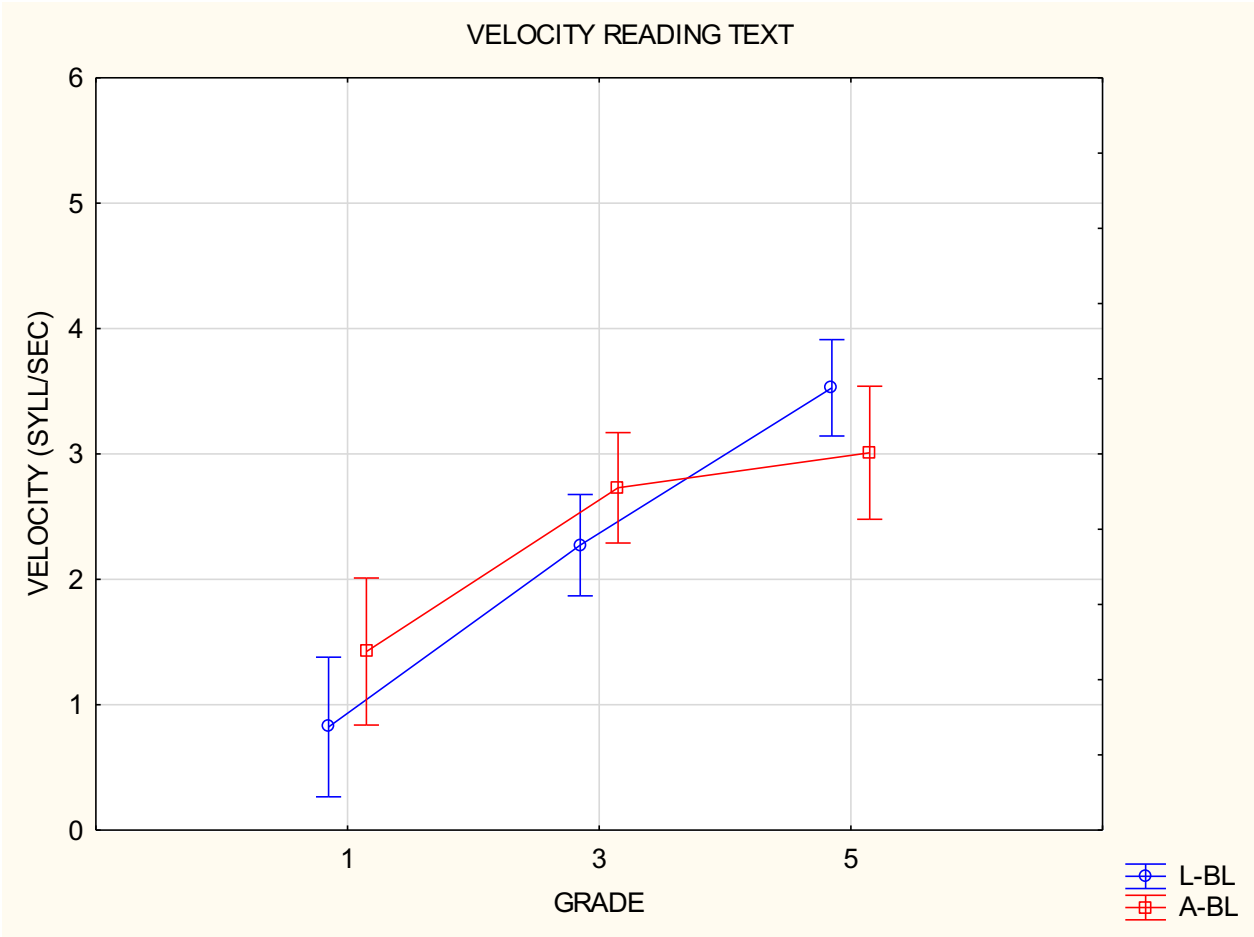



**Figure 11** shows raw scores at the reading rate measure in Grade 1, 3, and 5, of L-BL and A-BL bilingual children. Bars indicate a confidence interval at 0.95.

*Figure 12* shows the passage reading errors of the two Groups. Only the interaction Group by Grade was significant $F(2,80) = 3.21$, $\eta^2 = 0.1$, $p < 0.05$ because of the inverted trend of the Groups' performance in Grade 5. Post hoc Bonferroni revealed no further difference.

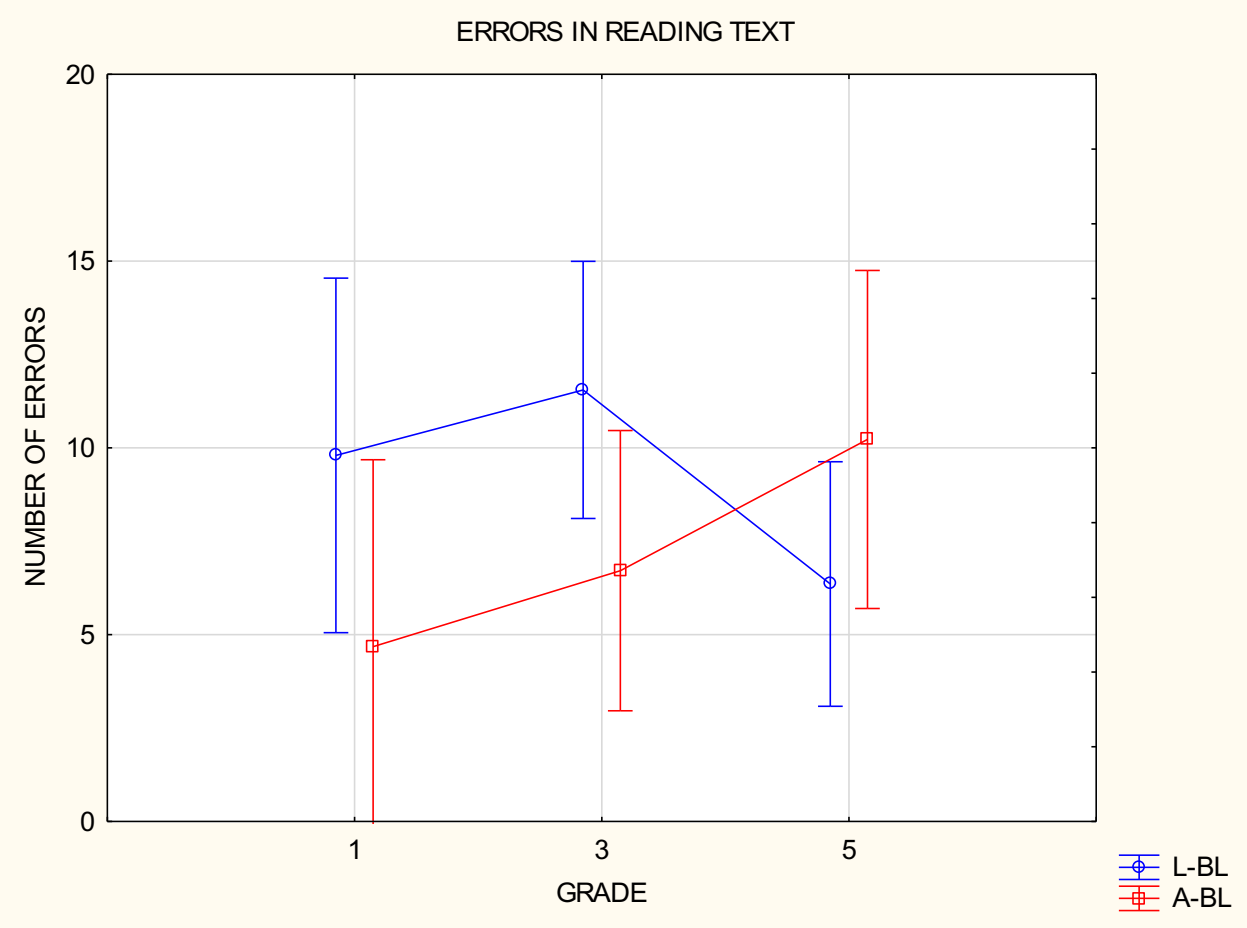



**Figure 12** shows raw scores at the reading accuracy measure in Grade 1, 3, and 5, of L-BL and A-BL bilingual children. Bars indicate a confidence interval at 0.95.

Finally, we analyzed reading comprehension. *Figure 13* illustrates the performance of the participants.

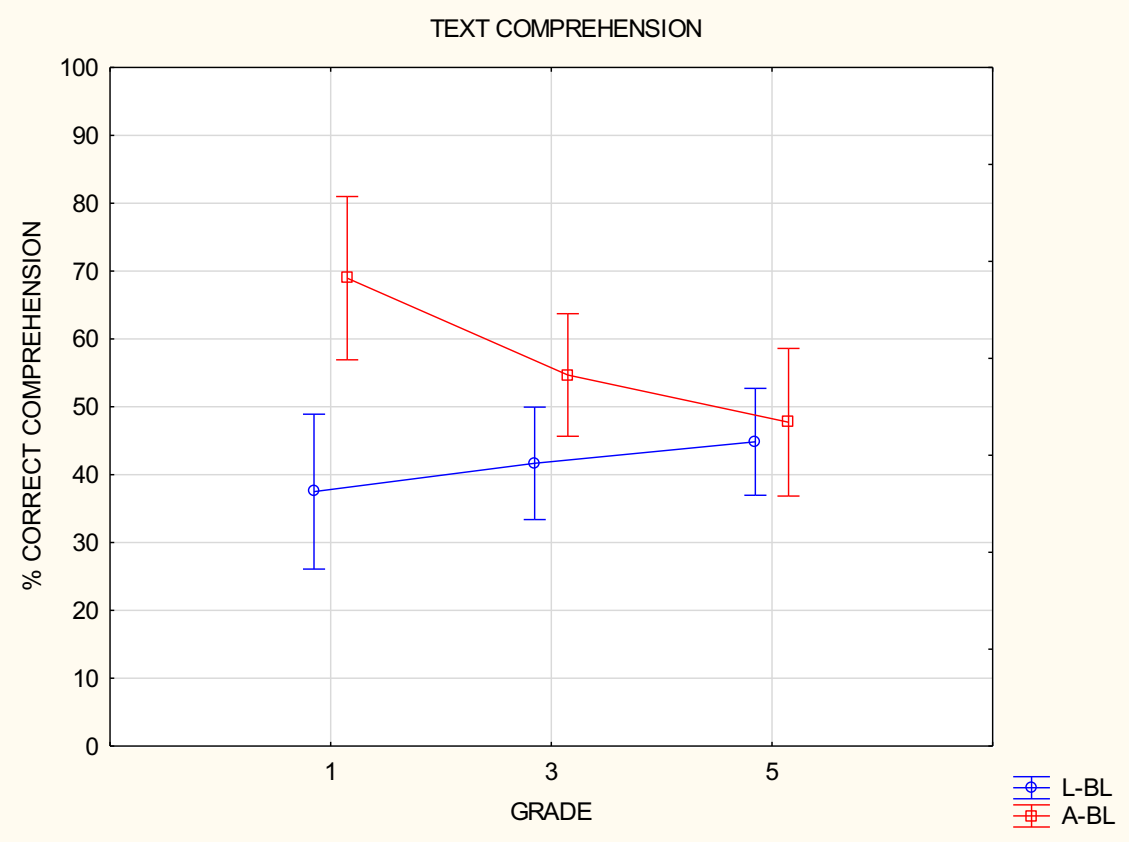



**Figure 13** shows raw scores in percentage at the reading comprehension in Grade 1, 3, and 5, of L-BL and A-BL bilingual children. Bars indicate a confidence interval at 0.95.

As can be seen, the two Groups' performance largely diverged in Grade 1, with A-BL being better than L-BL. In Grade 5, performance turned out to be quite similar instead. An effect of Group was found $F(1,80) = 14.66$, $\eta^2 = 0.2$, $p < 0.001$. The Interaction Group by Grade was also significant $F(2,80) = 3.56$, $\eta^2 = 0.1$, $p < 0.05$, because A-BL first-graders were different from L-BL first-graders. Thus, the effect of Group is due to first-graders.

***Short-term memory measure***

Lastly, we analyzed the performance at the pseudowords repetition task, as illustrated in *Figure 14*. A-BL Group seems to perform better than the L-BL peers, especially in Grade 1 and 3. Besides, the skill improves similarly from Grade 1 to Grade 3 in both Groups, with a slight downfall in A-BL fifth-graders, conversely.

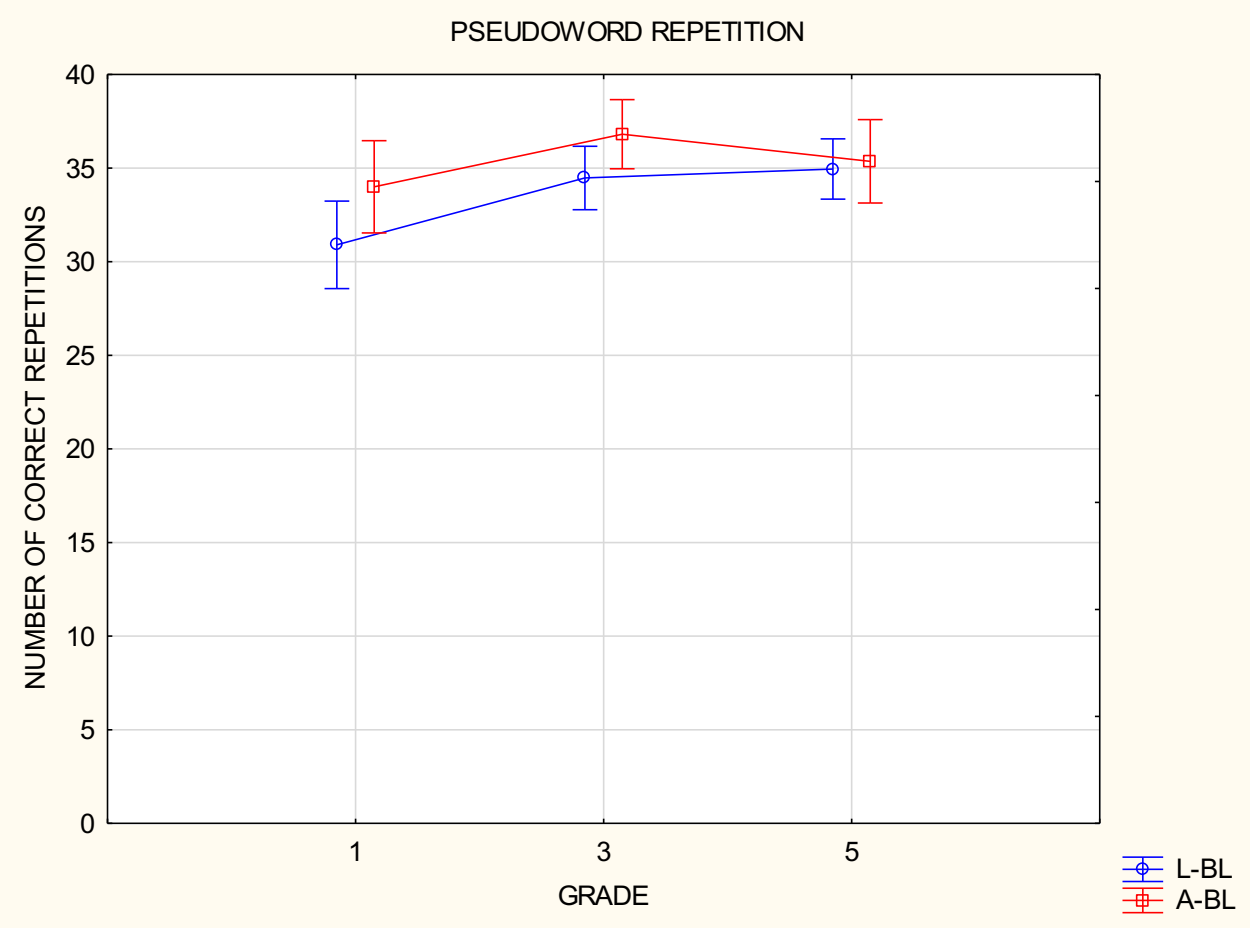



**Figure 14** shows raw scores (Max=40) at the pseudowords repetition task in Italian in Grade 1, 3, and 5, of L-BL and A-BL bilingual children. Bars indicate a confidence interval at 0.95.

The analyses confirmed these observations: an effect of Group $F(1,80) = 5.35$, $\eta^2 = 0.06$, $p < 0.05$, and of Grade $F(2,80) = 4.82$, $\eta^2 = 0.1$, $p < 0.05$ were found on *Pseudoword Repetition Task.* As for Grades, no difference was found with Bonferroni *post hocs*.

To summarize, biliterate-bilinguals with an alphabetic L1 were better at the short-term memory measure (pseudowords repetition), at reading comprehension – at least at the first stage of schooling – and at expressive vocabulary. However, the differences tend to disappear in grade fifth. The other skills considered did not show consistent differences, except for word reading speed tendencies in favor of biliterate children with Mandarin as their L1, but at the expenses of more errors, at least in grade first.

**Correlation among the different measures**

To conclude our analyses, we investigated relations among the performance at the different tasks in the two biliterate Groups (alphabetic-L1 and logographic-L1). We will report correlation coefficients of the two Groups, Alphabetic-BL and Logographic-BL, separately, as different correlations were found in the two Groups.

First, we present correlations among tasks in the Alphabetic Biliterate Group. As shown in *Table 9*, we found that *Raven* and vocabulary correlated with reading accuracy and working memory. Reading proficiency measures (velocity and accuracy) correlated with each other, text comprehension with accuracy. Working memory correlated with almost all the other measures.

| | Raven | Vocab | W_vel | W_acc | NW_vel | NW_acc | Text_vel | Text_acc | Text_compr | NW_rep |
|---|---|---|---|---|---|---|---|---|---|---|
| **Raven** | 1,0000 | 0,0044 | 0,2337 | -0,2910 | 0,2490 | -0,3322 | 0,2554 | -0,4346 | 0,2733 | 0,3843 |
| **Vocab** | 0,0044 | 1,0000 | 0,2470 | -0,2480 | 0,2491 | -0,3251 | 0,3028 | -0,4379 | 0,2036 | 0,3808 |
| **W_vel** | 0,2337 | 0,2470 | 1,0000 | -0,5135 | 0,8965 | -0,5953 | 0,9107 | -0,2542 | -0,0056 | 0,4858 |
| **W_acc** | -0,2910 | -0,2480 | -0,5135 | 1,0000 | -0,5320 | 0,8444 | -0,5252 | 0,4956 | -0,2137 | -0,4790 |
| **PW_vel** | 0,2490 | 0,2491 | 0,8965 | -0,5320 | 1,0000 | -0,6007 | 0,9049 | -0,2752 | -0,0698 | 0,4826 |
| **PW_acc** | -0,3322 | -0,3251 | -0,5953 | 0,8444 | -0,6007 | 1,0000 | -0,6453 | 0,6495 | -0,2803 | -0,4716 |
| **Text_vel** | 0,2554 | 0,3028 | 0,9107 | -0,5252 | 0,9049 | -0,6453 | 1,0000 | -0,3643 | 0,0981 | 0,5025 |
| **Text_acc** | -0,4346 | -0,4379 | -0,2542 | 0,4956 | -0,2752 | 0,6495 | -0,3643 | 1,0000 | -0,5952 | -0,4188 |
| **Text_compr** | 0,2733 | 0,2036 | -0,0056 | -0,2137 | -0,0698 | -0,2803 | 0,0981 | -0,5952 | 1,0000 | 0,0626 |
| **PW_rep** | 0,3843 | 0,3808 | 0,4858 | -0,4790 | 0,4826 | -0,4716 | 0,5025 | -0,4188 | 0,0626 | 1,0000 |

**Table 9** shows the correlation matrix of different measures tested of the Alphabetic-L1 biliterate Group. Significant correlations are marked in red ($p < 0.05$).

Second, we present correlations among tasks in the Logographic Biliterate Group. In this Group, almost all the measures correlated with each other, as can be seen in *Table 10.*

| | Raven | Vocab | W_vel | W_acc | NW_vel | NW_acc | Text_vel | Text_acc | Text_compr | NW_rep |
|---|---|---|---|---|---|---|---|---|---|---|
| **Raven** | 1,0000 | 0,2814 | 0,4701 | -0,3949 | 0,3774 | -0,3725 | 0,4765 | -0,0564 | 0,3947 | 0,3299 |
| **Vocab** | 0,2814 | 1,0000 | 0,6922 | -0,6713 | 0,6071 | -0,6033 | 0,6899 | -0,3710 | 0,5868 | 0,5134 |
| **W_vel** | 0,4701 | 0,6922 | 1,0000 | -0,6215 | 0,9256 | -0,5724 | 0,9459 | -0,3533 | 0,3347 | 0,4253 |
| **W_acc** | -0,3949 | -0,6713 | -0,6215 | 1,0000 | -0,5953 | 0,9118 | -0,6489 | 0,6113 | -0,4466 | -0,7451 |
| **PW_vel** | 0,3774 | 0,6071 | 0,9256 | -0,5953 | 1,0000 | -0,5128 | 0,9324 | -0,3444 | 0,3350 | 0,4137 |
| **PW_acc** | -0,3725 | -0,6033 | -0,5724 | 0,9118 | -0,5128 | 1,0000 | -0,6018 | 0,6746 | -0,4213 | -0,7509 |
| **Text_vel** | 0,4765 | 0,6899 | 0,9459 | -0,6489 | 0,9324 | -0,6017 | 1,0000 | -0,3940 | 0,3686 | 0,4745 |
| **Text_acc** | -0,0564 | -0,3710 | -0,3533 | 0,6113 | -0,3444 | 0,6746 | -0,3940 | 1,0000 | -0,2590 | -0,5844 |
| **Text_compr** | 0,3947 | 0,5868 | 0,3347 | -0,4466 | 0,3350 | -0,4213 | 0,3686 | -0,2590 | 1,0000 | 0,3738 |
| **PW_rep** | 0,3299 | 0,5134 | 0,4253 | -0,7451 | 0,4137 | -0,7509 | 0,4745 | -0,5844 | 0,3738 | 1,0000 |

**Table 10** shows the correlation matrix of different measures tested of the Logographic-L1 biliterate Group. Significant correlations are marked in red ($p < 0.05$).

## Discussions

The study aimed to investigate the effects on Italian literacy and literacy-related achievement of a biliteracy education in immigrant students. Biliteracy aims to sustain both the child's languages, supporting the minority language (L1) in its written form, to obtain some advantage in the majority language (Italian L2), the language of learning at mainstream school. Thus, we were primarily interested in verifying whether the Italian language and reading performance of biliterate immigrant primary school children differed significantly from that exhibited by immigrant peers, exposed only orally to L1 and to literacy at school only in Italian L2. To do so, language and reading performance in Italian of two groups of bilinguals, one biliterate and one monoliterate were compared, seeking for a possible effect of biliteracy.

Secondly, we analyzed reading development in both groups of bilinguals, and we predicted to find automatization of reading starting from Grade 3, regardless of their early or later exposure to

Italian L2 (Bellocchi *et al.*, 2016). To do so, we focused on when strategies for reading words and pseudowords start to diverge.

Third, we investigated the relationship between cognitive, linguistic, and Italian reading measures in both the biliterate and monoliterate groups. To address it, we performed correlations between the various skills evaluated.

Lastly, because our biliterate sample differed as to the script system of their first language (alphabetic *vs*. logographic), we analyzed the two biliterate groups separately, comparing the abilities tested and investigating correlations among the various abilities in these two groups, separately. Because Italian is an alphabetic language, we predicted a more effective transfer in the alphabetic-L1 group, with a better performance than the logographic-L1 group. In the rest of the discussion, we are going to tackle each study goal.

***Effects of biliteracy***

The present study explored the possible differences in cognitive skills between monoliterate and biliterate children. In this regard, results were partially in line with our predictions. As predicted, biliterate bilingual group had a better non-verbal intelligence performance than monoliterates, with a moderate effect size, i.e. the effect was relatively rubust. On the opposite, the latter were slightly better than biliterates at the short-term memory task (pseudowords repetition), but the effect size was low. The biliterate group might show a cognitive benefit (Raven) because they are exposed in more contexts to the two languages and have a more varied bilingual experience. In fact, monoliterate children have an exposure to the L1 in a limited number of domains (because of oral and in-home exposure only), and this may result in a relatively low degree of bilingual proficiency (Blom *et al.*, 2017). Conversely, biliterate bilinguals use both languages more frequently, even outside the home environment, and they learn to read and write in both languages. The more exposure to their two languages, the more they are proficient bilinguals. Antoniou, Grohmann, Kambanaros, and Katsos (2016), studying executive control in bilectal Cypriot Greek-Standard

Modern Greek and English-Greek multilingual children, compared to Modern Greek monolingual children, found that both groups of bilingual children exhibited an advantage over monolinguals, which emerged only after statistically controlling for their lower language proficiency. Thus, they show that the level of language proficiency moderates the emergence of cognitive advantages in bilinguals (see also Videsott, *et al.*, 2012; Weber, *et al.*, 2016). As Cummins (1992) claimed, continued development in two languages could enhance learners' cognitive development.

We found monoliterate to outperform biliterate children in short-term memory measure. Measuring working memory with a digit span task, Rothou *et al.* (2017) revealed that biliterate Albanian-Greek performed higher than their monoliterate peers. Our results are not in conflict with those of Rothou because two different tests were used, which measures different construct of memory (memory for pseudowords and memory for digits). Pseudowords repetition is indeed dependent on language experience (Chiat & Polišenská, 2016). Gathercole (2006) found evidence of a relationship between vocabulary and pseudowords repetition performance. Sorenson, Duncan, and Paradis (2016) found that vocabulary knowledge affected pseudowords repetition performance in groups of sequential bilingual children exposed to South Asian or Chinese languages. Thus, having monoliterates a possible broader experience and exposure to Italian than biliterates, this might have affected their performance at pseudowords repetition. In fact, monoliterate bilingual children were slightly better than biliterates at expressive vocabulary, even if the difference was not significant (and not present in grade 5).

In any case, although further investigation with a larger sample of participants and with a broader variety of verbal working memory tasks is needed, advantages in non-verbal general cognitive function may emerge from biliteracy.

***Bilterate outcomes on Italian literacy acquisition***

Regarding whether the Italian competence and literacy in immigrants Italian-L2 learners could benefit by learning to read and also write in the first language, we only found an advantage of

biliteracy in text comprehension, which this may be due to their higher experience with written texts in two languages. At the same time, we found that third grade biliterates produced more errors in reading a passage aloud than monoliterates. In this connection, we need to keep in mind that our study is a cross-sectional study and the performance drop in Grade 3 may be just due to sampling. Interestingly, by Grade 5, the two groups tended to behave more similarly. The advantage of biliteracy in text comprehension is in line with Rolstad *et al.* (2005), who demonstrated that Heritage or Dual-Language approach to bilingual education enhances reading achievements in L2, including reading comprehension. A possible explanation is the relationship between reading comprehension and non-verbal reasoning. The comprehension of a text is a complex ability that requires more than merely decoding skills. Successful reading comprehension requires processing at different levels. Readers must first have accurate word level processing – decoding and accessing a words meaning. Then, they must string the meanings of many words together to form a representation of the text, and understand their language's syntax in order to get the sentential meaning. Finally, it is necessary to link sentences and paragraphs, make inference and acquire a global meaning representation for a given text (Landi, Frost, Menc, Sandak & Pugh, 2013). Accurate interpretation requires the ability to use and manage the relevant syntactic and pragmatic information, to abstract, to make inferences, to connect pieces of information in a temporal and causal sequence, among the others (refer to Kintsch & Rawson 2005, for a review on this matter). These higher levels are related to abstract reasoning. Thus, it is likely that such an advantage could be linked to better performance of the biliterate group's abstract reasoning, as found in Raven's performance. Indeed, we also found that Raven contributes to comprehension. In line with this, Grant, Gottardo, and Geva (2011) also reported that nonverbal reasoning was significantly correlated with reading understanding in sequential Spanish-English bilinguals and simultaneous Portuguese-English bilinguals. Moreover, the fact that the biliterate children exercise their reading comprehension competence also in their first language could boost the second language, resulting from a transfer of those *higher-than-mere-language* abilities involved in the

skill, between the languages. Thus one possible explanation for the difference in reading comprehension performance between monoliterate and biliterate readers might be due to a higher ability in abstract reasoning.

The finding apparently conflicts with Rothou *et al.* (2017), as they found biliterate children having lower performance on the reading comprehension than their monoliterate peers (and also monolingual). However, Rothou *et al.* (2017) represent quite an exception, and the groups of biliterates and monoliterates tested did not differ in non-verbal reasoning skills, unlike our biliterate and monoliterate samples.

To conclude, we found that biliteracy enhanced reading comprehension whereas we did not find effects on other reading skills. To comprehend a text effectively is a fundamental skill for academic success, as it is the baseline for the *reading-to-learn* at school. The study of the school subjects requires more engagement and effort than just reading for pleasure and entertainment. As actively reading school text can be challenging for foreign students, to sustain biliteracy appears a good choice.

***Reading development in Italian***

To understand reading development of biliterate and monoliterate Italian-L2 bilinguals, we investigated whether they automatized reading around age 8 (Grade 3), as happens in monolingual children (Sulpizio & Colombo, 2013). In line with Bellocchi *et al.* (2016), we found that both groups reached mature decoding skills by Grade 3, as words were read faster and better than pseudowords. This is indicative that children were using different strategies to perform the two tasks: the lexical route to read words, and the phonological (or sub-lexical) route to read new words or pseudowords. According to Frith's (1985), children start to read using only the phonological route of grapheme-phoneme conversion. By grade 3, they are automatizing reading, *i.e.*, they read words recognizing them as a whole or as formed by pieces larger than single segment, and accessing their pronunciation directly through the lexicon (*e.g.*, Pritchard, Coltheart, Palethorpe, &

Castles, 2012). This reading strategies change is reflected in reading words and pseudowords: while words are read faster for a lexicality effect, pseudowords are treated as new words and read grapheme by grapheme, thus slower.

Even if our bilinguals were not uniformly exposed to oral Italian from age 3, this result was in line with a recent study of Guasti *et al.* (submitted), which revealed that dual-language children aged between 8-11 years, exposed to Italian from kindergarten, performed as good as monolingual children in terms of automatization of reading.

This finding also has a clinical impact. In Italy, Developmental Dyslexia is diagnosed starting from age 8. It is essential to ascertain when strategies for reading words and pseudowords start to diverge in this regard. To know that typically developing dual language students start to automatize around grade 3, as Italian monolinguals, may reduce over- or under-diagnosis of dyslexia in this population. In the case of poor reading performance, they could be delayed due to lower exposure to the second language. In this respect, to adopt a dynamic approach in assessing the dual-language child becomes of fundamental importance (see Caffrey, Fuchs & Fuchs, 2008, for a review).

***Correlations among cognitive, linguistic and reading tasks***

As for the relationships between the cognitive, linguistic, and Italian-L2 reading abilities in both the biliterate and monoliterate groups, we found that cognitive and language performance correlated with reading proficiency measures, as predicted. This result was in line with research that repeatedly demonstrated the contribution of cognitive and language skills to literacy development (*e.g.*, Gathercole *et al.*, 2003; White, 2019; Duff *et al.*, 2015. In particular, we found non-verbal fluid intelligence to correlate positively with almost all the other measures, except for passage reading accuracy. It should be noted that biliterate children outperformed their monoliterate peers on Raven scores, and they were outperformed by monoliterates on passage reading accuracy measure.

As IQ was found to differentiate the two bilingual groups (with biliterates outperforming monoliterates), we wanted to better understand the role of fluid intelligence and its interaction with reading and language measures. Grouping empirically the measures tested in three broad-principal factors of *Velocity* (including the reading velocity measures and the reading comprehension), *Vocabulary* (including vocabulary, pseudowords repetition, and the reading accuracy measures), and *Raven* (representing the IQ), we explored the correlation among them partialling out a component at a time. What emerged from this analysis is that IQ and vocabulary contributed to reading speed and comprehension, but we did not find a direct link between IQ and vocabulary. The relation between IQ and text comprehension is understandable since the latter requires abstract reasoning and inference abilities, among others, as discussed above. The association between IQ and reading speed is less clearly understandable. A higher IQ may be associated with more effective top-down processes that may be useful in reading quickly or that facilitate a predicting behaviour. It is, in fact, reasonable to think that IQ is implicated in tasks, such as reading fluently a text, which requires a high cognitive prowess. Lastly, the relationship between vocabulary and reading speed and comprehension is also understandable, since a larger vocabulary is responsible for higher reading speed and a higher comprehension while reading a text.

We also conducted a Principal Component Analysis to establish which variables accounted the most for data variability and we found that Raven contributed in different ways in the two groups: it turned out to weigh more for biliterate than for monoliterate group. Biliterate were found to outperform monoliterate peers in IQ, nevertheless, we could not prove a causal effect, as it could be simply casual.

***Biliteracy: when the two languages share or diverge on the writing system***

To conclude our study, we analyzed biliterate children's performance separately to verify the presence of a more effective cross-linguistic transfer between a minority-L1 that shares the same alphabetic writing system with Italian (Bialystok *et al.*, 2005; Geva *et al.*, 2000).

Concerning language and literacy achievement, we found differences between the L1-alphabetic group and the L1-logographic group at expressive vocabulary, reading comprehension, and working memory tasks, with biliterates with an alphabetic L1 being better than their L1-logographic peers. Among the other skills considered, we did not find consistent differences between the groups. Thus, the difference in terms of writing systems appeared to guarantee a more effective cross-linguistic transfer on some tasks in biliterate children with alphabetic L1. In particular, vocabulary showed benefit from the closeness of the languages involved, as biliterates obtained a better performance. In fact, besides similar orthographic systems, alphabetic languages also share phonological forms, which could yield an easier lexicon acquisition in the second language. Thus we suggest that bilinguals with two languages belonging to the same system could learn more easily new words in Italian-L2, facilitated by the language orthography and phonology proximity. That said, it follows that reading comprehension benefitted from larger vocabulary knowledge of the alphabetic-L1 biliterate group (see Spencer, Quinn & Wagner, 2014). Correlational and longitudinal studies have, in fact, underlined the influence of oral vocabulary on reading comprehension (*e.g.*, Rothou & Padeliadu 2015; Verhoeven, Leeuwe, & Vermeer 2011).

Lastly, there is extensive evidence that children's ability to repeat pseudowords is not entirely independent of language experience, as we already noticed, and is affected by the phonological proximity of the pseudowords to real words in the language. Thus children should be better able to repeat more wordlike pseudowords (Jones, Tamburelli, Watson, Gobet, & Pine, 2010; Messer, Leseman, Boom, & Mayo, 2010). Translating this evidence onto our research, a wider alphabetic-L1 children's vocabulary knowledge could have facilitated the pseudowords repetition task.

Thus, with regard to the properties investigated, we found an effective cross-linguistic transfer only in the alphabetic-L1 group, supporting the idea that students can transfer knowledge and skills from one language to another (Cummins, 1991).

***Correlations among measures in biliterates children***

Finally, we investigated the relationship between the various abilities in the two biliterate groups. We found that almost all the measures correlated with each other in the logographic group, confirming that both cognitive and lexical skills contribute to reading skill development (Gathercole *et al.*, 2003; Duff *et al.*, 2015). In the alphabetic biliterate group, we found something slightly different and less generalized. Working memory correlated with almost all the other measures, confirming its role in children's reading skills development (White, 2019). Moreover, pseudowords repetition performance was related to vocabulary knowledge, as it is in monolingual children (Gathercole, 2006). Raven and vocabulary correlated only with reading accuracy, among the reading proficiency measures. The link between vocabulary and reading words accurately is clear, as the lexical experience and knowledge contribute to the word recognition. It is possible that IQ is more implicated in word recognition in alphabetic-L1-L2 bilinguals, and that other factors – such as working memory or reading measures themselves – play a more relevant role in the other reading proficiency performance (rapidity and comprehension). Accuracy was found to correlate both with velocity and comprehension. Thus, the most accurate while reading, the more they read fluently and comprehend the text.

***Limitations***

The present study has some limitations, which can be used as a basis for future research. First, we had to renounce to include an Italian monolingual control group, because it was too unbalanced in size compared to the bilingual group. This was partly due to some difficulties we encountered in recruiting Italian mainstream schools, in part because of the Covid-19 sanitary emergency, which determined the schools' closure in Italy since February 2020. Thus, we could not compare biliterate and monoliterate bilingual students with Italian monolingual peers, as Rothou *et al.* (2017) did.. Even the bilingual samples were slightly unbalanced in terms of Grades, especially in the monoliterate group, and in terms of biliterate alphabetic *vs.* logographic samples, with Mandarin-

logographic children being more represented. Moreover, in the alphabetic-biliterate group, we included different languages together because they share the feature to have an alphabetic writing system. However, they differ for other morpho-syntactic or orthographic rules, determining their less or greater proximity to Italian. This may create some imbalances. Future research may consider more homogeneous groups, although this is often a difficult enterprise. Anyhow, these alphabetic minority languages mirror the principal presence of foreign ethnicity in Italian classes.

Secondly, the study has a cross-sectional design, and we are aware that a longitudinal design would be preferable for more valuable analysis.

Another limitation is that we did not assess bilinguals' native oral language, and thus we did not have information about L1 proficiency level. For the biliterate children, we did not investigate L1 literacy performance because no tools or qualified mother-tongue persons to collect data were available. Thus we could not verify the literacy outcome in the minority language, the distinctive role of oral competence in L1 reading ability, the relationships between and within languages, and changes in the relationships that may occur over time.

Lastly, as we were not able to estimate SES by questionnaires about academic parental level in some cases, we made SES estimation based on the geographical living area to get a homogeneous low-average SES sample. Even if we did our best to create translated versions of the questionnaire for each language, in organizing parents meeting with the teachers and some Italian-fluent foreign parents, we could not obtain all the questionnaires filled out.

Future research should more carefully document children's characteristics, such SES, the amount of exposure to the two languages, and changes in exposure over time. These *lacunae* fuel our constant hunger for knowledge and understanding of second language reading acquisition.

**Conclusions**

There are questions of considerable pedagogical importance, to better manage immigrants' difficulty at school, which have received little research attention in the Italian context. The current

study explored one possible strategy in supporting immigrant students' learning development in Italian-majority language at primary school: a biliteracy program aimed to acquire both L1-heritage and L2-Italian literacy in order to achieve better literacy outcome in the majority language (Italian).

The current study found biliterate bilinguals improve reading comprehension, among the literacy skills, which represent a fundamental ability for academic success. Concerning language and decoding skills in Italian, biliterates developed at comparable levels with their monoliterate peers. Thus, it seems that being able to read in both an L1-minority and Italian may improve reading comprehension, even though not literacy skills in general. However, no disadvantages emerged from the biliteracy mode, and both monoliterate and biliterate did not seem to face difficulties in decoding tasks. They both performed as Italian monolinguals, preferably using the lexical route to reading starting approximately from grade 3.

Furthermore, biliterate bilingual children appear to have advanced cognitive skills (Raven), which have been shown to contribute to reading development. Thus, biliteracy seems to provide some advantages over monoliteracy in some areas, although in other it seems to be the reverse (working memory skill, with monoliterates outperforming them).

Finally, within the biliterate groups, we found biliterates with an alphabetic L1 – as Italian is – to outperform their biliterate peers with a logographic L1 (Chinese Mandarin) on some task. We conclude that a cross-linguistic transfer of skills might be more effective in those bilinguals who couple two languages with the same written system.

Even if there might not be a wide biliteracy advantage when it comes to read, surely knowing more than one language, in both oral and written forms, provides significant advantages in some key skills for academic success. Moreover, it does not hamper literacy development in Italian. To conclude, updating the Italian educational policy in schools, implementing biliteracy route to learning remains a good starting point for a more inclusive and beneficial educational experience for children.

**References**

Abu-Rabia, S., & Siegel, L. S. (2002). Reading, Syntactic, Orthographic, and Working Memory Skills of Bilingual Arabic-English Speaking Canadian Children. *J Psycholinguist Res* 31, 661-678.

Alloway, T., Gathercole, S. E., Adams, A. M., Willis, C., Eaglen, R., & Lamont, E. (2005). Working Memory and Awareness as Predictors of Progress Towards Early Learning Goals at School Entry. *British Journal of Developmental Psychology*, 23, 417-426.

Antoniou, K., Grohmann, K. K., Kambanaros, M., & Katsos, N. (2016). The effect of childhood bilectalism and multilingualism on executive control. *Cognition*, 149, 18-30.

Baker, C. (2001). An Introduction to Bilingual Education. In *Foundations of Bilingual Education and Bilingualism* (3rd ed, pp. 181-202). Clevedon, England: Multilingual Matters.

Baker, C. (2001). Bilingual Education for Bilingualism and Biliteracy. In *Foundations of Bilingual Education and Bilingualism* (3rd ed, pp. 203-227). Clevedon, England: Multilingual Matters.

Baker, C. (2001). The Effectiveness of Bilingual Education. In *Foundations of Bilingual Education and Bilingualism* (3rd ed, pp. 228-243). Clevedon, England: Multilingual Matters.

Belacchi, C., Scalisi T. G., Cannoni E., & Cornoldi C. (2008). *CPM-Coloured Progressive Matrices. Standardizzazione Italiana*. *Manuale.* Giunti O.S. Organizzazioni Speciali, Firenze.

Bellocchi, S., Bonifacci, P., & Burani, C. (2016). Lexicality effect and stress assignment in bilingual children reading Italian as a second language. Bilingualism: Language and Cognition, 19(1), 89-105.

Berens, M., Kovelman, I., & Petitto, L. A. (2013). Learning to read in two languages: Should bilingual children learn reading in two languages at the same time or in sequence? Evidence of a bilingual reading advantage in children in bilingual schools from monolingual English-only homes**.** *Biling. Res. J*., 36(1), 35-60.

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, *9*(1), 43-61.

Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, Language Proficiency, and Learning to Read in Two Writing Systems. *Journal of Educational Psychology, 97*(4), 580-590.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525-531.

Blom, E., Boerma, T., Bosma, E., Cornips, L., & Everaert, E. (2017). Cognitive Advantages of Bilingual Children in Different Sociolinguistic Contexts. *Front. Psychol.* 8(552).

Bonifacci, P., & Tobia, V. (2016). Crossing Barriers: Profiles of Reading and Comprehension Skills in Early and Late Bilinguals, Poor Comprehenders, Reading Impaired, and Typically Developing Children. *Learning and Individual Differences*, 47, 17–26.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371-399.

Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The Predictive Validity of Dynamic Assessment: A Review. *The Journal of Special Education*, 41(4),254-270.

Chiat, S., & Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests: Effects of bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language, and Hearing Research,* 59, 1179-1189.

Ciccarelli, L. (1998). Comprensione del linguaggio, dei processi di elaborazione e memoria di lavoro: uno studio in età prescolare. *Unpublished doctoral dissertation, University of Padua*.

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International journal of language & communication disorders*, *43*(1), 1-40.

Compton, D. L. (2000). Modeling the Growth of Decoding Skills in First-Grade Children. *Scientific Studies of Reading*, 4(3), 219-259.

Cornoldi, C., & Carretti, B. (2016). *Prove MT-3-Clinica: La valutazione delle abilità di Lettura e Comprensione per la scuola primaria e secondaria di I grado.* Florence, Italy: Giunti Edu.

Cossu, G. (2013). *TNL - Test Neuropsicologico Lessicale per l'età evolutiva*. Florence, Italy: Hogrefe.

Cummins, J. (1979). Linguistic Interdependence and the Educational Development of Bilingual Children. *Review of Educational Research*, 49(2),222-251.

Cummins, J. (1989). Language and literacy acquisition in bilingual contexts. *Journal of multilingual and multicultural development,* 10(1),17-31.

Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (p. 70–89). Cambridge University Press.

Cummins, J. (1992). Bilingualism and Second Language Learning. *Annual Review of Applied Linguistics*, 13, 50-70.

Cummins, J. (2008). BICS and CALP: Empirical and Theoretical Status of the Distinction In: Street, B. & Hornberger, N. H. (Eds.) *Encyclopedia of Language and Education, 2nd Edition, Volume 2: Literacy.* (pp. 71-83). New York: Springer Science + Business Media LLC.

Daniels, P. T., & Bright, W. (1996). *The World's Writing Systems*. New York: Oxford University Press.

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: an example of publication bias? *Psychol. Sci.* 26, 99-107.

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, 116(15),7565-7574.

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes?. *Journal of child psychology and psychiatry, and allied disciplines*, 56(8), 848–856.

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: myth or reality? *Exp. Psychol.* 61, 234-251.

Engel de Abreu, P. M. J., Gathercole, S. E., & Martin, R. (2011). Disentangling the Relationship Between Working Memory and Language: The Roles of Short-Term Storage and Cognitive Control. *Learning and Individual Differences,* 21, 569-574.

Engel de Abreu, P. M., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012) Bilingualism enriches the poor: enhanced cognitive control in low-income minority children. *Psychol Sci*., 23(11),1364-1371.

Francis, D. J., Lesaux, N. K., & August, D. (2006). Language of Instruction. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (p. 365-413). Lawrence Erlbaum Associates Publishers.

Frith, U. (1985). Beneath the surface of developmental dyslexia. *Surface dyslexia* 32(1), 301-330.

Garraffa, M., Beveridge, M., & Sorace, A. (2015). Linguistic and Cognitive Skills in Sardinian-Italian Bilingual Children. *Frontiers in psychology*, *6*,1898.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: a test of phonological working memory. *Memory*, 2(2), 103-127.

Gathercole, S. E., Brown, L., & Pickering, S. J. (2003). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment levels. *Educational and Child Psychology, 20*(3), 109-122.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27,* 513-543.

Geva, E., & Siegel, L. S. (2000). Orthographic and cognitive factors in the concurrent development of basic reading skills in two languages. *Reading and Writing: An Interdisciplinary Journal, 12,* 1-30.

Geva, E. (2006). Second-Language Oral Proficiency and Second-Language Literacy. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (p. 123-139). Lawrence Erlbaum Associates Publishers.

Gottardo, A., Yan, B., Siegel, L. S., & Wade-Woolley, L. (2001). Factors related to English reading performance in children with Chinese as a first language: More evidence of cross-language transfer of phonological processing. *Journal of Educational Psychology, 93,* 530-542.

Grant, A., Gottardo, A., & Geva, E. (2011). Reading in English as a first or second language: The case of grade 3 Spanish, Portuguese, and English speakers. *Learning Disabilities Research & Practice, 26*(2), 67-83.

Grohmann, K. K. (2014). Towards Comparative Bilingualism. *Linguistic Approaches to Bilingualism*. 4.10.1075/lab.4.3.06gro.

Guasti, M. T., Costa, F., Chailleux, M., Pagliarini, E., Sarti, D., Granocchio, E., & Stucchi, N. (submitted). Diagnosing dyslexia in L2 Italian-speaking children: Are monolingual tools efficient? *Journal of research in reading*.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes Publishing Company.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychon. Bull. Rev.* 4, 625-658.

Hoff, E., Rumiche, R., Burridge, A., Ribot, K. M., & Welsh, S. N. (2014). Expressive vocabulary development in children from bilingual and monolingual homes: A longitudinal study from two to four years. *Early Childhood Research Quarterly*, *29*(4), 433-444.

Huang, H.-S., & Hanley, J. R. (1994). Phonological awareness and visual skills in learning to read Chinese and English. *Cognition, 54,* 73-98.

Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and frequency in Specific Language Impairment: Accuracy and error data from two nonword repetition tests. *Journal of Speech, Language, and Hearing Research*, *53,* 1642-1655.

Kaufman AS, Kaufman NL. Kaufman Brief Intelligence Test-II.Circle Pines, MN: American Guidance Service; 2004.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *Blackwell handbooks of developmental psychology. The science of reading: A handbook* (p. 209-226). Blackwell Publishing.

Landi, N., Frost, S. J., Menc, W. E., Sandak, R., & Pugh, K. R. (2013). Neurobiological bases of reading comprehension: Insights from neuroimaging studies of word level and text level processing in skilled and impaired readers. *Reading & writing quarterly: overcoming learning difficulties*, *29*(2), 145-167.

Messer, M. H., Leseman, P. P. M., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology, 105,* 306-323.

Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) (2020). *Gli alunni con cittadinanza non italiana A.S. 2018/2019* – Gestione Patrimonio Informativo e Statistica https://miur.gov.it.

Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor*. Amsterdam: John Benjamins.

Montrul, S. (2016). Dominance and proficiency in early and late bilingualism, in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 5-35.

Mumtaz, S., & Humphreys, G. (2001), The effects of bilingualism on learning to read English: evidence from the contrast between Urdu-English bilingual and English monolingual children. *Journal of Research in Reading*, 24,113-134.

Murineddu M., Duca V., & Cornoldi C. (2006), Difficoltà di apprendimento degli studenti stranieri. *Difficoltà di apprendimento*, 12(1), 49-70.

Nichols, E. S., Wild, C. J., Stojanoski, B., Battista, M. E., & Owen, A. M. (2020). Bilingualism Affords No General Cognitive Advantages: A Population Study of Executive Function in 11,000 People. *Psychological Science*, 31(5), 548-567.

Niolaki, G., & Masterson, J. (2012). Transfer effects in spelling from transparent Greek to opaque English in seven-to-ten-year-old children. *Bilingualism: Language and Cognition*, *15*(4), 757-770.

Paap, K. R. (2014). The role of componential analysis, categorical hypothesizing, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *J. Cogn. Psychol.* 26, 242–255.

Papastefanou, T., Powell, D., & Marinis, T. (2019). Language and Decoding Skills in Greek-English Primary School Bilingual Children: Effects of Language Dominance, Contextual Factors

and Cross-Language Relationships Between the Heritage and the Majority Language. *Front. Commun*., 4(65).

Perfetti, C. A., & Marron, M. A. (1998). Learning to Read: Literacy Acquisition by Children and Adults. In D. A. Wagner (Ed.), *Advances in Adult Literacy Research and Development* (pp. 89-138). Philadelphia, PA: Hampton Press.

Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., & Tan, L.-H. (2007). Reading in two writing systems: Accommodation and assimilation of the brain's reading network. *Bilingualism: Language and Cognition, 10*(2), 131146.

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1268.

Quiroga, T., Lemos-Britton, Z., Mostafapour, E., Abbott, R. D., & Berninger, V. W. (2002). Phonological awareness and beginning reading in Spanish-speaking ESL first graders: Research into practice. *Journal of School Psychology, 40*(1), 85-111.

Raven, J., Raven, J. C., & Court, J. H. (2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Section 3: The Standard Progressive Matrices.

Rolstad, K., Mahoney, K., & Glass, G. V. (2005) The Big Picture: A Meta-Analysis of Program Effectiveness Research on English Language Learners. *Educational Policy*, 19(4), 572-594.

Rothou, K., & Padeliady, S. (2015). Inflectional morphological awareness and word reading and reading comprehension in Greek. *Applied Psycholinguistics*, 36(4), 1007-1027.

Rothou, K. M., & Tsimpli, I. (2017). Biliteracy and reading ability inchildren who learn Greek as a second language. *International Journal of Bilingual Education and Bilingualism*, 23(8), 1036-1050.

Sampson, G. (2015). *Writing Systems*. Sheffield: Equinox.

Sartori, G., & Job, R. (2007). *DDE-2: Batteria per la valutazione della dislessia e della disortografia evolutiva-2*. Giunti OS, Organizzazioni Speciali.

Sears, C. (1998). *Second Language Students in Mainstream Classrooms: A Handbook for Teachers in International Schools*. Channel View Publications Ltd, Bristol (United Kingdom).

Shakkour, W. (2014) Cognitive Skill Transfer in English Reading Acquisition: Alphabetic and Logographic Languages Compared. *Open Journal of Modern Linguistics*, 4, 544-562.

Silva-Corvalán, C., & Treffers-Daller, J. (Eds.). (2015). *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Cambridge: Cambridge University Press.

Sorenson Duncan, T., & Paradis, J. (2016). English language learners' nonword repetition performance: The influence of age, L2 vocabulary size, length of L2 exposure, and L1 phonology. *Journal of Speech, Language, and Hearing Research, 59,* 39-48.

Spencer, M., Quinn, J. M., & Wagner, R. K. (2014). Specific reading comprehension disability: Major problem, myth, or misnomer? *Learning Disabilities Research & Practice*, *29*(1), 3-9.

Sulpizio, S., & Colombo, L. (2013). Lexical stress, frequency, and stress neighborhood effects in the early stages of Italian reading development. *Quarterly Journal of Experimental Psychology*, 66, 2073-2084.

Surrain, S., & Luk, G. (2019). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005-2015. *Bilingualism: Language and Cognition,* 22(2), 401-415.

Tong, F., Lara-Alecio, R., Irby, B., Mathes, P., & Kwok, O. (2008). Accelerating Early Academic Oral English Development in Transitional Bilingual and Structured English Immersion Programs. *American Educational Research Journal*, 45(4),1011-1044.

Valian, V. (2015). Bilingualism and cognition. *Biling. Lang. Cogn.* 18, 3-24.

van der Velde Kremin, L., Arredondo, M. M., Hsu, L., Satterfield, T., & Kovelman, I. (2019). The effects of Spanish heritage language literacy on English reading for Spanish-English bilingual children in the US. *International journal of bilingual education and bilingualism*, *22*(2), 192–206.

van Steensel, R. (2006). Relations between socio-cultural factors, the home literacy environment and children's literacy development in the first years of primary education. *Journal of Research in Reading, 29*(4), 367-382.

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary Growth and Reading Development across the Elementary School Years. *Scientific Studies of Reading*, 15(1), 8-25.

Vernice, M., Matta, M., Tironi, M., Caccia, M., Lombardi, E., Teresa Guasti, M., Sarti, D., & Lang, M. (2019). An Online Tool to Assess Sentence Comprehension in Teenagers at Risk for School Exclusion: Evidence From L2 Italian Students. *Frontiers In Psychology*, 10(2417), 1-13.

Videsott, G., Della Rosa, P. A., Wiater, W., Franceschini, R., & Abutalebi, J. (2012). How does linguistic competence enhance cognitive functions in children? A study in multilingual children with different linguistic competences, *Bilingualism: Language and Cognition,* 15(4), 884-895.

Weber, R. C., Johnson, A., Riccio, C. A., & Liew, J. (2016). Balanced bilingualism and executive functioning in children. *Bilingualism* 19, 425-431.

White, M. J. (2019). The development of English proficiency and working memory in 5–6 year old ELLs in their first year of formal education. *International Journal of Bilingual Education and Bilingualism*, DOI: 10.1080/13670050.2019.1571009.

Ziegler, J. C., & Goswami, U. (2005). Reading Acquisition, Developmenta Dyslexia, and Skilled Reading across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin, 131,* 3-29.

## Appendix

### *Monoliterate and Biliterate bilinguals*

***Table 1*** Descriptive statistics of monoliterate group in Grade 1

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 27 | 88,22 | 76,00 | 98,00 | 5,66 |
| **RAVEN** | 27 | 22,37 | 10,00 | 32,00 | 6,41 |
| **Vocabulary** | 27 | 23,56 | 5,00 | 39,00 | 9,22 |
| **Words velocity** | 27 | 0,84 | 0,35 | 1,79 | 0,41 |
| **Words accuracy** | 27 | 22,44 | 3,00 | 60,00 | 16,92 |
| **Pseudowords velocity** | 27 | 0,83 | 0,35 | 1,61 | 0,33 |
| **Pseudowords accuracy** | 27 | 12,26 | 1,00 | 30,00 | 8,82 |
| **Text velocity** | 27 | 0,99 | 0,38 | 2,55 | 0,59 |
| **Text accuracy** | 27 | 6,89 | 0,50 | 20,00 | 5,68 |
| **Text comprehension** | 27 | 4,41 | 1,00 | 10,00 | 1,95 |
| **Pseudowords repetition** | 27 | 33,11 | 26,00 | 39,00 | 3,62 |

***Table 2*** Descriptive statistics of monoliterate group in Grade 3

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 49 | 110,31 | 94,00 | 123,00 | 4,48 |
| **RAVEN** | 49 | 27,94 | 13,00 | 34,00 | 4,47 |
| **Vocabulary** | 49 | 30,88 | 7,00 | 43,00 | 7,53 |
| **Words velocity** | 49 | 2,57 | 0,50 | 4 | 0,78 |
| **Words accuracy** | 49 | 5,22 | 0,00 | 21,00 | 4,50 |
| **Pseudowords velocity** | 49 | 1,73 | 0,47 | 2,75 | 0,44 |
| **Pseudowords accuracy** | 49 | 5,14 | 0,00 | 18,00 | 3,98 |
| **Text velocity** | 49 | 2,79 | 0,48 | 7,17 | 1,20 |
| **Text accuracy** | 49 | 3,29 | 0,50 | 9,00 | 2,31 |
| **Text comprehension** | 49 | 4,90 | 0,00 | 11,00 | 2,68 |
| **Pseudowords repetition** | 49 | 35,84 | 23,00 | 40,00 | 2,83 |

***Table 3*** Descriptive statistics of monoliterate group in Grade 5

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 19 | 128,63 | 119,00 | 137,00 | 4,45 |
| **RAVEN** | 19 | 29,53 | 18,00 | 36,00 | 5,38 |

| | | | | | |
|---|---|---|---|---|---|
| **Vocabulary** | 19 | 24,79 | 8,00 | 84,00 | 16,84 |
| **Words velocity** | 19 | 3,26 | 1,56 | 4,84 | 0,97 |
| **Words accuracy** | 19 | 3,74 | 0,00 | 18,00 | 5,58 |
| **Pseudowords velocity** | 19 | 2,05 | 1,54 | 2,9 | 0,43 |
| **Pseudowords accuracy** | 19 | 3,42 | 0,00 | 17,00 | 4,82 |
| **Text velocity** | 19 | 3,32 | 1,54 | 5,01 | 1,05 |
| **Text accuracy** | 19 | 5,00 | 0,00 | 19,50 | 5,53 |
| **Text comprehension** | 19 | 5,05 | 1,00 | 9,00 | 2,41 |
| **Pseudowords repetition** | 19 | 37,21 | 34,00 | 40,00 | 1,51 |

***Table 4*** Descriptive statistics of biliterate group in Grade 1

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 19 | 84,11 | 79,00 | 92,00 | 4,28 |
| **RAVEN** | 19 | 28,16 | 20,00 | 35,00 | 4,63 |
| **Vocabulary** | 19 | 23,95 | 7,00 | 35,00 | 8,30 |
| **Words velocity** | 19 | 0,86 | 0,34 | 2,5 | 0,5 |
| **Words accuracy** | 19 | 19,26 | 1,00 | 70,00 | 19,89 |
| **Pseudowords velocity** | 19 | 0,82 | 0,39 | 2,15 | 0,43 |
| **Pseudowords accuracy** | 19 | 12,58 | 0,00 | 37,00 | 10,50 |
| **Text velocity** | 19 | 1,11 | 0,41 | 3,62 | 0,75 |
| **Text accuracy** | 19 | 7,38 | 0,00 | 21,50 | 7,16 |
| **Text comprehension** | 19 | 6,29 | 0,00 | 10,00 | 2,66 |
| **Pseudowords repetition** | 19 | 32,37 | 17,00 | 39,00 | 4,98 |

***Table 5*** Descriptive statistics of biliterate group in Grade 3

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 35 | 108,91 | 99,00 | 123,00 | 5,32 |
| **RAVEN** | 35 | 31,11 | 17,00 | 36,00 | 4,28 |
| **Vocabulary** | 35 | 28,51 | 10,00 | 43,00 | 7,93 |
| **Words velocity** | 35 | 2,23 | 0,7 | 3,9 | 0,78 |
| **Words accuracy** | 35 | 7,83 | 0,00 | 25,00 | 7,27 |
| **Pseudowords velocity** | 35 | 1,59 | 0,6 | 2,54 | 0,43 |
| **Pseudowords accuracy** | 35 | 7,77 | 0,00 | 24,00 | 6,07 |
| **Text velocity** | 35 | 2,48 | 0,93 | 4,41 | 0,97 |
| **Text accuracy** | 35 | 9,34 | 1,00 | 39,00 | 8,53 |
| **Text comprehension** | 35 | 5,71 | 2,00 | 11,00 | 2,37 |
| **Pseudowords repetition** | 35 | 35,54 | 27,00 | 40,00 | 3,15 |

***Table 6*** Descriptive statistics of biliterate group in Grade 5

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 32 | 130,34 | 121,00 | 152,00 | 6,14 |
| **RAVEN** | 32 | 32,03 | 18,00 | 36,00 | 3,92 |
| **Vocabulary** | 32 | 31,31 | 16,00 | 81,00 | 11,96 |
| **Words velocity** | 32 | 3,05 | 1,43 | 5,4 | 0,82 |
| **Words accuracy** | 32 | 4,44 | 0,00 | 18,00 | 4,17 |
| **Pseudowords velocity** | 32 | 2,1 | 1,19 | 3,45 | 0,52 |
| **Pseudowords accuracy** | 32 | 5,03 | 0,00 | 23,00 | 5,07 |
| **Text velocity** | 32 | 3,35 | 1,62 | 5,4390 | 0,92 |
| **Text accuracy** | 32 | 7,69 | 0,00 | 24,00 | 7,15 |
| **Text comprehension** | 32 | 5,50 | 0,00 | 9,00 | 2,20 |
| **Pseudowords repetition** | 32 | 35,09 | 25,00 | 40,00 | 3,65 |

### *L1-logographic-biliterates and L1-alphabetic-biliterates*

***Table 7*** Descriptive statistics of L1-logographic-biliterates in Grade 1

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 10 | 83,10 | 79,00 | 87,00 | 3,35 |
| **RAVEN** | 10 | 26,70 | 20,00 | 34,00 | 4,81 |
| **Vocabulary** | 10 | 19,40 | 7,00 | 32,00 | 8,82 |
| **Words velocity** | 10 | 0,67 | 0,34 | 1,10 | 0,27 |
| **Words accuracy** | 10 | 24,80 | 1,00 | 56,00 | 17,36 |
| **Pseudowords velocity** | 10 | 0,69 | 0,39 | 1,13 | 0,30 |
| **Pseudowords accuracy** | 10 | 14,20 | 1,00 | 36,00 | 10,82 |
| **Text velocity** | 10 | 0,823 | 0,43 | 1,48 | 0,37 |
| **Text accuracy** | 10 | 9,80 | 0,00 | 21,50 | 8,09 |
| **Text comprehension** | 10 | 4,50 | 0,00 | 8,00 | 2,27 |
| **Pseudowords repetition** | 10 | 30,90 | 17,00 | 37,00 | 5,99 |

***Table 8*** Descriptive statistics of L1-logographic-biliterates in Grade 3

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 19 | 109,26 | 99,00 | 123,00 | 5,62 |
| **RAVEN** | 19 | 31,05 | 17,00 | 36,00 | 4,74 |

| | | | | | |
|---|---|---|---|---|---|
| **Vocabulary** | 19 | 24,95 | 10,00 | 36,00 | 7,23 |
| **Words velocity** | 19 | 2,04 | 0,70 | 3,34 | 0,79 |
| **Words accuracy** | 19 | 9,42 | 1,00 | 25,00 | 7,67 |
| **Pseudowords velocity** | 19 | 1,54 | 0,60 | 2,56 | 0,50 |
| **Pseudowords accuracy** | 19 | 8,68 | 1,00 | 24,00 | 6,28 |
| **Text velocity** | 19 | 2,27 | 0,93 | 4,41 | 1,03 |
| **Text accuracy** | 19 | 11,55 | 1,00 | 39,00 | 9,38 |
| **Text comprehension** | 19 | 5,00 | 2,00 | 8,00 | 2,05 |
| **Pseudowords repetition** | 19 | 34,47 | 27,00 | 40,00 | 3,39 |

***Table 9*** Descriptive statistics of L1-logographic-biliterates in Grade 5

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 21 | 131,76 | 123,00 | 152,00 | 6,25 |
| **RAVEN** | 21 | 33,00 | 25,00 | 36,00 | 3,07 |
| **Vocabulary** | 21 | 30,86 | 16,00 | 44,00 | 7,46 |
| **Words velocity** | 21 | 3,29 | 1,84 | 5,41 | 0,83 |
| **Words accuracy** | 21 | 3,81 | 0,00 | 18,00 | 3,86 |
| **Pseudowords velocity** | 21 | 2,22 | 1,50 | 3,46 | 0,54 |
| **Pseudowords accuracy** | 21 | 4,24 | 0,00 | 12,00 | 3,62 |
| **Text velocity** | 21 | 3,53 | 2,21 | 5,44 | 0,84 |
| **Text accuracy** | 21 | 6,36 | 0,00 | 24,00 | 5,79 |
| **Text comprehension** | 21 | 5,38 | 2,00 | 9,00 | 2,13 |
| **Pseudowords repetition** | 21 | 34,95 | 25,00 | 40,00 | 3,93 |

***Table 10*** Descriptive statistics of L1-alphabetic-biliterates in Grade 1

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 9 | 85,22 | 79,00 | 92,00 | 5,09 |
| **RAVEN** | 9 | 29,78 | 25,00 | 35,00 | 4,09 |
| **Vocabulary** | 9 | 29,00 | 25,00 | 35,00 | 3,61 |
| **Words velocity** | 9 | 1,09 | 0,38 | 2,53 | 0,63 |
| **Words accuracy** | 9 | 13,11 | 2,00 | 70,00 | 21,68 |
| **Pseudowords velocity** | 9 | 0,97 | 0,43 | 2,15 | 0,52 |
| **Pseudowords accuracy** | 9 | 10,78 | 0,00 | 37,00 | 10,46 |
| **Text velocity** | 9 | 1,42 | 0,41 | 3,62 | 0,95 |
| **Text accuracy** | 9 | 4,68 | 1,00 | 17,50 | 5,10 |

| | | | | | |
|---|---|---|---|---|---|
| **Text comprehension** | 9 | 8,27 | 6,47 | 10,00 | 1,31 |
| **Pseudowords repetition** | 9 | 34,00 | 29,00 | 39,00 | 3,12 |

***Table 11*** Descriptive statistics of L1-alphabetic-biliterates in Grade 3

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 16 | 108,50 | 100,00 | 118,00 | 5,09 |
| **RAVEN** | 16 | 31,19 | 23,00 | 36,00 | 3,82 |
| **Vocabulary** | 16 | 32,75 | 22,00 | 43,00 | 6,68 |
| **Words velocity** | 16 | 2,47 | 1,01 | 3,90 | 0,74 |
| **Words accuracy** | 16 | 5,94 | 0,00 | 21,00 | 6,49 |
| **Pseudowords velocity** | 16 | 1,66 | 1,01 | 2,27 | 0,34 |
| **Pseudowords accuracy** | 16 | 6,69 | 0,00 | 22,00 | 5,82 |
| **Text velocity** | 16 | 2,73 | 1,26 | 4,29 | 0,85 |
| **Text accuracy** | 16 | 6,71 | 1,00 | 29,00 | 6,76 |
| **Text comprehension** | 16 | 6,56 | 3,00 | 11,00 | 2,50 |
| **Pseudowords repetition** | 16 | 36,81 | 32,00 | 40,00 | 2,34 |

***Table 12*** Descriptive statistics of L1-alphabetic-biliterates in Grade 5

| | N | Mean | Min | Max | Standard Dev. |
|---|---|---|---|---|---|
| **Age (months)** | 11 | 127,64 | 121,00 | 137,00 | 5,14 |
| **RAVEN** | 11 | 30,18 | 18,00 | 34,00 | 4,81 |
| **Vocabulary** | 11 | 32,18 | 16,00 | 81,00 | 18,18 |
| **Words velocity** | 11 | 2,59 | 1,43 | 3,67 | 0,60 |
| **Words accuracy** | 11 | 5,64 | 0,00 | 14,00 | 4,65 |
| **Pseudowords velocity** | 11 | 1,88 | 1,19 | 2,49 | 0,42 |
| **Pseudowords accuracy** | 11 | 6,55 | 0,00 | 23,00 | 7,05 |
| **Text velocity** | 11 | 3,01 | 1,62 | 4,86 | 1,01 |
| **Text accuracy** | 11 | 10,23 | 0,00 | 22,00 | 8,98 |
| **Text comprehension** | 11 | 5,73 | 0,00 | 9,00 | 2,41 |
| **Pseudowords repetition** | 11 | 35,36 | 30,00 | 40,00 | 3,20 |

## Chapter 3

# NARRATIVE SKILLS IN PRIMARY SCHOOL CHILDREN IN ITALIAN-ENGLISH IMMERSION PROGRAMS

### Introduction

Narration is the basis of our connate sociality. Among the oral language skills, the ability to produce a narration allows people to share fictional events, abstract ideas, and personal experiences, attributing meanings to them (Nelson, 2007) and organizing them in relevant episodes. For children, to narrate is essential to share their life experiences, something that has happened, to express their imagination, their emotional experiences, describe a plot, and organize their beliefs and knowledge. Narrations have, in fact, a steady chronological structure and directionality, with a series of events linked to each other: an onset, a development, and an end (Levorato & Nesi, 2001). Narrative skills can be defined as the ability to tell a cohesive and structured story (Bonifacci, Barbieri, Tomassini & Roch, 2018), which can be understood by an interlocutor. Moreover, the narrative ability has been shown to predict later development of literacy skills like reading, writing, text comprehension (*e.g.*, Ohakill & Cain, 2012; Pearson, 2002; Wellman, Lewis, Freebairn, Avrich, Hansen & Stein, 2011 for a longitudinal study), and of comprehension of the language of mathematics (*e.g.*, Bliss, McCabe & Miranda, 1998; McCabe, 1996; Walach, 2008). Thus, narrative skills are essential for success at school (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker & Walters, 2015), by providing experience in using extended, contextualized, cohesive discourse units and abstract texts that children will encounter in written texts (Hadley, 1998; Westby, 2005). One of the main tasks young children are facing in becoming literate is discovering the interrelationships between oral language and literacy (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker & Walters, 2012). The ability to tell a story could be defined as a bridge between oral language and literacy. It requires children to plan and produce contextualized and cohesive

discourse (Gagarina *et al.*, 2015), and discourse is the principal linguistic medium through which academic information is disseminated and acquired.

Narratives are considered an ecologically valid way to investigate children's communicative competence and language often used by researchers and clinicians (Botting, 2002). It offers a platform for a wide range of linguistic abilities (Hughes, McGillvray & Schmidek, 1997) in a relatively natural context. Narrative tasks address how language is expressed in everyday life and thus represent the relationship between different linguistic processing levels (Bonifacci *et al.*, 2018). Among multiple linguistic features, narratives allow parallel assessment of the capacities beyond the specifics of language, and of the linguistic specific features too – using relatively straightforward language samples (Heilmann, Miller & Nockerts, 2010).

Narrative assessment taps language-specific and language-general skills, making it a reliable instrument for bilingual children who speak more than one language. Narrative tasks might be less biased against dual-language children than norm-referenced assessment tools because the cognitive component could be tapping into language-general capacities (Paradis, Genesee, & Crago, 2011). Thus, narratives might be less susceptible to the disadvantages of tests standardized on typical monolingual performance (Simon-Cereijido & Gutiérrez-Clellen, 2009).

In the past few years, instruments for assessing bilingual children have been the focus of the *COST Action IS0804 Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment*. Within it, an instrument for the evaluation of narrative abilities across languages has been developed. Moreover, the analysis of narrative skills development has recently been examined more thoroughly in studies that targeted bilingual populations (*e.g.*, Roch, Florit & Levorato, 2016; Squires, Lugo-Neris, Peña, Bedore, Bohman & Gillam, 2014).

The following sections will describe the narrative task, its features, and its use among bilingual children. We will present the different assessment levels enclosed in a narrative task – namely, microstructure and macrostructure – and their respective relevance in the bilingual context.

We will explain the relevance of assessing narrative comprehension to obtain a complete picture of a child's abilities. Then we will describe the most common methods and procedures to collect and analyze narrative productions. Finally, we will conclude with the *Multilingual Assessment Instrument for Narratives (LITMUS-MAIN,* Gagarina, *et al.*, 2012), which has been used in the present study.

**Macrostructure and microstructure**

In this section, we will describe the two levels on which a narrative task taps, language-specific and language-general

When a child produces a story, s/he simultaneously integrates the various levels underlying the language, representing the narrative *microstructure*, while conceiving the complex narrative framework defined as *macrostructure*. Microstructure and macrostructure represent two distinct but interrelated levels bearing on narrative competence (Liles, Duffy, Merritt & Purcell, 1995; Pearson 2002) and thus they allow assessing children's language from two different perspectives. Microstructure focuses on the linguistic structures used and covers a wide range of linguistic elements, including length and complexity of utterances, vocabulary, and morphosyntactic aspects, among others (even code-mixing in bilingual populations). It follows that microstructure elements are language-specific and differ across languages (Gagarina *et al.*, 2015).

Macrostructure refers to the story's higher-order hierarchical organization, including episodic structure, contents, and *story grammar* components (Heilmann, Miller, & Nockerts, 2010). According to the story grammar model, all stories have a universal organizational pattern (Trabasso & Nickels, 1992): a setting, which defines the temporal and spatial background and the protagonists; an initial event; a plan to overcome the situation and the actions to implement it; the consequences of the actions; and the mental states of the protagonists as an initial reaction or as the outcome response (Stein & Glenn, 1979). In this regard, Shapiro e Hudson (1991) identified five staple components of a story: (i) the setting; (ii) the protagonist's goal; (iii) the attempt to pursue the

goal; (iv) the outcome (and the resolution of the event); (v) the emotional reactions. These are the fundamental elements of story episodes, and they are essentially the same, whatever is the language or cultural extraction (Berman & Slobin, 1994; McCabe & Bliss, 2005; Verhoeven, 2004). For this reason, macrostructure can be said to be language-independent (Pearson, 2002).

The ability to produce well-formed episodes in narratives indicates an understanding of the narrative scheme, causality relations, the need to justify plans and actions, and perspective-taking (Trabasso *et al.*, 1992; Trabasso & Rodkin, 1994). To produce cohesive and coherent stories is essential to interpret protagonists' goals, intentions, and emotions, which presupposes the awareness of others' state of mind. This ability involves cognitive processes implicated in interpreting intentionality and making inferences about factual and emotional aspects of stories (Westby, 2005), which are reflected in the Theory of Mind (ToM) abilities. The presence of mental state terms in narratives is a good indicator of story production's linguistic complexity (Lever & Sénéchal, 2011; Squires *et al.*, 2014). Moreover, it is a crucial aspect for the development of school-based discourse (Grazzani & Ornaghi, 2012) and complex syntax (Nippold, Ward-Lonergan, & Fanning, 2005). The use of metalinguistic (such as *say*) and metacognitive (such as *think*) verbs, of emotional adjectives (such as *sad, happy*) are usually more frequently present in narratives of literate children. It follows that more complex macrostructure elements could emerge only at a certain stage of cognitive and linguistic development in children.

In the following section, we will consider the narrative development in children.

**Narrative development in children**

Research on typically developing monolingual pre-school children has shown that, at 5 years of age, they have acquired a basic narrative macrostructure of stories (Trabasso, Stein, & Johnston, 1981), and can produce narratives that include more than two sequentially connected elements. However, their narratives often end suddenly or lack a proper story ending (Peterson & McCabe, 1983). At the same age, children start to include in their narratives some characters' mental states

and motivations, although this ability mostly emerges during primary school years (Stein *et al.*, 1979). Berman (1988), assessing monolingual pre-school and school-age children's narratives, found that the former had very variable productions, with stories generally poorly developed at micro- and macrostructure levels. In particular, they did not master more complex grammatical structures yet, nor they have a vocabulary competence, which allowed them to produce complete and complex stories. At about 6 years of age, at the beginning of schooling, children proved to have acquired the basic macrostructure of the story, even if they still partially lacked the ability to organize the general and sequential sense of the story, mainly focusing on concrete events or visual details they saw in the pictorial stimuli. Around second-to-third grade, at about 7 years of age, children are able to narrate a sequence of events that is purpose-oriented: stories typically include an initiating event, characters' goals and actions, and consequences. Later on, during primary school, children's ability to structure and organize narratives continues to develop, and older children include more and more narrative elements, mental states, and temporal and causal relations between events (Stein *et al*., 1979).

Studies on bilingual children have shown that – like their monolingual peers – they can produce narratives in both their languages (L1 and L2), with increasing complexity and length, passing from pre-school to school-age (Fiestas & Peña, 2004; Gutiérrez-Clennen, 2002; Pearson, 2002; Iluz-Cohen & Walters, 2012). This development involves both the microstructure and the macrostructure level and leads to enhancing of the narrative production. In a longitudinal study by Uccelli and Páez (2007) involving typically developing (SES) Spanish-English bilingual children from a low-socioeconomic status, the authors found that the macrostructure measures assessed in L1 and L2 significantly improved from kindergarten to primary school. The measures collected evaluated narrative productivity and quality (*i.e.,* story elements, story, sequence, and perspectives), which contributed to a global measure of narrative quality. Squires, Lugo-Neris, Peña, Bedore, Bohman, and Gilliam (2014) analyzed the macrostructure of stories told by children with typical development and with language impairment in Spanish (L1) and English (L2), in a retelling task

(retell a story with a picture stimuli sequence, see the section below about telling and retelling narrative paradigms). Results showed that bilinguals with typical development during the first grade of schooling included more references to characters, initiating events, and internal responses in both L1 and L2 compared to kindergarteners. Thus, bilingual children displayed the same developmental pattern as monolingual children in both languages, as far as macrostructure is concerned.

In the following section, we will first consider the assessment of the microstructure level in bilingual children.

**The assessment of microstructure in L2-learners**

The language assessment of bilingual children should not follow a standard route as for monolinguals. By this time, it is well known that bilingual vocabulary width – assessed in one language – is commonly less broad than that of monolingual peers. In a broad-scale study of bilingual Spanish-English learners and monolingual English in Miami, Oller, Pearson, and Cobo-Lewis (2007) found lower vocabulary scores for the bilinguals in both languages as a primary feature of such population. The authors affirmed that it was due to the "distributed" bilingual lexical knowledge, as bilinguals tend to know some words in one language but not in the other and *vice versa*. Vocabulary acquisition is, in fact, strictly linked to the context of exposure, and daily-life terms or home-routine terms are likely acquired in the family-language context (L1), whereas L2 vocabulary could leave the place for school-related terms. Moreover, bilingual lexis is typically studded by bilingual *phenomena* such *as code-mixing* or *switching*, which are physiological in bilingualism and have to be interpreted as a communicative strategy.

Paradis (2016), in a review on English development as a second language, argued that English Language Learners (ELLs) could take longer than three years to converge on monolingual norms and approach monolingual norms asynchronously across linguistic subdomains. In the initial stages of acquisition of English as a second language (L2), some typical pattern of error could be observed: frequent errors in verbal bound morphology (third singular person, past tense; *e.g., *John*

***walk*** *to school /*I* ***talk*** *to her yesterday*), abuse of General All-Purpose (GAP) verbs (*e.g., do, go, make, play*), redundant use of auxiliaries (*e.g., be, do*) as tense markers (*e.g.,*But if I* ***was*** *hurt my teeth*). Blom, Paradis, and Sorenson Duncan (2012) examined the verb morphological development of 14 English Language Learner (ELL) children with diverse first language (L1) backgrounds (Korean, Mandarin, Cantonese, Spanish, Romanian, Japanese, and Farsi). Children were evaluated longitudinally, from preschool- till school-age (5;4 to 7;4 age;months), in the use of third-person singular -s (*i.e.*, "*I walk; you walk; he walks*"). They found that at the beginning of the study children committed morphosyntactic errors consistently, and then gradually improved across ages (children's mean accuracy with third-person singular -*s* was at 84% by the end of the study). However, when children were grouped according to the typology of their L1, those with richly L1 verb inflectional paradigms reached monolingual accuracy norm even faster, compared to those with isolating L1s (with little or no verb inflection), thus proving the influence of the first language on developmental trajectories.

It is evident that some time of exposure is required for bilinguals to reach a monolingual level of competence and that some language fragility is part of a typical bilingual L2 development. Before a reasonable time of L2 exposure, tests evaluating vocabulary or morphosyntax – standardized on a monolingual population – produce poorly reliable information, difficult to interpret. In fact, the surface similarity between a bilingual profile in an early stage of L2 learning and the profile of a monolingual child with Developmental Language Disorder (DLD) could lead to over-identification of DLD in the bilingual population, and this has been a source of persistent concern (Bedore & Peña, 2008; Kohnert, 2010; Paradis *et al.*, 2011). Thus, variability in the language abilities of L2 learners presents challenges for clinical practice. Increased knowledge of English language learning development with and without DLD and evidence-based alternative assessment strategies can help overcome these challenges (Paradis, 2016). A dynamic assessment is a valid strategy, which allows to monitor any gain or improvement in L2 development over time (*e.g.,* after one year of English L2 exposure and an L1 featuring the presence of articles, one can

observe a correct use of determinative/indeterminate *the/a* articles in English, see Zdorenko & Paradis, 2011). As said, bilingual performance depends primarily on the quantity of time of exposure, and it is possible to obtain a distinct profile after at least three years of exposure in the absence of disorders (Vender, Garaffa, Sorace and Guasti, 2016). Moreover, as a DLD should manifest in both languages, it would be useful to evaluate both the languages (and their development), compare them, and better orientate a judgment. Unfortunately, this is often a not passable road because of the absence of comparable instruments to assess a specific ability.

The assessment of microstructure in L1 and L2, when collected with the same instrument and method, could offer an excellent base to collect information on different levels (*e.g.*, lexis, morphosyntax) and to compare errors across languages. Elicited narrative tasks synergistically solicit children's language skills, as they are asked to produce a story with its complexity, and are a more ecological instrument in providing a sample of the child's linguistic competence. Lastly, they allow appreciating children's progress, for both languages and across time. It might be useful to observe the bilingual child's developmental milestones, evaluate his evolution across time, his improvement potential. The analysis of narrative microstructure ability provides a clear idea of how typical bilingual development proceeds. In fact, in a longitudinal study by Squires *et al.* (2014), TD bilingual children improved microstructure measures from kindergarten to first grade more than their DLD peers in both languages.

Nevertheless, studies that examine the development of narrative microstructure features of bilingual children's languages are still relatively scant, and only L2 is usually assessed (Gagarina *et al.*, 2015). In a study by Bonifacci *et al.* (2018), sixty-four early bilingual preschoolers (mean age 4;8) – exposed to Italian since kindergarten – speaking different minority languages at home (L1) were compared to a group of Italian monolingual peers. Beyond some standardized measures, their linguistic abilities were also tested using a story-sequencing task. Even if it emerged that bilinguals and monolinguals differed in fewer measures when a more ecological task is adopted (a narrative task), they still differed in some lexical and morphological indexes. In fact, Bilinguals committed

more free morphemes errors, more lexical errors and produced a lower number of total words and fewer types and tokens. Narrative microstructure resulted in a more realistic language production measure; however, a bilingual child evaluated just on linguistic aspects of his production is still penalized – mainly when only L2 is assessed. Moreover, the dynamic assessment might have some limitations, as a dilated time needed to reach a complete assessment over time.

EEL children typically converge faster on monolingual norms on cognitive-linguistic measures, such as the narrative macrostructure. Thus, it is worth considering the ability of a bilingual to produce a narrative story grammar as a starting point in the evaluation process. We will tackle the macrostructure assessment in bilinguals in the following section.

**Bilingual assessment: the importance of macrostructure level**

As discussed above, narratives of L2 learners could contain several errors at the microstructure level. Several studies indicated that typical bilingual development (TD) children score lower than monolingual peers in terms of vocabulary and morphosyntactic abilities in their narrative productions (Berman *et al.*, 1994; Fiestas *et al.*, 2004; Pearson, 2002; Uccelli *et al.*, 2007). Some of the early clinical markers of DLD (*e.g.*, small vocabulary, lexical access difficulties, the omission of verb inflections, and the omission of auxiliary verbs) are also indicators of typical language development in bilingual children (Leonard, 1998). Thus, being a narrative weak in lexis and morphosyntax is a common feature of monolingual children with DLD (*e.g.*, Reilly, Losh, Bellugi &Wulfeck, 2004; Kit-Sum To, Stokes, Cheung & T'sou, 2010), it becomes difficult to disentangle DLD and bilingualism. The risk of misinterpreting some linguistic weakness is real.

Narratives of TD bilingual children are equivalent to the narrative information produced by their monolingual peers (Fiestas *et al.*, 2004; Uccelli *et al.*, 2007). Contrarily, it is well known that language impairment affects language processes along with specific properties of narrative discourse. In particular, DLDs' narrative output has been described as being little informative and organized (Tsimpli, Peristeri & Andreou, 2016). Monolingual DLDs also encounter difficulties at

the narrative macrostructure level, with stories poorly verbose (Schneider, Hayward & Dubé, 2006), difficulty maintaining a topic, event sequencing, and informativeness (Roth & Spekman, 1986; Olley, 1989), and in referencing of individuals and events. Thus, TD bilinguals' narratives should differ from monolingual and bilingual children with DLD concerning story complexity measures. Furthermore, DLD children – both bilingual and monolingual – should manifest the same difficulties in macrostructure. Cleave, Girolametto, Chen, and Johnson (2010) confirmed this assumption. They compared two groups of DLD children on some norm-referenced standardized language tests and narrative measures: one group comprised 12 Canadian dual language learners (English dominant with different minority languages; mean age in months 52.9); the other group, 14 monolingual English speakers (mean age in months 52.2). Children were assessed on two different narrative samples, elicited with different instruments (Renfrew Bus Story, Cowley & Glasgow, 1994; Edmonton Narrative Instrument, ENNI, Schenider, Dubé, & Hayward, 2005). Results showed that both DLD groups did not differ on narrative structure and complexity scores, differently from a standardized test, in which dual-language children achieved lower scores than monolinguals. The authors concluded that narrative tasks could be a more sensitive way to assess dual language learners with a developmental language disorder than language standardized tests, passible to more bias. Macrostructure has been shown to be a stable measure distinguishing children with and without DLD – regardless of bilingual or monolingual status – in numerous other studies, and this is valid already at the early stages of bilingual acquisition. Squires *et al.* (2014) showed that Spanish-English bilingual TD children, in a retelling task, outperformed a comparable group of children with language impairment on macrostructure measures in kindergarten and first grade. Moreover, TD bilingual children exhibited more extensive improvements in both languages for macrostructure, while children with language impairment made a limited improvement. Rezzonico, Chen, Cleave, Greenberg, Hipfner-Boucher, Johnson, Milburn, Pelletier, Weitzman, and Girolametto (2015) found significant differences on narrative measures between TD and DLD

children (monolingual and bilingual), regardless of their bilingual status, and despite differences in exposure to L2 (English) in the bilinguals. Moreover, both monolingual and bilingual TD children had strong macrostructure scores, which grew over six months (from Time 1 to Time 2). Comparable results were obtained in other studies, even for diverse first-language backgrounds (e.g., Pearson, 2002; see Paradis, Schneider & Duncan, 2013, for older children).

Narratives of TD bilingual children are also similar across their two languages. In the study conducted to develop the Multilingual Assessment Instrument for Narratives (MAIN), Gagarina *et al.* (2012) found that bilingual children built analog story structures in both their languages. They interpreted this result as an effect of a sharing of competencies between L1 and L2. As the mental representation of story grammar is a cognitive competence at the interface with language processes, it might develop independently and before the full achievement of L2 acquisition. Thus, the macrostructure transfer is primarily related to a conceptual level, available in the processing of L2 once a linguistic threshold has been reached (Bonifacci *et al.*, 2018; see also Pearson 2002; Schwartz and Shaul, 2013). These results suggest that the narrative measure of macrostructure should be taken into consideration when assessing bilingual children. This ability is not biased against children who have less experience with a particular language, supporting the identification of language impairment in a bilingual context (Boerma, Leseman, Timmermeister, Wijnen & Blom, 2016). Moreover, it allows a comparison between the two languages spoken by the child. In the absence of language impairment, a comparable macrostructure would be expected in the stories produced in the two languages.

The effectiveness of an inter-language connection at the macrostructure level was also observed in the study of Tsimpli *et al.* (2016). The authors tested the narrative abilities of four groups of school-age children (mean age 9 years, approximately), using the Greek version of the MAIN. Twenty-one Greek-speaking monolingual and 15 bilingual children with DLD were assessed, along with two comparable-size samples of monolingual and bilingual age-matched TD children. They found a bilingualism effect both between TD and DLD children: an advantage for

bilingual children – with and without DLD – was observed in developing and understanding the stories' global coherence compared to monolingual children. In this case, bilinguals appeared to benefit from some cognitive advantage, which compensated for DLD children's weaknesses and preserving narrative abilities at the macrostructure level. The authors suggested that bilingualism boosted the children's ability to abstract language-independent narrative properties, making narrative structure and coherence more accessible for bilingual children, even with DLD. As said, macrostructure is organized at a more abstract level, tightly linked to discourse representation. Despite the bilingual advantage found, Tsimpli *et al.* did not identify macrostructure as a disentangling tool for DLD in their population. They interpreted the positive impact of bilingualism on macrostructure as due to the older age of the participants – compared to previous studies – which permits greater ease to share abstract structures between the two languages. Thus, in Tsimpli *et al.*' (2016) study, narrative macrostructure would not be crucial in the identification of language impairment in bilingual older children. But this study represents an exception, as a narrative task analysed on the macro-level can support the assessment of bilinguals' language abilities at an earlier stage of their life, being generally useful for differential diagnosis.

To conclude the review, Boerma, Leseman, Timmermeister, Wijnen, and Blom (2016) pointed out that expressing a narrative's macrostructure is relatively independent of experience in a specific language, also understanding it. To produce and comprehend the coherence of a story is particularly valuable to identify language impairment in a bilingual context. We will discuss in detail narrative comprehension in the following section.

**Narrative comprehension**

Boerma *et al.* (2016), beyond narrative production, compared narrative comprehension skills of 5-to-6 years old mono- and bilingual children with and without DLD, using the MAIN (Gagarina *et al.* 2012). Bilingual children had Dutch as the majority language and different minority languages (with at least one parent native speaker of another language than Dutch). They found a negative

effect of DLD in both the mono- and bilingual groups, with children with DLD having weaker narrative comprehension (they found negative effects on all MAIN components). Furthermore, they did not observe bilingualism's effects: bilingual and monolingual children performed similarly on narrative comprehension, indicating that there was no bilingual disadvantage due to having received less input in Dutch. These findings suggested that children with DLD, either mono- or bilingual, might have relatively more difficulties with the ability to comprehend a narrative's macrostructure, which means understanding internal states, feelings, and intentions, and goal-directed behaviors of protagonists. In fact, children with DLD are reported to be delayed in understanding others' mental states (*e.g.*, Farrant, Fletcher & Maybery, 2006), supporting the relationship between language and ToM.

Several previous studies have shown that narrative production and comprehension are related. For instance, Cain (2003) demonstrated that comprehension skill was related to the coherence of narratives produced. Children who had difficulty understanding narrative texts were also more impoverished at producing structured, purposeful fictional narratives (see also Cain & Oakhill, 1996), supporting the idea that narrative production and comprehension share the ability to construct a coherent representation of the meaning of a story.

Together these findings indicate that the inclusion of comprehension within a narrative task could support the identification of DLD, making it a complete language assessment method for bilingual children and less biased than other tests tapping into language-specific knowledge (Paradis *et al.,* 2010).

In the following section, we will present the two principal modes to assess narrative production.

**Assessment modes: Telling and Retelling**

There are varied methods and procedures to collect and analyze narrative speech samples. Firstly, samples can be spontaneous or elicited. Personal experiences' narration fits more with a

spontaneous production, while fictional stories are produced more easily with elicitation methods. Among the narrative elicitation paradigms *telling* and *retelling* are the most represented. The telling mode consists in the generation of a story looking at a stimulus picture and based on an initial prompt. The retelling task offers a script model, besides pictures, and the child has to recall and tell the story he listened to, as detailed as possible. In both cases, using picture elicitation tasks encourages the child to talk.

In the context of research on narrative competence, there is no unanimous agreement on one task's preference rather than the other. The telling format is presumed to be more complicated since it demands a higher processing load. The child must generate his own story without the benefit of a prior script, selecting relevant information, their relations, and the hierarchical order of the episodes to narrate. Conversely, retelling seems to have lower demands because children can rely on the story script, stored in the memory. However, retelling just involves different processing domains and is more than just a repetition of the stimulus narrative (Gagarina *et al.*, 2015). Nelson (2007) defined the telling as a "purer" measure of narrative production, as it pays more attention to linguistic aspects. On the contrary, a child should be more focused on the comprehension and memorization of the script during a retelling performance (Gutiérrez-Clennen, 2002), as it involves story recall, verbal memory, and attention (Boudreau, 2007; Dodwell & Bavin, 2008; Graybeal, 1981). In fact, the two narrative practices should engage different cognitive abilities and processing modes, and they would not be equivalent (Meritt & Liles, 1989). Lever *et al.* (2011) suggested that telling might involve a constructive process, whereas story retelling is based on reconstructive and reinterpretation processes, through which characters, thoughts, and actions are reproduced. The authors compared performances at story telling and story retelling in a group of first-grade students and found that narratives produced in the retelling mode were significantly longer, richer in mental state references, and better structured than those elicited in the telling task. This was in line with some previous studies (Schneider & Dubé, 2005; Hayward, Gillam & Lien, 2007) that have found that retelled narratives often included more structural elements and were more detailed and more

grammatically accurate. Thus, even if the telling procedure may offer the child more freedom to use his/her imagination, performances are typically more exhaustive at retelling tasks (Roch *et al.*, 2016).

Nevertheless, the telling format may better reflect the child's lexis and provide more information about narrative formulation abilities than retelling (Schneider *et al.*, 2006). For this reason, specialists and researchers often use story telling to detect valuable information about children's language development, lexical knowledge, and morphosyntactic ability (Cleave *et al.*, 2010; Iluz-Cohen *et al.*, 2012). On its side, retelling provides information about how children modify and assimilate a story's vocabulary and grammatical structures and the content of the story (Gillam & Carlisle, 1997). Finally, in research, retelling offers control over certain aspects of the narrative, as length, complexity, and content of the story, thus allowing for a more congruent analysis of the performance (Hadley, 1998).

**The *Multilingual Assessment Instrument for Narratives***

This section will present the *Multilingual Assessment Instrument for Narratives (MAIN*, Gagarina, *et al.*, 2012).

The tool was designed to assess bilingual children's narrative production and comprehension skills from 3 to 9 years (later referred to as *Language Impairment Testing in Multilingual Settings: Multilingual Assessment Instrument for Narratives, LITMUS-MAIN*). It was developed within the *COST Action IS0804 Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment*, by the Working Group for Narrative and Discourse, to evaluate narrative abilities across languages. It can thus be used with a variety of languages and language combinations. The tool has been piloted in 17 languages (Afrikaans, Albanian, Croatian, Cypriot Greek, Dutch, English, Finnish, French, German, Greek, Hebrew, Italian, Lithuanian, Russian, Swedish, Turkish) and 14 language pairs with more than 500 monolingual and bilingual children.

The test aims to elicit narrative production in three modes: story telling, story retelling, and storytelling after listening to a model story. The latter was introduced as a variant of the retelling procedure, whereas the story model is similar in narrative structure yet different in content from the target story to be told (the child first sees and listen to a story, then sees and have to generate a new story, which is just similar to the model).

Each story script included in the instrument is designed with the same structure (when applicable, *i.e.*, retelling and model story): an initial setting statement which defines time and place, followed by three episodes. Each episode shares the same scaffolding: an internal state of the protagonist which determines his goal, a goal for the protagonist, an attempt by the protagonist to reach the goal, an outcome of his attempt, and another expression of internal state as a reaction. The congruence across the scripts set makes the stories comparable and allows investigating a child's performance cross-linguistically in a within-subject design.

The tool provides four six-picture sequences based on 3 to 9-year-old children's linguistic and cognitive abilities (Gagarina *et al.*, 2015). Two sets were designed for the telling mode (Baby Birds story and Baby Goats story), and two for the retelling or model story modes (Cat story and Dog story). The pictorial sequences were designed to portray clearly and explicitly content components and actions, with the same details and introduction of protagonists, their interaction and relation with the background and foreground, and comparable onset, development, and conclusion of the storyline. Thus, the parallelism of storylines for the different sets of pictures guarantees comparability across the four narratives. Each picture sequence portrays three short episodes in order to provide three opportunities for a child to produce the story structure elements targeted for macrostructure analysis. Thus, pictorial content is congruent with scripts, and all the four stories – *Baby Birds, Baby Goats, Cat*, and *Dog* – are matched for the number of main protagonists, the structure of the plot, internal states, and general actions performed by the protagonists (see Appendix section for an example of picture stimuli).

As mentioned above, the stimulus scripts were designed to be comparable in both macrostructure and microstructure information. Identical macrostructure components accommodate different formulations (in different languages) of similar lexical items. Thus, the test allows for cross-language analyses on both levels, and the scoring system is the same for all the languages tested so that bilingual children can be assessed in both of their languages identically and reliably.

The story's microstructure was designed to be as similar as possible across stories and languages, with the same logical sequence of utterances and many other linguistic features. Conversely to macrostructure, which is coded and scored with a fixed taxonomy, the microstructure can be analyzed with no preclusions, and narrative productions could be assessed from different microstructure perspectives. Nevertheless, Gagarina *et al.* (2015) indicated ten features as an initial basis for the microstructure analysis of LITMUS-MAIN narratives. They include measures of narrative length and lexis (*e.g.,* the total number of word tokens *(Token),* number of different words *(Type))*, and measures of syntactic complexity and discourse cohesion (*e.g., the* proportion of subordinating or coordinating constructions).

Macrostructure features number and sequence assessed are identical across the four versions of the stimulus scripts and languages. Each story is formed by two setting elements referring to time and place (*e.g., once upon a time/one day; on a bush/near a tree*) and three episodes, each of which includes, per protagonist:

- An initial internal state term;
- A goal;
- An attempt;
- An outcome;
- An internal state term as a reaction (the consequence).

These story structure components form the structural complexity of the story. Analysis of structural complexity provides information about the child's narrative development level and allows comparison across languages (Gagarina *et al.*, 2015). Moreover, by including internal state terms in

the coding and scoring, LITMUS-MAIN provides information about the cognitive-linguistic interface.

*Low complexity* implies the elicitation of two elements, namely, one attempt and one outcome; *medium complexity* implies two elements, namely, one goal and one attempt or one outcome; *high complexity* implies the production of a complete episode, with one goal, attempt, and outcome. Internal state terms assessment consider the mention of state verbs, including perceptual verbs (*e.g., see, look, notice),* physiological state terms (*e.g., thirsty, hungry*), emotion terms (*e.g., sad, happy, angry, worried*), experiential expressions (*e.g., surprised, startled*), belief-mental verbs (*e.g., think, know, decide*), and linguistic verbs (*e.g., say, shout*).

Macrostructure features and internal state terms are also examined in comprehension. LITMUS-MAIN comprehends nine open-ended questions for each story, three of which tap the protagonist's goals, whereas the others target the internal state terms (*e.g., "How does the dog feel?"/"How do you think the cat feels?"*), and their rationale *(e.g., "Why do you think that the dog is feeling...?").*

A set of background questions completes the assessment to evaluate the acquisition conditions and the quality and quantity of the input to the child in both of her/his languages.

Further details on administering, scoring, and coding the task will be given in the material section.

**Aim and hypothesis**

The present study is cross-sectional research that aimed to explore an Italian-English immersion Program's outcomes at the narrative level. As said, the narrative competence assessment is considered a more ecologically way to understand a child's communicative competence (Botting, 2002), especially in a bilingual context. Thus, this study completes the first study of the present thesis on the Italian-English immersion program, providing more data about bilingual children's language use in a relatively natural mode.

Firstly, we verified whether bilingual children showed similar narrative macrostructure and comprehension skills in Italian (L1) compared to their Italian monolingual peers in grade 1, 3, and 5. At these higher-order measures, we did not expect to find differences between the two matched TD groups: bilinguals and monolinguals should behave similarly in producing and comprehend the story grammar elements (Fiestas *et al.*, 2004; Uccelli *et al.*, 2007). Moreover, we expected to find improvements across the three grades at least in some Italian macrostructure parameters in both bilingual and Italian monolingual group (Stein *et al*., 1979; Squires *et al*., 2014).

Unfortunately, we could not compare the same scores in English with an English monolingual control group. We only observed the development across grades of story production and comprehension in English (L2): as long as bilingual students become literate, they should produce more complex and coherent stories, with more mentions to protagonists' goals, intentions, and emotions. Thus, we expected to find improvements across the three grades.

The third aim was to verify whether bilingual children in the immersion program performed similarly to the Italian monolingual group at the microstructure level in Italian. To do so, we compared Italian-English TD bilingual students to Italian monolingual on narrative microstructure scores. As Italian is the dominant-majority language of the bilingual sample, and we already discussed the no-disadvantage effect of bilingual education on the L1 (see Chapter 1 of the present thesis), we expected to find no significant differences between monolingual and bilingual across the Italian microstructure measures. Moreover, we could expect improved performance across the three grades, hand in hand with language development.

To observe the bilinguals' development of microstructure in English, we analyzed the microstructure parameters performance across Grade 1, 3, and 5. Even in this case, we expected to find children improving by grade. In particular, as it could be necessary up to three years to converge on monolingual norms for EELs (Paradis, 2016), we expected to find an improvement for the majority of the measures, especially from grade 1 to grade 3. Considering the morphology, we contemplated to find more errors at the bound morphology (*e.g.*, the omission of *–s* third singular

person, of *–ed* past tense, hyper-regularization of irregular single past forms, use of auxiliaries as tense markers), compared to free morphology, especially in the first- and third-graders (because at initial stages of acquisition of English as L2). Nevertheless, as Italian is a richly inflected language and could influence the L2 development, we should find gradual improvement across grades (Blom *et al.*, 2012), and we could expect to find no difference between bound and free morphology errors by grade 5.

Finally, we analyzed the relationship between the Italian and English performance at macrostructure and microstructure within the bilingual sample. Following Gagarina *et al.* (2012), who found evidence of a transfer of competencies for the macrostructure representation between L1 and L2, we expected to find correlations between macrostructure measures, which means, bilingual children performing similarly across Italian and English. Regarding microstructure, we could expect some correlations among some measures related to the lexical level, but not at the morphosyntactic level. As said, we could expect more morphology errors in English L2 than in Italian L1, thus no relation between the measures. Moreover, utterances might be longer in Italian than in English, partially because English is a second language that children are still learning and partially because of the intrinsic features of synthesis in the English syntax. For the lexical level, we could expect a transfer of knowledge from L1 to L2, where more talkative children in L1 are more talkative in L2, too (Cummins, 1991).

In sum, our expectations for the specific research questions addressed were the following:

6) We did not expect bilingual and monolingual children to show any differences in Italian macrostructural and comprehension indexes, and we expected to find improvements across the three grades.
7) We expected to find improvements across the three grades in English macrostructure and comprehension measures;

8) We did not expect to find differences between bilingual and monolingual children in Italian microstructural indexes of linguistic production and we expected to find improvements across the three grades;
9) We expected to find improvements across the three grades in English microstructure measures, especially from grade 1 to grade 3 at the bound morphology. Moreover we expected to find no difference between bound and free morphology errors by grade 5;
10) We expected to find correlations between Italian and English macrostructure measures and some correlation at the microstructure level: in the lexical skills used to produce a narrative but not in the morphosyntax.

## Material and methods

### Participants

Participants were the same students recruited for the present thesis's first study (see chapter 1 on Italian-English immersion program). With respect to the other study, we missed some of the children due to a few narrative recordings' damage. Thus, the final sample of the present study comprised a total of 163 children. The bilingual group consisted of 98 children, attending Grade 1 ($N$ = 39), Grade 3 ($N$ = 31) and Grade 5 ($N$ = 24), whereas the Italian monolingual group included 69 children attending Grade 1 ($N$ = 21), Grade 3 ($N$ = 32) and Grade 5 (N = 16).

To recap, a group of early bilingual pupils (attending a simultaneous Italian-English immersion program) and a group of Italian monolingual peers were included in the study. All children were typically developed and matched on SES (estimated as medium-high). The sample was fairly balanced for sex, with 89 participants (54.3% of the sample) being female and 75 being male (45.7%).

*Table 1* shows details regarding the two groups of bilinguals and monolinguals.

| **Grades** | **Bilinguals** ($N = 121$) | **Monolinguals** ($N = 71$) |
|---|---|---|
| **Grade 1** | $N = 43$ | $N = 21$ |
| Mean age in months | $M = 82.6$ ($SD = 4.1$) | $M = 87.3$ ($SD = 4.1$) |
| Females | $N = 18$ | $N = 12$ |
| Males | $N = 21$ | $N = 9$ |
| **Grade 3** | $N = 31$ | $N = 32$ |
| Mean age in months | $M = 104.4$ ($SD = 3.9$) | $M = 109$ ($SD = 4.7$) |
| Females | $N = 18$ | $N = 18$ |
| Males | $N = 13$ | $N = 14$ |
| **Grade 5** | $N = 24$ | $N = 16$ |
| Mean age in months | $M = 129.3$ ($SD = 5.2$) | $M = 128.19$ ($SD = 4.12$) |
| Females | $N = 15$ | $N = 8$ |
| Males | $N = 19$ | $N = 8$ |

**Table 1** shows the number of subjects, the average ages in months, and the number of females and males of the following subgroups: bilingual first-graders, bilingual third-graders, bilingual fifth-graders, monolingual first-graders, monolingual third-graders, and monolingual fifth-graders.

As evident, the bilingual and monolingual subgroups in Grades 1 and 3 differed significantly according to their age. Monolingual first-graders were significantly older ($t(58) = 4.20$, $p < 0.001$) than bilingual first-graders and monolingual third-graders were significantly older ($t(61) = 4.17$, $p < 0.001$) than their bilingual peers. As we explained in the first study of the present thesis, and we will briefly resume in the following section, they were tested in different periods, which explains the significant age differences. It will be taken into account in the statistical analysis.

## Materials and procedures

Ethical approval was obtained by the Ethics Committee of the University of Milano – Bicocca (n. 280).

Bilinguals and (partially) monolinguals were tested in different periods: bilinguals were tested between April and June (during the final part of the school year); monolingual first- and third-graders were tested in September (at the beginning of the following school year), having just begun their second and their fourth year of elementary school, respectively.

Before the beginning of the narrative task administration, all parents gave informed consent. Children were tested individually, monolinguals only in Italian, bilinguals in Italian and English, in

a quiet room in their school during school hours. For bilinguals, the tasks were presented in two separate sessions on two different days, and they were first tested in Italian and then in English.

Children's narrative abilities were examined through story retelling tasks, with wordless picture stories from the MAIN (Gagarina *et al*., 2012). We selected the story "Cane" (*Dog*) for Italian and the story "Cat" for English (the latter only for bilingual children). Each participant was presented with an elicitation picture set (one in Italian and one in English for bilinguals, which differentiates for contents: setting, characters, actions, but share the same story structure). The stories were presented on a computer screen with audio input via headphones. On the screen, the child first had to fictitiously choose a story from one of three colored envelopes (which enclose the same story), thus motivating the child in the task as much as possible. The pictures representing the story were presented throughout a PowerPoint® presentation, initially showing the entire story with all the six pictures, then in pairs of two. Pictures and stories were presented with identical timing for exposure duration and transitions for all languages. The child was then asked to retell the story to the experimenter at his best, while pictures were shown again on the screen; the retelling task was followed by nine open-ended comprehension questions on the story content.

An Italian mother-tongue and a proficient English person gave instructions for the tasks – as many times as necessary – in Italian and English, respectively. Stimulus scripts that children had to listen for retelling were recorded by native speakers and presented to the participants as audio recordings through headphones. Scripts are exhaustive and include an integral macrostructure, microstructure, and internal state features of narrative discourse. Comprehension questions were read aloud by the experimenter, and children had to answer orally.

During task administration, the child's narrative production and comprehension answers were recorded using an audio recorder tool. After a session has ended, the children's narratives were transcribed by the experimenter. Transcription was carried out manually using CHILDES/CHAT (MacWhinney, 2000) system. Two autonomous experimenters checked all the

transcriptions. If some uncertainties in the transcription emerged, it was discussed, agreed upon, and resolved by the two experimenters, thus assuring an adequate transcription reliability level.

### *Coding and scoring*

The children's narratives and comprehension answers were coded manually using the MAIN scoring sheet (Gagarina *et al*., 2012). Narrative productions were coded both for macrostructure and microstructure parameters, the first based on story grammar, the second referring to linguistic features used in constructing a coherent discourse in a specific language (Gagarina, Klop, Tsimpli & Walters, 2016). For macrostructure, the MAIN scoring rubric (Gagarina *et al*., 2012) was adopted; for microstructure, we created a coding rubric, and we calculated the scores in the CHAT file using the CLAN software (MacWhinney, 2000). For the analysis, we excluded utterances that were not fully intelligible or complete. Entire sentences in the non-target language (*code-switching*) were included in the macrostructure analyses but excluded from the microstructure analyses. Single words within a sentence in the non-target language (*code-mixing*) were considered in both analyses. The two experimenters shared coding agreement on nearly 30% of randomly selected narratives, both for Italian and English (thus reaching approximately 60% of agreement on coding).

Scores and coding are presented below in detail.

### *Macrostructure scores*

General *macrostructure elements score* was given by the total of macrostructure elements produced by the child. For each element – the setting (time and place), the character's initial mental state, the goal (G), the attempt (A), the outcome (O), and the character's mental state as a consequence – the child obtained 1 point. Different but compatible formulations for the same macrostructure components are accepted. The maximum score for macrostructure components production is fixed at 17 points for each story. Two macrostructure elements scores were obtained for each child: one for Italian and one for English. Alongside the total macrostructure score, we calculated how often a

child produced single macrostructure elements in the story (that is, the total number of single G, A, and O).

Besides these general and single quantitative measures of story production, according to the MAIN scoring rubric (Gagarina et al., 2012), we also obtained two macrostructure complexity scores: the mental state score and the complexity score. The *mental state score* separately accounts for the total number of mental state terms produced by the child in the narrative, resulting in a maximum score of 6 for each story set (each story script provides two mental states for each episode). Two mental state scores were obtained for each child: one for Italian and one for English. The *complexity score* of story production counts the relations among the macrostructure elements of goal, attempt, and outcome, for each episode produced. Thus, it includes a measure of how often a child produces partial (AO), incomplete (GA/GO), and complete (GAO) event sequences. The child is assigned 1 point for each mention of at least two elements within each episode of the story, namely, one attempt (A) and one outcome (O) for *low complexity*, and one goal (G) and one attempt (A) or one goal (G) and one outcome (O) for *medium complexity* (these story element combinations are considered more complex because the children mention the characters' goals). Thus, the score can range with different complexity scores depending on the different macrostructure elements combination the child produced (to make an example, the child could produce 1 GO and 2 AO). One point of *high complexity* was assigned when the child produced a complete episode, with the sequence of the three relevant elements: goal, attempt, and outcome (GAO). Only 1 point of complexity was assigned for the episode (no other low or medium complexity combinations were considered). Thus, it was coded only the highest level of complexity the child reached for each episode. Complexity scores (low, medium, and high) could range between 0 and 3. Two macrostructure complexity scores were calculated, one for each of the two languages (Italian and English).

***Comprehension score***

The comprehension score is calculated as the total of correct answers to the comprehension questions administered after the retelling task. Nine open-ended questions investigate the comprehension of mental states and goals of the story protagonists. The MAIN scoring sheet provides orienteering correct answers that the child could give for each question in order to assign 1 point. Different but compatible answers are accepted as correct. The maximum comprehension score is 9. Two comprehension scores were obtained, one for Italian and one for English.

***Microstructure scores***

To assess microstructure complexity in children's narrative productions, we arranged a scoring rubric, which targeted lexical and morphosyntactic levels.

For *the lexical level*, we scored the following set of parameters:

- Token: total number of correct words pronounced by the child. It comprises words phonetically inaccurate yet excludes phonologic paraphasia (phoneme or syllable substitution/omission/inversion/insertion within a target word), hesitations, reformulations, and unintelligible words;
- Type: number of different words within the story production. It considers all repeated words (even inflected forms of the same verbs, nouns, or adjectives) only once. It is calculated by subtracting all repeated words from the number of tokens;
- Type/Token Ratio (TTR): it is calculated by the ratio between the number of non-repeated words (Type) and the number of correct words (Token). It is considered an index of lexical diversity as it mirrors the range of vocabulary for the narrative production, independently of speech sample size (Richards, 1987);

To analyze the *morphosyntactic level*, we scored the following set of indexes:

- Mean Length of Utterances (MLU): it is calculated by dividing the total number of tokens by the number of sentences. To identify the limits of each sentence in the

speech (*i.e.*, when a sentence closes), hierarchical criteria were followed, as suggested by Marini, Andreetta, del Tin, and Carlomagno (2011). When the acoustic criterion (a pause/break between two sentences) was not apparent, semantic (the sentence enclose an independent semantic unit), grammatical (the sentence is grammatically complete), or phonological criterion (the sentence is delimited by a false start) steered the sentences segmentation. This logic consents to identify utterances with good reliability, even in speech with different characteristics. MLU is indicative of the average length of sentences that the child produces;

- Proportion of errors and omission in free and bound morphemes on tokens: errors of commission and omission in free and bound morphemes were counted. The proportion is calculated over tokens (total number of words), then transformed into a percentage. The lower the values of this ratio correspond to a smaller number of errors and omissions of the free and bound morphology and *vice versa*. The rubric coded free morphemes errors or omissions related to articles, prepositions, pronouns (for English), and a general category for other kinds of errors (e.g., adverbs or conjunctions). For bound morphology, we considered errors of commission and omission borne by verbs (*e.g.*, the omission of *–s* third singular person, of *–ed* past tense, hyper-regularization of irregular single past forms, use of auxiliaries as tense markers), nouns, adjectives (*e.g.*, omission or erroneous insertion of *–s* plural) and a general category for other kinds of errors. We are aware that this way of calculating proportions is not the most accurate and the use of obligatory contexts of use of a given morpheme would have been more appropriate. However, for reason of time, we decided to have a rough vision of the errors at this moment.

**Statistical analyses**

To compare the Italian narratives of the bilingual and the Italian monolingual groups in Grade 1, 3, and 5, an analysis of variance was performed, with *Group* (bilingual and monolingual) and *Grade* (first, third and fifth) as independent variables. Macrostructure and microstructure scores were fixed as dependent variables, and testing time as a covariate (only in the comparison between monolingual and bilingual children), thus controlling for different ages of children. Significant main effects and interactions will be reported. Bonferroni *post hocs* were used to make post hoc comparisons.

In order to analyze bilingual's English development across grades, an ANOVA was performed, with *Grade* (first, third, and fifth) and macrostructure and microstructure scores fixed as factors.

## Results

### *Comparison between groups on the Italian macrostructure and comprehension*

To answer the first research question, we examined whether Italian-English TD bilingual school-aged students performed similarly to their monolingual peers on narrative macrostructure and comprehension measures. We found very few effects that we are going to present below.

*Figure 1* represents children's performances at *total macrostructure score*. As it is evident, there is an improvement across grades. Indeed, only the effect of Grade $F(2,156) = 3.8$, $\eta^2 = 0.04$, $p < 0.05$[8] was significant, with Grade 1 differing from Grade 3 and 5 (Bonferroni *post hoc* $p < 0.05$).

[8] In our analyses, we report the Eta-squared ($\eta^2$), which is a measure of effect size for use in ANOVA. It ranges between 0 and 1. Effect sizes follow this interpretation: .01 ~ small; .06 ~ medium; >.14 ~ large.

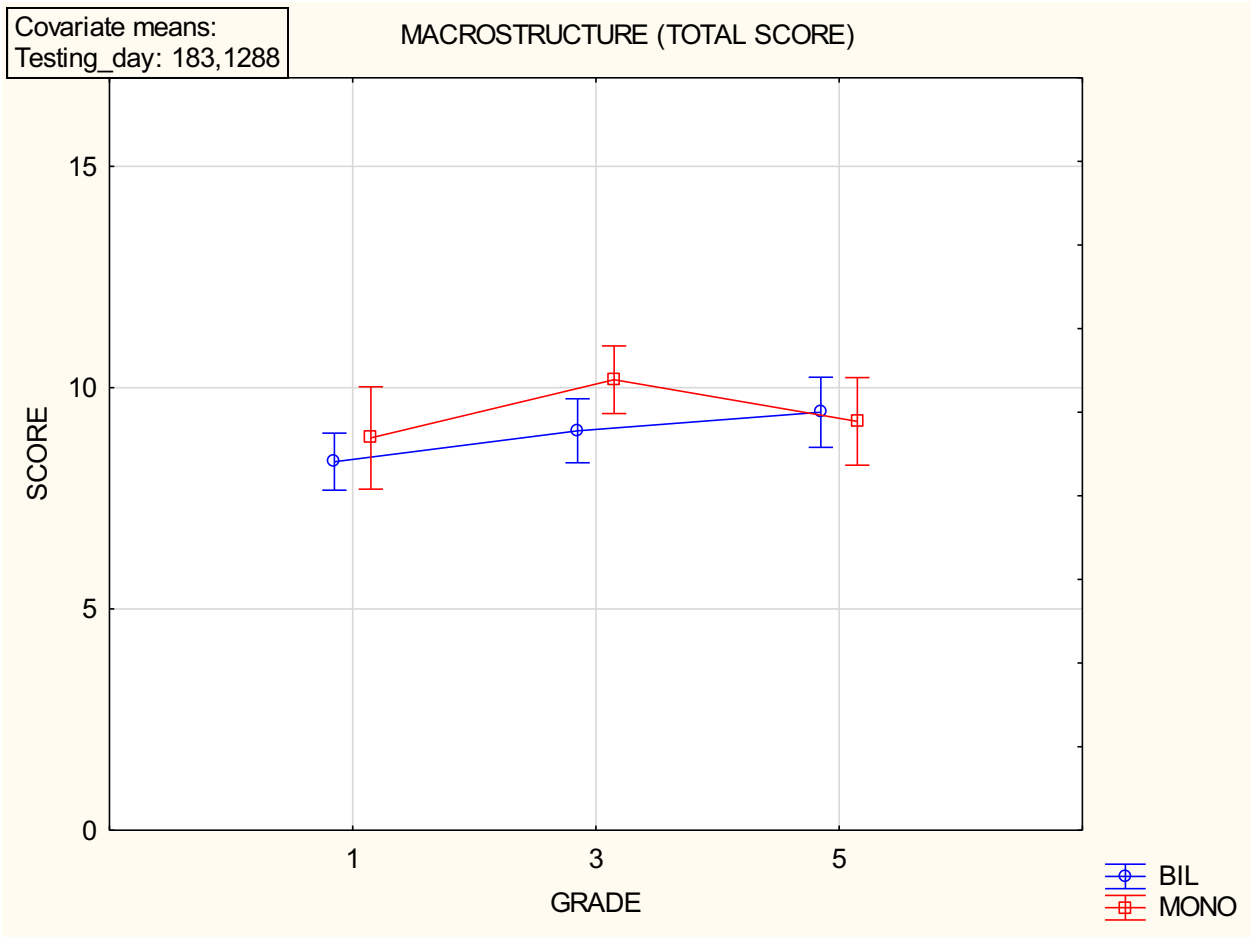



**Figure 1** shows the average *total macrostructure* scores (MAX=17) in Italian in Grade 1, 3, and 5. Bars indicate a confidence interval at 0.95.

Analyzing the frequencies of single macrostructure elements in the story (total number of single G, A, and O), we found effects only for the *Goal* parameter. *Figure 2* shows the number of goals produced by monolingual and bilingual children. As evident, monolingual third-graders were more productive in the protagonist's intentions.

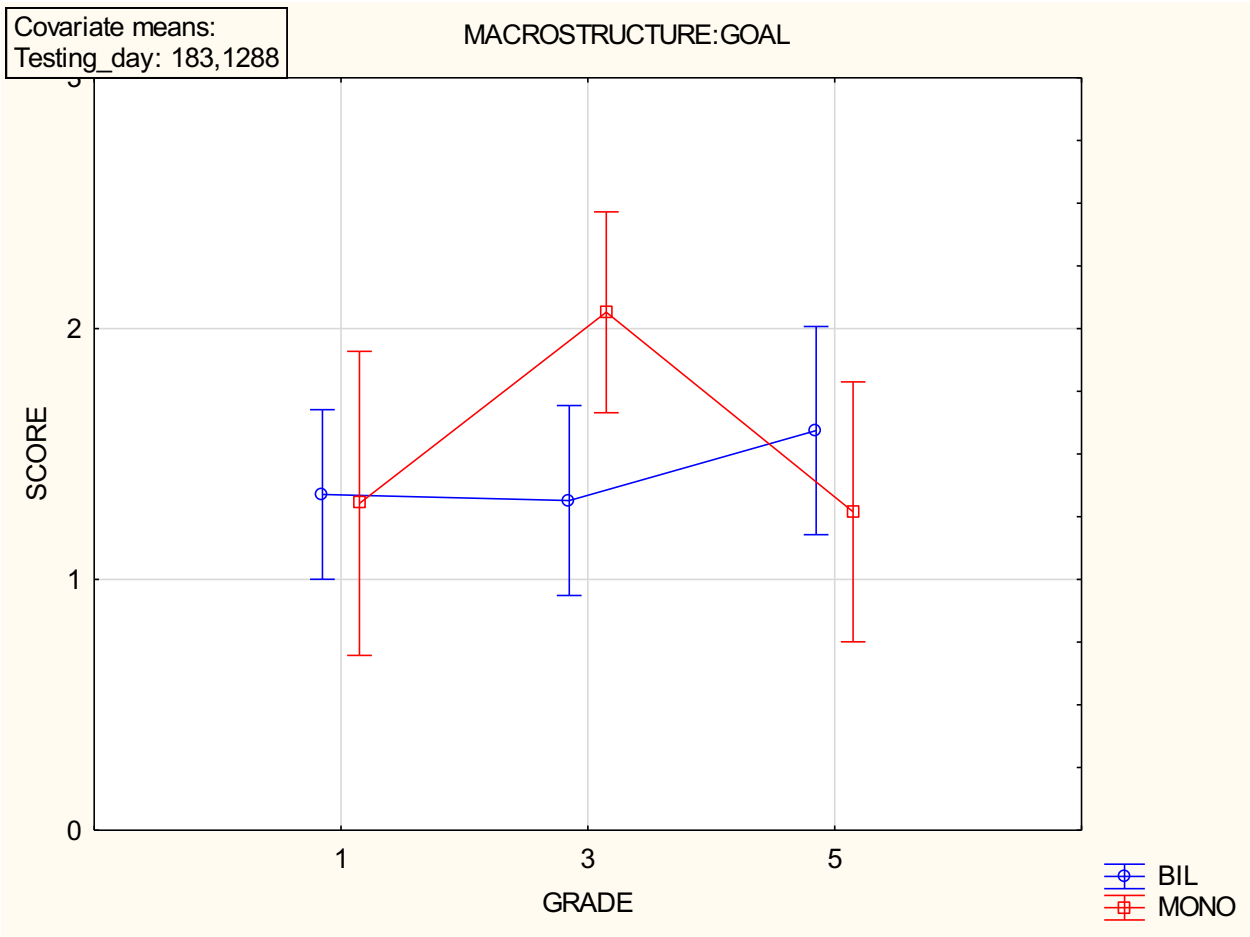



**Figure 2** shows average *Goal* scores (MAX=3) in Italian in Grade 1, 3, and 5 of monolingual and bilingual children. Bars indicate a confidence interval at 0.95.

Only the Interaction Group*Grade was found to be significant $F(2,156) = 4.06$, $\eta^2 = 0.04$, $p < 0.05$ because, in the monolingual group, first graders differed from third graders (Bonferroni *post hoc* $p < 0.005$).

Considering macrostructure complexity scores, we only found effects on the *medium complexity* score (GA/GO combination). *Figure 3* shows the number of GA or GO produced by monolingual and bilingual children. The same groups' performance as for Goal score is observable.

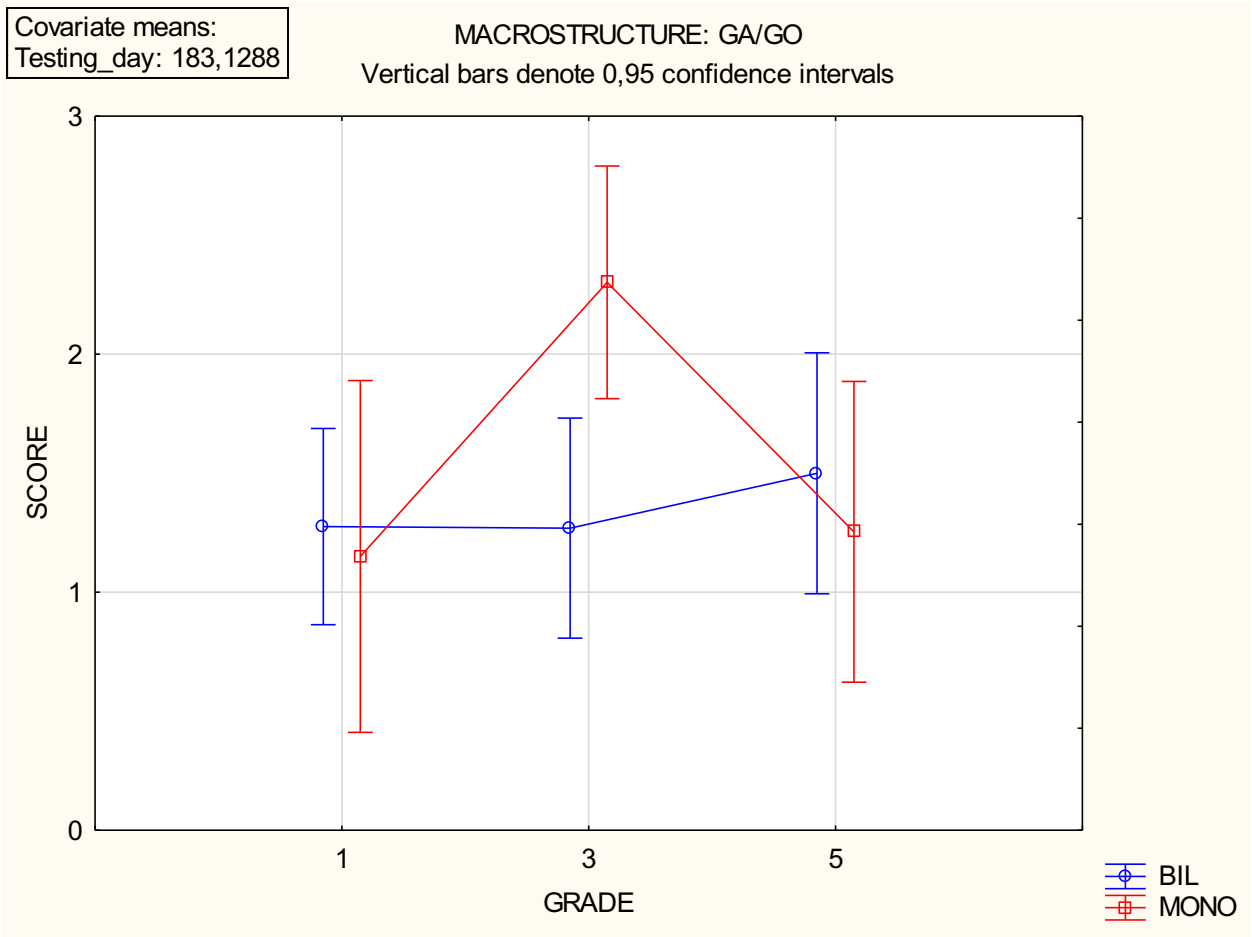



**Figure 3** shows average *GA/GO* scores (MAX=3) in Italian in Grade 1, 3, and 5 of monolingual and bilingual children. Bars indicate a confidence interval at 0.95.

We found an effect of Grade $F(2,156) = 3.3$, $\eta^2 = 0.04$, $p < 0.01$, and Interaction Group*Grade $F(2,156) = 4.7$, $\eta^2 = 0.05$, $p < 0.05$, because Grade 1 differed from Grade 3 in the monolingual group (Bonferroni *post hoc* $p < 0.001$). Thus, the effect of grade is due to monolingual children.

*Figure 4* illustrates the use of mental terms, showing an improvement across grades. Statistical analyses confirms the effect of Grade, $F(2,156) = 3.14$, $\eta^2 = 0.03$, $p < 0.05$. Bonferroni *post hoc* revealed that first grade is different from third and fifth grades ($p < 0.05$).

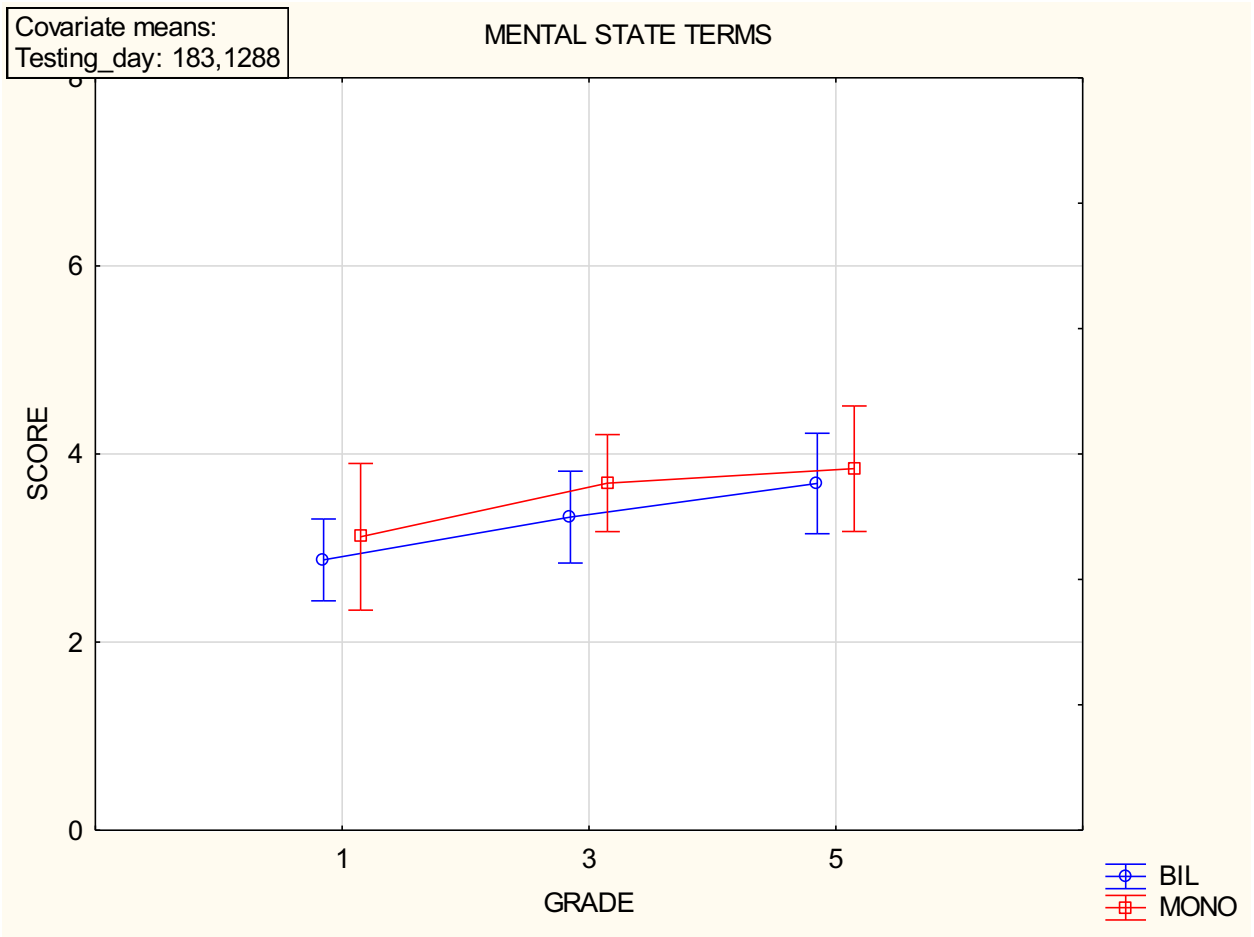



**Figure 4** shows average mental state scores (MAX=6) in Italian in Grade 1, 3, and 5 of monolingual and bilingual children. Bars indicate a confidence interval at 0.95.

In conclusion, at the level of macrostructure, the two groups (mono and bilingual children) do not differ. An effect of grades is typically found with an improvement from grade 1 to 3 on total macrostructure and use of mental state. Some effects for single macrostructure elements and complexity were found only in the monolingual group.

***English macrostructure and comprehension in the bilingual group***

To observe the development of story production and comprehension in English (L2), we examined the bilingual group's English macrostructure and narrative comprehension performance across Grade 1, 3, and 5. Here we found more effects than in Italian among the various macrostructure scores that we are going to present below. No effect on comprehension was found.

As evident from *Figure 5,* children improved sharply across Grades at *total macrostructure score* in English. We found, in fact, effect of Grade $F(2,91) = 10.8$, $\eta^2 = 0.19$, $p < 0.001$, with Grade 1 differing from Grades 3 and 5 (Bonferroni *post hoc* $p < 0.001$).

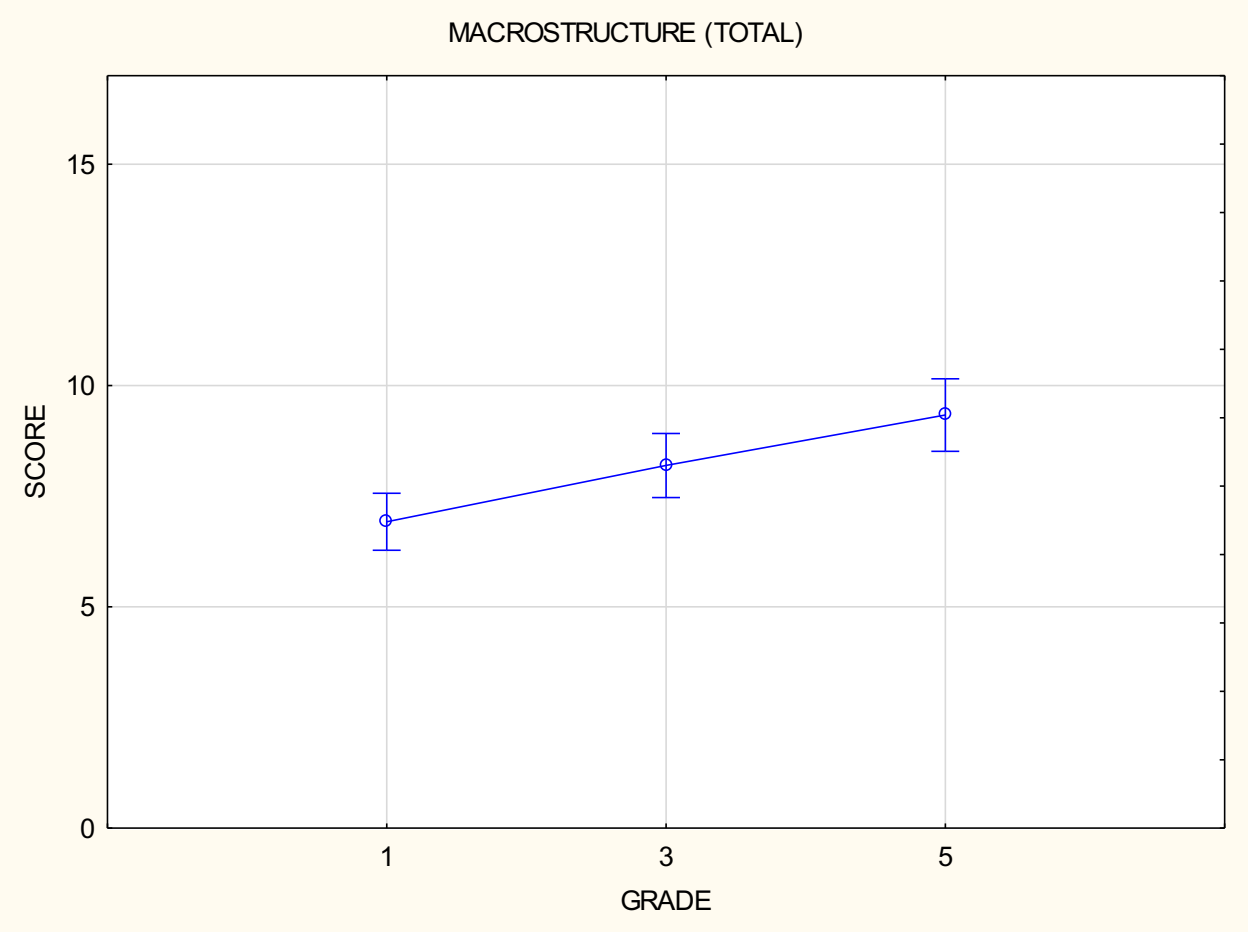



**Figure 5** shows the average *total macrostructure scores* in English of the bilingual children in Grade 1, 3, and 5. Bars indicate a confidence interval at 0.95.

An improvement across Grades was also present at the *Attempt* score, as evident in *Figure 6*. The analyses confirmed the effect of Grade $F(2,91) = 10.8$, $\eta^2 = 0.19$, $p < 0.001$, with Grade 1 differing from Grades 3 and 5 (Bonferroni *post hoc* $p < 0.05$).

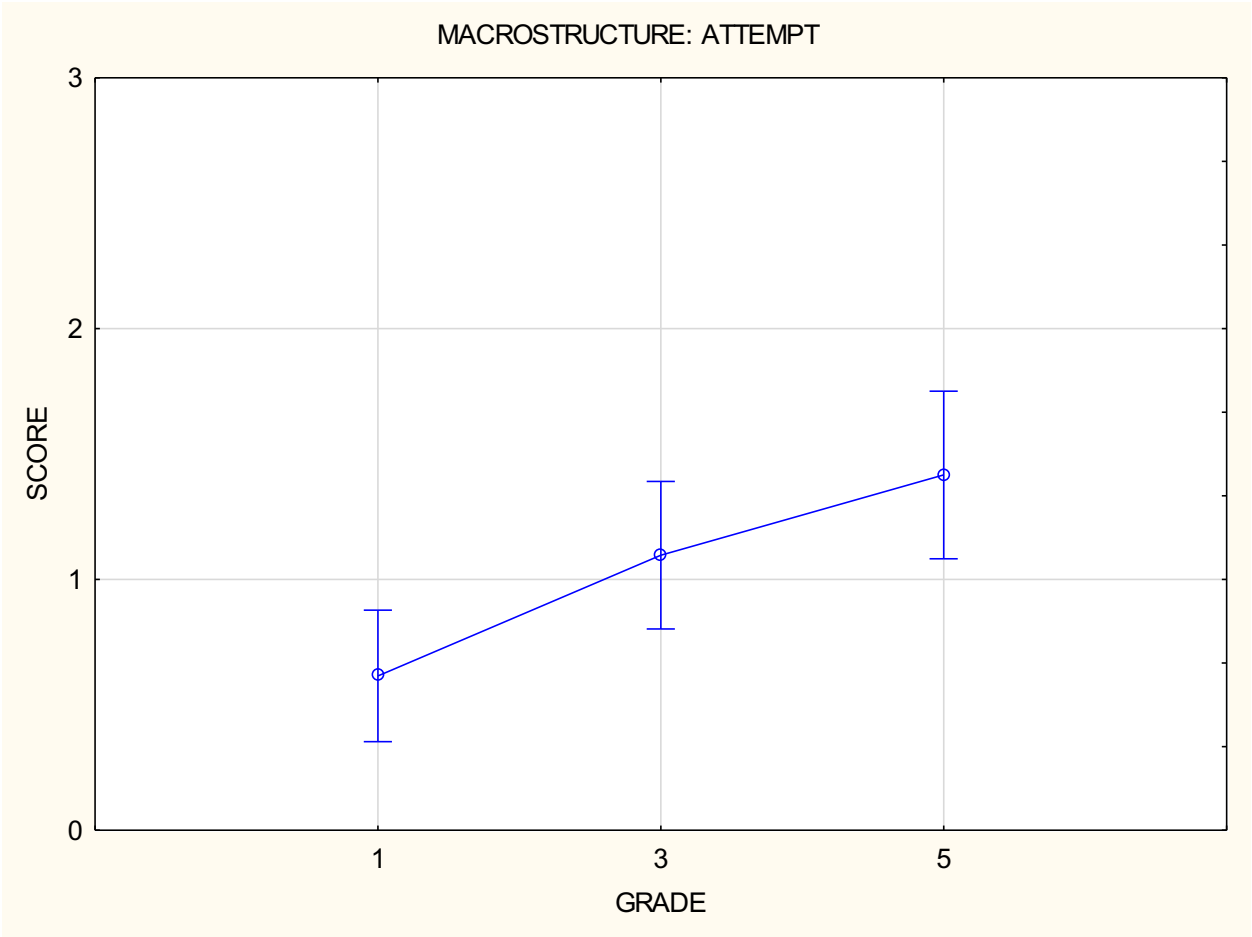



**Figure 6** shows average *Attempt* scores in English of the bilingual children in Grade 1, 3, and 5. Bars indicate a confidence interval at 0.95.

An improvement across Grades was also observed on some English macrostructure complexity scores illustrated in *Figures 7, 8*, and *9*, which represent the *low complexity* scores (AO), the *high complexity* scores (GAO), and the *mental state* scores, respectively. For AO, the analyses confirmed the effect of Grade $F(2,91) = 10.4$, $\eta^2 = 0.18$, $p < 0.001$, with Grade 1 differing from Grades 3 and 5 (Bonferroni *post hoc* $p < 0.01$).

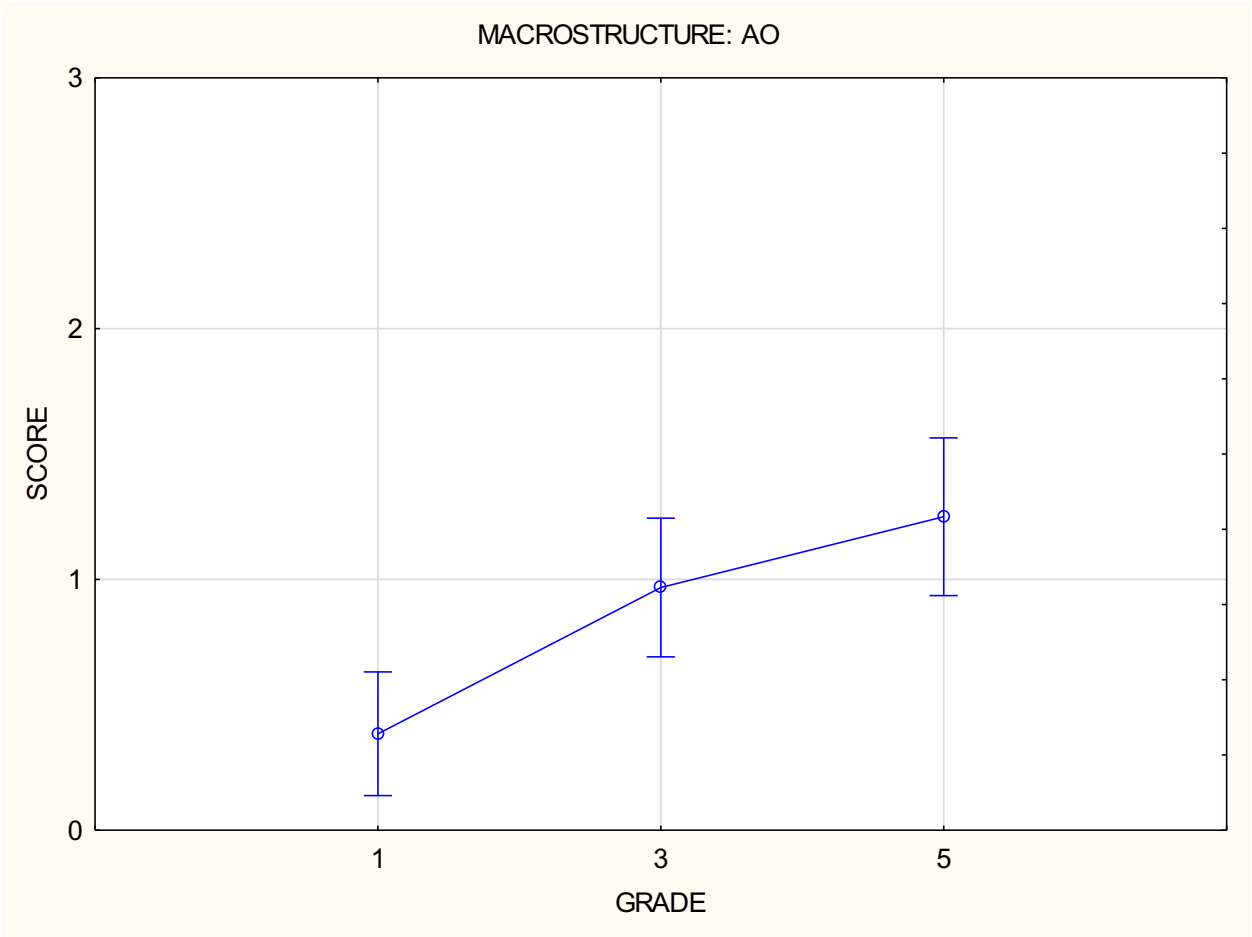



**Figure 7** shows average *AO* scores in English of the bilingual children in Grade 1, 3, and 5. Bars indicate a confidence interval at 0.95.

For GAO, the analyses confirmed the effect of Grade $F(2,91) = 4.45$, $\eta^2 = 0.08$, $p < 0.05$, with Grade 1 differing from Grade 5 (Bonferroni *post hoc* $p < 0.05$).

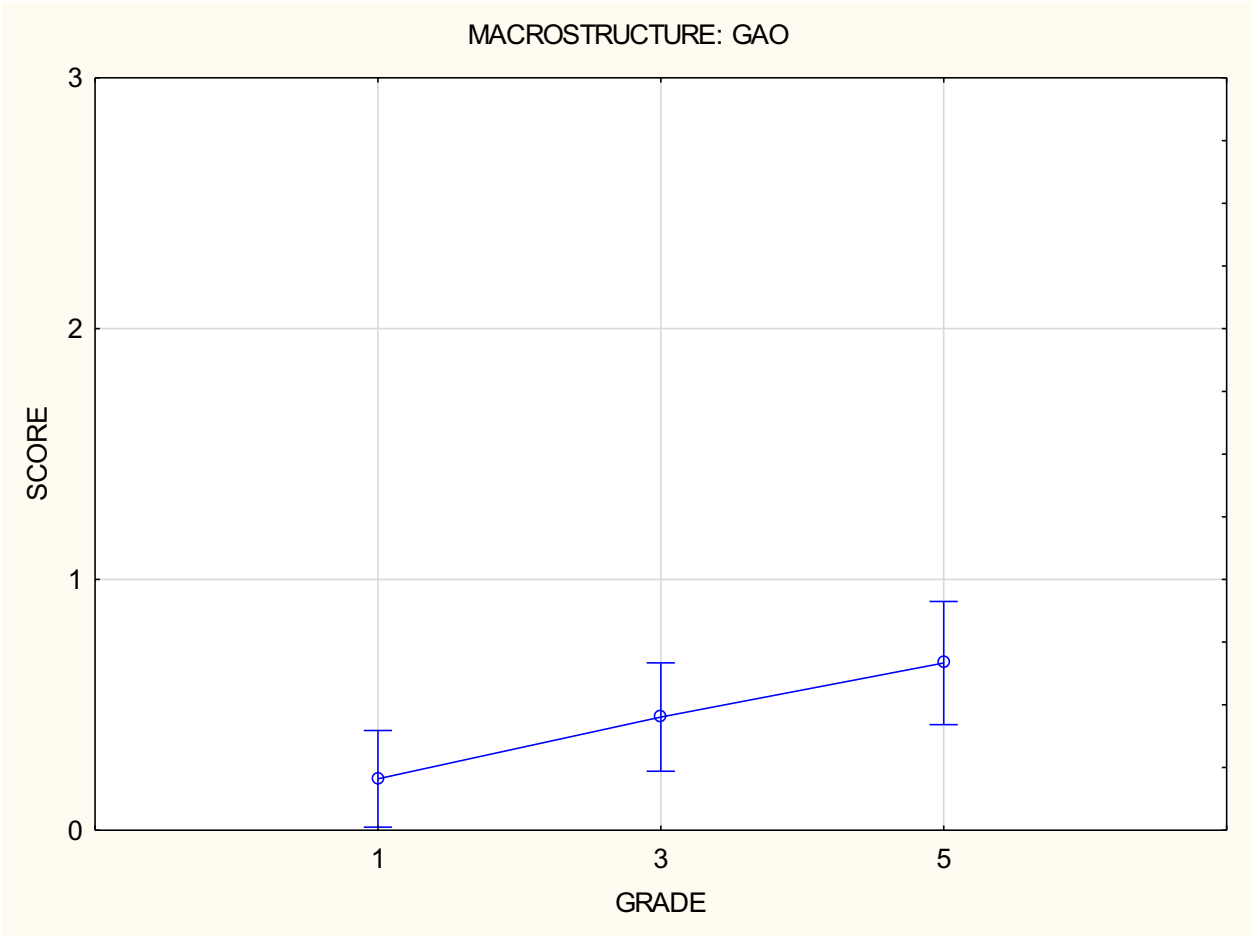



**Figure 8** shows average *GAO* scores in English of the bilingual children in Grade 1, 3, and 5. Bars indicate a confidence interval at 0.95.

For the use of mental state terms, the analyses confirmed the effect of Grade $F(2,91) = 8.04$, $\eta^2 = 0.15$, $p < 0.001$, with Grade 1 differing from Grade 5 (Bonferroni *post hoc* $p < 0.001$).

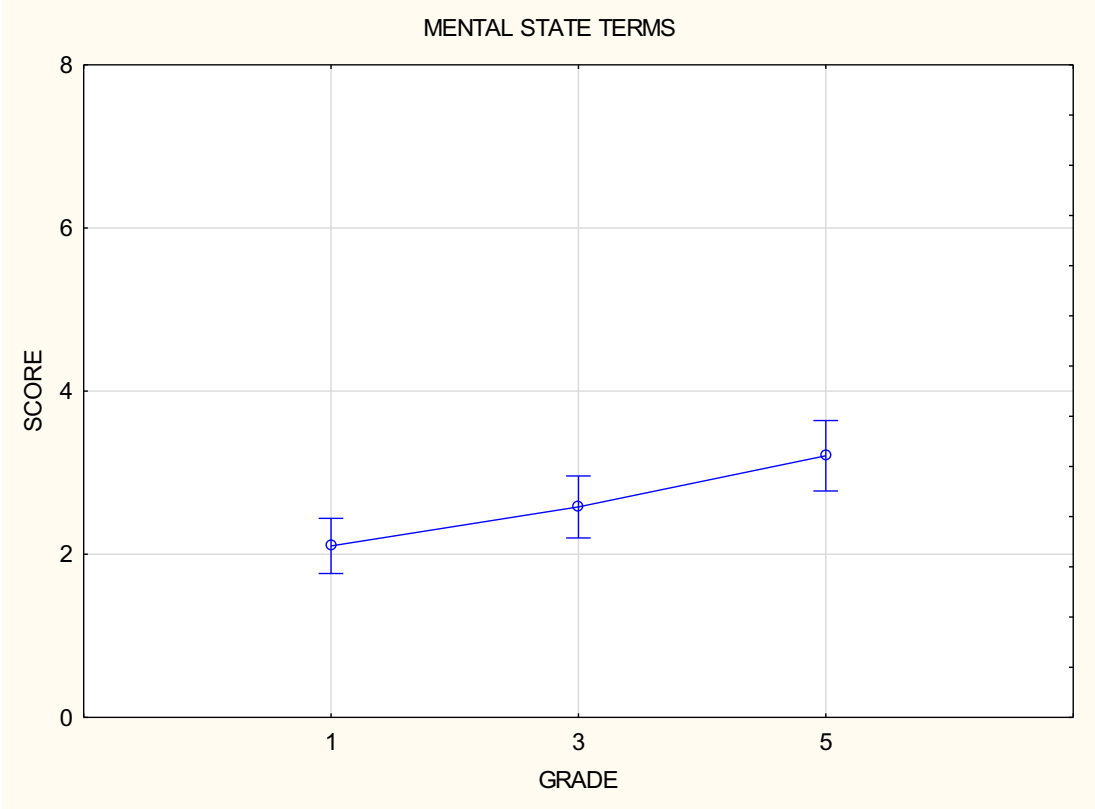



**Figure 9** shows average *Mental State* scores in English of the bilingual children in Grade 1, 3, and 5. Bars indicate a confidence interval at 0.95.

To conclude, a general improvement across grades in the measures of macrostructure was noticeable not only in Italian, as seen before, but also in English in the bilingual children.

***Comparison between monolingual and bilingual on the Italian microstructure***

Considering microstructure, we compared Italian monolingual to Italian-English TD bilingual students on narrative microstructure scores selected for the present study: lexical parameters (token, type, and type/token); morphosyntactic parameters, namely MLU and percentage of errors in the use of free (articles, prepositions, pronouns, and other categories) and bound morphology (verbs, nouns, and adjectives and other categories' suffixes). *Tables 2* and *3* report the means and standard deviations of the various microstructure parameters for bilingual and monolingual children, respectively, in the three grades.

| Measure | Grade 1 Mean (SD) | Grade 3<br>Mean (SD) | Grade 5<br>Mean (SD) |
|---|---|---|---|
| % MorphErr | 0.02 (0.01) | 0.01 (0.01) | 0.004 (0.005) |
| Type | 52.2 (9.2) | 60.1 (8.4) | 59.7 (9.1) |
| Token | 89.6 (21.5) | 111.8 (21.5) | 110.1 (21.9) |
| Type/token | 0.59 (0.07) | 0.54 (0.05) | 0.55 (0.5) |
| MLU | 10.6 (2.4) | 12.6 (2.7) | 13.6 (3) |

**Table 2** shows means and standard deviations of percentage of morphosyntax errors (%MorphErr), type of different items (Type), token used (Token), type/token ratio (Type/token) and Mean length of utterance (MLU) in Italian by monolingual children.

| Measure | Grade 1<br>Mean (SD) | Grade 3<br>Mean (SD) | Grade 5<br>Mean (SD) |
|---|---|---|---|
| %MorphErr | 0.02 (0.01) | 0.01 (0.01) | 0.009 (0.001) |
| Type | 49.1 (8.9) | 58.1 (8.6) | 60.2 (8.7) |
| Token | 88.4 (21.7) | 108.7 (18) | 111.2 (22) |
| Type/token | 0.57 (0.08) | 0.54 (0.06) | 0.55 (0.7) |
| MLU | 14 (11.7) | 13 (2.8) | 13.9 (2.6) |

**Table 3** shows means and standard deviations of percentage of morphosyntax errors (%Err), type of different items (Type), token used (Token), type/token ratio (Type/token) and Mean length of utterance (MLU) in Italian by bilingual children.

Here we report the lexical and morphosyntactic parameters in which we found effects.

At the lexical level, we found an effect of Grade both on *Token* (2,156) = 16, $\eta^2 = 0.17$, $p < 0.001$ and on *Type* F(2,156) = 15, $\eta^2 = 0.16$, $p < 0.001$. Bonferroni *post hoc* showed that first-graders differed from third- and fifth-graders ($p < 0.001$), and in both cases an improvement is observed in third and fifth grades. An effect of Grade was also found in the proportion *Type/Token* F(2,156) = 5, $\eta^2 = 0.06$, $p < 0.01$, with Grade 1 differing from Grade 3 (Bonferroni *post hoc* $p < 0.05$). Children in

Grade 1 used more different words in proportion to the total words produced than Grade 3 and 5. We will discuss this piece of evidence in the discussions.

Then, we analyzed the morphosyntactic level. First of all, we must notice that the percentage of commission and omission errors in morphosyntax is meager, attested around 2% in Grade 1 and decreasing further in grade 5 (proportion of errors has been calculated by dividing errors over total number of token). An effect of Grade in the percentage of errors $F(2,156) = 14.6$, $\eta^2 = 0.16$, $p < 0.001$. Bonferroni *post hoc* showed that first-graders differed from third- and fifth-graders ($p < 0.01$). An effect of testing day was also found $F(1,156) = 14$, $\eta^2 = 0.08$, $p < 0.001$. No effect was found for the MLU, in line with the observation that MLU is not a sensitive measure for being older than age 3 (Rondal, Ghiotto, Bredart, & Bachelet, 1987; Devescovi & Pizzuto, 1995).

***English microstructure development in the bilingual group***

We analyzed the bilinguals' English linguistic competence across the microstructure parameters scored on their narrations to observe the development across Grade 1, 3, and 5. *Table 4* reports the means and standard deviations of the various English microstructure parameters for bilingual children, respectively, in the three grades.

| Measure | Grade 1<br>Mean (SD) | Grade 3<br>Mean (SD) | Grade 5<br>Mean (SD) |
|---|---|---|---|
| **%Err free m.** | 0.04 (0.03) | 0.03 (0.02) | 0.015 (0.01) |
| **%Err bound m.** | 0.1 (0.04) | 0.07 (0.04) | 0.03 (0.02) |
| **Type** | 33.7 (8.4) | 44.9 (7.4) | 51.5 (8.2) |
| **Token** | 73.4 (22.5) | 98.8 (19.4) | 115.7 (24) |
| **Type/token** | 0.47 (0.08) | 0.46 (0.06) | 0.50 (0.06) |
| **MLU** | 9.7 (3.12) | 13.5 (2.7) | 15.9 (3.3) |

**Table 4** shows means and standard deviations of percentage of errors in free morphology (%Err free m.), percentage of errors in bound morphology (%Err bound m.), type of different items (Type), token used (Token), type/token ratio (Type/Token) and Mean length of utterance (MLU) in English by bilingual children.

Here we present the lexical and morphosyntactic indexes in which we found effects. Considering the lexical level, we only found an effect of Grade on *Token* $F(2,91) = 29.1$, $\eta^2 = 0.39$, $p < 0.001$ and on *Type* $F(2,91) = 39.2$, $\eta^2 = 0.46$, $p < 0.001$. Bonferroni *post hoc* confirmed that, in both cases, all three Grades differed from each other ($p < 0.05$), with an improvement across grades.

As regard to the morphosyntactic level, we found an effect of Grade in all of the measures. The effect was significant in *Mean Length of Utterances* (*MLU*) $F(2,89) = 32.9$, $\eta^2 = 0.42$, $p < 0.001$, with all three Grades differing from each other (Bonferroni *post hoc* $p < 0.05$). A repeated measure Anova with the type of error (free and bound morphemes) as within subject factor and grade as a between subject factor revealed a significant effect of grade $F(2,91) = 27.9$, $\eta^2 = 0.38$, $p < 0.001$, an effect of type of error $F(2,91) = 103.4$, $\eta^2 = 0.53$, $p < 0.001$ and an interaction of grade by type of error $F(2,91) = 10.1$, $\eta^2 = 0.18$, $p < 0.001$. Bonferroni *post hocs* showed that the proportions of bound and free morphemes errors differ in first and third grade ($p < 0.001$), but not in fifth grade. In fact, a higher proportion of bound morphemes errors was found. In addition, the three grades differ from each other in terms of proportions of errors, with fewer errors from first, to third and fifth grade, as evident in *Figure 10*.

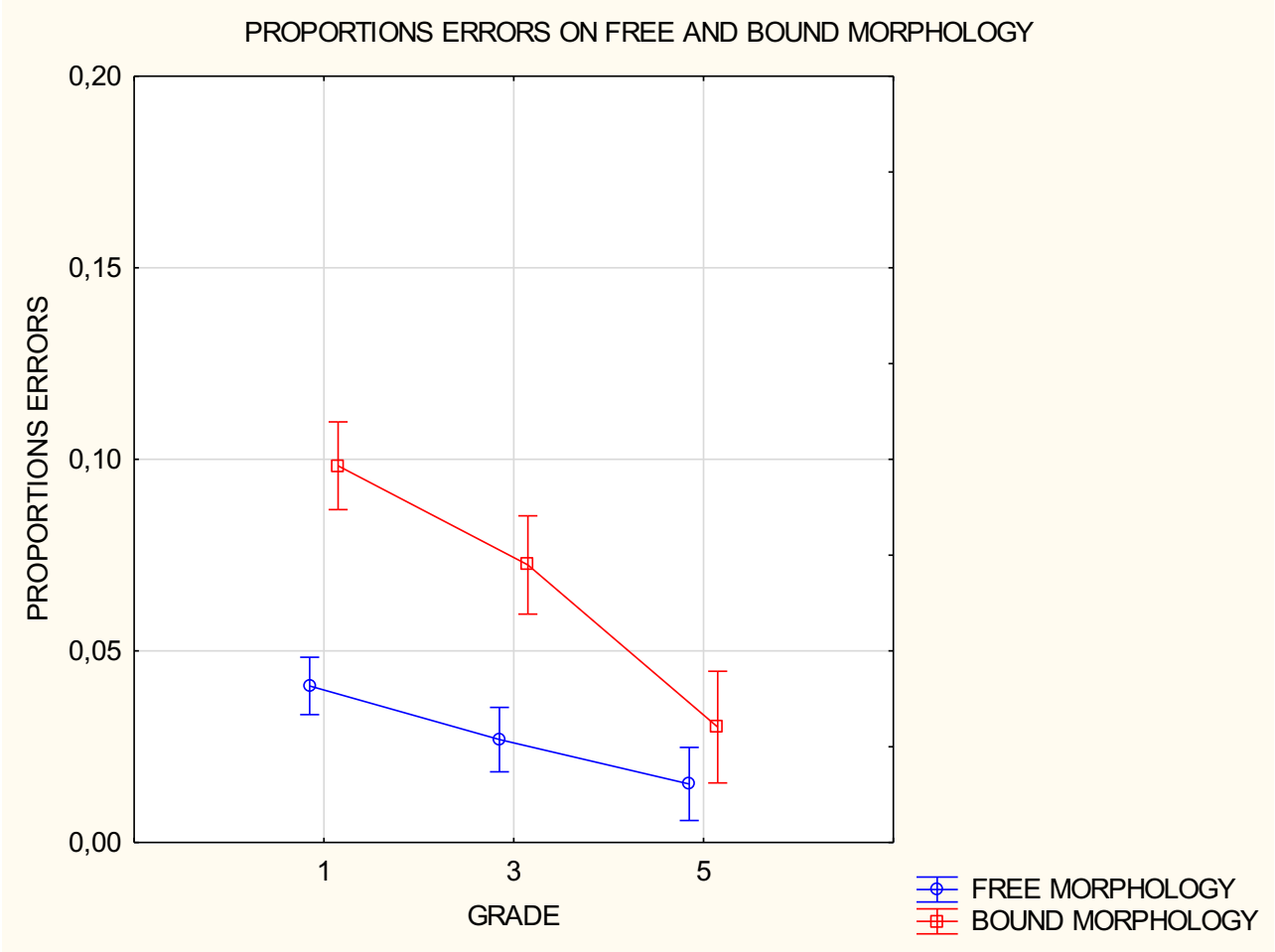



**Figure 10** shows bilinguals' average percentage scores of *errors and omissions in free morphemes/Tokens* (blue line and *bound morphemes/Tokens* (red line) in English in Grade 1, 3, and 5. Bars indicate a confidence interval at 0.95.

## Correlations between Italian and English macrostructure and microstructure in the bilingual group

To verify whether the bilinguals' performance in Italian was related to English, we carried out two correlation analyses, one considering macrostructure measures and the other microstructure measures. Thus, performances on each measure were correlated in the two languages.

Total macrostructure scores in Italian and English correlated $r = 0.44$, $p < 0.05$: those who produced a more rich-and-informative story in Italian also did in English; mental state scores also correlated $r = 0.49$, $p < 0.05$, thus reflecting a linguistic complexity skill in both the children's languages. Both the total macrostructure and the mental state scores in Italian correlated with the English comprehension $r = 0.27$ and $r = 0.22$, respectively ($p < 0.05$): a good ability in story

production in the first language is mirrored in the ability to comprehend a story macrostructure in a second language.

As far as microstructure measures are concerned, a significant positive correlation was found between *Token* $r = 0.58$, *Type* $r = 0.61$, and *Type/Token* $r = 0.32$ ($p < 0.05$) in Italian and English. Interestingly, a significant correlation was also found between Italian Type and English MLU $r = 0.46$ ($p < 0.05$) and between Italian Token and English Type $r = 0.53$ ($p < 0.05$): those who uses more and more diversified words in Italian, knows more words in English and are able to compose longer sentences in English.

**Discussions**

The study aimed to investigate the outcomes of an Italian-English immersion program at the narrative level. Narrative competence mirrors communicative competence ecologically and both at linguistic and cognitive levels (microstructure and macrostructure) (Hughes *et al.*, 1997). Thus, we gained more data about bilingual children's language development within an immersion program, integrating the results of the first study on Italian-English immersion program (chapter 1). Moreover, analyzing both macrostructure and microstructure, we obtained proof of the value of macrostructure in assessing bilinguals, and we detected typical patterns of language development in a TD English learning bilingual population.

In the following sections, we will discuss the macrostructure and microstructure separately, and the relations between languages in the bilingual group.

***Macrostructure***

The macrostructure and comprehension measures in Italian of bilinguals in grade 1, 3, and 5, educated in 50:50 Italian-English immersion programs (with Italian as L1 and majority language) were compared to that of Italian monolingual peers. Following Uccelli *et al.*, 2007, we expected to find similar bilingual and monolingual children's skills, both in producing and comprehending the

story grammar elements. Moreover, we expected to find improvements across the three grades (Stein *et al*., 1979).

The results confirmed our expectations concerning the comparison between the groups, with bilinguals and monolinguals performing similarly.

Concerning the development of children's performance across grades, we found effects on the following Italian measures: total macrostructure, goal, medium complexity (the combination of goal with attempt or outcome), and mental state scores. In total macrostructure, we found an improvement in performance across grades, even though it was significant only from grade 1 to grades 3 and 5. In goal and GA/GO scores we found an effect of grade only in the monolingual group, from grade 1 to grade 3. Bilingual showed a slight improvement even if it did not turn out to be significant. Considering mental state production, we found an improvement across grades, as expected. Once again, it was significant from grade 1 to grade 3 and 5, probably because of the cross-sectional design of the study.

Considering the bilingual group's English macrostructure development, we found some effects, reflecting an improvement across grades in English as a second language. We found improvement in the total macrostructure score, attempt score, AO score (low complexity score), GAO score (high complexity score), and mental state score. With the exceptions of GAO and mental state scores, where the improvement was significant only between first- and fifth-graders, the significant improvements were detected between grade 1 and grades 3 and 5 (with no significant improvement from grade 3 to grade 5), as for Italian macrostructure. Once again, it might be a consequence of the cross-sectional design.

As Stein *et al*. (1979) stated, children's ability to structure and organize narratives develops during primary schooling, including increasing narrative elements, temporal and causal relations between events, and mental states. This ability reflects the maturation of cognitive processes implicated in interpreting intentionality and making inferences about factual and emotional aspects of stories (Westby, 2005), and this is valid both for monolingual and bilingual school-aged children.

In general, analyzing the macrostructure narrative production level, the results proved that immersion education does not disadvantage the language development of bilingual children and that the immersion program remains a reliable education system (Costa, Guasti & Sharley, 2020).

We did not find a difference in comprehension between groups, but neither an improvement across grades in the bilingual group, both in Italian and English. As for English, this might be related to a weakness in English-L2 vocabulary, which has been shown to be smaller with respect to monolinguals (Hoff, Rumiche, Burridge, Ribot & Welsh, 2014). This is further plausible in the case of bilingual children learning English as a second language in an Italian speaking country where Italian represents the majority language outside school. As for Italian, even if the bilinguals' vocabulary size did not turn out to be statistically different from that of the monolingual group (see the first study of the thesis), vocabulary from Grade 3 to Grade 5 did not grow sharply in the bilingual group, and this might be related to the absence of improvement in the narrative comprehension. Nevertheless, we did not find an improvement either in the monolingual group, which might be due to the study's design.

***Microstructure***

We compared the performance in Italian of bilinguals, at the microstructure level, to that of Italian monolingual peers. As demonstrated in the present thesis's first study, we did not find a disadvantage effect of bilingual education on the L1. Moreover, Italian is the dominant-majority language of the bilingual sample, as they were all born in Italy from Italian families, and they live in Italy. These considered, we did not expect to find a difference between the two groups of children at the Italian microstructure level. Considering the development, we did expect to find improvements across the three grades, with older children performing better than the younger.

Considering the comparison between bilinguals and monolinguals, the results were in line with expectation and with the results found in the first study (chapter 1): no significant differences were found across the microstructure measures.

As far as microstructure development across grades is concerned, we found effects on all the lexical measures: token, type, and their ratio (type/token). As for macrostructure, the improvement was not linear across the three grades in token and type measures, being relevant only from grade 1 to grades 3 and 5. This was not totally in line with our expectations, but being the study a cross-sectional one, the result reflects a possible weakness of the Grade 5 group, and it is in line with previous results. Even in the type/token ratio, the effect was present only from first- to third-graders, but this time first-graders outperforming third-graders. In fact, children in grade 1 seemed to produce more diversified words, with respect to third-graders, but it was relative to the total number of words produced. First-graders produced stories with fewer different words in absolute, but they also produced few words in total; thus, the proportion resulted higher in average. Nevertheless, in the case of the lexical diversity index, a little improvement was noticeable from grade 3 to grade 5, even though it did not turn out to be significant.

At the morphosyntactic level, we mostly found an improvement from grade 1 to grades 3 and 5 in the proportion of errors of commission and omission in free and bound morphemes on tokens, even if the percentage of errors was very low since grade 1. Thus, we can conclude that a bilingual education in an Italian-English immersion program does not invalidate the linguistic development in Italian L1 (Costa *et al.*, 2020).

Considering the bilinguals' English L2 microstructure development, we expected to find children improving across grades in most measures. Results confirmed our expectations as we found improvement across the three grades in token and type measures, in MLU, and the proportion of errors of commission and omission in free and bound morphemes on tokens. As predicted, we found more English-bound morphology errors compared to free morphology, especially in grade 1 and 3. This behavior was in line with a typical English-L2 learning process, where TD children at initial stages of English acquisition tend to commit frequent errors in the use of bound morphology (Brown, 1973). The most frequent errors bilingual English-learners committed were, in fact, the omission of *–s* third singular person, of *–ed* past tense, the regularization of irregular single past

forms, or the use of auxiliaries as tense markers. By grade 5, the use of bound morphology has improved, with fewer errors and no difference between bound and free morphology errors and omissions. This is compatible with the idea that there is positive influence from Italian L1 – a richly inflected language – on the English L2 development, as found in a study by Blom *et al.* (2012) with children speaking richly inflected languages.

In conclusion, analyzing the narrative level, results proved that the immersion education does not disadvantage the language development of bilingual children in both their languages.

**Correlations between Italian and English in the bilingual group**

To verify the relationship between the Italian and English narrative skills in the bilingual group we carried out correlation analyses. As predicted by Gagarina *et al.* (2012), we expected to find evidence of a transfer of competencies for the macrostructure representation between L1 and L2, with positive correlations among some macrostructure measures. Bilingual children performed similarly across Italian and English at the total macrostructure scores and the mental state scores. This similarity reflected that children who produced a more rich and informative story, with a fair linguistic complexity, in Italian also did in English. Interestingly, both the Italian scores correlated with the English comprehension score, suggesting that a good ability in story production in L1 supports the ability to comprehend a story macrostructure in L2.

Considering microstructure, we expected to find correlations between Italian and English at least at the lexical level, resulting from a transfer of knowledge from L1 to L2 (Cummins, 1991). Results confirmed our expectations, as we found correlations in token, type, and type/token ratio measures in Italian and English.

As bilingual children are learners of English as a second language, they committed more errors at the morphology level in English than in Italian – bound morphology in particular – especially in the initial stages of their L2 learning, as found in other works on early second language acquisition (*e.g.*, Blom *et al.*, 2012). We expected longer utterances in Italian than in

English, partially because English is a second language, partially because of the intrinsic features of synthesis in the English syntax. Thus it was plausible to not find correlations at the morphosyntactic level as we did. Interestingly, a significant correlation was found between Italian Type and English MLU and between Italian Token and English Type, suggesting that children who use more and more diversified words in Italian know more words in English and can compose longer sentences in English.

### *Limitations*

The current study has some limitations. Firstly, the study has a cross-sectional design, which could justify some partially unexpected results on the development across grades, with a Grade 5 sample's weakness. Moreover, the study did not include an English monolingual control group; thus, a comparison between bilinguals and English monolinguals' English performance was not possible, although it would be interesting. It could have enabled us to verify the development outcome also in the second language of education. Lastly, we performed low detailed analyses of morphology for microstructure indexes to have a first idea of children's development. Future studies should consider implementing more detailed analyses, to have a more accurate vision of the development of morphosyntactic abilities.

## Conclusions

The present study completed the results of the first study of the present thesis on Italian-English immersion program, providing data about bilingual children's language use at the narrative level, as it is considered a more natural communicative competence (Botting, 2002).

Considering the comparison between bilinguals and monolinguals, our predictions were all confirmed by the present study results. We did not find differences between bilingual and Italian monolingual children in both macrostructure and microstructure measures.

In Italian, bilingual primary school students showed improvements both at some macrostructure and microstructure measures. Nevertheless, we mostly found improvements from

grade 1 to grade 3, and a downfall in Grade 5 performance, but this could be attributed to the study's sampling. We detected the same development trend in English macrostructure, demonstrating a similar pattern of development across the two languages. A more linear development from grade 1 to grade 5 was found at the English microstructure level. All considered, we could plausibly hypothesize a positive transfer of skills from one language to the other, where Italian could support the same ability in English.

We also proved that errors in the English bound morphology are typical of initial stages of English acquisition and have not to be interpreted as a signal of language difficulties. Italian bilingual children learning English as a second language, in fact, improved their English morphology competence by grade 5.

Finally, we found some positive correlations between the bilinguals' Italian and English performances both at the macrostructure and lexical microstructure, confirming the effect of a transfer from one language to the other.

We can conclude that even exploring the narrative level of competence, a *simultaneous Italian-English immersion program* does not hinder the Italian L1, and fosters the right development in both the languages of education.

**References**

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism, 11(1), 1-29.*

Berman, R. (1988). On the ability to relate events in narrative. *Discourse Processes*, *11*, 469-497.

Berman, R. A., & Slobin, D. I. (1994). Narrative structure. In R. A. Berman and D. I. Slobin (Eds.), *Relating events in narrative: A cross-linguistic developmental study* (pp. 39-84). New York/London: Psychology Press.

Bliss, L., McCabe, A., & Miranda, E. (1998). Narrative assessment profile: Discourse analysis for school-age children. *Journal of Communication Disorders, 31*, 347-363.

Blom, E., Paradis, J., & Soreson Duncan, T. (2012). Effects of input properties, vocabulary size and L1 on the development of third person singular –s in child L2 English. Language Learning, 62, 965-994.

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: implications for clinical practice. *Int J Lang Commun Disord,* 51(6),626-638.

Bonifacci, P., Barbieri, M., Tomassini, M., & Roch, M. (2018) In few words: Linguistic gap but adequate narrative structure in preschool bilingual children. *Journal of Child Language*, 45(1), 120-147.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, 18, 1-21.

Brown, R. (1973). *A first language: The early stages*. London: George Allen & Unwin.

Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology*, 21, 335-351.

Cain, K., & Oakhill, J. (1996). The nature of the relationship between comprehension skill and the ability to tell a story. *British Journal of Developmental Psychology, 14*, 187-201.

Cleave, P. L., Girolametto, L. E., Chen, X., & Johnson, C. J. (2010). Narrative abilities in monolingual and dual language learning children with specific language impairment. *Journal of Communication Disorders*, 43(6), 511-522.

Costa, F., Guasti, M. T., & Sharley, S. (2020). Double Literacy Effects on Language and Reading Skills in Italian–English Primary School Children. *Journal of Monolingual and Bilingual Speech* 2(2), 185-218.

Cowley, J., & Glasgow, C. (1994). *The Renfrew Bus Story: Language Screening by Narrative Recall*. Circle Pines, MN: American Guidance.

Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70-89). Cambridge, England: Cambridge University Press.

Devescovi, A., & Pizzuto, E., (1995). Lo sviluppo grammaticale. In Sabbadini, G., (eds), Manuale di neuropsicologia dell'età evolutiva, Bologna: Zanichelli.

Farrant, B., Fletcher, J., & Maybery, M. T. (2006). Specific language impairment, theory of mind, and visual perspective taking: evidence for simulation theory and the developmental role of language. *Child Development*, 77, 1842-1853.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech and Hearing Services in Schools*, 35, 155-168.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives, *ZAS Papers in Linguistics* 56. Berlin: ZAS. (Includes manual and MAIN materials for assessment, available online at http://www.zas.gwz-berlin.de/zaspil56.html).

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. 2015. Assessment of Narrative Abilities in Bilingual Children. In S. Armon-Lotem, J. de Jong & N. Meir (eds.) *Methods for assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol, UK: Multilingual Matters.

Gagarina, N., Klop, D., Tsimpli, I., & Walters, J. (2016). Narrative abilities in bilingual children. *Applied Psycholinguistics*, 37(1), 11-7.

Gillam, R. B., & Carlisle, R. (1997). Oral reading and story retelling of students with specific language impairment. *Language, Speech, and Hearing Services in Schools, 28*, 30-42.

Gutiérrez-Clellen, V. F., 2002. Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education, 13(2*), 175-197.

Hadley, P. A. (1998). Language sampling protocols for eliciting text-level discourse. *Language, Speech, and Hearing Services in Schools, 29*, 132-147.

Hayward, D. V., Gillam, R. B., & Lien, P. (2007). Retelling a script-based story: Do children with and without language impairments focus on script and story elements? *American Journal of Speech-Language Pathology, 16*, 235-245.

Heilmann, J., Miller, J. F., & Nockerts, A. (2010). Sensitivity of narrative organization measures using narrative retells produced by young school- age children. *Language Testing*, 27, 603-626.

Hoff, E., Rumiche, R., Burridge, A., Ribot, K. M., & Welsh, S. N. (2014). Expressive vocabulary development in children from bilingual and monolingual homes: A longitudinal study from two to four years. *Early Childhood Research Quarterly*, *29*(4), 433-444.

Hughes, D., McGillvray, L., & Schmidek, M. (1997). *Guide to narrative language: Procedures for assessments*. Eau Claire, WI: Thinking Publications.

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition, 15*(1), 58-74.

Kit-Sum To, C., Stokes, S. F., Cheung, H., & T'sou, B. (2010). Narrative Assessment for Cantonese-Speaking Children. *Journal of Speech, Language and Hearing Research, 53(3)*, 648-669.

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders*, 43, 456-473.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Lever, R., & Sénéchal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology, 108*, 1-24.

Levorato, M. C., & Nesi, B. (2001). *Imparare a comprendere e produrre testi*. In Camaioni, L., *Psicologia dello sviluppo del linguaggio*. Il Mulino, Bologna.

Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech and Hearing Research, 38*, 415-425.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Marini, A., Andreetta, S., del Tin, S., & Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. Aphasiology 25(11), 1372–92.

McCabe, A. (1996). Evaluating narrative discourse skills. In K. Cole, P. Dale and D. Thal (Eds.), *Assessment of communication and language* (pp. 121-142). Baltimore, MD: Paul H. Brookes.

McCabe, A., & Bliss, L. S. (2005). *Narratives from Spanish-speaking children with impaired and typical language development.* Imagination, Cognition and Personality, 24, 331-346.

Meritt, D. D., & Liles, B. Z. (1989). Narrative analysis: Clinical applications of story generation and story retelling. *Journal of Speech and Hearing Disorders*, 54, 438-447.

Nelson, K. (2007). *Young minds in social worlds: Experience, meaning and memory.* Harvard University Press, Cambridge, MA.

Nippold, M. A., Ward-Lonergan, J. M., & Fanning, J. L. (2005). Persuasive writing in children, adolescents, and adults: A study of syntactic, semantic, and pragmatic development. *Language, Speech, and Hearing Services in Schools*, 36, 125-138.

Ohakill, J. V., & Cain, K. (2012). The precursors of reading comprehension and word reading in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, 6, 91-121.

Olley, L. (1989). Oral narrative performance of normal and language impaired school aged children. *Australian Journal of Human Communication Disorders, 17*, 43-65.

Oller, D., Pearson, B., & Cobo-Lewis, A. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, 28(2), 191-230.

Paradis, J., Genesee, F., & Crago, M. (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning* (2nd ed.). Baltimore, MD: Brookes.

Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating children with language impairment among English-language learners from diverse first-language backgrounds. *Journal of Speech, Language, and Hearing Research*, 56, 971-981.

Paradis J. (2016). The Development of English as a Second Language With and Without Specific Language Impairment: Clinical Implications. *J Speech Lang Hear Res,* 59(1), 171-82.

Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In Oller, D. K., & Eilers, R. E., *Language and Literacy in Bilingual Children*. Multilingual Matters, Clevedon, 135-174.

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narratives.* New York: Plenum Press.

Reilly, J., Losh, M., Bellugi, U., & Wulfeck, B. (2004). "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language, 88*, 229-247.

Rezzonico, S., Chen, X., Cleave, P. L., Greenberg, J., Hipfner-Boucher, K., Johnson, C. J., Milburn, T., Pelletier, J., Weitzman, E., & Girolametto, L. (2015). Oral narratives in monolingual and bilingual preschoolers with SLI. *Int J Lang Commun Disord*, 50(6), 830-41.

Richards, M. (1987). Type/token ratios: What do they really tell us? *Journal of Child Language*, 14(2), 201-9.

Roch, M., Florit, E., & Levorato, C. (2016). Narrative skills in Italian-English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, 37, 49-67.

Rondal, J. A., Ghiotto, M., Bredart, S., & Bachelet, J. F. (1987). Age-relation, reliability and grammatical validity of measures of utterance length. *J Child Lang*, 14(3), 433-46.

Roth, F. P., & Spekman, N. J. (1986). Narrative discourse: Spontaneously generated stories of learning-disabled and normally achieving students. *Journal of Speech and Hearing Disorders, 51,* 8-23.

Schneider, P., & Dubé, R. V. (2005). Story presentation effects on children's retell content. *American Journal of Speech-Language Pathology, 14*(1), 52-60.

Schneider, P., Dubé, R. V., & Hayward, D (2005). The Edmonton Narrative Norms Instrument.

Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Journal of Speech Pathology and Audiology,* 30(4), 224-238.

Schwartz, M., & Shaul, Y. (2013). Narrative development among language-minority children: the role of bilingual versus monolingual preschool education. *Language, Culture and Curriculum,* 26(1), 36-51.

Shapiro, L. R., & Hudson, J. A. (1991). *Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives.* Developmental Psychology, 27, 960-964.

Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexicon and grammar. Applied Psycholinguistics 30(2), 315–37.

Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gilliam, R. B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language and Communication Disorders*, 49, 60-74.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing: Multidisciplinary perspectives* (Vol. 2, pp. 53-120). Hillsdale, NJ: Erlbaum.

Trabasso, T., Stein, N. L., & Johnson, N. S. (1981). Children's knowledge of events: A causal structure. In G. Bower (Ed.), *Learning and motivation* (Vol. 15, pp. 237-282). New York: Academic Press.

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. Discourse Processes, 15, 249-275.

Trabasso, T., & Rodkin, P. C. (1994). Knowledge of goals/plans:A conceptual basis for narrating "Frog where are you?". In R. A. Berman and D. I. Slobin (Eds.), *Relating events in narrative: A cross linguistic developmental study* (pp. 85-106). Mahwah, NJ: Lawrence Erlbaum Associates.

Tsimpli, I., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics*, 37(1), 195-216.

Uccelli, P., & Paéz, M. M. (2007). Narrative and vocabulary development on bilingual children from kindergarten to first grade: developmental changes and associations among English and Spanish skills. *Language, Speech and Hearing Services in Schools*, 38, 225-236.

Vender, M., Garraffa, M., Sorace, A., & Guasti, M. T. (2016). How early L2 children perform on Italian clinical markers of SLI: A study of clitic production and nonword repetition. *Clinical Linguistics and Phonetics*, *30*(2), 150-169.

Verhoeven, L. (2004). *Bilingualism narrative and construction.* In Stromqvist, S., & Verhoeven, L., *Relating Events in Narrative: Typological and Contextual Perspectives.* Lawrence Erlbaum, Hillsdale, NJ, 435-454.

Walach, G. P. (2008). *Language intervention for school-age students: Setting goals for academic success*. St. Louis, MO: Mosby.

Wellman, R. L., Lewis, B. A., Freebairn, L. A., Avrich, A. A., Hansen, A. J., & Stein, C. M. (2011). Narrative ability of children with speech sound disorders and the prediction of later literacy skills. *Language, Speech, and Hearing Services in Schools*, 42(4), 561-57.

Westby, C. E. (2005). Assessing and facilitating text comprehension problems. In H. Catts and A. Kamhi (Eds.), *Language and reading disabilities* (pp. 157-232). Boston, MA: Allyn & Bacon.

Zdorenko, T., & Paradis, J. (2012). Articles in child L2 English: When L1 and L2 acquisition meet at the interface. *First Language*, 32(1-2), 38-62.

## Appendix

***Pictures used in the study for the story retelling tasks* (from the *Multilingual Assessment Instrument for Narratives, MAIN*, Gagarina, *et al.*, 2012).**

**Figure 1** Il cane (Dog) stimulus pictures (Italian story retelling).

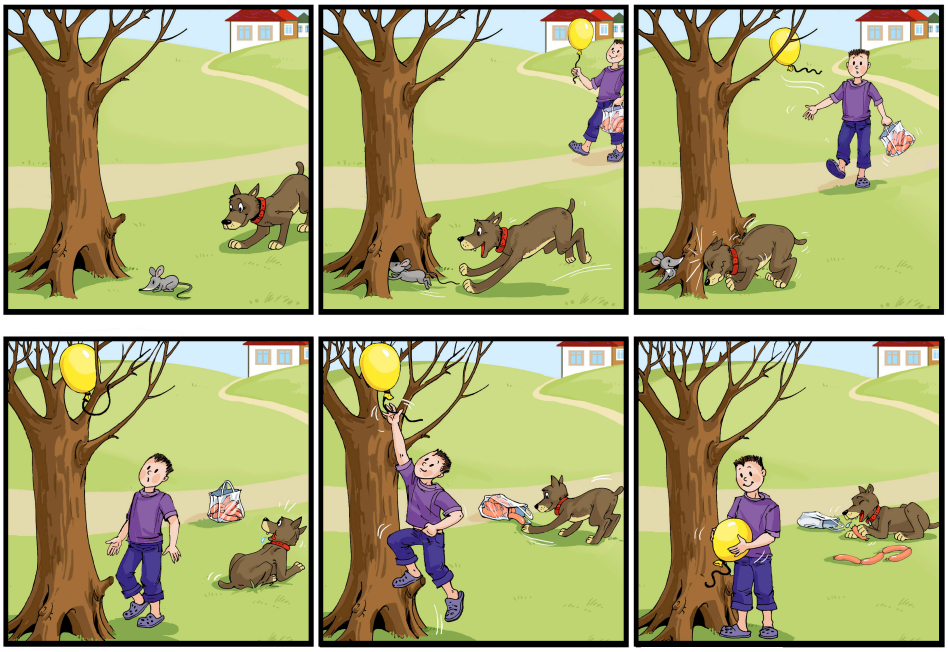
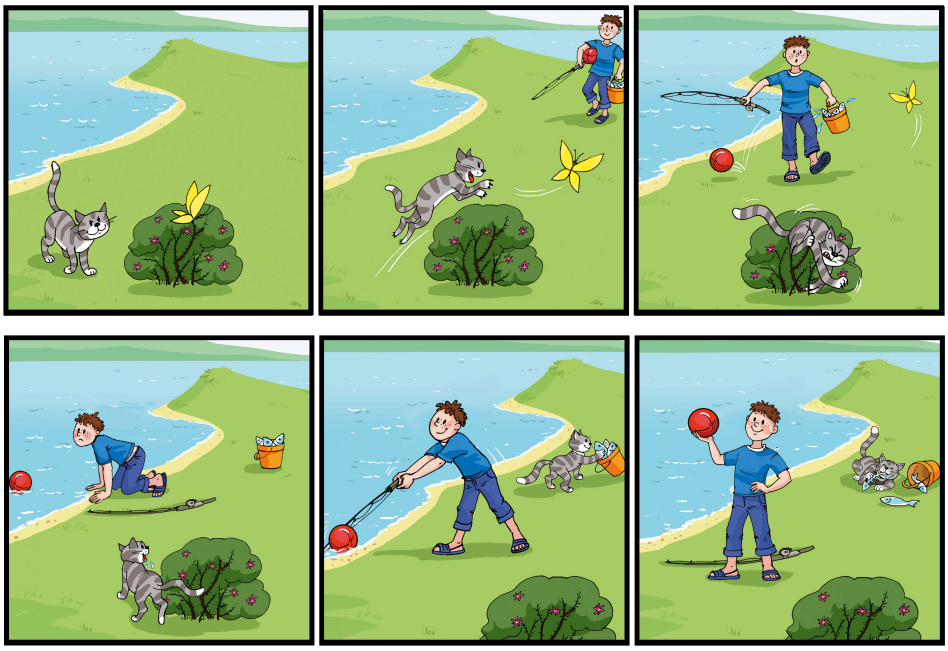


**Figure 2** Cat stimulus pictures (English story retelling).



## Conclusions

The present thesis aimed at exploring different contexts in which children may grow up bilingual in Italy.

We firstly explored the context of bilingual education in school, where a child is *immersed* in two languages simultaneously and to the same extent. In particular, we considered simultaneous (or 50:50) Italian-English immersion program, the most represented and coveted bilingual program in Italian bilingual private primary schools. The program foresees that Italian and English are education vehicles of contents, and not only academic subjects. Thus, children learn the two languages by being consistently and continuously exposed to the languages themselves.

As many misbeliefs still surround bilingual education, we wanted to explore the Italian-English immersion educational outcomes in primary school children, from the language, reading, cognitive, and narrative point of view. Children were early sequential bilingual, exposed to English as a second language (L2) by the age of three in kindergarten. The aims of the first and third cross-sectional studies of the present thesis were to verify (1) whether learning to read in Italian (L1) is delayed when simultaneously learning to read in English (L2); (2) whether language and reading skills in English improve across the primary school years; (3) whether the bilinguals' Italian language and reading attainment correlate to their English performance; (4) whether oral language and working memory abilities correlate with reading proficiency measures (Kovelman, Baker & Petitto, 2008; Alloway & Alloway 2010); (5) whether bilinguals show an advantage on working memory; (6) whether bilingual children showed similar narrative macrostructure, comprehension and microstructure skills in Italian (L1), compared to Italian monolingual peers; (7) whether narrative competence in Italian and in English develop in a parallel way across grades at the macro- and microstructure levels in bilingual children; finally, (8) whether Italian narrative macrostructure and microstructure competence in bilinguals correlate with same skills in English (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker & Walters, 2012).

Our expectations were: (1) to find no disadvantage in Italian language, reading and narrative skills in the bilingual group, compared to Italian monolingual peers; (2) to find improvements across the grades in Italian and English language, reading and narrative measures; (3) to find correlations between the language and reading skills in Italian and English, suggesting a positive transfer from Italian to English; (4) to find correlations between language and reading abilities in both the languages, supporting the idea that oral language supports reading development; (5) to find no bilingual advantage on memory tasks, but a positive correlation between memory skills and reading proficiency measures; to find correlations between Italian and English narrative macrostructure measures and some correlation at the microstructure level (at the lexical level).

The results confirmed all our predictions concerning language, reading, and narrative measure. In line with the literature (*e.g.,* Genesee & Jared, 2008; Lindholm-Leary & Hernández, 2011; Berens, Kovelman & Petitto, 2013; Bialystok, 2018; Costa, Sharley & Guasti, 2019; Uccelli & Paéz, 2007), early sequential bilingual Italian-English students attending a simultaneous immersion program at primary school are not disadvantaged or delayed on their language, reading, and narrative development in Italian (L1) compared to Italian monolinguals, and they improve in both Italian and English across grades, consolidating their skills. In particular, the narrative competence, which is considered a more ecological and natural way to understand children's communicative and language skills (Botting, 2002), showed improvements across grades both at the macrostructure (Stein & Glenn, 1979) and microstructure levels.

We found positive correlations between the bilinguals' Italian and English performances in reading, supporting the evidence of a presumable transfer of skills from Italian L1 – a language with a transparent orthography – to English L2, which has an opaque orthography (D'Angiulli & Serra, 2001; Bialystok, Luk & Kwan, 2005; Montanari, 2013; Blom, Paradis & Sorenson Duncan, 2012). Correlations between skills in the two languages were also found in language measures; for example, we found some positive correlations in narrative abilities between the two languages, both

at the macrostructure and lexical microstructure, in line with Gagarina *et al.* (2012) and Cummins (1991).

Exploring the relationship between oral language abilities and reading measures, we found that language abilities in one language correlate with reading abilities in the same language. This finding supports the idea that oral language supports reading development and is consistent with Kovelmann *et al.*'s (2008) results. These authors showed that the age of first oral bilingual exposure impacts literacy development at age 8 and found strong associations between language proficiency and reading in bilingual children.

In line with Ratiu and Azuma (2015), and Lehtonen, Soveri, Laine, Järvenpää, de Bruin, and Antfolk (2018), we found that working memory skills are not boosted by a bilingual condition, which does not provide an advantage on this ability. However, all the memory tasks showed to improve across grades in both Italian and English, and they show to predict reading performance in both languages (Swanson & Beebe-Frankenberger, 2004; White, 2019).

In the light of our findings, we can conclude that, when early exposed to English as a second language, children in a simultaneous Italian-English immersion program are not disadvantaged; they acquire a good proficiency in both the languages and develop across grades. In line with Costa, Guasti, and Sharley (2020), a bilingual immersion program represents a reliable education system.

Learning English as a second language gives access to many social and job possibilities for our children's future adults. Bilingual education should be fostered in modern Italian society and should be an available educational opportunity for all children. In this direction, our final aim was to make education stakeholders, parents, and all the public community aware of bilingual language and academic outcomes.

The second bilingual background we explored is the education of bilingual children in Italy with a minority first language. They are mostly represented by immigrant children and are usually referred to as *Heritage bilinguals.* This bilingual population learns the first-minority language at home – or when one, in the cultural communities – mostly orally, but rarely achieves a full L1

development. They learn Italian only once they enter Italian schools, often in its oral and written form at the same time. It follows that it is harder for them to acquire Italian literacy compared to Italian mother-tongue peers.

We investigated for the first time the language, reading, and cognitive outcomes of the biliteracy education at primary school, where heritage bilinguals learn to read and write both in Italian-L2 (at Italian mainstream school) and the minority-L1 (mostly in weekend classes organized by the community). Children were sequential bilingual, exposed to Italian as a second-majority language (L2). They were of immigrant origin, living stably in Italy and starting primary school since grade 1. We took into account most represented minority languages in Italy: Mandarin Chinese, Arabic, Spanish, Ukraine, Polish and Russian, among the others. Biliteracy initiatives have started to arise in Italy on the wave of evidence that oral and written support to the minority-L1 could enhance the majority-L2 literacy acquisition (*e.g.*, Papastefanou, Powell, & Marinis 2019; van der Velde Kremin, Arredondo, Hsu, Satterfield, & Kovelman, 2019), but they are still few.

The aims of the second cross-sectional study of the present thesis were to verify (1) whether bilingual-biliterate children gain some advantage compared to bilingual-monoliterate children in cognitive, vocabulary, and reading in Italian (L2); (2) whether both bilingual-biliterate and bilingual-monoliterate automatized reading around age 8 (grade 3), as in monolingual children development, diverging in reading strategies for words and pseudowords (Bellocchi, Bonifacci, & Burani 2016); (3) whether there are correlations between the various cognitive, linguistic, and reading measures, in both groups of bilinguals.

Our expectations were: (1) to find a cognitive advantage in the biliterate group compared to monoliterates, because of their higher bilingual proficiency (Blom, Boerma, Bosma, Cornips & Everaert, 2017; Weber, Johnson, Riccio & Liew, 2016) or at least to find no clear-cut disadvantage; (2) to find similar performance at expressive vocabulary between the bilingual groups, as biliteracy should not influence the vocabulary width in Italian L2; to find a possible beneficial effect of biliteracy in reading performance, compared to monoliteracy (Papastefanou, *et al.,* 2019); (3) to

find both bilingual groups reaching mature decoding skills by Grade 3, with words read faster and better than pseudowords (Bellocchi, *et al.*, 2016); (4) lastly, to find cognitive and language performance to correlate with reading proficiency measures (Gathercole, Brown & Pickering, 2003; Duff, Reen, Plunkett & Nation, 2015).

Biliterate bilinguals showed better performance than monoliterates at non-verbal intelligence (as evaluated with Raven), in line with Weber, *et al.*, 2016. Monoliterate bilinguals were better at short-term memory (as evaluated with pseudowords repetition) instead, in line with Sorenson, Duncan, and Paradis (2016). Concerning language, we did not find a difference in expressive vocabulary between biliterate and monoliterate children, as expected. Reading expectations were not fully satisfied: biliterates developed at comparable levels with their monoliterate peers except for reading comprehension, in which they were better. Thus, the cross-sectional study would suggest that biliteracy enhances reading comprehension, a fundamental ability for academic success (in line with Rolstad, Mahoney & Glass, 2005) even though not literacy skills in general. However, no disadvantages emerged from the biliteracy mode on decoding skills, as both groups start using the lexical route to reading from grade 3, as expected (Bellocchi, *et al.*, 2016). Moreover, this result is in line with a recent study on automatization of reading in dual-language children of Guasti, Costa, Chailleux, Pagliarini, Sarti, Granocchio & Stucchi (submitted), even if our bilinguals were not uniformly exposed to oral Italian from age 3.

Finally, we found that cognitive and language performance correlated with reading proficiency measures, in line with literature that supports the contribution of cognitive and language skills to literacy development (Gathercole *et al.*, 2003; Duff *et al.*, 2015).

To conclude, even if there might not be a comprehensive advantage in all the reading measures, educating heritage bilinguals children also in their L1 enhances some crucial skills for their academic success, such as fluid intelligence, which contribute to reading development, and reading comprehension, the baseline for the *reading-to-learn* school subjects. Moreover, it does not hamper literacy development in Italian.

Undeniably biliteracy provides some advantages in some critical skills over monoliteracy. Thus, sustaining biliteracy route to learning appears an excellent choice to be implemented in Italian educational policy.

In Italy, bilingualism, and bilingual education, are general terms, which enclose children of heterogeneous extractions, backgrounds, needs, and goals. Bilingual immersion programs for monolinguals and bilingual children's biliteracy education are significantly different from each other (Byalistok, 2018). At the same time, they are both effective and row together towards a more inclusive society and a beneficial educational experience for future generations of children.

**References**

Bellocchi, S., Bonifacci, P., & Burani, C. (2016). Lexicality effect and stress assignment in bilingual children reading Italian as a second language. Bilingualism: Language and Cognition, 19(1), 89-105.

Berens, M. S., Kovelman, I., Petitto, L. A. (2013). Should bilingual children learn reading in two languages at the same time or in sequence? *Biling Res J.*, 1;36(1), 35-60.

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, *9*(1), 43-61.

Bialystok, E. (2018). Bilingual education for young children: review of the effects and consequences. *Int. J. Biling. Educ. Biling*., 21(6), 666-679.

Blom, E., Paradis, J., & Soreson Duncan, T. (2012). Effects of input properties, vocabulary size and L1 on the development of third person singular –s in child L2 English. Language Learning, 62, 965-994.

Blom, E., Boerma, T., Bosma, E., Cornips, L., & Everaert, E. (2017). Cognitive Advantages of Bilingual Children in Different Sociolinguistic Contexts. *Front. Psychol.* 8(552).

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, 18, 1-21.

Costa F., Sharley, S. & Guasti M. T. (2019). Does English delay learning to read Italian in primary school children? *Rivista di grammatica generativa*.

Costa, F., Guasti, M. T., & Sharley, S. (2020). Double Literacy Effects on Language and Reading Skills in Italian–English Primary School Children. *Journal of Monolingual and Bilingual Speech* 2(2), 185-218.

Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70-89). Cambridge, England: Cambridge University Press.

D'Angiulli, A., & Serra, E. (2001). The development of reading in English and Italian bilingual children. *Applied Psycholinguistics*, 22, 479-507.

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes?. *Journal of child psychology and psychiatry, and allied disciplines*, 56(8), 848–856.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U. & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives, *ZAS Papers in Linguistics* 56. Berlin: ZAS. (Includes manual and MAIN materials for assessment, available online at http://www.zas.gwz-berlin.de/zaspil56.html).

Gathercole, S. E., Brown, L., & Pickering, S. J. (2003). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment levels. *Educational and Child Psychology, 20*(3), 109-122.

Genesee, F., Jared, D. (2008). Literacy development in early French immersion programs. *Canadian Psychology/Psychologie Canadienne, 49*(2), 140-147.

Guasti, M. T., Costa, F., Chailleux, M., Pagliarini, E., Sarti, D., Granocchio, E., & Stucchi, N. (submitted). Diagnosing dyslexia in L2 Italian-speaking children: Are monolingual tools efficient? *Journal of research in reading*.

Kovelman, I., Baker, S. A., & Petitto, L. A. (2008). Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism: Language and Cognition*, *11*(2), 203-223.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018) Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychol Bull*., 144(4), 394-425.

Lindholm-Leary, K.J., Hernández, A. (2011). Achievement and language proficiency of Latino students in dual language programmes: native English speakers, fluent English/previous ELLs, and current ELLs. *Journal of Multilingual and Multicultural Development*, 32(6), 531-545.

Montanari S. (2013). A Case Study of Bi-literacy Development among Children Enrolled in an Italian–English Dual Language Program in Southern California. *International Journal of Bilingual Education and Bilingualism*, 17, 509-525.

Papastefanou, T., Powell, D., & Marinis, T. (2019). Language and Decoding Skills in Greek-English Primary School Bilingual Children: Effects of Language Dominance, Contextual Factors

and Cross-Language Relationships Between the Heritage and the Majority Language. *Front. Commun*., 4(65).

Ratiu, I., & Azuma, T. (2015). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology*, *27*(1), 1-11.

Rolstad, K., Mahoney, K., & Glass, G. V. (2005) The Big Picture: A Meta-Analysis of Program Effectiveness Research on English Language Learners. *Educational Policy*, 19(4), 572-594.

Sorenson Duncan, T., & Paradis, J. (2016). English language learners' nonword repetition performance: The influence of age, L2 vocabulary size, length of L2 exposure, and L1 phonology. *Journal of Speech, Language, and Hearing Research, 59,* 39-48.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing: Multidisciplinary perspectives* (Vol. 2, pp. 53-120). Hillsdale, NJ: Erlbaum.

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The Relationship Between Working Memory and Mathematical Problem Solving in Children at Risk and not at Risk for Math Disabilities. *Journal of Education Psychology,* 96, 471-491.

Uccelli, P. & Paéz, M. M. (2007). Narrative and vocabulary development on bilingual children from kindergarten to first grade: developmental changes and associations among English and Spanish skills. *Language, Speech and Hearing Services in Schools*, 38, 225-236.

van der Velde Kremin, L., Arredondo, M. M., Hsu, L., Satterfield, T., & Kovelman, I. (2019). The effects of Spanish heritage language literacy on English reading for Spanish-English bilingual children in the US. *International journal of bilingual education and bilingualism*, *22*(2), 192–206.

Weber, R. C., Johnson, A., Riccio, C. A., & Liew, J. (2016). Balanced bilingualism and executive functioning in children. *Bilingualism* 19, 425-431.

White, M. J. (2019). The development of English proficiency and working memory in 5–6 year old ELLs in their first year of formal education. *International Journal of Bilingual Education and Bilingualism*, DOI: 10.1080/13670050.2019.1571009.